

NIST



WTEC Panel Report on

INTERNATIONAL ASSESSMENT OF RESEARCH AND DEVELOPMENT IN SIMULATION-BASED ENGINEERING AND SCIENCE

Sharon C. Glotzer (Chair)
Sangtae Kim (Vice Chair)
Peter T. Cummings
Abhijit Deshmukh
Martin Head-Gordon
George Karniadakis
Linda Petzold
Celeste Sagui
Masanobu Shinozuka



World Technology Evaluation Center, Inc.
4800 Roland Avenue
Baltimore, Maryland 21210



ABSTRACT

This WTEC panel report assesses the international research and development activities in the field of Simulation-Based Engineering and Science (SBE&S). SBE&S involves the use of computer modeling and simulation to solve mathematical formulations of physical models of engineered and natural systems. SBE&S today has reached a level of predictive capability that it now firmly complements the traditional pillars of theory and experimentation/observation. As a result, computer simulation is more pervasive today – and having more impact – than at any other time in human history. Many critical technologies, including those to develop new energy sources and to shift the cost-benefit factors in healthcare, are on the horizon that cannot be understood, developed, or utilized without simulation. A panel of experts reviewed and assessed the state of the art in SBE&S as well as levels of activity overseas in the broad thematic areas of life sciences and medicine, materials, and energy and sustainability; and in the crosscutting issues of next generation hardware and algorithms; software development; engineering simulations; validation, verification, and uncertainty quantification; multiscale modeling and simulation; and SBE&S education. The panel hosted a U.S. baseline workshop, conducted a bibliometric analysis, consulted numerous experts and reports, and visited 59 institutions and companies throughout East Asia and Western Europe to explore the active research projects in those institutions, the computational infrastructure used for the projects, the funding schemes that enable the research, the collaborative interactions among universities, national laboratories, and corporate research centers, and workforce needs and development for SBE&S.

The panel found that SBE&S activities abroad are strong, and compete with or lead the United States in some strategic areas. Both here and abroad, SBE&S is changing the way disease is treated, the way surgery is performed and patients are rehabilitated, and the way we understand the brain; changing the way materials and components are designed, developed, and used in all industrial sectors; and aiding in the recovery of untapped oil, the discovery and utilization of new energy sources, and the way we design sustainable infrastructures. Data-intensive and data-driven applications were evident in many countries. Achieving millisecond timescales with molecular resolution for proteins and other complex matter is now within reach due to new architectures and algorithms. The fidelity of engineering simulations is being improved through inclusion of physics and chemistry. There is excitement about the opportunities that petascale computers will afford, but concern about the ability to program them. Because fast computers are now so affordable, and several countries are committed to petascale computing and beyond, what will distinguish us from the rest of the world is our ability to do SBE&S better and to exploit new architectures we develop before those architectures become ubiquitous. Inadequate education and training of the next generation of computational scientists and engineers threatens global as well as U.S. growth of SBE&S. A persistent pattern of subcritical funding overall for SBE&S threatens U.S. leadership and continued needed advances, while a surge of strategic investments in SBE&S abroad reflects recognition by those countries of the role of simulation in advancing national competitiveness and its effectiveness as a mechanism for economic stimulus.

There are immediate opportunities to strengthen the U.S. capability for SBE&S through strategic investments in industry-driven partnerships with universities and national laboratories; new and sustained mechanisms for supporting R&D in SBE&S; and a new, modern approach to educating and training the next generation of researchers in high performance computing, modeling and simulation for scientific discovery and engineering innovation. Key findings in the three thematic domain areas of the study and in the crosscutting areas and technologies that support SBE&S reinforce these overarching findings in specific ways.

WORLD TECHNOLOGY EVALUATION CENTER, INC. (WTEC)

R. D. Shelton, President
Michael DeHaemer, Executive Vice President
Geoffrey M. Holdridge, Vice President for Government Services
David Nelson, Vice President for Development
V. J. Benokraitis, Assistant Vice President
L. Pearson, Project Manager
Grant Lewison (Evaluametrics, Ltd.), Advance Contractor, Europe
Gerald Hane (Globalvation, Inc.), Advance Contractor, Asia
Patricia M.H. Johnson, Director of Publications
Patricia Foland, Research Associate
Halyna Paikoush, Event Planner and Administrative Support

ACKNOWLEDGEMENTS

We at WTEC wish to extend our gratitude and appreciation to the panelists for their valuable insights and their dedicated work in conducting this international benchmarking study of R&D in simulation-based engineering and science. We wish also to extend our sincere appreciation to the advisory board, the presenters at the U.S. baseline workshop, and to the panel's site visit hosts for so generously and graciously sharing their time, expertise, and facilities with us. For their sponsorship of this important study, our thanks go to the National Science Foundation (NSF), the Department of Energy (DOE), the National Aeronautics and Space Administration (NASA), the National Institutes of Health (NIH), the National Institute of Standards and Technology (NIST), and the Department of Defense (DOD). We believe this report provides a valuable overview of ongoing R&D efforts in simulation-based engineering and science that can help scientists and policymakers effectively plan and coordinate future efforts in this important field.

R. D. Shelton

Copyright 2009 by WTEC. The U.S. Government retains a nonexclusive and nontransferable license to exercise all exclusive rights provided by copyright. This document is sponsored by the National Science Foundation (NSF) and other agencies of the U.S. Government under an award from NSF (ENG-0739505) to the World Technology Evaluation Center, Inc. The Government has certain rights in this material. Any writings, opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the United States Government, the authors' parent institutions, or WTEC. A selected list of available WTEC reports and information on obtaining them is on the inside back cover of this report.

FOREWORD

We have come to know that our ability to survive and grow as a nation to a very large degree depends upon our scientific progress. Moreover, it is not enough simply to keep abreast of the rest of the world in scientific matters. We must maintain our leadership.¹

President Harry Truman spoke those words in 1950, in the aftermath of World War II and in the midst of the Cold War. Indeed, the scientific and engineering leadership of the United States and its allies in the twentieth century played key roles in the successful outcomes of both World War II and the Cold War, sparing the world the twin horrors of fascism and totalitarian communism, and fueling the economic prosperity that followed. Today, as the United States and its allies once again find themselves at war, President Truman's words ring as true as they did more than a half-century ago. The goal set out in the Truman Administration of maintaining leadership in science has remained the policy of the U.S. Government to this day. For example, the top goal of the NSF² is to, "foster research that will advance the frontiers of knowledge, emphasizing areas of greatest opportunity and potential benefit and establishing the nation as a global leader in fundamental and transformational science and engineering."

The United States needs metrics for measuring its success in meeting this goal. That is one of the reasons that the National Science Foundation (NSF) and many other agencies of the U.S. Government have supported the World Technology Evaluation Center (WTEC) for the past 20 years. While other programs have attempted to measure the international competitiveness of U.S. research by comparing funding amounts, publication statistics, or patent activity, WTEC has been the most significant public domain effort in the U.S. Government to use peer review to evaluate the status of U.S. efforts in comparison to those abroad. Since 1989, WTEC has conducted over 60 such assessments in a wide variety of fields including advanced computing, nanoscience and nanotechnology, biotechnology, and advanced manufacturing.

The results have been extremely useful to NSF and other agencies in evaluating ongoing research programs and in setting objectives for the future. WTEC studies also have been important in establishing new lines of communication and identifying opportunities for cooperation between U.S. researchers and their colleagues abroad, thus helping to accelerate the progress of science and technology within the international community. WTEC is an excellent example of cooperation and coordination among the many agencies of the U.S. Government that are involved in funding research and development: almost every WTEC study has been supported by a coalition of agencies with interests related to the particular subject.

As President Truman said over 50 years ago, our very survival depends upon continued leadership in science and technology. WTEC plays a key role in determining whether the United States is meeting that challenge, and in promoting that leadership.

Michael Reischman
Deputy Assistant Director for Engineering
National Science Foundation

¹ Remarks by the President on May 10, 1950, on the occasion of the signing of the law that created the National Science Foundation. *Public Papers of the Presidents* 120: 338.

² Investing in America's Future: Strategic Plan FY2006-2011, Arlington: NSF 06-48.

TABLE OF CONTENTS

Abstract.....	ii
Acknowledgements.....	iii
Foreword.....	v
Preface.....	xi
Executive Summary.....	xiii
 1. Introduction	
<i>Sharon C. Glotzer</i>	
Background and Scope.....	1
Methodology.....	3
Overview of the Report.....	8
References.....	12
 2. Life Sciences and Medicine	
<i>Linda Petzold</i>	
Introduction.....	15
Molecular Dynamics.....	16
Systems Biology.....	17
Biophysical Modeling.....	20
Summary of Key Findings.....	23
References.....	24
 3. Materials Simulation	
<i>Peter T. Cummings</i>	
Introduction.....	27
Current State of the Art in Materials Simulation.....	28
Materials Simulation Code Development.....	29
Materials Simulation Highlights.....	30
Summary of Key Findings.....	35
References.....	36
 4. Energy and Sustainability	
<i>Masanobu Shinozuka</i>	
Introduction.....	39
SBE&S Research Activities in North America.....	41
SBE&S Research Activities in Asia.....	43
Research Activities in Europe.....	48
Conclusions.....	51
References.....	52
 5. Next-Generation Architectures and Algorithms	
<i>George Em Karniadakis</i>	
Introduction.....	53
High-Performance Computing Around the World.....	55
New Programming Languages.....	59
The Scalability Bottleneck.....	60
Summary of Findings.....	62
Acknowledgements.....	64
References.....	64

6. Software Development*Martin Head-Gordon*

Introduction	65
Role of Universities, National Laboratories, and Government	66
Software Life Cycle – Managing Complexity.....	67
Supercomputing Software Versus Software for MidRange Computing.....	69
World Trends in Simulation Software Development	71
Comparative Aspects of Funding for Applications Software Development	74
Emerging Opportunities in Software Development	75
Summary of Findings.....	76
References	77

7. Engineering Simulations*Abhijit Deshmukh*

Introduction	79
Engineering Simulation Highlights.....	80
Summary of Key Findings	86
Comparison of U.S. and Worldwide Engineering Simulation Activities	87
References	88

8. Verification, Validation, and Uncertainty Quantification*George Em Karniadakis*

Introduction	89
Effects of Uncertainty Propagation.....	91
Methods	94
Industrial View	98
Summary of Findings.....	99
References	100

9. Multiscale Simulation*Peter T. Cummings*

Introduction	103
Current State of the Art in Multiscale Simulation.....	106
Multiscale Simulation Highlights	107
Summary of Key Findings	109
References	110

10. Big Data, Visualization, and Data-Driven Simulations*Sangtae Kim*

Introduction	113
Particle Physics Research: Petabytes Per Second.....	114
Big Data in Life Sciences Research.....	115
Big Data in Industry: Enterprise-Scale Knowledge Integration.....	116
Big Data – The Road Ahead for Training and Education.....	117
Summary of Key Findings	117
References	118

11. Education and Training*Celeste Sagui*

Introduction	119
Where the United States Stands.....	120

How Other Countries Compare.....	125
Case Study: The University of Stuttgart—A Success Story.....	134
Conclusions	136
References	137
A. Appendix A. Biographies of Panelists and Advisors.....	139
B. Appendix B. Site Reports—Asia.....	149
C. Appendix C. Site Reports—Europe	235
D. Appendix D. Survey Questionnaire.....	374
E. Appendix E. Bibliometric Analysis of Simulation Research	377
<i>Grant Lewison</i>	
F. Appendix F. Glossary	393

PREFACE

In 1946 in Philadelphia, ENIAC, the first large-scale electronic computer in the United States, was constructed and set to work toward calculating the path of artillery shells – a simulation using numerical calculations and based on the equations of physics. It was needed and it was promising, but it was also perceived as a tool too expensive and complicated to reach wide use. Thomas J. Watson, the chairman of IBM, has been widely quoted for remarking in 1943: "I think there is a world market for maybe five computers."

More than sixty years later, we find applications of predictive computer simulation all around us. When storms arise, two- and three-dimensional views are generated, their most likely paths are predicted, and even the uncertainty of the prediction is shown. Uncannily realistic images are conjured for movies, television, and video games. Consumer goods from cars to potato chips are designed, produced, and tested. Guiding wise decision-making, computer simulation is now used to assess stability of our energy networks, safety of industrial processes, collective behavior of people and economies, and global changes in climate.

What changed to bring us to this point? Almost everything:

- Computers were auditorium-sized, multiton facilities based on vacuum tubes, then transistors. They have become diverse in size and use: large supercomputers, multipurpose personal tools, and dedicated, embedded processing units, all based on thumbnail-sized microelectronic chips.
- Computer programs were hard-wired with no means of storage. They have become intricately crafted constructions of logic, created line-by-line or with drag-and-drop logic symbols, stored in memory chips and small rotating disks.
- Because reprogramming required rewiring, input was physical and output was exclusively by printed text. Input now comes from keyboards, mice, sensors, and voice recognition, and results appear on screens, from audio speakers and headsets, and even as three-dimensional objects that are sculpted or deposited by computer control.
- Computers initially were dedicated to a small group of users and were unconnected to each other. The invention of computer networks and then Berners-Lee's invention of the World Wide Web transformed the connectedness of computers and the people using them, as well as commerce, medicine, information access, technology, and intellectual disciplines.

Computer-based simulation has become a crucial part of the present infrastructure, central to applying these advances to the conduct of scientific research and engineering practice. To examine the breadth, depth, and implications of this change, a blue-ribbon panel on "Simulation-Based Engineering Science" was commissioned in 2005, led by Prof. J. Tinsley Oden of the University of Texas at Austin and sponsored by the Engineering Directorate of the National Science Foundation. In their May 2006 report,¹ they proposed SBES as being a new discipline in engineering science, in which modern computational methods, computational thinking, and devices and collateral technologies are combined to address problems far outside the scope of traditional numerical methods. They proposed that advances in SBES offer hope of resolving a variety of fundamental and complex problems in engineering science that affect discovery, health, energy, environment, security, and quality of life of all peoples. At the same time, their view of SBES was of being an interdisciplinary field, lying at the intersection of several disciplines and positioned where many of the important developments of the next century will be made. The basic building blocks of SBES are computational and applied mathematics, engineering science, and computer science, but the fruition of SBES requires the symbiotic development and the enrichment of these fields

¹ Oden, J.T., T. Belytschko, J. Fish, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and A. Yip. 2006. *Simulation-based engineering science: Revolutionizing engineering science through simulation.*, Arlington, VA: National Science Foundation. http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.

by a host of developing technologies in data acquisition and management, imaging, sensors, and visualization. Advances in SBES will require advances in all of these component disciplines.

The present, broader study on “Simulation-Based Engineering and Science” (SBE&S) was commissioned to investigate these points through an assessment of international activities in the field. The methodology incorporated bibliographic research, examination of relevant literature, and in particular, site visits and personal contacts with leaders in the field around the world. Through these means, the intent was to identify specific technical strengths and weaknesses; sites where the most advanced activities are occurring; what research challenges and apparent roadblocks exist; and what opportunities are just appearing.

One important change was that “SBES” was broadened to “SBE&S” to represent its multidisciplinary nature more effectively. By considering SBE&S as the overarching topic, the study also explored impacts that computer simulation is having on pure, curiosity-driven science, on problem-driven science, and on engineering development, design, prediction, decision-making, and manufacturing.

These goals resonated within NSF and with other agencies. NSF’s Engineering Directorate was the lead sponsor of this study, joined by programs from NSF’s Mathematics and Physical Sciences Directorate, the Department of Defense, the Department of Energy, the National Aeronautics and Space Administration, the National Institute for Biomedical Imaging and Bioengineering, the National Library of Medicine, and the National Institute of Standards and Technology. We selected the World Technology Evaluation Center, Inc. (WTEC) to arrange bibliometrics, trip logistics, and diverse aspects of aiding the study and its report preparation.

We are grateful for all these commitments and for the commitment and diligence of the study participants. They include the study panel, the advisory panel, and the gracious hosts of our site visits, who all contributed thoughtful, informative analysis and discussion. Our goal is now to use these insights to enhance and guide SBE&S activities throughout the nation.

P. R. Westmoreland, K. P. Chong, and C. V. Cooper
National Science Foundation
January 2009

EXECUTIVE SUMMARY

BACKGROUND

Simulation-Based Engineering and Science (SBE&S) involves the use of computer modeling and simulation to solve mathematical formulations of physical models of engineered and natural systems. Today we are at a “tipping point” in computer simulation for engineering and science. Computer simulation is more pervasive today – and having more impact – than at any other time in human history. No field of science or engineering exists that has not been advanced by, and in some cases transformed by, computer simulation. Simulation has today reached a level of predictive capability that it now firmly complements the traditional pillars of theory and experimentation/observation. Many critical technologies are on the horizon that cannot be understood, developed, or utilized without simulation. At the same time, computers are now affordable and accessible to researchers in every country around the world. The near-zero entry-level cost to perform a computer simulation means that anyone can practice SBE&S, and from anywhere. Indeed, the world of computer simulation is becoming flatter every day. At the same time, U.S. and Japanese companies are building the next generation of computer architectures, with the promise of thousand-fold or more increases of computer power coming in the next half-decade. These new massively manycore computer chip architectures will allow unprecedented accuracy and resolution, as well as the ability to solve the highly complex problems that face society today. Problems ranging from finding alternative energy sources to global warming to sustainable infrastructures, to curing disease and personalizing medicine, are *big* problems. They are complex and messy, and their solution requires a partnership among experiment, theory, and simulation working across all of the disciplines of science and engineering. There is abundant evidence and numerous reports documenting that our nation is at risk of losing its competitive edge. Our continued capability as a nation to lead in simulation-based discovery and innovation is key to our ability to compete in the 21st century.

To provide program managers in U.S. research agencies and decision makers with a better understanding of the status, trends and levels of activity in SBE&S research abroad, these agencies sponsored the WTEC International Assessment of R&D in Simulation-Based Engineering and Science. Sponsors included

- National Science Foundation
- Department of Energy
- National Institutes of Health (NIBIB, National Library of Medicine)
- NASA
- National Institute of Standards and Technology
- Department of Defense

The study was designed to gather information on the worldwide status and trends in SBE&S research and to disseminate this information to government decision makers and the research community. A panel of experts reviewed and assessed the state of the art as well as levels of activity overseas in the broad thematic areas of SBE&S in life sciences and medicine, materials, in energy and sustainability; and in the crosscutting issues of next generation hardware and algorithms; software development; engineering simulations; validation, verification, and uncertainty quantification; multiscale modeling and simulation; and education. To provide a basis for comparison, the panel hosted a U.S. Baseline workshop at the National Science Foundation (NSF) on November 1-2, 2007. Following the workshop, a WTEC panel of U.S. experts visited 59 sites in Europe and Asia involved in SBE&S research. Information gathered at the site visits, along with individual conversations with experts, access to reports and research publications, and a bibliometric analysis, provided the basis for the assessment. An advisory panel provided additional feedback to the study panel.

In this executive summary, we first present the panel’s overarching findings concerning major trends in SBE&S research and development, threats to U.S. leadership in SBE&S, and opportunities for gaining or reinforcing the U.S. lead through strategic investments. We then highlight the most notable findings in each of the three

thematic areas and each of the crosscutting areas of the study. Throughout this summary, key findings are shown in bold italics, with additional text provided to support and amplify these findings.

MAJOR TRENDS IN SBE&S RESEARCH AND DEVELOPMENT

Within the thematic areas of the study, *the panel observed that SBE&S is changing the way disease is treated and surgery is performed, the way patients are rehabilitated, and the way we understand the brain; the way materials and components are designed, developed, and used in all industrial sectors; and aiding in the recovery of untapped oil, the discovery and utilization of new energy sources, and the way we design sustainable infrastructures.* In all of these areas, there were ample examples of critical breakthroughs that will be possible in the next decade through application of SBE&S, and in particular through the following four major trends in SBE&S:

Finding 1: Data-intensive applications, including integration of (real-time) experimental and observational data with modeling and simulation to expedite discovery and engineering solutions, were evident in many countries, particularly Switzerland and Japan.

Modeling and simulation of very large data sets characterized some of the most cutting-edge examples of SBE&S in life sciences and medicine, especially with regards to systems biology. Big data already plays a prominent role in elementary particle physics, climate modeling, genomics, and earthquake prediction. The trend will become even more prominent globally with petascale computing.

Finding 2: Achieving millisecond timescales with molecular resolution for proteins and other complex matter is now within reach using graphics processors, multicore CPUs, and new algorithms.

Because this is the time scale on which many fundamental biomolecular processes occur, such as protein folding, this represents an enormous breakthrough and many opportunities for furthering our understanding of processes important for biomedical and life science applications. These same speed-ups will further revolutionize materials simulation and prediction, allowing the level of resolution needed for materials design as well as modeling chemical and physical processes such as those relevant for biofuels, batteries, and solar cells.

Finding 3: The panel noted a new and robust trend towards increasing the fidelity of engineering simulations through inclusion of physics and chemistry.

Although the panel did not see much evidence of sophisticated approaches towards validation and verification, uncertainty quantification, or risk assessment in many of the simulation efforts, which would complement the use of physics-based models in engineering simulations, there was widespread agreement that this represents a critical frontier area for development in which the United States currently has only a slight lead.

Finding 4: The panel sensed excitement about the opportunities that petascale speeds and data capabilities would afford.

Such capabilities would enable not just faster time to solution, but also the ability to tackle important problems orders-of-magnitude more complex than those that can be investigated today with the needed level of predictiveness. Already, inexpensive graphics processors today provide up to thousand-fold increases in speed on hundreds of applications ranging from fluid dynamics to medical imaging over simulations being performed just a year or two ago.

THREATS TO U.S. LEADERSHIP IN SBE&S

The panel identified three overarching issues that specifically threaten U.S. leadership in SBE&S.

Finding 1: The world of computing is flat, and anyone can do it. What will distinguish us from the rest of the world is our ability to do it better and to exploit new architectures we develop before those architectures become ubiquitous.

First and foremost, many countries now have and use HPC – *the world is flat!* A quick look at the most recent Top 500 list of supercomputers shows that Japan, France, and Germany in particular have world-class resources. They also have world class faculty and students and are committed to HPC and SBE&S for the long haul. In terms of hardware, Japan has an industry-university-government roadmap out to 2025 (exascale), and Germany is investing nearly US\$1 billion in a new push towards next-generation hardware in partnership with the European Union. It is widely recognized that it is relatively inexpensive to start up a new SBE&S effort, and this is of particular importance in rapidly growing economies (e.g. India, China, Finland). Furthermore, already there are more than 100 million NVIDIA graphics processing units with CUDA compilers distributed worldwide in desktops and laptops, with potential code speedups of up to a thousand-fold in virtually every sector to whomever rewrites their codes to take advantage of these new general programmable GPUs.

Aggressive, well-funded initiatives in the European Union may undermine U.S. leadership in the development of computer architectures and applied algorithms. Examples of these initiatives include the Partnership for Advanced Computing in Europe (PRACE) which is a coalition of 15 countries and led by Germany and France, and based on the ESFRI Roadmap (ESFRI 2006); TALOS – Industry-government alliance to accelerate HPC solutions for large-scale computing systems in Europe; and DEISA – Consortium of 11 leading European Union national supercomputing centers to form the equivalent of the U.S. TeraGrid. There is also some flux, with some alliances dissolving and new consortia being formed. Already, the European Union leads the United States in theoretical algorithm development, and has for some time; these new initiatives may further widen that lead and create new imbalances.

Finding 2: Inadequate education and training of the next generation of computational scientists threatens global as well as U.S. growth of SBE&S. This is particularly urgent for the United States; unless we prepare researchers to develop and use the next generation of algorithms and computer architectures, we will not be able to exploit their game-changing capabilities.

There was grave concern, universally voiced at every site in every country including the United States, that today's computational science and engineering students are ill-prepared to create and innovate the next generation of codes and algorithms needed to leverage these new architectures. Much of this appears to arise from insufficient exposure to computational science and engineering and underlying core subjects beginning in high school and undergraduate and continuing through graduate education and beyond. Increased topical specialization beginning with graduate school was noted as a serious barrier to a deep foundation in simulation and supporting subjects. There is a clear gap in the preparation students receive in high performance computing and what they need to know to develop codes for massively parallel computers, let alone petascale and massively multicore architectures. Worldwide, students are not learning to “program for performance.” Nearly universally, the panel found concern that students use codes primarily as black boxes, with only a very small fraction of students learning proper algorithm and software development, in particular with an eye towards open-source or community code development. Students receive no real training in software engineering for sustainable codes, and little training if any in uncertainty quantification, validation and verification, risk assessment or decision making, which is critical for multiscale simulations that bridge the gap from atoms to enterprise. Moreover, in many countries, the very best computational science and engineering students leave research for finance companies (this was, until recently, a particularly huge problem in Switzerland). Despite the excitement about manycore and petascale computing, the panel noted universal concern that the community is not prepared to take advantage of these hardware breakthroughs because the current generation of algorithms and software must be rethought in the context of radically new architectures that few know how to program.

Finding 3: A persistent pattern of subcritical funding overall for SBE&S threatens U.S. leadership and continued needed advances amidst a recent surge of strategic investments in SBE&S abroad that reflects recognition by those countries of the role of simulations in advancing national competitiveness and its effectiveness as a mechanism for economic stimulus.

It is difficult to draw a precise boundary around funding specific to SBE&S whether globally or within the United States due to its interdisciplinary nature. Nevertheless we can consider the total funding for the entire information technology/computer science/computational science/cyberinfrastructure landscape (NITRD in the case of the United States) and identify within this large space specific, high-profile programs in which simulations play a central, enabling role (e.g. SCIDAC, ITR, Cyber-enabled Discovery, etc.). This exercise allows an “apples to apples” comparison with activities around the world, such as the notable growth spurts in SBE&S funding in China and Germany.

As noted in Oden (2006), the pattern of subcritical funding for SBE&S is masked in the overall budget across the landscape for information and computational technology (\$3.5 billion NITRD FY2009 request) (OMB 2008). The situation is particularly dire in light of the role of SBE&S in achievement of national priorities in energy, health, and economic development and alarming when viewed in comparison with the panel’s findings of investments targeting SBE&S across the globe. The panel found that the level of SBE&S investments abroad reflect the recognition of the strategic role of simulations in advancing national competitiveness as well as a tacit recognition of their effectiveness as an economic stimulus mechanism. Consequently, funding for SBE&S is surging in both peer OECD partners (e.g., Germany and Japan) as well as in rapidly developing economies (e.g., China). Evidence for this is exemplified by the following observations:

- In Germany, specific and focused investments in SBE&S are patterned along the recommendations in the 2006 NSF blue ribbon panel report on SBES (Oden 2006) as part of the 20+% year-on-year increase in funding for research. As a consequence of this new funding, Germany already exhibits many of the innovative organizational and collaborative structures deemed to be the most promising for advancing SBE&S in the context of energy, medicine, and materials research. The panel observed extensive restructuring of universities to enable more interdisciplinary work and strong university-industry partnerships.
 - One example of such a partnership is between the Fraunhofer IWM and Karlsruhe University. The IWM has an annual budget of \$16 million/yr, with 44% coming from industry, and 50% supporting research in SBE&S.
- Funding for SBE&S in China is robust and an integral part of the national strategy for upgrading the R&D infrastructure of the top and mid- tier universities. Most recently, it has become apparent that China’s SBE&S funding within the context of a \$billion-plus increase in funding for universities, is an integral element of their economic stimulation strategy.
 - Although China is not yet a strong U.S. competitor in SBE&S, their “footprint” is changing rapidly. China contributes 13% of the world’s output in simulation papers, second to the United States at 26% (from 2003-2005) and third to the 12 leading Western European countries combined (32%) and growing fast (although they publish in less than first-tier journals and the papers are generally cited less often). China also had the highest relative commitment to simulation research over the decade 1996-2005, with 50% more papers than expected on the basis of its overall output (the US published about 10% fewer papers than expected on this basis). The panel found nonuniform quality overall, but saw many high quality examples on par with the best examples anywhere. There was a palpable sense of a strategic change in the direction of funding and research towards innovation, combined with a clearly articulated *recognition by industry and the Chinese government that innovation requires simulation*. China’s S&T budget has doubled every 5 years since 1990, with 70% of the funding going to the top 100 universities, which produce 80% of all PhDs, support 70% all graduates, 50% of all international students, and 30% of all undergraduates. There is a specific recognition by the Vice Minister of Education of the need to train a new generation of “computationally-savvy” students with solid training in analytical thinking, and substantial new funds to do this in programs under his control. One example is the 211 Fund, which is the equivalent of roughly US\$1 billion/year; all new projects must have an integrated simulation component.
- For Japan, in addition to the expected level of funding for the successor to the Earth Simulator (Life Simulator), the “accomplishment-based funding” for systems biology including SBE&S activities was a noteworthy finding of the SBE&S funding landscape in Japan. Japan leads the US in bridging physical systems modeling to social-scale engineered systems, in particular by the introduction of large-scale agent-based simulations.

- The United Kingdom has had a decade-plus experience with community-based code development for SBE&S in the form of the highly successful Collaborative Computational Projects (CCP). Although the program is now limited in scope to maintenance of codes, its multiple contributions to the standard tool set of the global SBE&S infrastructure far exceed the scale of the investment, suggesting that this mechanism merits further consideration both by the United States and the global community.
- Although outside the scope of the panel's trips to Europe, China, and Japan, recent investments in R&D and in SBE&S by both Singapore and Saudi Arabia are particularly noteworthy. With substantial involvement by U.S. faculty and enormous financial investment, Singapore has reinvented their R&D infrastructure and is carrying out world-class research in many strategic areas, including new programs in computational science and engineering. Stunning investments by Saudi Arabia in the King Abdullah University of Science and Technology (KAUST) (with \$20B endowment) has generated substantial interest around the world, with some of the best faculty from the United States and Europe leading many of their major activities. A multimillion dollar IBM Blue Gene P supported by IBM provides access to supercomputing resources for KAUST researchers and students. In particular, there is substantial concern that KAUST, with its state-of-the-art facilities and generous graduate scholarships, will be able to attract away the best and brightest Asian students, who have for many years come to the United States for graduate study and remained to become leaders in high-tech industries. It is particularly interesting to note that Applied Mathematics and Computational Science, the core of SBE&S, is one of KAUST's very first graduate programs.
- The panel found healthy levels of SBE&S funding for internal company projects in our visits to the industrial R&D centers, underscoring industry's recognition of the cost-effectiveness and timeliness of SBE&S research. The mismatch *vis a vis* the public-sector's investment in SBE&S hinders workforce development. Notably, the panel saw many examples of companies (including U.S. auto and chemical companies) working with E.U. groups rather than U.S. groups for "better IP agreements."
- Because SBE&S is often viewed within the United States more as an enabling technology for other disciplines, rather than a discipline in its own right, investment in and support of SBE&S is often not prioritized as it should be at all levels of the R&D enterprise. As demonstrated in the 2005 PITAC report (Benioff 2005), the panel found that investment in computational science in the United States and the preparation of the next generation of computational researchers remains insufficient to fully leverage the power of computation for solving the biggest problems that face us as a nation going forward.

OPPORTUNITIES FOR THE UNITED STATES TO GAIN OR REINFORCE LEAD IN SBE&S THROUGH STRATEGIC RESEARCH AND INVESTMENTS

The panel identified several specific ways in which the United States could gain or reinforce its lead in SBE&S, and these are described in detail in the report. Three of the most important findings in terms of opportunities are:

Finding 1: There are clear and urgent opportunities for industry-driven partnerships with universities and national laboratories to hardwire scientific discovery and engineering innovation through SBE&S. This would lead to new and better products, as well as development savings both financially and in terms of time. This idea is exemplified in a new National Academies' report on Integrated Computational Materials Engineering (ICME) (NRC 2008), which found a reduction in development time from 10-20 yrs to 2-3 yrs with a concomitant return on investment of 3:1 to 9:1.

Finding 2: There is a clear and urgent need for new mechanisms for supporting R&D in SBE&S. Particular attention must be paid to the way that we support and reward the long-term development of algorithms, middleware, software, code maintenance and interoperability.

For example, the panel found that *community code development projects are much stronger within the European Union than the United States, with national strategies and long-term support.* Many times the panel was told that the United States is an "unreliable partner" in these efforts due to our inability to commit for longer than typically three years at a time. Both this perception and the reality means that the United States has little influence over the direction of these community codes and at the same time is not developing large codes of its own. As one alarming example of this, six of the top seven electronic structure codes for materials physics today come from the European Union, despite the fact that the original science codes are from the United States (and for which the

United States won the Nobel Prize), and the scientific advances underlying the codes are distributed evenly between the United States and European Union. Indeed, the U.S. lead in scientific software applications has decreased steadily since 1998. One consequence of this lack of involvement and influence is that, in some cases, codes may not be available to U.S. researchers. As one critical example, a major new code for electronic transport from Spain/United Kingdom is effectively off limits to U.S. defense labs.

The panel found universal agreement (although the disparity is greatest in the United States) that *investment in algorithm, middleware, and software development lags behind investment in hardware, preventing us from fully exploiting and leveraging new and even current architectures. This disparity threatens critical growth in SBE&S capabilities needed to solve important worldwide problems as well as many problems of particular importance to the U.S. economy and national security.*

A related issue is evidenced lack of support and reward for code development and maintenance. It is widely recognized within the SBE&S community that the timescale to develop large complex code often exceeds the lifetime of a particular generation of hardware. Moreover, although great scientific advances achieved through the use of a large complex code is highly lauded, the development of the code itself often goes unrewarded. The United Kingdom, which once led in supporting long-term scientific software development efforts, does not provide the support it once did. The panel found worldwide recognition that progress in SBE&S requires crossing disciplinary boundaries, which is difficult in many countries due to outdated university structures. The United States is perceived to be a leader in pulling together interdisciplinary teams for large, complex problems, with the caveats noted above.

Finding 3: There is a clear and urgent need for a new, modern approach to educating and training the next generation of researchers in high performance computing specifically, and in modeling and simulation generally, for scientific discovery and engineering innovation.

Particular attention must be paid to teaching fundamentals, tools, programming for performance, verification and validation, uncertainty quantification, risk analysis and decision making, and programming the next generation of massively multicore architectures. At the same time, students must gain deep knowledge of their core discipline.

KEY STUDY FINDINGS

Here we briefly summarize the panel's main findings in each of the three primary thematic domains and the crosscutting areas. Each of these findings as well as additional observations are discussed in detail in the succeeding chapters of this report. Many of the findings underpinning individual crosscutting areas echo those of the thematic domains, giving rise to occasional but telling redundancies that serve to amplify the importance of many crosscutting technologies and issues to the application of simulation to particular science and engineering domains.

Thematic Area: Life sciences and medicine

There is an unprecedented opportunity in the next decade to make game-changing advances in the life sciences and medicine through simulation.

Finding 1: Predictive biosimulation is here.

New techniques for modeling variability in patient physiology, as well as new software platforms, lower-cost computational resources, and an explosion of available data, have fueled recent advances in “predictive biosimulation”—the *dynamic simulation of biological systems*. This is a major advance in the use of SBE&S in the medical field, moving beyond simple chemistry and molecular structure. Predictive biosimulation has been used to support the development of new therapeutic drugs, meeting the pharmaceutical industry's need for new approaches to deal with the complexity of disease and the increased risk and expense of R&D.

- Several factors are likely to accelerate adoption of biosimulation technology in the pharmaceutical industry. The FDA, through its Critical Path Initiative, has recognized the importance of modeling and simulation and is committing more resources toward making biosimulation part of the evaluation process for new therapies. Academic groups and companies are increasingly specializing in the field and building active collaborations

with industry. Pharmaceutical companies are creating and expanding internal systems biology groups that incorporate biological modeling.

- Modeling of whole organs and complete physiological systems is becoming possible.
 - Modeling of complete cardiovascular systems provides information and insight both for medical research and for the evaluation of surgical options.
 - Modeling provides information both for cancer research and for laser surgery.
 - Neuromuscular biomechanics seeks fundamental understanding of the mechanisms involved in the production of movement and is motivated by opportunities to improve treatments for cerebral palsy.
 - Biomechanical simulation is used in surgical planning and design of physical therapies.
 - Rapidly increasing amounts of data, new algorithms, and fast computers make modeling the brain in full experimental detail a realistic goal for the foreseeable future.

Finding 2: Pan-SBE&S synergy argues for a focused investment of SBE&S as a discipline.

The exciting possibilities of predictive biosimulation depend on a collective set of advances in SBE&S capabilities, a finding in common with the other two themes. This dependence buttresses the notion of a focused advancement of SBE&S capabilities as a discipline. For predictive biosimulation to realize its potential impacts in life sciences and medicine will require:

- Large, focused multidisciplinary teams. Much of the research in this area requires iteration between modeling and experiment.
- Integrated, community-wide software infrastructure for dealing with massive amounts of data, and addressing issues of data provenance, heterogeneous data, analysis of data, and network inference from data.
- High-performance algorithms for multiscale simulation on a very large range of scales and complexities.
- High-performance computing and scalable algorithms for multicore architectures. In particular, petascale computing can finally enable molecular dynamics simulation of macromolecules on the biologically interesting millisecond timescale.
- Techniques for sensitivity and robustness analysis, uncertainty analysis, and model (in)validation.
- Visualization techniques, applicable to massive amounts of data, that can illustrate and uncover relationships among data.
- Appropriately trained students who are conversant in both the life sciences and SBE&S and can work in multidisciplinary teams.

Finding 3: Worldwide SBE&S capabilities in life sciences and medicine are threatened by lack of sustained investment and loss of human resources.

World-class research exists in SBE&S in molecular dynamics, systems biology, and biophysical modeling throughout the United States, Europe, and Japan, and, in some areas, China. The United States has historically played a leadership role in a number of research areas that have driven the development of new algorithms, architectures, and applications for molecular dynamics (MD) simulations. However, this role is now endangered by a lack of sustained investment at all federal agencies, and by the loss of PhD students and faculty to other disciplines and to the private sector.

- The quality of leading researchers within this field in the United States is comparable to that of leading researchers in Europe and Japan.
- Infrastructure (access to computing resources and software professionals) and funding models to support ambitious, visionary, long-term research projects are much better in Europe and Japan than in the United States.
- Funding models and infrastructure to support multi-investigator collaborations between academia and industry are much more developed in Europe than in the United States.
- Support for the development of community software for biosimulation is much stronger in Europe and Japan than in the United States in recent years, following the decline in DARPA funding in this area.

Thematic Area: Materials

SBE&S has long played a critical role in materials simulation. As experimental techniques for synthesizing and characterizing new materials become increasingly sophisticated, the demand for theoretical understanding and design guidelines is driving a persistently increasing need for materials simulation.

Finding 1: Computational materials science and engineering is changing how new materials are discovered, developed, and applied, from the macroscale to the nanoscale.

- This is particularly so in nanoscience, where it is now possible to synthesize a practically infinite variety of nanostructured materials and combine them into new devices or systems with complex nanoscale interfaces. The experimental tools of nanoscience—such as scanning probes, neutron scattering, and various electron microscopies—all require modeling to understand what is being measured. Hence, the demand for high-fidelity materials simulation is escalating rapidly, and the tools for verification of such simulations are becoming increasingly available. The nanoscale is particularly ripe for SBE&S since both experiment and simulation can access the same length and time scales.

Finding 2: World-class research in all areas of materials simulation is to be found in the United States, Europe, and Asia; the United States leads in some, but not all, of the most strategic of these.

- Algorithm innovation takes place primarily in the United States and Europe, some in Japan, with little activity to date in this area in China.
- There is a rapid ramping-up of materials simulation activities in some countries, particularly China and Germany.
 - Some of these efforts in Germany involve close partnerships between industry, government, and academia (e.g. through the Fraunhofer Institutes). New efforts are sprouting against a backdrop of an extraordinary revival of science in Germany, tied to recent increases in the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, the equivalent of the U.S. National Science Foundation) budget. In the most recent year, DFG funding jumped over 20%, following on three years of increases in the 5–10% range.
 - Historically China has been mostly concerned with applications; however, as part of the Chinese government's current emphasis on supporting creativity and innovation in both the arts and sciences, one can anticipate increased activity in this area in China.

Finding 3: The United States' ability to innovate and develop the most advanced materials simulation codes and tools in strategic areas has been eroding and continues to erode.

- The United States is at an increasingly strategic disadvantage with respect to crucial, foundational codes, as it has become increasingly reliant on codes developed by foreign research groups. Political considerations can make those codes effectively unavailable to U.S. researchers, particularly those in Department of Defense laboratories.
- Many large codes, both open source and non-open source, require collaboration among large groups of domain scientists, applied mathematicians, and computational scientists. However, there is much greater collaboration among groups in materials code development in Europe compared to the United States. There appear to be several reasons for this:
 - The U.S. tenure process and academic rewards systems suppress collaboration.
 - Funding, promotion, and awards favor high-impact science (publications in *Nature* and *Science*, for example), while the development of simulation tools is not considered to be high-impact science. Yet, these tools (which can take many years to develop) are often the key factor in enabling the high-impact science.
- The utility of materials simulation codes for practical application is threatened by the lack of standards for interoperability of codes. The CAPE-OPEN effort undertaken in the chemical process simulation (computer-aided process engineering, or CAPE) field to develop interoperability standards in that field (<http://www.colan.org>) is an example of an approach that could benefit the materials community.
- The current practice of short-term (less than five years, typically three-year cycles) funding is insufficient for the United States to develop key materials simulation codes that will run on the largest U.S. computing

platforms and on future many-core processors. Such codes will be needed to explore the next generation of materials and to interpret experiments on the next generation of scientific instruments. Longer term funding (minimum five years) of multidisciplinary teams is needed.

- Training is a critical issue worldwide, as undergraduate students in the physical and chemical sciences and engineering domains that feed into the computational materials discipline are increasingly illiterate in computer programming. The situation is perhaps most acute in the United States.

Thematic Area: Energy and Sustainability

The search for alternative energy sources and cleaner fuels, and for long-term sustainable infrastructures, is one of the greatest and most serious challenges facing the planet over the next decade. The role for SBE&S is enormous and urgent. Broad and extensive simulation-based activities in energy and sustainability exist in all regions. However, the scarcity of appropriately trained students in the U.S. (relative to Europe and Asia) is seen as a substantial bottleneck to progress.

Finding 1: In the area of transportation fuels, SBE&S is critical to stretch the supply and find other sources.

- Recognition that future energy sources for transportation cannot be substituted for traditional sources in the short term is motivating the simulation of oil reservoirs to optimize recovery, refinery processes to optimize output and efficiency, and the design of more fuel-efficient engines, including diesel and hybrid engines.
- The use of SBE&S in oil production and optimizing the fossil-fuel supply chain has traditionally been led by the United States. The panel found significant activity in Europe, including new funding models and leveraging strengths in development of community codes for applications to this important problem.

Finding 2: In the discovery and innovation of alternative energy sources – including biofuels, batteries, solar, wind, nuclear – SBE&S is critical for the discovery and design of new materials and processes.

- Nuclear energy is a major source for electricity in the United States and abroad. Rapid progress and increased investment in SBE&S-driven materials design for nuclear waste containment are urgently needed. France leads the world in nuclear energy and related research (e.g., containment).
- Tax incentives for green sources of energy, despite lower energy density, are driving extensive R&D efforts for more efficient wind energy and tidal energy systems, especially in Europe.
- Alternative fuels such as biodiesel are constituted from a complex mixture of plant-derived organic chemicals. Optimal engine combustion designs require the thermophysical properties of such complex mixtures, and in view of the myriad possible compositions faced by the experimentalist, simulations provide the only viable option.

Finding 3: Petascale computing will allow unprecedented breakthroughs in sustainability and the simulation of ultra-large-scale sustainable systems, from ecosystems to power grids to whole societies.

- Simulations to evaluate the performance of large-scale and spatially distributed systems such as power transmission systems that are subject to highly uncertain natural hazards (earthquake, hurricanes, flood) hold important promise for the design of sustainable infrastructures and for disaster prediction and prevention.
- Japan is pushing forward the modeling of human behavior through massive agent-based simulations enabled by their Life Simulator, the successor to Japan's Earth simulator.
- It is well recognized that even with technological breakthroughs in viable alternative sources of energy, the lack of ease of substitution due to a lack of critical infrastructure (e.g. diesel stations for diesel engines) and public acceptance is a major challenge. The modeling and simulation of integration of alternative sources into the power grid is a growing area, revealing technological issues as well as how the public adopts the alternative sources, behaves, etc.
- Simulation-based engineering is being used to create "smart" infrastructures.
- The United States leads in modeling of large-scale infrastructure systems, but the gap is closing, particularly in Asia.

Crosscutting Issues: Next Generation Architectures and Algorithms

Finding 1: The many orders-of-magnitude in speedup required to make significant progress in many disciplines will come from a combination of synergistic advances in hardware, algorithms, and software, and thus investment and progress in one will not pay off without concomitant investments in the other two.

In the last decade, large investments have been made for developing and installing new computer architectures that are now breaking the petaflop barrier, and the exascale era will be here within ten years. However, recent investments in mathematical algorithms and high-quality mathematical software are lagging behind. Specifically, on the algorithmic front, advances in linear solvers, high-order spatio-temporal discretization, domain decomposition, and adaptivity are required to match the new manycore and special-purpose computer architectures and increase the “effective” performance of these expensive supercomputers.

Finding 2: The United States leads both in computer architectures (multicores, special-purpose processors, interconnects) and applied algorithms (e.g., ScaLAPACK, PETSC), but aggressive new initiatives around the world may undermine this position. Already, the European Union leads the United States in theoretical algorithm development, and has for some time.

In particular, the activities of the Department of Energy laboratories have helped to maintain this historic U.S. lead in both hardware and software. However, the picture is not as clear on the development of *theoretical* algorithms; several research groups across Europe (e.g., in Germany, France, and Switzerland) are capitalizing on new “priority programs” and are leading the way in developing fundamentally new efforts for problems on high dimensions, on $O(N)$ algorithms, and specifically on fast linear solvers related to hierarchical matrices, similar to the pioneering work in Europe on multigrid in the 1980s. Initiatives such as the “Blue Brain” at EPFL and “Deep Computing” at IBM-Zurich (see site reports in Appendix C) have set ambitious goals far beyond what the U.S. computational science community can target at present. It is clear that barriers to progress in developing next-generation architectures and algorithms are increasingly on the theoretical and software side and that the return on investment is more favorable on the theoretical and software side. This is manifested by the fact that the “half-time” of hardware is measured in years, whereas the “half-time” of software is measured in decades and theoretical algorithms transcend time.

Finding 3: Although the United States currently leads in the development of next-generation supercomputers, both Japan and Germany have been, and remain, committed over the long-term to building or acquiring leadership-class machines and capability, and China is now beginning to develop world-class supercomputing infrastructure.

Crosscutting Issues: Scientific and Engineering Simulation Software Development

Modern software for simulation-based engineering and science is sophisticated, tightly coupled to research in simulation models and algorithms, and frequently runs to millions of lines of source code. As a result, the lifespan of a successful program is usually measured in decades, and far surpasses the lifetime of a typical funding initiative or a typical generation of computer hardware.

Finding 1: Around the world, SBE&S relies on leading edge (supercomputer class) software used for the most challenging HPC applications, mid-range and desktop computing used by most scientists and engineers, and everything in between.

Simulation software is too rich and too diverse to suit a single paradigm. Choices as disparate as software targeting cost-effective computer clusters versus leading edge supercomputers, or public domain software versus commercial software, choices of development tools etc, are mapped across the vast matrix of applications disciplines. The best outcomes seem to arise from encouraging viable alternatives to competitively co-exist, because progress driven by innovation occurs in a bottom-up fashion. Thus strategic investments should balance the value of supporting the leading edge (supercomputer class) applications against the trailing vortex (mid-range computing used by most engineers and scientists).

Finding 2: Software development leadership in many SBE&S disciplines remains largely in U.S. hands, but in an increasing number of areas it has passed to foreign rivals, with Europe being particularly resurgent in software for mid-range computing, and Japan particularly strong on high-end supercomputer applications. In some cases, this leaves the United States without access to critical scientific software.

In many domains of SBE&S, the US relies on codes developed outside the US, despite having led previously in the theoretical research that formed the intellectual basis for such codes. A major example of that is in first principles codes for materials, the theoretical framework for which the United States won the 1998 Nobel Prize in Chemistry. Just one decade later, the WTEC panel learned of American defense researchers denied access to foreign-developed electronic transport software as a matter of principle, with no equivalent U.S. codes available as a substitute.

Finding 3: The greatest threats to U.S. leadership in SBE&S come from the lack of reward, recognition and support concomitant with the long development times and modest numbers of publications that go hand-in-hand with software development; the steady erosion of support for first-rate, excellence-based single-investigator or small-group research in the United States; and the inadequate training of today's computational science and engineering students – the would-be scientific software developers of tomorrow.

Fundamentally, the health of simulation software development is inseparable from the health of the applications discipline it is associated with. Therefore the principal threat to U.S. leadership comes from the steady erosion of sustained support for first-rate, excellence-based single-investigator or small-group research in the United States. A secondary effect that is specific to software development is the distorting effect that long development times and modest numbers of publications have on grant success rates. Within applications disciplines, it is important to recognize and reward the value of software development appropriately, in balance with the direct exploration of phenomena. Software development benefits industry and society through providing useful tools too expensive, long-term and risky to be done as industrial R&D, trains future scientists and engineers to be builders and not just consumers of tools, and helps to advance the simulation state of the art. Future investments in software development at the applications level are best accomplished as part of re-invigorating American physical and biological science generally. More specific investments in simulation software can be justified on the basis of seeking targeted leadership in areas of particular technological significance, as discussed in more detail in chapters on opportunities in new energy, new materials, and the life sciences.

Crosscutting Issues: Engineering Simulation

Simulation and modeling are integral to every engineering activity. While the use of simulation and modeling has been widespread in engineering, the WTEC study found several major hurdles still present in the effective use of these tools.

Finding 1: Software and data interoperability, visualization, and algorithms that outlast hardware obstruct more effective use of engineering simulation.

Interoperability of software and data is a major hurdle resulting in limited use of simulation software by non-simulation experts. Commercial vendors are defining de facto standards, and there is little effort beyond syntactic compatibility. Codes are too complicated to permit any user customization, and effective workflow methods need to be developed to aid in developing simulations for complex systems.

- In most engineering applications, algorithms, software and data/visualization are primary bottlenecks. Computational resources (flops and bytes) were not limiting factors at most sites. Lifecycle of algorithms is in the 10-20 years range, whereas hardware lifecycle is in the 2-3 years range. Visualization of simulation outputs remains a challenge and HPC and high-bandwidth networks have exacerbated the problem.

Finding 2: Links between physical and system level simulations remain weak. There is little evidence of atom-to-enterprise models that are coupled tightly with process and device models and thus an absence of multi-scale SBE&S to inform strategic decision-making directions.

Experimental validation of models remains difficult and costly, and uncertainty quantification is not being addressed adequately in many of the applications. Models are often constructed with insufficient data or physical measurements, leading to large uncertainty in the input parameters. The economics of parameter estimation and model refinement are rarely considered, and most engineering analyses are conducted under deterministic settings. Current modeling and simulation methods work well for existing products and are mostly used to understand/explain experimental observations. However, they are not ideally suited for developing new products that are not derivatives of current ones.

Finding 3: Although U.S. academia and industry are, on the whole, ahead (marginally) of their European and Asian counterparts in the use of engineering simulation, pockets of excellence exist in Europe and Asia that are more advanced than U.S. groups, and Europe is leading in training the next generation of engineering simulation experts.

- Examples of pockets of excellence in engineering simulation include Toyota, Airbus, and the University of Stuttgart.
- European and Asian researchers rely on the United States to develop the common middleware tools, whereas their focus is on application-specific software.
- The transition from physical systems modeling to social-scale engineered systems in the United States lags behind the Japanese, who are, e.g., modeling the behavioral patterns of six billion people using the Life Simulator.
- European universities are leading the world in developing curricula to train the next generation of engineering simulation experts, although the needed combination of domain, modeling, mathematical, computational and decision-making skills is still rare.

Crosscutting Issues: Validation, Verification, and Uncertainty Quantification

The NSF SBES report (Oden 2006) stresses the need for new developments in V&V and UQ in order to increase the reliability and utility of the simulation methods in the future at a profound level. A report on European computational science (ESF 2007) concludes that “without validation, computational data are not credible, and hence, are useless.” The aforementioned National Research Council report on integrated computational materials engineering (ICME) (NRC 2008) states that, “Sensitivity studies, understanding of real world uncertainties, and experimental validation are key to gaining acceptance for and value from ICME tools that are less than 100 percent accurate.” A clear recommendation was reached by a recent study on Applied Mathematics by the U.S. Department of Energy (Brown 2008) to “significantly advance the theory and tools for quantifying the effects of uncertainty and numerical simulation error on predictions using complex models and when fitting complex models to observations.”

The data and other information the WTEC panel collected in its study suggests that there are a lot of “simulation-meets-experiment” types of projects but no systematic effort to establish the rigor and the requirements on UQ and V&V that the cited reports have suggested are needed.

Finding 1: Overall, the United States leads the research efforts today, at least in terms of volume, in quantifying uncertainty, mostly in computational mechanics; however, there are similar recent initiatives in Europe.

- Specifically, fundamental work in developing the proper stochastic mathematics for this field, e.g., in addressing the high-dimensionality issue, is currently taking place in Germany, Switzerland, and Austria.

In the United States, the DOD Defense Modeling and Simulation Office (DMSO) has been the leader in developing V&V frameworks, and more recently the Department of Energy has targeted UQ through the ASCI program and its extensions (including its current incarnation, PSAAP). The ASCI/PSAAP program is focused on computational physics and mechanics problems, whereas DMSO has historically focused on high-level systems engineering, e.g., warfare modeling and simulation-based acquisition. One of the most active groups in V&V and UQ is the Uncertainty Estimation Department at Sandia National Labs, which focuses mostly on quantifying

uncertainties in complex systems. However, most of the mathematical developments are taking place in universities by a relatively small number of individual researchers.

Finding 2: Although the U.S. Department of Defense and Department of Energy have been leaders in V&V and UQ efforts, they have been limited primarily to high-level systems engineering and computational physics and mechanics, respectively, with most of the mathematical developments occurring in universities by small numbers of researchers. In contrast, several large European initiatives stress UQ-related activities.

There are currently no funded U.S. national initiatives for fostering collaboration between researchers who work on new mathematical algorithms for V&V/UQ frameworks and design guidelines for stochastic systems. In contrast, there are several European initiatives within the Framework Programs to coordinate research on new algorithms for diverse applications of computational science, from nanotechnology to aerospace. In Germany, in particular, UQ-related activities are of utmost importance at the Centers of Excellence; for example, the University of Stuttgart has a chaired professorship of UQ (sponsored by local industry), and similar initiatives are taking place in the Technical University of Munich and elsewhere.

Finding 3: Existing graduate level curricula, worldwide, do not teach stochastic modeling and simulation in any systematic way.

Indeed, very few universities offer regular courses in stochastic partial differential equations (SPDEs), and very few textbooks exist on numerical solution of SPDEs. The typical graduate coursework of an engineer does not include advanced courses in probability or statistics, and important topics such as design or mechanics, for example, are taught with a purely deterministic focus. Statistical mechanics is typically taught in Physics or Chemistry departments that may not fit the background or serve the requirements of students in Engineering. Courses in mechanics and other disciplines that emphasize a statistical description *at all scales* (and not just at the small scales) due to both intrinsic and extrinsic stochasticity would be particularly effective in setting a solid foundation for educating a new cadre of simulation scientists.

Crosscutting Issues: Multiscale Modeling and Simulation

True multiscale simulation has long been a goal of SBE&S. It naturally arises as a desirable capability in virtually every application area of SBE&S. It particularly arises when the notion of design and/or control comes into play, since in this case the impact of changes in problem definition at the larger scales needs to be understood in terms of the impact on structures at the smaller scales. For example, in materials, the question might be how changes in the desirable performance characteristics of a lubricant translate into differences in molecular structure of the constituent molecules, so that new lubricants can be synthesized. In biology, the question might be what needs to happen at the cell-cell interaction level in order to control the spread of cancer. In energy and sustainability, the goal might be to understand how setting emissions targets for carbon dioxide at power plants in the United States may affect the growth and biodiversity of South American rain forests. Answering all of these questions requires a combination of downscaling and upscaling, and so fall in the realm of multiscale simulation.

Finding 1: Multiscale modeling is exceptionally important and holds the key to making SBE&S more broadly applicable in areas such as design, control, optimization, and abnormal situation modeling. However, although successful examples exist within narrow disciplinary boundaries, attempts to develop general strategies have not yet succeeded.

Finding 2: The lack of code interoperability is a major impediment to industry's ability to link single-scale codes into a multiscale framework.

Finding 3: Although U.S. research in multiscale simulation today is on a par with Japan and Europe, it is diffuse, lacking focus and integration, and federal agencies have not traditionally supported the long-term development of codes that can be distributed, supported, and successfully used by others. This contrasts starkly with efforts in Japan and Europe, where large, interdisciplinary teams are supported long term to distribute codes either in open-source or commercial form.

- The current standing of the United States in multiscale modeling and simulation is due in part to U.S. leadership in high-performance computing resources. This is because to validate a multiscale simulation methodology it is necessary to perform the simulation at full detail enough times to provide a validation data set, and this may require enormous computational resources. Thus, continued pre-eminence in leadership-class computing may underpin future U.S. leadership in multiscale modeling and simulation.
 - One concern noted by the panel is that access to the largest computational facilities (such as the emerging petascale platforms) may be difficult for a project whose stated aim is to run a series of petaflop-level simulations in order to develop and validate a multiscale modeling methodology. Access to petaflop-level resources today is generally focused on solving a small set of problems, as opposed to developing the methodologies that will enable the solution of a broader range of problems.

Crosscutting Issues: Big Data, Visualization, and Data-driven Simulation

The role of big data and visualization in driving new capabilities in SBE&S is pronounced and critical to progress in the three thematic areas of this report. We can also interpret the “big data” challenges faced by the SBE&S community as a recursive resurfacing of the CPU-memory “data bottleneck” paradigm from the chip architecture scale to societal scale. The locus of large-scale efforts in data management correlates by discipline and not by geographical region.

Finding 1: The biological sciences and the particle physics communities are pushing the envelope in large-scale data management and visualization methods. In contrast, the chemical and material science communities lag in prioritization of investments in the data infrastructure.

- In the biological sciences, in both academic laboratories and industrial R&D centers, there is an appreciation of the importance of integrated, community-wide infrastructure for dealing with massive amounts of data, and addressing issues of data provenance, heterogeneous data, analysis of data and network inference from data. There are great opportunities for the chemical and materials communities to move in a similar direction, with the promise of huge impacts on the manufacturing sector.

Finding 2: Industry is significantly ahead of academia with respect to data management infrastructure, supply chain, and workflow.

- Even within a given disciplinary area, there is a notable difference between industrial R&D centers and academic units, with the former placing significantly more attention to data management infrastructure, data supply chain, and workflow, in no small part due to the role of data as the foundation for intellectual property (IP) assets. In contrast, most universities lack a campus-wide strategy for the ongoing transition to data-intensive research and there is a widening gap between the data infrastructure needs of the current generation of graduate students and the capabilities of the campus IT infrastructure. Moreover, industrial firms are particularly active and participate in consortia to promote open standards for data exchange – a recognition that SBE&S is not a series of point solutions but an integrated set of tools that form a workflow engine. Companies in highly regulated industries, e.g., biotechnology and pharmaceutical companies, are also exploring open standards and data exchange to expedite the regulatory review processes for new products.

Finding 3: Big data and visualization capabilities are inextricably linked, and the coming “data tsunami” made possible by petascale computing will require more extreme visualization capabilities than are currently available, as well as appropriately trained students who are proficient in data infrastructure issues.

- Big data and visualization capabilities go hand in hand with community-wide software infrastructure; a prime example is in the particle physics community and the Large Hadron Collider infrastructure networking CERN to the entire community in a multi-tier fashion. Here, visualization techniques are essential given the massive amounts of data, and the rarity of events (low signal to noise ratio) in uncovering new science.

Finding 4: Big data, visualization and dynamic data-driven simulations are crucial technology elements in numerous “grand challenges,” including the production of transportation fuels from the last remaining giant oil fields.

- Global economic projections for the next two decades and recent examples of price elasticity in transportation fuels suggest that the scale of fluctuation in reservoir valuations would be several orders of magnitude beyond the global spending in SBE&S research.

Crosscutting Issues: Education and Training

Education and training of the next generation of computational scientists and engineers proved to be the number one concern at nearly all of the sites visited by the panel. A single, overarching major finding can be summarized as follows:

Finding: Continued progress and U.S. leadership in SBE&S and the disciplines it supports are at great risk due to a profound and growing scarcity of appropriately trained students with the knowledge and skills needed to be the next generation of SBE&S innovators.

This is particularly alarming considering that, according to numerous recent reports,

- The U.S. lead in most scientific and engineering enterprises is decreasing across all S&E indicators;
- The number of U.S. doctorates both in the natural sciences and in engineering has been surpassed by those in the European Union and by those in Asia.
- Countries in the European Union and Japan are scampering to recruit foreign students to make up for the dwindling numbers of their youths. This is driving fierce competition for international recruiting. The United States is sitting at exactly the “fertility replacement rate” and only a small percentage of its population joins scientific and engineering professions. China and India, on the other hand, are likely to become increasingly strong SBE&S competitors.

The panel observed many indicators of increased focus on SBE&S education in China and in Europe:

- The European Union is investing in new centers and programs for education and training in SBE&S—all of an interdisciplinary nature. New BSc and MSc degrees are being offered in SBE&S through programs that comprise a large number of departments. In many cases, a complete restructuring of the university has taken place in order to create, for instance, MSc degrees, graduate schools, or international degrees in simulation technology.
- European and Asian educational/research centers are attracting more international students (even from the United States). Special SBE&S programs (in English) are being created for international students. The United States is seeing smaller increases in international enrollment compared to other countries. This is due not only to post-9/11 security measures but also—and increasingly—to active competition from other countries.

All countries, however, share many of the same challenges when it comes to SBE&S education and training:

- Students are trained primarily to run existing codes rather than to develop the skills in computational mathematics and software engineering necessary for the development of the next generation of algorithms and software. There are pitfalls in interdisciplinary education such as computational science and engineering, including a tradeoff between breadth and depth. In order to solve “grand challenge” problems in a given field, solid knowledge of a core discipline, in addition to computational skills, are crucial.
- Demand exceeds supply. There is a huge demand in the European Union and Asia for qualified SBE&S students who get hired immediately after their MSc degrees by industry or finance: there is both collaboration and competition between industry and academia. This phenomenon in the United States may be tempered in the future due to the current economic situation. However, one might argue that the need for increased physics-based modeling and simulation in macroeconomics and financial economics, and in industry in general as it cuts costs while preserving the ability for innovation, has never been more urgent.

The WTEC Panel on Simulation-Based Engineering and Science

S.C. Glotzer (Chair), S.T. Kim (Vice Chair), P.T. Cummings, A. Deshmukh, M. Head-Gordon, G. Karniadakis, L. Petzold, C. Sagui and M. Shinozuka

April 2009

REFERENCES

- Benioff, Marc. R. and Edward D. Lazowska, Co-chairs. 2005. *Computational science: Ensuring America's competitiveness*. President's Information Technology Advisory Committee (PITAC).
http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf
- European Computational Science Forum of the European Science Foundation (ESF). 2007. *The Forward Look Initiative. European computational science: The Lincei Initiative: From computers to scientific excellence*. Information available online: <http://www.esf.org/activities/forward-looks/all-current-and-completed-forward-looks.html>.
- European Strategy Forum for Research Infrastructures (ESFRI). 2006. *European roadmap for research infrastructures: Report 2006*. Luxembourg: Office for Official Publications of the European Communities. Available online: ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf
- Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science Foundation blue ribbon panel on simulation-based engineering science*. Arlington, VA: National Science Foundation. Available online: http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.
- OMB. 2008. Agency NITRD Budget by program component area FY 2008 budget estimate and budget requests. President's FY 2009 Budget (NITRD Supplement).
- National Research Council (NRC), 2008. National Academy of Engineering Report of Committee on Integrated Computational Materials Engineering (ICME). National Academies Press, Washington, D.C.

ADDITIONAL READING

- Atkins, D.M., National Science Foundation, 2006. NSF's cyberinfrastructure vision for 21st century discovery. Version 7.1. <http://nsf.gov>.
- Bramley, R., B. Char, D. Gannon, T. Hewett, C. Johnson, and J. Rice. 2000. Enabling technologies for computational science: Frameworks, middleware and environments. In *Proceedings of the Workshop on Scientific Knowledge, Information, and Computing*, E. Houstis, J. Rice, E. Gallopoulos, and R. Bramley, eds., Boston: Kluwer Academic Publishers, pp. 19-32.
- Committee on Science, Engineering, and Public Policy (COSEPUP) [a joint unit of the National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine]. 2007. *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology. Washington, DC: National Academies Press. Available online: http://www.nap.edu/catalog.php?record_id=11463#orgs.
- Council on Competitiveness. 2004. *Final Report from the high performance users conference: Supercharging U. S. innovation and competitiveness*, Washington, D.C..
- Coveney, P. et al. 2006. U.K. Strategic framework for high-end computing.. <http://www.epsrc.ac.uk/CMSWeb/Downloads/Other/2006HECStrategicFramework.pdf>
- Department of Defense (DoD). 2002. Report on high performance computing for the national security community, http://WWW.HPCMO.HPC.mil/Htdocs/DOCUMENTS/04172003_hpc_report_unclass.pdf.
- Department of Defense. 2000. Defense Science Board, Report of the Defense Science Board Task Force on DoD supercomputing needs, <http://www.acq.osd.mil/dsb/reports/dodssupercomp.pdf>.
- Department of Energy. 2000. Office of Science, Scientific discovery through advanced computing.
- Executive Office of the President. 2004. Office of Science and Technology Policy, Federal plan for high-end computing: report of the High-End Computing Revitalization Task Force (HECRTF).
- Interagency Panel on Large Scale Computing in Science and Engineering (The Lax Report). 1982. *Report of the panel on large scale computing in science and engineering*. Department of Defense and the National Science Foundation, in cooperation with the Department of Energy and the National Aeronautics and Space Administration.
- Joy, W., and K. Kennedy, eds. 1999. President's Information Technology Advisory Committee (PITAC), *Report to the President: Information technology research: Investing in our future*. Available online: <http://www.nitrd.gov/pitac/report/>.

- Joseph, E., A. Snell, and C. G. Willard. 2004. Council on Competitiveness Study of U.S. HPC Users. White Paper. http://www.compete.org/pdf/HPC_Users_Survey.pdf.
- Kaufmann, N. J., C. G. Willard, E. Joseph, and D. S. Goldfarb 2003. Worldwide high performance systems technical computing census. IDC Report No. 62303.
- Keyes, D., P. Colella, T.H. Dunning, Jr., and W.D. Gropp, eds. 2003. *A science-based case for large-scale simulation*, vol. 1. Washington, DC: Office of Science, U.S. Department of Energy (DOE). Available online: <http://www.pnl.gov/scales/>.
- . 2004. *A science-based case for large-scale simulation*, vol. 2. Washington, DC: DOE Office of Science. Available online: <http://www.pnl.gov/scales/>.
- National Research Council (NRC). 2005. Computer Science and Telecommunications Board, Getting up to speed: The future of supercomputing. National Academies Press, Washington, D.C.
- National Research Council (NRC), 2008. The potential impact of high-end capability computing on four illustrative fields of science and engineering. National Academies Press, Washington, D.C.
- National Science Foundation. 1993. NSF Blue Ribbon Panel on High Performance Computing. *From desktop to teraflop: Exploiting the U.S. lead in high performance computing*.
- National Science Foundation. 1995. *Report of the task force on the future of the NSF supercomputer centers program*.
- National Science Foundation. 2003. *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue ribbon advisory panel on cyberinfrastructure, 2003*.
- President's Council of Advisors on Science and Technology. 2004. *Report to the President: Sustaining the nation's innovation ecosystems, information technology, manufacturing, and competitiveness*.
- Society for Industrial and Applied Mathematics (SIAM) Working Group on Computational Science and Engineering (CSE) Education. 2001. Graduate education in computational science and engineering. *SIAM Review* 43(1):163-177.
- Workshop and Conference on Grand Challenges Applications and Software Technology. 1993. Pittsburgh, Pennsylvania.

xxx

CHAPTER 1

INTRODUCTION

Sharon C. Glotzer

BACKGROUND AND SCOPE

The impetus for this study builds upon many national and international reports highlighting the importance and promise of computational science and engineering. From the 1982 US interagency study on computational science now known as the Lax Report (Lax 1982) to the 2006 U.S. National Science Foundation Blue Ribbon panel study on simulation-based engineering science known as the Oden report (Oden 2006) to the 2008 National Research Council Reports on High End Capability Computing for Science & Engineering (NRC 2008a) and Integrated Computational Materials Engineering (NRC 2008b), critical advances in such fields as medicine, nanotechnology, aerospace and climate prediction made possible by application of high-end computing to the most challenging problems in science and engineering have been well documented. Yet, as demonstrated in the 2005 report of the President's Information Technology Advisory Committee (PITAC) (Benioff 2005), investment in computational science in the United States and the preparation of the next generation of computational researchers remains insufficient to fully leverage the power of computation for solving the biggest problems that face us as a nation going forward. As the United States and Japan continue to invest in the next generation of computer architectures, with the promise of thousand-fold or more increases of computer power coming in the next half-decade, an assessment of where the United States stands relative to other countries in its use of, commitment to, and leadership in computer simulation for engineering and science is therefore of urgent interest. That is the focus of this report.

Today we are at a "tipping point" in computer simulation for engineering and science. Computer simulation is more pervasive today – and having more impact – than at any other time in human history. No field of science or engineering exists today that has not been advanced by, and in some cases utterly transformed by, computer simulation. The Oden report (Oden 2006) summarizes the state of simulation-based engineering science today: *"...advances in mathematical modeling, in computational algorithms, and the speed of computers and in the science and technology of data-intensive computing have brought the field of computer simulation to the threshold of a new era, one in which unprecedented improvements in the health, security, productivity, and competitiveness of our nation may be possible."* The report further argues that a host of critical technologies are on the horizon that cannot be understood, developed, or utilized without simulation methods. Simulation has today reached a level of predictiveness that it now firmly complements the traditional pillars of science and engineering, namely theory and experimentation/observation. Moreover, there is abundant evidence that computer simulation is critical to scientific discovery and engineering innovation. At the same time, computers are now affordable and accessible to researchers in every country around the world. The near-zero entry-level cost to perform a computer simulation means that anyone can practice SBE&S, and from anywhere. Indeed, the world of computer simulation is flattening, and becoming flatter every day. In that context, it is therefore meaningful to examine the question of what it means for any nation, the United States in particular, to lead in SBE&S.

The next decade is shaping up to be a transformative one for SBE&S. A new generation of massively multicore computer chip architectures is on the horizon that will lead to petaflop computing, then exaflop computing, and beyond. These architectures will allow unprecedented accuracy and resolution, as well as the ability to solve the highly complex problems that face society today. The toughest technological and scientific problems facing society, from finding alternative energy sources to global warming to sustainable infrastructures, to curing disease and personalizing medicine, are *big* problems. They are complex and messy, and their solution requires a partnership among experiment, theory, and simulation working across all of the disciplines of science and engineering. Recent reports like *Rising Above the Gathering Storm* (COSEPUP 2007) argue that our nation is at risk of losing its competitive edge. Our continued capability as a nation to lead in simulation-based discovery and innovation is key to our ability to compete in the 21st century.

Yet because it is often viewed more as an enabling technology for other disciplines, rather than a discipline in its own right, investment in and support of computational science and engineering is often not prioritized as it should be at all levels of the R&D enterprise. The 2005 PITAC (Benioff 2005) report writes, “*The universality of computational science is its intellectual strength. It is also its political weakness. Because all research domains benefit from computational science but none is solely defined by it, the discipline has historically lacked the cohesive, well-organized community of advocates found in other disciplines. As a result, the United States risks losing its leadership and opportunities to more nimble international competitors. We are now at a pivotal point, with generation-long consequences for scientific leadership, economic competitiveness, and national security if we fail to act with vision and commitment.*”

To provide program managers in U.S. research agencies as well as researchers in the field a better understanding of the status and trends in SBE&S R&D abroad, in 2007 the National Science Foundation (NSF), in cooperation with the Department of Energy (DOE), NASA, National Institutes of Health (NIH), National Library of Medicine, the National Institute of Standards and Technology (NIST), and the Department of Defense (DOD), sponsored this *WTEC International Assessment of Simulation-Based Engineering and Science*. The study was designed to gather information on the worldwide status and trends in SBE&S research and to disseminate it to government decision makers and the research community. The study participants reviewed and assessed the state of the art in SBE&S research and its application in academic and industrial research. Questions of interest to the sponsoring agencies addressed by the study included:

- Where are the next big breakthroughs and opportunities coming from in the United States and abroad, and in what fields are they likely to be?
- Where is the United States leading, trailing, or in danger of losing leadership in SBE&S?
- What critical investments in SBE&S are needed to maintain U.S. leadership, and how will those investments impact our ability to innovate as a nation?

Simulation-based engineering and science is a vast field, spanning the physical, chemical, biological, and social sciences and all disciplines of engineering, and it would be impossible to cover all of it in any meaningful way. Instead, the study focused on SBE&S and its role in three primary thematic areas that will be transformative in the health and prosperity, and in the economic and national security, of the US. The thematic areas studied are:

- Life sciences and medicine
- Materials
- Energy and sustainability

In turn, these domains are ones in which SBE&S promises to have a transformative effect in the near term provided that continued and, in many cases, expedited progress occur in the following crosscutting areas, which were identified in the Oden report (Oden 2006) and included in this study:

- Next generation architectures and algorithms
- Scientific and engineering simulation software development
- Engineering simulation
- Validation, verification and uncertainty quantification

- Multiscale modeling and simulation
- Big data, visualization, and data-driven simulation
- Education and training

As with all WTEC reports, this report will present findings that can be used as guidance for future recommendations and investments.

METHODOLOGY

The agency sponsors recruited the study chair, Sharon C. Glotzer, Professor of Chemical Engineering, Materials Science and Engineering, Physics, Applied Physics, and Macromolecular Science and Engineering at the University of Michigan, Ann Arbor, in early 2007. On March 13, 2007 an initial meeting was held at NSF headquarters in Arlington, VA with the agency sponsors, WTEC representatives, and the study chair, to establish the scope of the assessment. WTEC then recruited a panel of U.S. experts (see Table 1.1).¹ The assessment was initiated by a kickoff meeting on July 10, 2007 at NSF headquarters in Arlington, Virginia. Participants discussed the scope of the project and the need for a North American baseline workshop, candidate sites in Europe and Asia for panel visits, the overall project schedule, and assignments for the final report. WTEC also recruited, at the request of the chair, a team of advisors (Table 1.2) to provide additional input to the process. Brief biographies of the panelists and advisors are presented in Appendix A.

Table 1.1. Panel Members

Panelist	Affiliation
Sharon C. Glotzer (Panel Chair)	University of Michigan, Ann Arbor
Sangtae Kim (Vice Chair)	Morgridge Institute for Research
Peter Cummings	Vanderbilt University and Oak Ridge National Laboratory
Abhijit Deshmukh	Texas A&M University
Martin Head-Gordon	University of California, Berkeley
George Karniadakis	Brown University
Linda Petzold	University of California, Santa Barbara
Celeste Sagui	North Carolina State University
Masanobu Shinozuka	University of California, Irvine

Table 1.2. Study Advisors

Panelist	Affiliation
Jack Dongarra	University of Tennessee and Oak Ridge National Laboratory
James Duderstadt	University of Michigan, Ann Arbor
J. Tinsley Oden	University of Texas, Austin
Gilbert S. Omenn	University of Michigan, Ann Arbor
Tomás Diaz de la Rubia	Lawrence Livermore National Laboratory
David Shaw	D. E. Shaw Research and Columbia University
Martin Wortman	Texas A&M University

¹ The panel is grateful to panelist Masanobu Shinozuka who, despite being unable to travel to the U.S. Baseline Workshop and the international trips due to an injury, participated in the initial and final workshops at NSF and contributed the chapter on Energy & Sustainability in the final report.

The panelists, sponsors, and WTEC convened a U.S. Baseline Workshop on November 20-21, 2007 at the Hilton Hotel in Arlington, VA to report on the current status of US R&D in simulation-based engineering & science. Table 1.3 lists the speakers and the titles of their presentations.

The speakers at the U.S. Baseline Workshop were selected by the panelists with input from the advisors and program managers, and with an eye towards complementing the expertise and knowledge of the study panel and emphasizing the thematic areas and crosscutting issues chosen for the study. The workshop included ample time for discussion. To kick off the workshop, Michael Reischman of NSF and Duane Shelton of WTEC provided welcoming remarks, discussed the importance of simulation-based engineering and science to NSF, and provided the context for the study. Study chair Sharon Glotzer presented an overview of the planned study and study process. Panel advisor J. Tinsley Oden of the University of Texas, Austin, who chaired the NSF Blue Ribbon Panel Study on Simulation-Based Engineering Science, provided a summary of that study.

Following this introduction, sessions were organized to target the specific themes and crosscutting issues chosen for the study. In the Materials session, chaired by panelist Peter Cummings, Emily Carter of Princeton University discussed the status and challenges of quantum-mechanics (QM) based materials simulations. She provided a summary of open source and commercial QM-based materials simulation codes being developed and/or used today, demonstrating the dearth of QM code development in the United States today relative to other countries and discussing the implications of this. Tom Arsenlis of Lawrence Livermore National Laboratory presented a talk for Tomás Diaz de la Rubia on the path towards petaflop/s computing for materials, national security, energy and climate applications. Arsenlis demonstrated several achievements by the DOE labs in pushing the boundaries of large scale computing. In the session on Energy and Sustainability, chaired by study vice-chair Sangtae Kim, Roger Ghanem of the University of Southern California described how computer simulation is used to predict complex interactions in urban development and design strategies for mitigating unwanted interactions. Lou Durlofsky of Stanford University described the use of computer simulation for optimizing oil and gas production and CO₂ sequestration, and discussed challenges and opportunities for the United States. Sangtae Kim presented a talk prepared by Rick Mifflin of Exxon-Mobil on high performance computing and U.S. competitiveness in the practice of oil reservoir simulation, demonstrating the enormous possibilities that remain for predicting high yield sites with increased use of simulation-based engineering and science. In the Life Sciences session, chaired by panelist Linda Petzold, David Shaw of D.E. Shaw Research demonstrated enormous speedups now possible in molecular dynamics simulation of biomolecules from critical advances in hardware, algorithms, and software. These recent advances now make possible the simulation of biological processes on the relevant time scales over which they occur. Alex Bangs of Entelos described the use of predictive biosimulation in pharmaceutical R&D, a major growth area for simulation.

In the session on Multiscale Simulation, chaired by panelist Peter Cummings, Ioannis Kevrekidis of Princeton University and W.K. Liu of Northwestern University described the status of method development for bridging time and length scales in complex science and engineering problems, comparing and contrasting efforts in the United States and abroad. In the session on Engineering Design, chaired by panelist Abhi Deshmukh, Troy Marusich of Third Wave Technologies discussed how advances in computational capabilities are now permitting the use of physics-based models in simulations to improve machining. Karthik Ramani of Purdue University described the use of simulation-based engineering and science for engineering design. Panelist George Karniadakis provided a review of, and led a discussion on, validation, verification and uncertainty quantification. He identified issues and challenges and reviewed U.S. efforts in this area. In the session on Big Data, chaired by study vice-chair Sangtae Kim, Fran Berman of the San Diego Supercomputer Center discussed the enormous data challenges associated with petascale (and beyond) simulations. Ernst Dow of Eli Lilly discussed trends in visualization both in the United States and abroad and the importance of visualization in interpreting and exploring very large data sets.

In the session on Next Generation Algorithms and High Performance Computing, chaired by panelist Linda Petzold, Jack Dongarra of the University of Tennessee presented an overview of high performance computing and its evolution over the years. He compared the power of laptop computers today with room-sized supercomputers of a decade earlier, and discussed future trends in HPC, including the great opportunities and challenges accompanying manycore chip architectures. Bill Gropp of Argonne National Laboratory and David

Keyes of Columbia University also discussed architecture trends and described their implications for algorithms. Keyes described the critical importance of algorithms in achieving the many orders of magnitude in speed up needed to tackle some of the most challenging problems in science and engineering.

The session on Software Development was chaired by panelist Martin Head-Gordon of the University of California, Berkeley. In this session, Dennis Gannon of Indiana University described the use of high performance computing for storm prediction and tornado forecasting, and C.C. Chen of the Aspen Institute discussed high impact opportunities in simulation-based product and process design and development. The final session of the Baseline Workshop focused on Education and was chaired by panelist Celeste Sagui. J. Tinsley Oden of the University of Texas, Austin, discussed the need for new organizational structures and approaches within universities to support computational science and engineering education. The session sparked perhaps the most passionate and lengthy discussions of the workshop, on the topic of educating the next generation of computational scientists and engineers. This is a theme that arose throughout the study as one of great concern around the world, as discussed later in this report.

A bibliometric analysis (see Appendix E), combined with input from the panelists, advisors, and participants of the Baseline Workshop and other experts, provided guidance in selecting European and Asian sites to visit. Sites were selected as representative of what were perceived or benchmarked to be the most productive and/or most reputable and/or highest impact research labs and groups outside of the United States in the various areas of SBE&S comprising the study. Because the field is so broad and both time and resources finite, it was impossible to visit every leading group or research lab, or to visit every country where major rapid growth in research, including computational science, is underway (most notably Singapore and Saudi Arabia). We note that more sites were visited for this study than in any previous WTEC study. The panel therefore relied on its own knowledge of international efforts, as well as informal conversations with colleagues and briefings heard in different venues, to complement what we learned in the site visits.

The international assessment phase of the WTEC study commenced in December 2007 with visits to the 21 sites in China and Japan shown in Figures 1.1, 1.2, and Table 1.4. To focus the discussions in the various laboratories, the host engineers and scientists were provided with a set of questions (Appendix D) compiled by the panel in advance of the visits. While the discussion did not necessarily follow the specific questions, they provided a general framework for the discussions. During its visit to China, the WTEC panel was privileged to attend a symposium on SBE&S organized by Dalian University, at which approximately 75 faculty and students heard presentations from a dozen faculty members in China whose research groups are actively exploiting SBE&S. Our hosts both in Asia and Europe demonstrated a wide range of computational research applied to many important problems in science and engineering. Talks discussed research in bioinformatics, computational nanomechanics, computational fluid dynamics, offshore oil drilling, composite materials, automotive design, polymer science and software engineering, as examples. The Asia trip concluded with a meeting in Nagoya, Japan on December 8, 2007, in which the panelists reviewed and compared their site visit notes. A week of visits to the 38 European sites listed in Figure 1.3 and Table 1.5 commenced on February 24, 2008, and concluded with a wrap-up meeting in Frankfurt, Germany on March 1, 2008.

The panel reconvened for a final workshop at NSF on April 21, 2008, to present its findings and conclusions. Presentations focused on the three thematic domains of life sciences and medicine; materials; energy and sustainability; and on the seven crosscutting issues identified above.

Table 1.3. Speakers at the U.S. Baseline Workshop

Name	Affiliation	Presentation Title
Michael Reischman	NSF	Importance of Field to NSF
Duane Shelton	WTEC	Context of Study
Sharon Glotzer (Chair)	University of Michigan	Introduction and Study Overview
J. Tinsley Oden	University of Texas at Austin	The NSF Blue Ribbon Panel Report on SBES
Peter Cummings	Vanderbilt University	Materials (Session Introduction)
Emily Carter	Princeton University	Status & Challenges in Quantum Mechanics Based Simulations of Materials Behavior
Tomás Diaz de la Rubia (Presented by Tom Arsenlis)	Lawrence Livermore National Laboratory	Toward Petaflops Computing for Materials, National Security, Energy, and Climate Applications
Sangtae Kim (Vice Chair) for M. Shinozuka	Morgridge Institute for Research	Energy & Sustainability (Session Introduction)
Roger Ghanem	University of Southern California	Sustainability by Design: Prediction and Mitigation of Complex Interactions in the Urban Landscape
Lou Durlofsky	Stanford University	Computational Challenges for Optimizing Oil and Gas Production and CO ₂ Sequestration
R.T. (Rick) Mifflin (Presented by Sangtae Kim)	Exxon Mobil	High-Performance Computing and U.S. Competitiveness in the Practice of Reservoir Simulation
Linda Petzold	University of California, Santa Barbara	Life Sciences (Session Introduction)
David Shaw	D. E. Shaw Research	Molecular Dynamics Simulations in the Life Sciences
Alex Bangs	Entelos	Predictive Biosimulation in Pharmaceutical R&D
Peter Cummings	Vanderbilt University	Multiscale Simulation (Session Introduction)
Ioannis Kevrekidis	Princeton University	Equation-Free Modeling and Computation for Complex/Multiscale Systems
W.K. Liu	Northwestern University	SBE&S Approach to Analysis and Design of Microsystems: From a Dream to a Vision to Reality
Abhi Deshmukh	Texas A&M University	Engineering Design (Session Introduction)
Troy Marusich	Third Wave Technologies	Improved Machining via Physics-Based Modeling
Karthik Ramani	Purdue University	Leveraging Simulations for Design: Past-Present-Future
George Karniadakis	Brown University	Validation, Verification, and Uncertainty Quantification (Session Introduction and Review)
Sangtae Kim (Vice Chair)	Morgridge Institute for Research	Big Data (Session Introduction)
Fran Berman	San Diego Supercomputer Center	Simulation and Data
Ernst Dow	Eli Lilly	Big Data and Visualization
Linda Petzold	University of California, Santa Barbara	Next-Generation Algorithms/High-Performance Computing (Session Introduction)
Jack Dongarra	University of Tennessee	An Overview of High Performance Computing and Challenges for the Future
William D. Gropp	Argonne National Laboratory	Architecture Trends and Implications for Algorithms
David Keyes	Columbia University	Attacking the Asymptotic Algorithmic Bottleneck: Scalable Solvers for SBE&S
Martin Head-Gordon	University of California, Berkeley	Software Development (Session Introduction)
Dennis Gannon	Indiana University	Software for Mesoscale Storm Prediction: Using Supercomputers for On-the-Fly Tornado Forecasting
Chau-Chyun Chen	Aspen Institute	High Impact Opportunities in Simulation-Based Product and Process Design and Development
Celeste Sagui	North Carolina State University	Education (Session Introduction)
J. Tinsley Oden	University of Texas at Austin	Goals and Barriers to Education in SBE&S

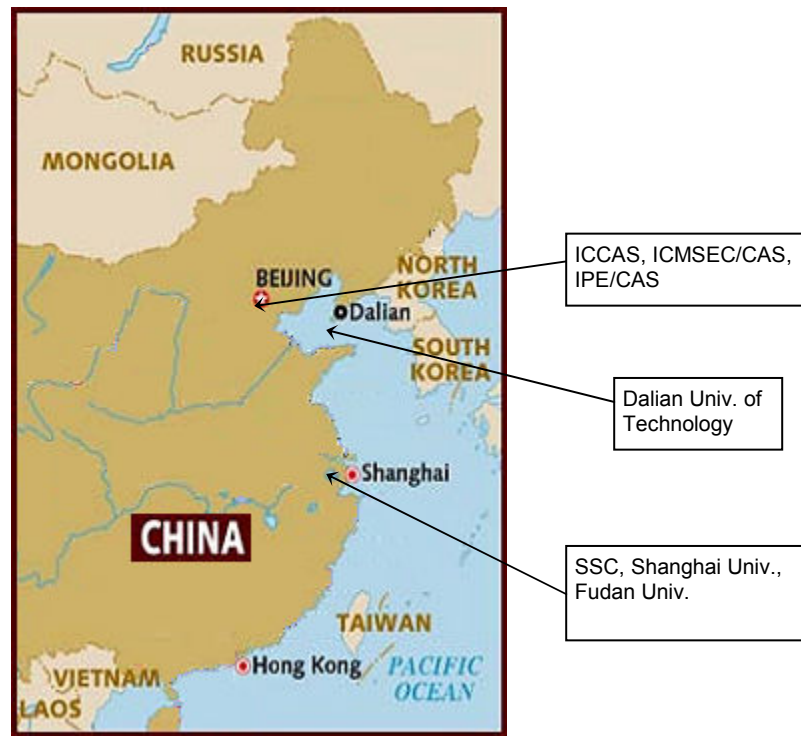


Figure 1.1. Sites visited in China.

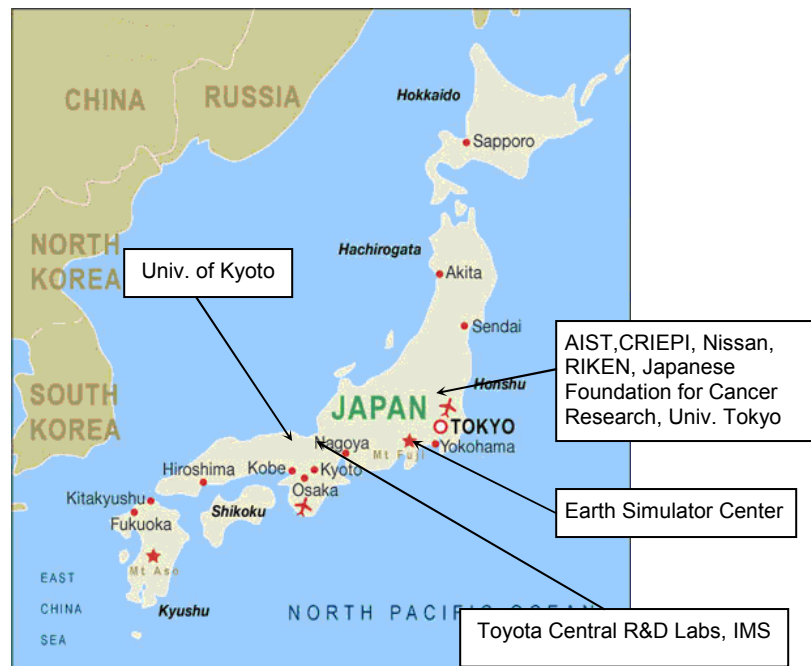


Figure 1.2. Sites visited in Japan.

Table 1.4. Sites Visited in Asia

#	Country	Site	#	Country	Site
1	China	Institute of Process Engineering, Institute of Computational Mathematics and Scientific/Engineering Computing (ICMSEC-CAS)	12	Japan	Central Research Institute- Electric Power Industry (CRIEPI)
2	China	ICCAS, Institute of Chemistry (Joint Lab for Polymer and Materials Science)	13	Japan	Japan Agency for Marine Earth Science and Technology, Earth Simulator Center (ESC)
3	China	Dalian University of Technology, Department of Engineering Mechanics	14	Japan	Japanese Foundation for Cancer Research
4	China	Dalian University of Technology, School of Automobile Engineering	15	Japan	Kyoto University, Department of Synthetic Chemistry and Biological Chemistry
5	China	Fudan University, Yulang Yang Research Group	16	Japan	Nissan
6	China	Peking University, Computer Science and Engineering	17	Japan	Institute for Molecular Science (IMS)
7	China	Shanghai University, School of Material Science and Engineering	18	Japan	Mitsubishi Chemical Group
8	China	Shanghai University, High Performance Computer Center	19	Japan	RIKEN Next-Generation Supercomputer R&D Center
9	China	Tsinghua University, Engineering Mechanics	20	Japan	Toyota Central Research Laboratory
10	Japan	Advanced Industrial Science and Technology (AIST), Research Institute for Computational Science (RICS)	21	Japan	University of Tokyo
11	Japan	AIST, National Institute of Materials Science (NIMS)			

OVERVIEW OF THE REPORT

This final report closely follows the outline of the final workshop held in April 2008. Following this introductory chapter, we address three major thematic domains in which SBE&S is having unprecedented impact and investigate SBE&S research internationally in those domains as compared with the United States. First, Linda Petzold discusses in Chapter 2 how advances in SBE&S are contributing to breakthroughs in the biomedical and life sciences arena. The simulation of biological processes on relevant time and length scales, and of whole organs and complete physiological systems, is becoming possible as we move towards petascale computing. Predictive biosimulation now supports the development of new therapeutic drugs, meeting the pharmaceutical industry's need for new approaches to deal with the complexity of disease and the increased risk and expense of research and development. Biomechanical simulation is used in surgical planning and design of physical therapies, and simulation of the brain is a realistic near-term goal.

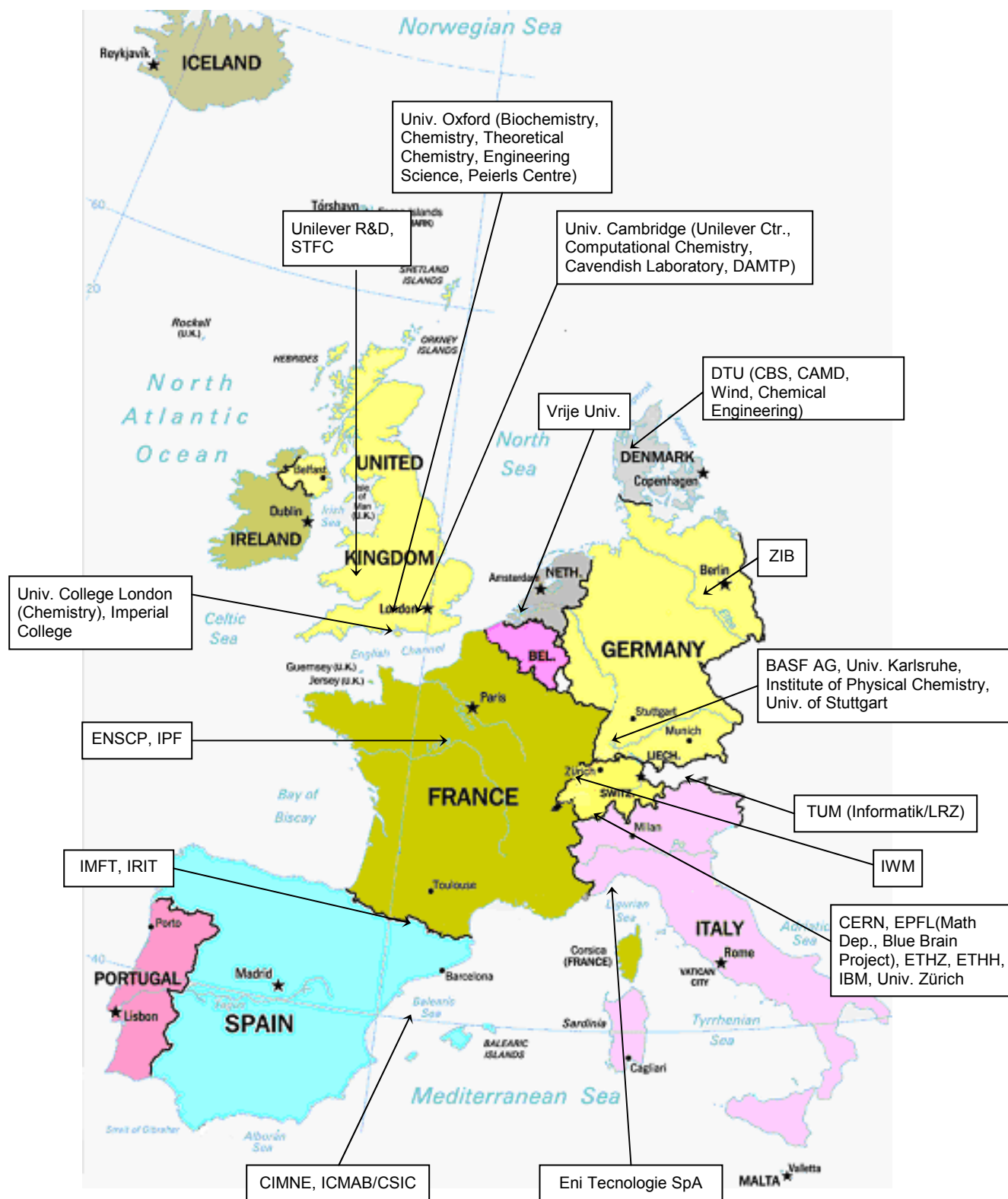


Figure 1.3. Sites visited in Europe.

Table 1.5. Sites Visited in Europe

#	Country	Site	#	Country	Site
1	Denmark	Technical University of Denmark (DTU), Center for Biological Sequence Analysis, Systems Biology Department (CBS)	20	Switzerland	European Organization for Nuclear Research (CERN)
2	Denmark	DTU, Fluid Mechanics (Wind)	21	Switzerland	Ecole Polytechnique Fédérale de Lausanne (EPFL), Blue Brain Project
3	Denmark	DTU, Physics, Center for Applied Materials Design (CAMD)	22	Switzerland	EPFL (Math and Milan Polytechnic-Quarteroni)
4	Denmark	DTU, Chemical Engineering	23	Switzerland	Swiss Federal Institute of Technology in Zurich (ETHZ) and Hönggerberg (ETHH)
5	France	École Nationale Supérieure de Chimie de Paris (ENSCP)	24	Switzerland	IBM Zürich Research Laboratory (ZRL)
6	France	Institut Français du Pétrole (IFP), Rueil-Malmaison	25	Switzerland	Zürich Universität (Department of Physical Chemistry)
7	France	Institute of Fluid Mechanics (IMFT)	26	United Kingdom	University of Cambridge, Department of Applied Mathematics and Theoretical Physics (DAMTP)
8	France	Institut de Recherche en Informatique (IRIT)	27	United Kingdom	University of Cambridge, Theory of Condensed Matter Group (Cavendish Laboratory)
9	Germany	BASF AG, Ludwigshafen am Rhein	28	United Kingdom	University of Cambridge, Centre for Computational Chemistry
10	Germany	Fraunhofer Institute for the Mechanics of Materials (IWM), Freiburg	29	United Kingdom	University of Cambridge, Unilever Center
11	Germany	Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)	30	United Kingdom	Science Technology Facilities Council (STFC) (Daresbury Laboratory)
12	Germany	Technical University of Munich and LRZ	31	United Kingdom	Imperial College London, (Rolls Royce UTC/ Thomas Young Centre)
13	Germany	University of Karlsruhe Institute of Physical Chemistry	32	United Kingdom	Unilever Research Laboratory, Bebington
14	Germany	University of Stuttgart, Institute of Thermodynamics and Thermal Process Engineering	33	United Kingdom	University College (Chemistry)
15	Italy	Eni Tecnologie SpA	34	United Kingdom	University of Oxford (Biochemistry, Chemistry, Engineering Science - Fluid Mechanics, Physics)
16	Netherlands	Vrije Universiteit, Molecular Cell Physiology	35	United Kingdom	University of Oxford, Department of Biochemistry
17	Netherlands	Vrije Universiteit, Theoretical Chemistry	36	United Kingdom	University of Oxford, Department of Engineering Science
18	Spain	Universitat Autònoma de Barcelona, Institut de Ciència de Materials de Barcelona (ICMAB-CSIC)	37	United Kingdom	University of Oxford, Rudolf Peierls Centre for Theoretical Physics
19	Spain	CIMNE (International Center for Numerical Methods in Engineering)	38	United Kingdom	University of Oxford, Physical and Theoretical Chemistry Laboratory

In Chapter 3, Peter Cummings describes advances and applications of SBE&S in materials R&D outside the United States. Computational materials science and engineering is changing how new materials are discovered, developed, and applied. The convergence of simulation and experiment at the nanoscale is creating new opportunities, acceleration of materials selection into the design process through SBE&S is occurring in some industries, and full-physics simulations are a realistic goal with the increasing power of next generation computers.

In Chapter 4, Masanobu Shinozuka describe how SBE&S is being used to solve outstanding problems in sustainability and energy. Around the world, SBE&S is aiding in the recovery of untapped oil, the discovery and utilization of new energy sources such as solar, wind, nuclear and biofuels, and the predictive design of sustainable infrastructures. Simulation of multi-billion-agent systems for research in sustainable infrastructures is a realistic goal for petascale computing.

In the succeeding chapters, we present the panel's findings concerning key crosscutting issues that underlie progress and application of SBE&S to important problems in all domains of science and engineering, including the three thematic areas above. In Chapter 5, George Karniadakis highlights trends in new high performance architectures and algorithms driving advances in SBE&S around the world. Massively multicore chip architectures will provide the foundation for the next generation of computers from desktops to supercomputers, promising sustained performance of a petaflop for many scientific applications in just a few years, and exaflops beyond that. Petaflop and exaflop computers under development in the United States and in Japan will be based mainly on concurrency involving, respectively, hundreds of thousands and millions of processing elements with the number of cores per processor doubling every 18-24 months. This disruptive change in computing paradigms will require new compilers, languages, and parallelization schemes as well as smart, efficient algorithms for simplifying computations.

Martin Head-Gordon discusses issues in scientific software development in the United States and abroad in Chapter 6. Advances in software capabilities have revolutionized many aspects of engineering design and scientific research – ranging from computer-aided design in the automotive industry to the development of the chlorofluorocarbon replacements that have helped to begin reversing stratospheric ozone depletion. The significance of these end-use examples is why government support for software development is a good strategic investment – particularly when pursued consistently over the long term.

In Chapter 7, Abhijit Deshmukh describes advances in engineering simulations. Simulation and modeling are integral to every aspect of engineering. Examples of the use of simulation in engineering range from manufacturing process modeling, continuum models of bulk transformation processes, structural analysis, finite element models of deformation and failure modes, computational fluid dynamics for turbulence modeling, multi-physics models for engineered products, system dynamics models for kinematics and vibration analysis to modeling and analysis of civil infrastructures, network models for communication and transportation systems, enterprise and supply chain models, and simulation and gaming models for training, situation assessment, and education.

In Chapter 8 George Karniadakis discusses validation, verification and uncertainty quantification (UQ) for SBE&S. Despite its importance throughout SBE&S, the field of UQ is relatively underdeveloped. However, there is intense renewed interest in modeling and quantifying uncertainty, and in verification and validation (V&V) of large-scale simulations. While *verification* is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model, *validation* is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. Rapid advances in SBE&S capabilities that will allow larger and more complex simulations than ever before urgently require new approaches to V&V and UQ.

Peter Cummings discusses multiscale methods for SBE&S in Chapter 9. Multiscale methods research involves the development and application of techniques that allow modeling and simulation of phenomena across disparate time and length scales that, in many cases, span many orders of magnitude. True multiscale simulations, with both automated upscaling and adaptive downscaling, are still rare because there has been no broadly

successful general multiscale modeling simulation methodology developed to date. The need for such methodology, however, continues to grow, and holds the key to making SBE&S more broadly applicable in areas such as design, control, optimization, and abnormal situation modeling.

Sangtae Kim describes issues and opportunities for data-intensive applications of SBE&S, including visualization, in Chapter 10. The life sciences and particle physics communities are pushing the envelope of data-intensive science. As computational capabilities continue to grow, big data, visualization and dynamic data-driven simulations are crucial technology elements in the solution of problems ranging from global climate change to earthquake prediction to the production of transportation fuels from the last remaining oil fields. However, insufficient cyberinfrastructure investments at university campuses particularly in the United States are hampering progress.

Finally, in Chapter 11 Celeste Sagui discusses the urgent global need for educating and training the next generation of researchers and innovators in SBE&S. From university campuses to industrial employers and government research facilities, the difficulty in finding talent with breadth and depth in their discipline as well as competency in scientific computation was the number one issue for concern identified by this panel. New paradigms for computational science education are emerging and are highlighted in this chapter.

Appendix A contains biographies of the panelists and advisors. Appendixes B and C include trip reports for each of the sites visited in Asia and Europe during the international assessment. The survey questions for host organizations are listed in Appendix D. A bibliometric analysis of simulation research is presented in Appendix E. Finally, a glossary of terms is provided in Appendix F. Additional information and documentation for all phases of the WTEC International Assessment of R&D in Simulation-Based Engineering and Science are available on the WTEC website at <http://www.wtec.org/sbes>.

REFERENCES

- Atkins, D.M., National Science Foundation, 2006. NSF's cyberinfrastructure vision for 21st century discovery. Version 7.1. <http://nsf.gov>.
- Benioff, Marc. R. and Edward D. Lazowska, Co-chairs. 2005. *Computational science: Ensuring America's competitiveness*. President's Information Technology Advisory Committee (PITAC). http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf.
- Bramley, R., B. Char, D. Gannon, T. Hewett, C. Johnson, and J. Rice. 2000. Enabling technologies for computational science: Frameworks, middleware and environments. In *Proceedings of the Workshop on Scientific Knowledge, Information, and Computing*, E. Houstis, J. Rice, E. Gallopoulos, and R. Bramley, eds., Boston: Kluwer Academic Publishers, pp. 19-32.
- Challenges in High-End Computing. 2007. <http://www.epsrc.ac.uk/CMSWeb/Downloads/Other/ChallengesHEC.pdf>
- Committee on Science, Engineering, and Public Policy (COSEPUP) [a joint unit of the National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine]. 2007. *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology. Washington, DC: National Academies Press. Available online: http://www.nap.edu/catalog.php?record_id=11463#orgs.
- Council on Competitiveness. 2004. *Final Report from the High Performance Users Conference: Supercharging U. S. Innovation and Competitiveness*, Washington, D.C..
- Coveney, P. et al. 2006. U.K. Strategic framework for high-end computing.. <http://www.epsrc.ac.uk/CMSWeb/Downloads/Other/2006HECStrategicFramework.pdf>
- Department of Defense (DoD). 2002. Report on high performance computing for the national security community, http://WWW.HPCMO.HPC.mil/Htdocs/DOCUMENTS/04172003_hpc_report_unclass.pdf.
- Department of Defense. 2000. Defense Science Board, Report of the Defense Science Board Task Force on DoD supercomputing needs, <http://www.acq.osd.mil/dsb/reports/dodssupercomp.pdf>.
- Department of Energy. 2000. Office of Science, Scientific discovery through advanced computing.

- European Computational Science Forum of the European Science Foundation (ESF). 2007. *The Forward Look Initiative. European computational science: The Lincei Initiative: From computers to scientific excellence*. Information available online: <http://www.esf.org/activities/forward-looks/all-current-and-completed-forward-looks.html>.
- European Strategy Forum for Research Infrastructures (ESFRI). 2006. *European roadmap for research infrastructures: Report 2006*. Luxembourg: Office for Official Publications of the European Communities. Available online: ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf
- Executive Office of the President. 2004. Office of Science and Technology Policy, Federal plan for high-end computing: report of the High-End Computing Revitalization Task Force (HECRTF).
- Interagency Panel on Large Scale Computing in Science and Engineering (The Lax Report). 1982. *Report of the Panel on Large Scale Computing in Science and Engineering*. Department of Defense and the National Science Foundation, in cooperation with the Department of Energy and the National Aeronautics and Space Administration.
- Joy, W., and K. Kennedy, eds. 1999. President's Information Technology Advisory Committee (PITAC), *Report to the President: Information technology research: Investing in our future*. Available online: <http://www.nitrd.gov/pitac/report/>.
- Joseph, E., A. Snell, and C. G. Willard. 2004. Council on Competitiveness Study of U.S. HPC Users. White Paper. http://www.compete.org/pdf/HPC_Users_Survey.pdf.
- Kaufmann, N. J., C. G. Willard, E. Joseph, and D. S. Goldfarb. 2003. Worldwide high performance systems technical computing census. IDC Report No. 62303.
- Keyes, D., P. Colella, T.H. Dunning, Jr., and W.D. Gropp, eds. 2003. *A science-based case for large-scale simulation*, vol. 1. Washington, DC: Office of Science, U.S. Department of Energy (DOE). Available online: <http://www.pnl.gov/scales/>.
- . 2004. *A science-based case for large-scale simulation*, vol. 2. Washington, DC: DOE Office of Science. Available online: <http://www.pnl.gov/scales/>.
- National Research Council (NRC). 2005. Computer Science and Telecommunications Board, Getting up to speed: The future of supercomputing. National Academies Press, Washington, D.C.
- National Research Council (NRC), 2008. National Academy of Engineering Report of Committee on Integrated Computational Materials Engineering (ICME). National Academies Press, Washington, D.C.
- National Research Council (NRC), 2008. The Potential Impact of High-End Capability Computing on Four Illustrative Fields of Science and Engineering. National Academies Press, Washington, D.C.
- National Science Foundation. 1993. NSF Blue Ribbon Panel on High Performance Computing. *From desktop to teraflop: Exploiting the U.S. lead in high performance computing*.
- National Science Foundation. 1995. *Report of the task force on the future of the NSF supercomputer centers program*.
- National Science Foundation. 2003. *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue ribbon advisory panel on cyberinfrastructure, 2003*.
- Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science Foundation blue ribbon panel on simulation-based engineering science*. Arlington, VA: National Science Foundation. Available online: http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.
- President's Council of Advisors on Science and Technology. 2004. *Report to the President: Sustaining the nation's innovation ecosystems, information technology, manufacturing, and competitiveness*.
- Society for Industrial and Applied Mathematics (SIAM) Working Group on Computational Science and Engineering (CSE) Education. 2001. Graduate education in computational science and engineering. *SIAM Review* 43(1):163-177.
- Workshop and Conference on Grand Challenges Applications and Software Technology. 1993. Pittsburgh, Pennsylvania.

CHAPTER 2

LIFE SCIENCES AND MEDICINE

Linda Petzold

INTRODUCTION

There is an unprecedented opportunity in the next few decades to make major advances in the life sciences and medicine through simulation. The mapping of the genome, together with high-throughput experimental techniques, imaging techniques, and the power of today's algorithms and computers, have the imminent potential to revolutionize the biological and medical sciences.

Some of the most important questions in the fields of biology, chemistry, and medicine remain unsolved as a result of our limited understanding of the structure, behavior, and interaction of biologically significant molecules. Many of these questions could in principle be answered if it were possible to perform detailed atomic-level molecular dynamics (MD) simulations of sufficient accuracy and duration. Important biological phenomena that might otherwise be explored using such simulations, however, occur over time scales on the order of a millisecond—about three orders of magnitude beyond the reach of the longest current MD simulations. Moreover, it remains unclear, because current simulations are not long enough to tell, whether the molecular mechanics force fields on which such simulations are typically based would be sufficiently accurate to fully elucidate these phenomena or if more complex force fields would need to be employed.

Modeling and simulation have been in use for some time in the pharmaceutical industry. Initial applications have focused mainly on chemistry and molecular structure. However, new techniques for modeling variability in patient physiology, as well as new software platforms, lower-cost computational resources, and an explosion of available data, have fueled recent advances in “predictive biosimulation”—the dynamic simulation of biological systems. Predictive biosimulation has been used to support the development of new therapeutic drugs, meeting the pharmaceutical industry's need for new approaches to deal with the complexity of disease and the increased risk and expense of research and development.

Several factors are likely to accelerate adoption of biosimulation technology in the industry. The FDA, through its Critical Path Initiative, has recognized the importance of modeling and simulation and is starting to commit more resources toward making biosimulation part of the evaluation process for new therapies. Academic groups and companies are increasingly specializing in the field and building active collaborations with industry. Pharmaceutical companies are creating and expanding internal systems biology groups that incorporate biological modeling. The potential impacts for systems biology in the pharmaceutical industry are huge. An average of \$1.4 billion and 14 years are spent on the development of a successful drug (Bangs 2007). The failure rate for new drugs is greater than 90%. Meanwhile, annual spending on new drugs increases year after year, while drug approvals are decreasing. Modeling and simulation can assist in drug development through identifying targets of intervention; determining the relationships for dose, exposure, and response; and assessing development strategies and trial designs in populations. Using a population of virtual patients, clinical variability can be explored and risk analysis undertaken. Another big variability amenable to modeling and simulation is the

heterogeneity of causes and mechanisms of nearly all clinical diagnoses. The groups of patients who might benefit from a drug or who could be at risk may be able to be identified through these technologies.

Modeling of whole organs and complete physiological systems is becoming possible. Modeling of complete cardiovascular systems provides information and insight both for medical research and for the evaluation of surgical options. Modeling provides information both for cancer research and for laser surgery. Neuromuscular biomechanics seeks fundamental understanding of the mechanisms involved in the production of movement and is motivated by opportunities to improve treatments for cerebral palsy. Biomechanical simulation is used in surgical planning and design of physical therapies. Rapidly increasing amounts of data, combined with high-performance algorithms and computers, have made modeling the brain in full experimental detail a realistic goal for the foreseeable future.

Where are the computational bottlenecks? Many of the models are multiscale, requiring new algorithms. Of critical importance are the need to deal with potentially large amounts of heterogeneous data and the need to communicate the source, the uncertainties, and the conditions under which the data were obtained. As models become more and more complex, they will require high-performance computing. This work is highly multidisciplinary—it requires a team of researchers, each of whom is expert in one discipline but knowledgeable in others. U.S. graduate schools are currently producing relatively few such PhD graduates. In the following, we examine the status of the areas of molecular dynamics, systems biology, and biophysical modeling in more detail.

MOLECULAR DYNAMICS

Molecular dynamics simulations can be used to model the motions of molecular systems, including proteins, cell membranes, and DNA, at an atomic level of detail. A sufficiently long and accurate MD simulation could allow scientists and drug designers to visualize for the first time many critically important biochemical phenomena that cannot currently be observed in laboratory experiments, including the folding of proteins into their native three-dimensional structures, the structural changes that underlie protein function, and the interactions between two proteins or between a protein and a candidate drug molecule, as shown in Figure 2.1 (Brooks and Case 1993; Frenkel and Smit 2001; Karplus and McCammon 2002; Schlick et al. 1999). Such simulations could in principle answer some of the most important open questions in the fields of biology and chemistry, and they have the potential to make substantial contributions to the process of drug development (Shaw et al. 2007).

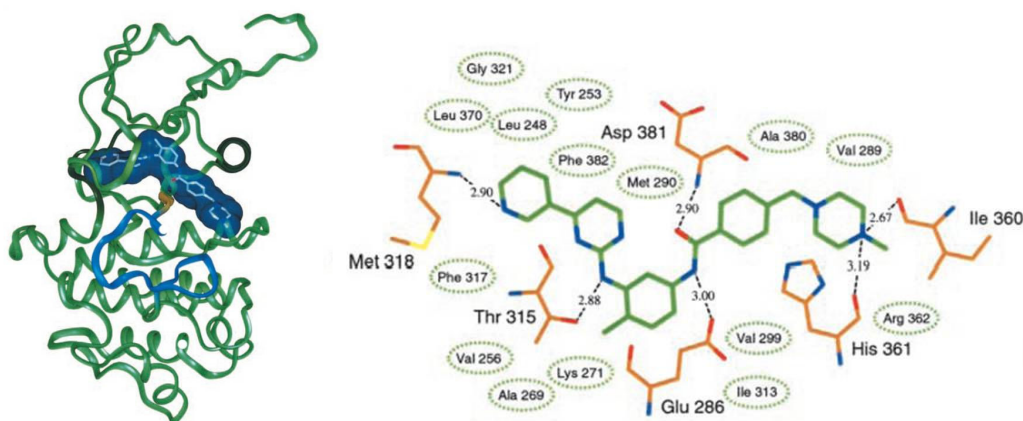


Figure 2.1. Binding of drugs to their molecular targets (Nagar et al. 2002).

A classical MD computation simulates the motion of a collection of atoms over a period of time according to the laws of Newtonian physics. The chemical system might consist of a protein and its surrounding solvent environment (which typically includes water and various ions), and might include other types of molecules such as lipids, carbohydrates, nucleic acids, or drug molecules. The entire chemical system occupies a small volume, typically tens of angstroms on a side, filled with tens or hundreds of thousands of atoms (Adcock and

McCammon 2006). The computation strains the resources of even the largest, most advanced computers; the vast majority of such MD simulations have been less than a microsecond in length, whereas many of the most important biological processes, such as protein folding, interactions between proteins, binding of drugs to their molecular targets, and the dynamics of conformational changes, often occur over time scales on the order of a millisecond.

To accomplish a millisecond simulation will require the use of a massive number of arithmetic processing units. Because of the global nature of the inter-atomic force calculations, this will necessitate a great deal of inter-processor communication, which must be very carefully managed in order to maintain efficiency. Further complicating the path to the millisecond timescale is the fact that current simulations are not long enough to tell whether commonly used force field approximations will be sufficiently accurate on longer timescales. Some known deficiencies of the most commonly used force fields include the omission of electrostatic polarization effects, the lack of explicitly represented hydrogen bonds, the use of simplified harmonic and trigonometric models to model the energetics of covalent bonds, and nonphysical repulsive Van der Waals terms. Some steps are already being taken to improve these approximations, but there may ultimately be a need to develop new biophysical theory.

There has been much recent progress in this area, and there is a huge opportunity for impact when simulation technologies reach the millisecond timescale. Current MD codes include AMBER (Case et al. 2005), CHARMM (Brooks et al. 1983), GROMACS (University of Groningen, 1993, The Netherlands, generally regarded as the fastest commonly used single-processor code,) and the parallel MD codes NAMD (University of Illinois, Urbana-Champaign, 1996, USA), LAMMPS (Sandia National Labs, 1999, USA), Blue Matter (IBM Research, 2002, USA), and Desmond (D. E. Shaw Research, 2006, USA). The greatest performance gains may ultimately be achieved using special-purpose MD machines, which have included the Delft MD Processor (Technische Universiteit Delft, 1988, The Netherlands), FASTRUN (Columbia University, 1990, USA), MD Engine (Fuji-Xerox, 1997, Japan), MD-GRAPE (University of Tokyo/RIKEN, 1996, Japan), and Anton (D. E. Shaw Research, 2008, USA). The recent arrival of GPGPU (general purpose graphics processing units) computing is likely to revolutionize MD simulation because the underlying algorithms in MD require fine-grained parallelism, which is ideally suited to these architectures. NAMD may be the first open-source MD code in use by the biosimulation community to exploit GPGPUs, as the code is currently being ported to NVIDIA processors using CUDA.

The United States has historically played a leadership role in a number of research areas that have driven the development of new algorithms, architectures, and applications for MD simulations. However, this role is now endangered by funding limitations at the National Institutes of Health (NIH), the National Science Foundation (NSF), the Defense Advanced Research Projects Agency (DARPA), and other federal agencies, and by the loss of PhD students and faculty to other disciplines and to the private sector.

SYSTEMS BIOLOGY

Systems biology has been defined by a previous WTEC study as “the understanding of network behavior, and in particular their dynamic aspects, which requires the utilization of mathematical modeling tightly linked to experiment” (Cassman et al. 2005, 1). Systems biology bridges molecular biology and physiology. At the core of this field is the focus on networks, where the goal is to understand the operation of systems, as contrasted with the reductionist tradition in biology of understanding the component parts. This involves a variety of approaches, including the identification and validation of networks, the creation of appropriate datasets, the development of tools for data acquisition and software development, and the use of modeling and simulation software in close linkage with experiment, often done to understand dynamic processes. The potential for systems biology is huge, both for advancing the understanding of fundamental biological processes, and for enabling predictive simulation in pharmaceutical research and development. Systems biology is particularly relevant to research in cancer and in diabetes, where it offers the potential of identifying targets from novel pathways. The WTEC panelists visited several institutions that are moving the systems biology field forward in significant ways.

Systems Biology Institute, Japan

One of the world's leading groups in systems biology is the Systems Biology Institute (SBI) in the Department of Systems Biology at the Japan Foundation for Cancer Research in Tokyo (see site report, Appendix C). This group is led by Dr. Hiroaki Kitano, who is often referred to as the “father of systems biology.” This group is well known for its software as well as for its research results. Its current focus is on the development of experimental data and software infrastructure. The software infrastructure includes Systems Biology Markup Language (SBML); Systems Biology Graphical Notation (SBN) (Kitano et al. 2005); CellDesigner (Funahashi et al. 2003); and Web 2.0 Biology, designed for the systematic accumulation of biological knowledge (see Figure 2.2). Development of systems biology as a field requires an extensive software infrastructure. Recognizing that it is difficult to publish software, the merit system in this lab values software contributions as well as publications. Biological systems under investigation include cancer robustness, type-2 diabetes, immunology, infectious diseases, metabolic oscillation, cell cycle robustness, and signaling network analysis. The experimental infrastructure under development includes the gTOW assay described below, microfluidics, and tracking microscopy.

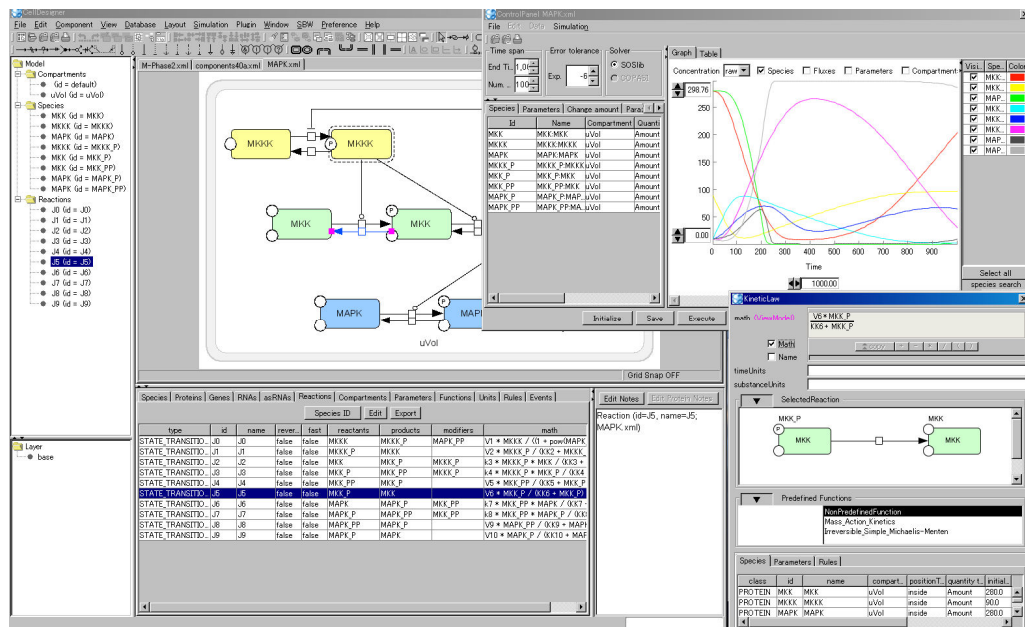


Figure 2.2. CellDesigner is a structured diagram editor for drawing gene-regulatory and biochemical networks. Networks are drawn based on the process diagram, with a graphical notation system proposed by Kitano, and are stored using the Systems Biology Markup Language (SBML), a standard for representing models of biochemical and gene-regulatory networks. Networks are able to link with simulation and other analysis packages through Systems Biology Workbench (SBW) (<http://www.systems-biology.org/cd/>).

Key issues in cellular architecture that are being investigated at the SBI include the implications of robustness vs. fragility on cellular architecture. Some systems have to be unstable to be robust (for example, cancer). Points of fragility in a system are important to the determination of drug effectiveness. SBI researchers use the robustness profile to reveal principles of cellular robustness, to refine computer models, and to find therapeutic targets. They use yeast (budding yeast and fission yeast) as the model organism and cell cycle as the model system. The models are currently ODE (ordinary differential equation) models. The limits of parameters are used as indicators of robustness; for example, how much can you increase or decrease parameters without disrupting the cell cycle? *This is a different type of uncertainty analysis than most of the current work in the United States.*

Experimental methods are required that can comprehensively and quantitatively measure parameter limits. The SBI has developed Genetic Tug of War (gTOW), an experimental method to measure cellular robustness. gTOW introduces extra copies of each gene to see how much it changes the cell cycle. There are implications of this technology for drug development.

Vrije University, The Netherlands

The theme of the systems biology center BioCentrum Amsterdam at Vrije Universiteit, led by Professor Hans Westerhoff, another worldwide leader in systems biology, is “to cure a disease, one must cure the network.” Professor Westerhoff heads a transnational research group on Systems Biology that includes the Manchester Centre for Integrative Systems Biology (MCISB) in the Manchester Interdisciplinary BioCentre (MIB) and the BioCentrum Amsterdam. He is also the Director of the University of Manchester Doctoral Training Centre in Integrative Systems Biology (supported by the UK’s Biotechnology and Biological Sciences Research Council [BBSRC] and the Engineering and Physical Sciences Research Council [EPSRC]). Here, a focus on the development of software infrastructure is also apparent.

BioCentrum Amsterdam is a driver behind the Silicon Cell program (Westerhoff 2001; <http://www.siliconcell.net>; see also site report in Appendix D). This is an initiative to construct computer replicas of parts of living cells from detailed experimental information. The Silicon Cells can then be used in analyzing functional consequences that are not directly evident from component data but arise through contextual dependent interactions. Another prominent effort at the center is in network-based drug design. The premise is that the recent decline in drug discovery may be due to the focus on single molecules, neglecting cellular, organellar, and whole-body responses. A specific research effort focuses on the treatment of African sleeping sickness using trypanosome glycolysis as a drug target. This has entailed development of an ODE model of trypanosome glycolysis and sensitivity analysis to identify those reactions that can best control the flux. Using the model, it was found that the network behavior was complicated by the fact that gene-expression response can either counteract or potentiate the primary inhibition.

Technical University of Denmark

The Center for Biological Sequence Analysis (CBS) at the Technical University of Denmark (see site report, Appendix D) is home to another world-class effort in systems biology. Over the last decade, the CBS has produced a large number of bioinformatics tools. The codes are very popular, as evidenced by the number of page-views to the CBS webpages (over 2 million a month). One of the major challenges in this field is the integration of data. The amount of data generated by biology is exploding. CBS has a relational data warehouse comprising 350+ different databases. In order to handle data integration of 120 terabyte size, CBS has developed its own integration tool. Ultimately, CBS researchers aim to develop data integration systems that can talk to one another. The implications for both scientific research and its culture are clearly expressed in the December 2005 *Nature* editorial “Let data speak to data”: “Web tools now allow data sharing and informal debate to take place alongside published papers. But to take full advantage, scientists must embrace a culture of sharing and rethink their vision of databases.”

The CBS group is moving towards other frontiers, the “disease interactomes.” This involves the use of text mining to relate proteins to diseases, or the identification of new disease gene candidates by the phenomic ranking of protein complexes (Lage et al. 2007). This new research leads to new types of “biobanks” with new computational challenges: (1) finding disease genes, and their “systemic” properties, in cases where the environment also plays a major role, and (2) extracting information from complex, messy “databases” and registries across countries. The combining of medical informatics with bioinformatics and systems biology represents a new trend in disease gene finding and phenotype association, which can also include social and behavioral levels. Linking medical informatics with bioinformatics and systems biology requires bridging the gap between the molecular level and the phenotypic clinical levels, and linking two exponentially growing types of computer-accessible data: biomolecular databases and their clinical counterparts. This necessitates development of a systems biology that recognizes a changing environment.

U.S. Systems Biology Efforts

Major efforts in systems biology exist also in the United States, with comparable impacts. Systems Biology Markup Language (Hucka et al. 2004), a key development that enables the sharing of models, was developed jointly by the SBI and the California Institute of Technology (Caltech). The Complex Pathway Simulator (COPASI) (Hoops et al. 2006), a software application for simulation and analysis of biochemical networks, is a collaborative project led by P. Mendes of Virginia Polytechnic Institute (Virginia Tech) and the University of Manchester (UK), and U. Kummer of Heidelberg University (Germany). Major modeling and experimental efforts exist at the Massachusetts Institute of Technology (MIT); the Institute for Systems Biology in Seattle; Caltech; the University of Washington; and the University of California's Berkeley, San Diego, and Santa Barbara campuses, among other U.S. universities. BioSPICE, an initiative that DARPA began in 2001 and ran for five years (Kumar and Feidler 2003; <http://biospice.lbl.gov/home.html>), was the impetus behind a tremendous surge in systems biology research and software development in the United States.

BIOPHYSICAL MODELING

Traditionally, physiology has been concerned with the integrative function of cells, organs, and whole organisms. As more detailed data has become available at the molecular level, it has become more difficult to relate whole organ function to the underlying detailed mechanisms that exploit this molecular knowledge. Organ and whole organism behavior needs to be understood at both a systems level and in terms of subcellular function and tissue properties.

Linking the vast amounts of information ranging from the genome to medical imaging poses many challenges for SBE&S. It requires the development of databases of information at all spatial and temporal scales (see Figure 2.3), the development of models and algorithms at a wide range of scales, the mathematical and software technology for linking the models so that fine-scale information is accessible by coarser models, and the effective use of petascale high-performance computers.

International Physiome Project

A great deal of work in this area has been taking place under the auspices of the Physiome Project (see <http://en.wikipedia.org/wiki/Physiome>). The Physiome Project is a worldwide effort focused on compiling laboratories into a single, self-consistent framework. Research initiatives related to Physiome include the Wellcome Trust Physiome Project, the IUPS (International Union of Physiological Sciences) Physiome Project, the EuroPhysiome Project, and the Physiome project of the NSR (National Simulation Resource) at the University of Washington Department of Bioengineering (Hunter, Robbins, and Noble 2002). The Life Science core focus of the Riken next-generation supercomputer center in Japan (see site report, Appendix C) center on the multiscale modeling of the virtual human, from genes to proteins, to cells, to tissues and organs.

In groundbreaking work associated with Physiome, a mathematical model of the heart has been developed (Smith et al. 2001) that encompasses the anatomy and cell and tissue properties of the heart and is capable of revealing the integrated physiological function of the electrical activation, mechanics, and metabolism of the heart under a variety of normal and pathological conditions. This model has been developed over the last twenty years as a collaborative effort by research groups at the University of Auckland, the University of Oxford, and the University of California, San Diego. A model of the lungs, including gas transport and exchange, pulmonary blood flow, and soft tissue mechanics, is being undertaken as a collaboration between universities in the UK, the United States, and New Zealand (Howatson, Pullan, and Hunter 2000).

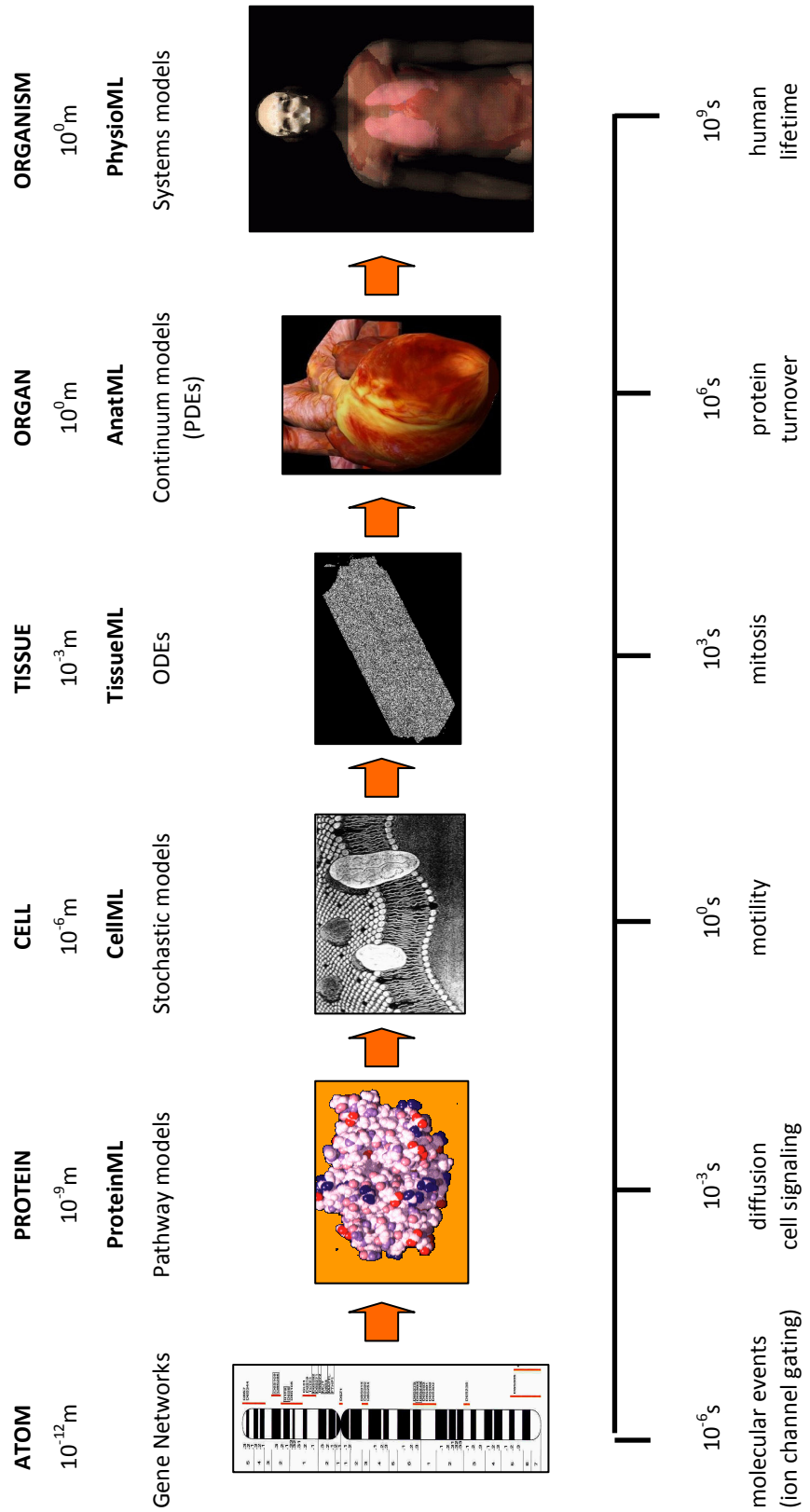


Figure 2.3. Spatial (top) and temporal (bottom) scales encompassed by the Human Physiome Project. Markup languages (PhysioML, AnatML, TissueML, CellML) are defined for each spatial level. The types of mathematical model appropriate to each spatial scale are also indicated. (Hunter et al. 2002; figure courtesy of P. Hunter).

EPFL Arterial Map

At Ecole Polytechnique Fédérale de Lausanne (EPFL), there is a project led by Professor A. Quarteroni for extracting geometry from raw data to get a 3D “map” of the arterial system (see site report, Appendix D). These reconstructions are then turned into a finite element structure that provides the basis for detailed simulations of blood flow. Particular attention has been paid to difficulties in attaching different length and time scales (the multiscale problem) to the cardiovascular system.

EPFL Blue Brain Project

Perhaps the ultimate frontier in biophysical modeling is to understand, in complete experimental detail, the structure and function of the brain. An impressive effort is underway in this area at the EPFL, in the Blue Brain project directed by neuroscientist Henry Markram. Named after the IBM Blue Gene supercomputer it relies on, the Blue Brain project (see site report, Appendix D) has the goal of modeling, in every detail, the cellular infrastructure and electrophysiological interactions within the cerebral neocortex, which represents about 80% of the brain and is believed to house cognitive functions such as language and conscious thought. The project started out to reverse-engineer the neocortex. This involved collecting experimental data for different types of cells, types of electrical behavior, and types of connectivity. The model includes 10,000 neurons, 340 types of neurons, 200 types of ion channels, and 30 million connections.

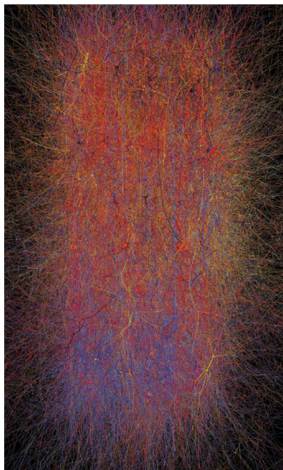


Figure 2.4. A visualization of the neocortical column model with 10,000 neurons. In false colors, the membrane voltage for all simulated compartments is shown, indicating the activity of the network. Image © 2009 by BBP/EPFL.

EPFL’s Blue Brain model is a detailed multiscale model, related to variables as measured in a laboratory. It is an enormous system of ordinary differential equations, modeling electrochemical synapses and ion channels. The main objective of the model is to capture electrical behavior. The model has been implemented in such a way that it is easy to modify and add to it. The current capability is a faithful *in silico* replica at the cellular level of a neocortical column of a young rat (Figure 2.4). A goal is to capture and explain diversity observed in the lab: electrical, morphological, synaptic, and plasticity. Building a database of experimental results is a major part of this effort. Some of the long-term objectives are to model a normal brain and a diseased brain to identify what is going wrong. Another objective is to evaluate emergent network phenomena. A demonstration was given in which the model replicated slice experiments that exhibit network-scale phenomena. In this way, it may someday be possible to do *in silico* pharmacology experiments.

The Blue Brain Project models 10,000 cells involving approximately 400 compartments per neuron with a cable-equation, Hodgkin-Huxley model. There are 30 million dynamic synapses. The computing platform uses a 4 rack IBM BlueGene/L with a 22.4 Tflop peak and 8192 processors. Additionally, an SGI Prism Extreme is used for visualization. The project features an impressive integration of experiment, modeling, high-performance computing, and visualization.

U.S. Biophysical Modeling Efforts

In the United States, biophysical modeling is being undertaken at the Stanford University Simbios Center for Vascular Biomechanics Research (Taylor and Draney 2004; <http://simbios.stanford.edu/index.html>), and at Brown University. Grinberg and colleagues (2008) review the issues involved in order to make detailed 3D simulations of blood flow in the human arterial tree feasible. Figure 2.5 illustrates the major vessels and bifurcations. A straightforward approach to the detailed computation is computationally prohibitive even on petaflop computers, so a three-level hierarchical approach based on the vessel size is utilized, consisting of a

Macrovascular Network (MaN), a Mesovascular Network (MeN), and a Microvascular Network (MiN). A multiscale simulation coupling MaN-MeN-MiN and running on hundreds of thousands of processors on petaflop computers is projected to be possible and will require no more than a few CPU hours per cardiac cycle, hence opening up the possibility of simulation studies of cardiovascular diseases, drug delivery, perfusion in the brain, and other pathologies.

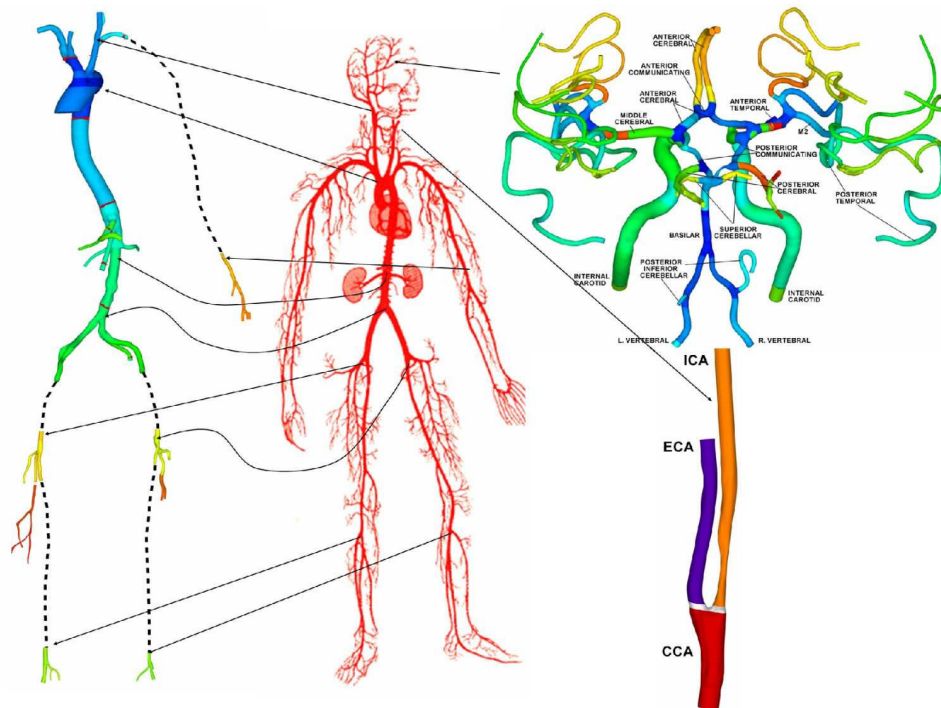


Figure 2.5. 3D model of major vessels and bifurcations of the human arterial tree, reconstructed at Brown University from a set of computed tomographic (CT), digital subtraction angiography (DSA) CT, and magnetic resonance angiography (MRA) images. Colors represent different parts of the model. Left: Aorta and adjacent arteries. Right top: cranial arterial network. Right bottom: carotid artery (Grinberg et al. 2008).

SUMMARY OF KEY FINDINGS

The role of SBE&S in life sciences and medicine is growing very rapidly and is critical to progress in many of these areas. For SBE&S to realize its potential impacts in life sciences and medicine will require:

- Large, focused multidisciplinary teams. Much of the research in this area requires iteration between modeling and experiment.
- Integrated, community-wide software infrastructure for dealing with massive amounts of data, and addressing issues of data provenance, heterogeneous data, analysis of data, and network inference from data.
- High-performance algorithms for multiscale simulation on a very large range of scales and complexities.

- High-performance computing and scalable algorithms for multicore architectures. In particular, petascale computing can finally enable molecular dynamics simulation of macromolecules on the biologically interesting millisecond timescale.
- Techniques for sensitivity and robustness analysis, uncertainty analysis, and model invalidation.
- Visualization techniques, applicable to massive amounts of data, that can illustrate and uncover relationships between data.
- Appropriately trained students who are conversant in both the life sciences and SBE&S and can work in multidisciplinary teams. This was mentioned by hosts at virtually all the WTEC panelists' site visits as the biggest bottleneck to progress in the field.

What is the status of SBE&S capabilities in life sciences and medicine worldwide?

- WTEC panelists are aware of world-class research in SBE&S in molecular dynamics, systems biology, and biophysical modeling in the United States, Europe, and Japan, and—in some specific areas, for example on increasing the potency of natural medicines—in China.
- The quality of leading researchers in the United States is comparable to that of leading researchers in Europe and Japan.
- Infrastructure (access to computing resources and software professionals) and funding models to support ambitious, visionary, long-term research projects are much better in Europe and Japan than in the United States.
- Funding models and infrastructure to support multi-investigator collaborations between academia and industry are much more developed in Europe than in the United States.
- Support for the development of community software is much stronger in Europe and Japan than in the United States in recent years, following the decline in DARPA funding in this area.

REFERENCES

- Adcock, S.A., and J.A. McCammon. 2006. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* 106:1589–1615.
- Bangs, A. 2007. *Predictive biosimulation in pharmaceutical R&D*. Presentation at the WTEC Workshop, Review of Simulation-Based Engineering and Science Research in the United States, Arlington, VA, 1–2 November.
- Brooks, B.R., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, 1983. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* 4:187–217.
- Brooks, C.I. and D.A. Case 1993. Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.* 93:2487–2502.
- Case, D., T. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. Woods. 2005. The Amber biomolecular simulation programs. *J. Comp. Chem.* 26:1668–1688.
- Cassman, M., A. Arkin, F. Doyle, F. Katagiri, D. Lauffenburger, and C. Stokes. 2005. *WTEC panel report on international research and development in systems biology*. Baltimore, MD: World Technology Evaluation Center, Inc.
- Frenkel, D., and B. Smit. 2001. *Understanding molecular simulation; From algorithms to applications*, 2nd ed. London: Academic Press.
- Funahashi, A., M. Morohashi, H. Kitano, and N. Tanimura. 2003. CellDesigner: A process diagram editor for gene-regulatory and biochemical networks. *Biosilico* 1(5):159–162.
- Grinberg, L., T. Anor, J. Madsen, A. Yakhot, and G.E. Karniadakis. 2008. Large-scale simulation of the human arterial tree. *Clinical and Experimental Pharmacology and Physiology*, to appear.
- Hoops, S., S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. 2006. COPASI — a Complex Pathway Simulator. *Bioinformatics* 22:3067–3074.
- Howatson, M., A.J. Pullan, and P.J. Hunter. 2000. Generation of an anatomically based three-dimensional model of the conducting airways. *Ann. Biomed. Eng.* 28:793–802.

- Hucka, M., A. Finney, B.J. Bornstein, S.M. Keating, B.E. Shapiro, J. Matthews, B.L. Kovitz, M.J. Schilstra, A. Funahashi, J.C. Doyle, and H. Kitano. 2004. Evolving a lingua franca and associated software infrastructure for computational systems biology: The Systems Biology Markup Language (SBML) project. *Systems Biology* 1(1):41–53.
- Hunter, P., P. Robbins, and D. Noble. 2002. The IUPS human physiome project. *Eur. J. Physiol* 445:1–9.
- Karplus, M., and J.A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* 9:646–652.
- Kitano, H., A. Funahashi, Y. Matsuoka, and K. Oda. 2005. Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology* 23: 961–966.
- Kumar, S.P., and J.C. Feidler. 2003. BioSPICE: A computational infrastructure for integrative biology. *OMICS: A Journal of Integrative Biology*. 7(3):225–225.
- Lage, K., E.O. Karlberg, Z.M. Storling, P.I. Olason, A.G. Pedersen, O. Rigina, A.M. Hinsby, Z. Tumer, F. Pociot, N. Tommerup, Y. Moreau, and S. Brunak 2007. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology* 25:309–316.
- Lehrer, J. 2008. Out of the blue. *Seed Magazine* March 3. Available online: http://www.seedmagazine.com/news/2008/03/out_of_the_blue.php.
- Nagar, B., W. Bornmann, P. Pellicena, T. Schindler, D. Veach, W.T. Miller, B. Cloarkson, and J. Kuriyan. 2002. Crystal structures of the kinase domain of c-Abl in complex with the small molecular inhibitors PD173955 and Imatinib (STI-571). *Cancer Research* 62:4236–4243.
- Nature editorial. 2005. “Let data speak to data.” *Nature* 438:531.
- Schlick, T., R.D. Skeel, A.T. Brunger, I.V. Kale, J.A. Board, J. Hermans. and K. Schulte. 1999. Algorithmic challenges in computational molecular biophysics. *J. Comput. Phys.* 151(1):9–48.
- Shaw, D.E., M. Deneroff, R. Dror, J. Kuskin, R. Larson, J. Salmon, C. Young, B. Batson, K. Bowers, J. Chao, M. Eastwood, J. Gagliardo, J. Grossman, C. R. Ho, D. Ierardi, I. Kolossvary, J. Klepeis, T. Layman, C. McLeavey, M. Moraes, R. Mueller, E. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. Wang. 2007. “Anton, a special-purpose machine for molecular dynamics simulation.” In *Proceedings of the 34th Annual International Symposium on Computer Architecture* (ACM ISCA 07), June 09–13, 2007. New York, NY: ACM. 1–12.
- Smith, N.P., P.J. Mulquiney, M.P. Nash, C.P. Bradley, D.P. Nickerson, and P.J. Hunter. 2001. Mathematical modeling of the heart: Cell to organ. *Chaos, Solitons and Fractals* 13:1613–1621.
- Taylor, C.A., and M.T. Draney. 2004. Experimental and computational methods in cardiovascular fluid mechanics. *Annual Review of Fluid Mechanics* 36: 197–231.
- Westerhoff, H. 2001. The silicon cell, not dead but live! *Metabolic Engineering* 3(3):207–210.

CHAPTER 3

MATERIALS SIMULATION

Peter T. Cummings

INTRODUCTION

The discovery and adoption of new materials has, in extreme cases, been revolutionary, changed lives and the course of civilization, and led to whole new industries. Even our archeological classifications of prehistory (the stone, bronze, and iron ages) are named by the predominant materials of the day; the evolution of these materials (and their use in tools) is related to the predominant dwellings and societies of these ages (caves, farmsteads, and cities, respectively). By analogy, we might refer to recent history (the last half-century) as the plastics age followed by the silicon age. Today, we recognize a number of classes of materials, including but not limited to

- smart/functional materials (whose physical and chemical properties are sensitive to, and dependent on, changes in environment, such as temperature, pressure, adsorbed gas molecules, stress, and pH)
- structural materials
- electronic materials
- nanostructured materials
- biomaterials (materials that are biological in origin, such as bone, or are biomimetic and substitute for naturally occurring materials, such as materials for the replacement and/or repair of teeth)

Materials today have ubiquitous application in many industries (chemicals, construction, medicine, automotive, and aerospace, to name a few). Clearly, materials play a central role in our lives, and have done so since antiquity.

Given the importance of materials, understanding and predicting the properties of materials, either intrinsically or in structures, have long been goals of the scientific community. One of the key methodologies for this, materials simulation, encompasses a wide range of computational activities. Broadly defined, materials simulation is the computational solution to a set of equations that describe the physical and chemical processes of interest in a material. At the most fundamental level (often referred to as “full physics”), the equation to be solved is the Schrödinger equation (or its equivalent). If quantum degrees of freedom are unimportant, and atoms/molecules can be assumed to be in their ground state, then the forces between atoms/molecules can be described by simple functions (called forcefields), and the equations to be solved are Newton’s equations or variants thereof (yielding molecular dynamics simulations). At longer time scales and larger spatial scales, mesoscale simulation methods such as dissipative particle dynamics (DPD) become the method of choice, while at scales near and at the macroscale, finite-element methods solving energy/mass/momentum partial differential equations are appropriate.

Full-physics materials simulations, when performed even on modest-scale structures (e.g., nanoparticles with several thousand atoms), have the ability to saturate with ease the largest available supercomputers (the so-called leadership-class computing platforms). Thus, full-physics (or first-principles) materials simulation of nanoscale

and larger structures are among the applications driving the current progression toward petascale, to be followed by exascale, computing platforms.

In practice, materials simulators rarely have enough compute cycles available, even on petascale computers, to address at their desired level of accuracy the complex materials modeling problems that interest them. As a consequence, materials simulations almost always involve some tradeoff between accuracy (reflected in the methodology), size (number of atoms, or electrons, in the case of *ab initio* calculations), and duration. For example, first-principles methods, such as density functional theory (DFT), and molecular-orbital-based methods, such as Hartree-Fock (HF) and Møller-Plesset (MP) perturbation theory, range in their system-size dependence from N^3 to N^7 , where N is the number of nuclei or outer-shell electrons. Computation times are clearly linear in the number of time steps simulated (or total simulated time); however, the number of time steps needed to bring a system to equilibrium can vary nonlinearly with the size of the system (e.g., structural relaxation times of linear polymers vary as the chain length squared). Thus, it is clear that materials simulations can rapidly escalate into computational grand challenges.

From among the available methods, the materials simulator must select the appropriate theory/method for given phenomenon and material; must be aware of the approximations made and how they effect outcome (error estimates); and ensure that the chosen methodology gives the right answer for the right reason (correct physics, correct phenomenon) (Carter 2007). The wide range of methods available—even at a single level of description, such as the *ab initio* (also referred to as first-principles or full-physics) level—combined with limits on their applicability to different problems makes choosing the right theory a significant challenge. Additionally, for the novice user, or for the applied user interested primarily in pragmatically performing the most accurate and reliable calculation within available resources, there is little help available to navigate the minefield of available methods.

The situation is complicated even further by a lack of consensus or communal approach in the U.S. quantum chemistry and materials modeling communities. Unlike some other computational science communities, such as the climate modeling community, that have recognized the value of coming together to develop codes collaboratively and have coalesced around a limited number of community codes, the U.S. quantum chemistry and materials modeling communities remain divided, with many groups developing their own in-house codes. Such closed codes are inherently difficult to validate. There are exceptions. For example, in the biological simulation community, several codes—AMBER (see Amber Online), CHARMM (see Charmm Online), and NAMD (see NAMD Online)—have emerged with significant funding from the National Institutes of Health (NIH) as community codes. The complexity of biological simulations has almost made this a necessity. The biological simulation community understands the strengths and weaknesses of the forcefields used in these codes, and the accuracy of the codes (for a given force field) has been subject to numerous consistency checks. The differences between these biological simulation codes often reduces to issues such as parallel performance and the availability of forcefield parameters for the system of interest.

This chapter addresses some of the issues and trends related to materials simulation identified during the WTEC panel's international assessment site visits. It begins with a brief discussion of the state of the art, then provides examples of materials simulation work viewed during the panel's international visits, and concludes with the findings of the panel related to materials simulation.

CURRENT STATE OF THE ART IN MATERIALS SIMULATION

Materials simulation is an exceptionally broad field. To fully review the current state of the field is beyond the scope of this chapter; instead it provides references to recent reviews. For first principles methods, excellent overviews are provided by Carter (Carter 2007; Huang and Carter 2008); other aspects of hard materials modeling are reviewed by Chatterjee and Vlachos (2007) and Phillpot et al. (Phillpot, Sinnott, and Asthagiri 2007). Reviews of modeling methods for soft materials (i.e., polymers and polymer composites) are provided by Glotzer and Paul (2002) Theodorou (2007), Likos (2006), and Zeng and colleagues (2008).

Hattel (2008) reviews combining process-level simulation with materials simulation methods, where applicable, for engineering design. Integration of materials simulation methods into design, termed integrated computational materials engineering (ICME), is the subject of a 2008 report from the National Academy of Science and the National Academy of Engineering (NRC 2008) that documents some of the achievements of ICME in industry and the prospects for the future of this field. Through documented examples, the report makes the case that investment in ICME has a 3:1 to 9:1 return in reduced design, manufacturing, and operating costs as a result of improved product development and faster process development. As just one example, the Virtual Aluminum Casting methodology (Allison et al. 2006) developed at Ford Motor Company resulted in a 15–25% reduction in product development time, substantial reduction in product development costs, and optimized products, yielding a return of investment greater than 7:1.

Clearly, materials simulation methods, particularly when integrated with engineering design tools, can have significant economic impact. Equally significant is the role of materials simulation in national security; for example, materials simulation lies at the heart of the National Nuclear Security Administration's (NNSA's) mission to extend the lifetime of nuclear weapons in the U.S nuclear stockpile. Through ICME modeling on some of the world's largest supercomputers, the NNSA Advanced Simulation and Computing (ASC) program helps NNSA to meet nuclear weapons assessment and certification requirements through computational surrogates for nuclear testing and prediction of weapon performance.

MATERIALS SIMULATION CODE DEVELOPMENT

Given the potential value of materials simulation, both for fundamental science (understanding the molecular origin of properties and dynamics) and for applications (as evidenced by the 2008 NRC report), and the ever-increasing complexity of systems that scientists and engineers wish to understand and model, development of materials simulation methods and tools is an activity engaged in worldwide. As noted in the introduction, within the United States (apart from notable exceptions described below) methods/tools development is largely individual-group-based. There are a number of factors that could account for this: (1) the tenure and promotion system in U.S. universities encourages individual achievement over collaboration during the first half-decade of a faculty member's career, thus establishing for many faculty a career-long focus on developing codes in competition with other groups; (2) funding agencies have tended not to fund code development activities *per se*, but rather to fund the solution of problems, a by-product of which may be a new method, tool, or code; (3) a corollary of this is that the person-power needed to perfect, distribute, and maintain tools/codes is generally not supported by funding agencies; and (4) within the United States, turning an in-house academic code into a commercial products is regarded as a successful outcome. Academic codes that have been commercialized are the original sources of almost all of the commercial materials and modeling codes available today. One might argue that the lack of federal agency funding for code development encourages an academic group to commercialize its code to create a revenue stream to fund distribution, support, and further development.

By contrast, in Japan and in Europe, there have been a number of community-based codes developed, many of which end up as open-source projects. In Japan, the multiscale polymer modeling code OCTA (<http://octa.jp>) was the result of a four-year multimillion-dollar government-industry collaboration to produce a set of extensible public-domain tools to model polymeric systems over many scales. In the UK during the 1970s, the predecessor of the Engineering and Physical Sciences Research Council (EPSRC) established Collaborative Computational Projects (CCPs, <http://www.ccp.ac.uk/>) with headquarters at Daresbury Laboratory to develop community-based codes for a wide range of simulation and physical sciences modeling applications (Allan 1999). Among the codes developed by the CCPs are the atomistic simulation package DL_POLY (developed within CCP5), the electronic structure code GAMESS-UK (developed within CCP1), and the ab initio surface-properties-calculation program CASTEP (developed within CCP3). CASTEP is now marketed commercially worldwide by Accelrys (see Accelrys Online); thus, U.S. academic researchers must pay for using CASTEP, whereas in the UK, academic users have free access to CASTEP as a result of the role of CPP3 in its development. Other codes developed in European academic and/or government laboratory groups that are routinely used in the United States for materials modeling include VASP (see Vasp Online), CPMD (see Cpmd Consortium Online), SIESTA (see Siesta Online), and ESPRESSO (see Quantum Espresso Online). The development of community-based codes is evidently one area in which the United States trails other countries, particularly Japan and the EU.

Although fewer in number, there have been examples of codes developed within the United States, but in general, the circumstances under which they have been developed are unusual. For example, the development of the quantum chemistry program NWCHEM (Kendall et al. 2000) was funded innovatively as part of the *construction* budget for the Environmental Molecular Science Laboratory (EMSL) at the Department of Energy's (DOE's) Pacific Northwest National Laboratory (PNNL). Maintaining the support for NWCHEM beyond the EMSL construction phase has proved challenging. The popular LAMMPS molecular dynamics (MD) code (Plimpton 1995) was developed at Sandia National Laboratory (SNL) under a cooperative research and development agreement (CRADA) between DOE and Cray Research to develop codes that could take advantage of the Cray parallel computers of the mid-1990s; as the code is no longer supported by the CRADA or any other federal funding, its developer, Steve Plimpton, has placed it in the public domain under the GNU public license (<http://www.gnu.org>), and there is now a reasonably active open-source development project for LAMMPS. Part of the attraction of LAMMPS is its high efficiency on large parallel machines; additionally, the existence of converters from AMBER to/from LAMMPS and CHARMM to/from LAMMPS mean that biological simulations can be performed efficiently on large machines using LAMMPS as the MD engine but making use of the many analysis tools available for AMBER and CHARMM. These are exceptions rather than the rule, and in each case there are unique circumstances surrounding the development of the code.

The United States is headquarters to several of the commercial companies in materials modeling, such as Accelrys, Schrodinger, Spartan, and Gaussian; however, the viability of commercial ventures for chemicals and materials modeling has long been a subject of debate, largely driven by questions about whether the worldwide commercial market for chemicals and materials modeling is currently large enough to sustain multiple commercial ventures.

To understand how the different models for developing codes have evolved in the United States on the one hand, and Japan and Europe on the other, the WTEC panel made this subject one of the focuses of its site visits. One of the clear differences that emerged is funding. Japanese and European agencies actively fund large collaborative projects that have as their focus and outcome new simulation methodologies and codes. A secondary factor may be the tenure and promotion system in Japan and Europe, where faculty members are normally not subject to a long tenure tracks and often join the department and group in which they obtained their PhDs. While philosophically the WTEC panelists believe that the U.S. system, which makes such inward-focused hiring practices almost impossible, is superior in general (and is increasingly being adopted outside the United States), it is nevertheless evident that positive side effects of the Japanese/European system are continuity and maintaining of focus over the many years needed to develop a complex materials simulation capability.

MATERIALS SIMULATION HIGHLIGHTS

The WTEC study panel visited a number of sites in Asia and Europe conducting materials simulations research, some of which are highlighted below. These highlights are not intended to be comprehensive—the panel visited more materials simulation sites than those listed below—nor is inclusion in this report meant to suggest that these sites conducted higher-quality research than other sites visited; rather, the intent is to choose a selection of sites that are representative of the range of activities observed during the WTEC study.

Mitsubishi Chemical

The Mitsubishi Chemical Group Science and Technology Research Center (MCRC), headquartered in Yokohama, Japan, is the corporate research center for Mitsubishi Chemical Corporation (MCC) one of three subsidiaries of Mitsubishi Chemical Holdings Company (MCHC). The majority of the research conducted by MCRC (~90%) is focused on healthcare and performance materials businesses.

MCRC staff use a variety of tools to design, construct, and operate optimal chemical and materials manufacturing processes. Approximately 90% of the tools used are commercial software packages (Aspen Tech and gPROMS for process-level modeling, STAR-CD and FLUENT for computational fluid dynamics). The panel's hosts stated that MCC management has become comfortable with relying on predictions of simulation studies for design and optimization of chemical and materials manufacturing processes. In order to achieve its design goals, MCRC

needs to couple these different codes together to create a multiscale model of an intended process. It achieves this goal by writing its own in-house codes to provide coupling between the commercial codes. MCRC supports an internal molecular modeling effort on polymers (illustrated in Figure 3.1), which is linked to the process modeling efforts. For equilibrium properties (thermodynamics and structure), MCRC researchers have developed a reliable capability to relate molecular structure and properties, based on a combination of the molecular self-consistent polymer reference interaction site model (SC-PRISM) and self-consistent field (SCF) theories.

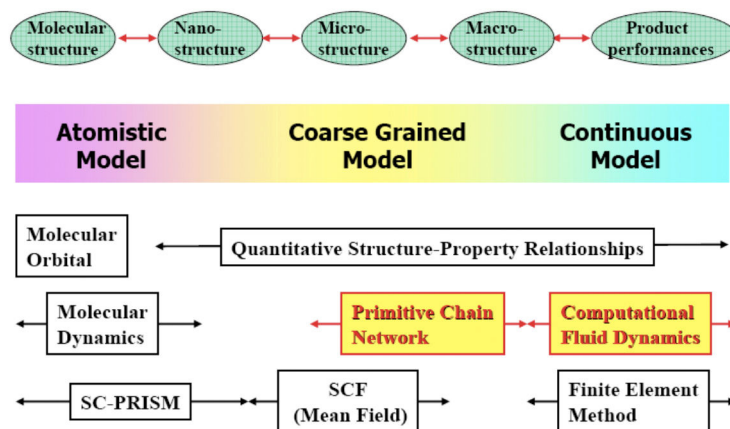


Figure 3.1. Mitsubishi multiscale modeling approach for polymers.

Of particular interest from a materials simulation point of view is the 17-member Computational Science Laboratory (CSL) headed by Shinichiro Nakamura. Using CSL-built in-house clusters of various sizes, the largest containing 800 CPUs, and the TSUBAME machine (<http://www.gsic.titech.ac.jp>), the CSL group uses *ab initio* and other methods for product design and solving process problems. In recent years, they have designed a robust (non-photodegrading) dye for use in printing (Kobayashi et al. 2007), designed a high-quantum-yield yttrium oxysulfide phosphor for use in television sets (Mikami and Oshiyama 1998; Mikami and Oshiyama 1999; Mikami and Oshiyama 2000; Mikami et al. 2002), found an effective additive to improve the performance of Li-ion batteries (Wang et al. 2001; Wang et al. 2002) and developed a new methodology for protein nuclear magnetic resonance (NMR) (Gao et al. 2007) based on calculating the NMR shift using the fragment molecular orbital (FMO) methodology developed by Kazuo Kitaura for large-scale *ab initio* calculations. CSL researchers have contributed to community-based *ab initio* codes (Gonze et al. 2002). The CSL performs fundamental, publishable research, as well as proprietary research that directly benefits MCC and MCHC.

Toyota Central R&D Labs, Inc.

Established in 1960 to carry out basic research for Toyota Group companies, the Toyota Central R&D Labs (TCRDL) performs a wide range of modeling activities. In the materials area, TCRDL is developing functional materials, including metallic composites, organic-inorganic molecular composites, and high-performance cell materials, as well as developing plastics and rubber recycling technologies. Funding for these activities is drawn from the Toyota Group corporations, and projects are selected from the technical needs of the Toyota Group and from proposals developed by the TCRDL staff. There are significant simulation efforts at TCRDL focused on the chemistry of combustion in internal combustion engines, on the multiscale modeling of fuel cells, on the performance of catalysts in oxidizing carbon monoxide, and on predicting the viscosity of synthetic lubricants. Most of the simulation work is performed on stand-alone workstations. However, if more computing power is needed, TCRDL has an NEC SX-5 on site and has mechanisms in place to buy computer time at the Earth Simulator. In terms of CPU time, TCRDL spends the most time in performing materials science simulations to uncover material structure-property relationships, followed in decreasing order by time spent on structural simulations of vibrations and acoustics, magneto-electric simulations in support of LIDAR (Light Detection and Ranging) applications, and CFD/combustion simulations of internal combustion engines.

One particularly noteworthy project is TCRDL's effort to predict the viscosity of synthetic lubricants. Molecular dynamics simulations of the viscosity of known molecules with small simulation volumes could reproduce the trends observed in the laboratory, but the absolute value of the viscosity predicted by the initial simulation was incorrect by orders of magnitude because of the difference in the film thicknesses and shear rates of the experiment versus the simulation. Large-scale molecular dynamics simulations were then conducted on the Institute for Molecular Science SR11000 in which the film thickness and shear rate in the simulation was comparable to the experimental measurements, and the two came into agreement.

TCRDL researchers propose to address in the future two computational materials grand challenges: to perform first-principles molecular dynamics simulations of fuel cells to understand the mechanisms of hydrogen atom transport within those systems, and to understand the nature of carbon-oxygen bonding and resonances in the vicinity of catalyzing surfaces.

Joint Laboratory of Polymer Science and Materials, Institute of Chemistry, Chinese Academy of Sciences

Founded in 2003 by Charles Han, formerly of the U.S. National Institute of Standards and Technology (NIST), the Joint Laboratory of Polymer Science and Materials (JLPSM) of the Institute of Chemistry, Chinese Academy of Sciences (ICCAS), has as its goal to unify the experimental, theoretical, and computational spectrum of polymer science and technology activities ranging from atomistic scale, to mesoscale materials characterization, to macroscale applied polymer rheology and processing. Between 1998 and 2004, the labs comprising the JLPSM published 1038 papers, were granted 121 patents, received four China science prizes, and presented 73 invited lectures at international conferences. For the JPPSM, 60–70% of student funding comes from the National Natural Science Foundation of China (the Chinese equivalent of the U.S. National Science Foundation), with almost none coming from the ICCAS. This is a different model than that used in the past. Increased funding is leading to 20–30% growth per year in students and faculty. Examples of materials simulations at the JLPSM include

- molecular dynamics simulations of polyelectrolytes in a poor solvent (Liao, Dobrynin, and Rubinstein 2003a and 2003b)
- free-energy calculations on effects of confinement on the order-disorder transition in diblock copolymer melts
- organic electronic and photonic materials via *ab initio* computations using VASP, charge mobility for organic semiconductors (Wang et al. 2007)
- organic and polymer light-emitting diodes (OLEDs and PLEDs) (Peng, Yi, and Shuai 2007)
- soft matter at multiple length scales, including molecular dynamics and coarse-grained models
- molecular dynamics-Lattice Boltzmann hybrid algorithm, microscopic, and mesoscale simulations of polymers
- theory of polymer crystallization and self-assembly

Several of the JLPSM researchers develop their own codes and in some cases have distributed them to other research groups.

One notable large activity is overseen by Charles Han. The project focuses on integrated theory/simulation/experiment for multiscale studies of condensed phase polymer materials. It is funded with 8 million RMB⁴ over four years, and has 20 investigators, with roughly 100,000 RMB per investigator. This is a unique and challenging program, inspired by the Doi project (<http://octa.jp>) in Japan, but it goes substantially beyond that project by closely integrating experiment, both as a validation tool and to supply needed input and data for models.

⁴ At the time of writing, \$US 1 (USD) is approximately 7 Chinese yuan (RMB). RMB is an abbreviation for renminbi, the name given to the Chinese currency, for which the principal unit is the yuan.

Materials Simulation Code Development in the UK

The UK has a relatively long tradition of developing materials simulation codes. Of particular note are the Collaborative Computational Projects (CCPs, <http://www.ccp.ac.uk>) hosted at Daresbury Laboratory (see Daresbury Laboratory site report for a full listing of the CCPs). The CCPs consist of Daresbury staff along with university collaborators whose joint goal is the development and dissemination of codes and training for their use. The majority, but not all, of the CCPs are involved in materials simulation. “CCP5”—the CCP for Computer Simulation of Condensed Phases—has produced the code DL_POLY, a popular molecular dynamics simulation package. Other codes produced or maintained by the CCPs are GAMESS-UK (a general purpose *ab initio* molecular electronic structure program, <http://www.cfs.dl.ac.uk/>), CASTEP (a density-functional-theory-based code to calculate total energies, forces optimum geometries, band structures, optical spectra, and phonon spectra, <http://www.castep.org>), and CRYSTAL (a Hartree-Fock and density-functional-theory-based code for computing energies, spectra and optimized geometries of solids, <http://www.crystal.unito.it>). CASTEP can also be used for molecular dynamics with forces calculated from density functional theory (DFT). Recent changes in the funding landscape in the UK has resulted in the CCPs being responsible primarily for code maintenance and distribution; they no longer have the resources to support significant new code development.

CASTEP (Cambridge Sequential Total Energy Package) was primarily developed by Mike Payne of Cambridge University. With coworkers Peter Haynes, Chris-Kriton Skylaris, and Arash A. Mostofi, Payne is currently developing a new code, ONETEP (order-N electronic total energy package, <http://www.onetep.soton.ac.uk>) (Haynes et al. 2006). ONETEP achieves order-N scaling (i.e., computational cost that scales linearly, $O(N)$, in the number of atoms), whereas DFT methods typically scale as the cube of the system size, $O(N^3)$, and molecular orbital methods as $O(N^5) - O(N^7)$. Examples are provided by Skylaris (Skylaris et al. 2008), and include the application to DNA shown in Figure 3.2.

As valuable as the CCPs have been, there are concerns for the future. The Cambridge researchers expressed a concern echoed by others, namely that of the declining programming skills of incoming graduate students. This is being addressed to some degree by the educational programs at centers for scientific computing established at the universities of Warwick (<http://www2.warwick.ac.uk>) and Edinburgh (<http://www.epcc.ed.ac.uk>), and workshops hosted by the CCPs at Daresbury Laboratory. Additionally, there was considerable anecdotal evidence that the UK’s Engineering and Physical Sciences Research Council (<http://www.epsrc.ac.uk>) is funding the CCPs primarily for code maintenance and training, not for code development. A long-term problem for the development of materials modeling codes is that the expectation of funding agencies and university administrators for “sound-bite science,” publishable in the highest-profile journals. Software engineering is incremental by nature, and these “sound-bite science” expectations mitigate against materials software engineering receiving its due credit. Finally, it is too early to tell whether the reorganization of UKs science laboratories under the new Science and Technology Facilities Council (<http://www.scitech.ac.uk/>) will be beneficial for the development of materials modeling codes.

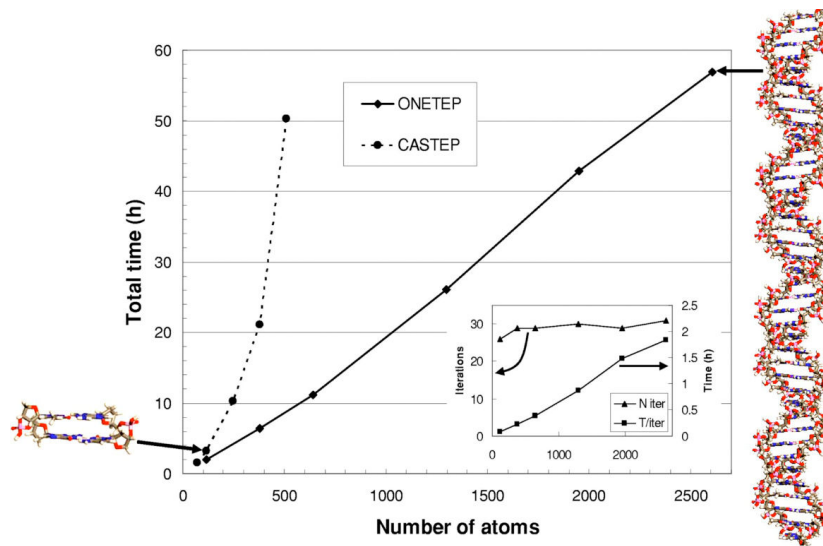


Figure 3.2. Linear scaling of ONETEP total energy calculation on DNA compared to CASTEP. Total time for the calculation is plotted against number of atoms. Calculations are performed on a 64-processor computer (from Skylaris et al. 2008).

Fraunhofer Institute for the Mechanics of Materials

The Fraunhofer Institute for Mechanics of Materials (Institut Werkstoffmechanik, IWM), with locations in Freiburg and Halle/Saale, Germany, is one of the Fraunhofer Institutes that constitute the largest applied research organization in Europe, with a research budget of €1.3 billion and 12,000 employees in 56 institutes. The institutes perform their research through “alliances” in Microelectronics, Production, Information and Communication Technology, Materials and Components, Life Sciences, Surface Technology and Photonics, and Defense and Security. Financing of contract research is by three main mechanisms: institutional funding, public project financing (federal, German Länder, EU, and some others), and contract financing (industry). The IWM has 148 employees (with another 75 based at its other campus in Halle), and a €15.5 million budget, with 44% of the budget coming from industry and 25–30% from the government. The IWM has seen significant growth in recent years (10% per year), a figure that is currently constrained by the buildings and available personnel. At the IWM, 50% of the funding supports modeling and simulation (a number that has grown from 30% five years ago). The IWM has 7 business units: (1) High Performance Materials and Tribological Systems, (2) Safety and Assessment of Components, (3) Components in Microelectronics, Microsystems and Photovoltaics, (4) Materials-Based Process and Components Simulation, (5) Components with Functional Surfaces, (6) Polymer Applications, and (7) Microstructure-Based Behavior of Components.

The IWM has a world-class effort in applied materials modeling. For example, it has developed microscopic, physics-based models of materials performance and then inserted these subroutines into FEM codes such as LSDYNA, ABAQUS, and PEMCRASH to develop multi-physics-based automobile crash simulations. The IWM enjoys strong collaborations with German automobile companies, which support this research. Other research areas include modeling of materials (micromechanical models for deformation and failure, damage analysis), simulation of manufacturing processes (pressing, sintering, forging, rolling, reshaping, welding, cutting), and simulation of components (prediction of behavior, upper limits, lifetime, virtual testing).

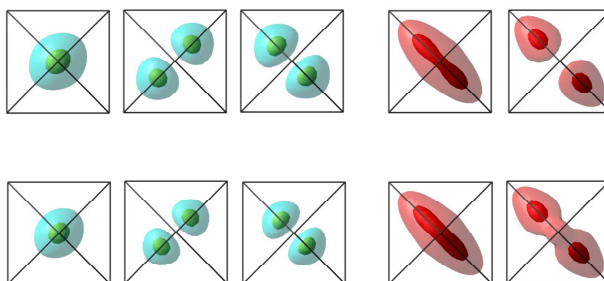


Figure 3.3. Lowest 3 electron wave functions and highest 2-hole wave functions of InAs (top row) and InGaAsN (bottom row) quantum dots grown on GaAs substrate calculated using Daresbury kkpw code (Tomić 2006).

The IWM employs mostly externally developed codes for its simulations (with some customized in-house codes that couple to the externally developed codes). All the codes run on parallel machines. At the largest (non-national) scale, the Fraunhofer institutes have a shared 2000-node cluster, of which capacity the IWM uses approximately 25%.

IWM staff find it challenging to find materials scientists with a theoretical background strong enough to become effective simulators. The typical training is in physics, and materials knowledge must then be acquired through on-site training. Currently it is difficult to obtain staff skilled in microstructure-level simulation, as they are in high demand by industry.

Energy Applications of Materials at Daresbury Laboratory

The researchers located in the materials CCPs at Daresbury Laboratory have an overarching materials goal of developing new materials, and modeling existing materials, for energy applications. Examples of energy-relevant materials under study at Daresbury Laboratory include fuel cells, hydrogen storage materials, nuclear containment systems, photovoltaics, novel electronics and spintronics systems (including quantum dots (Tomić 2006), and solid state battery electrodes. Specific properties of interest include structure and composition, thermodynamics and phase stability, reaction kinetics and dynamics, electronic structure (correlation), and dynamics. The toolkit for studying these properties at Daresbury Laboratory consists of the CCP codes DL_POLY, CASTEP, CRYSTAL, and a parallel implementation of the kppw continuum electronic structure code (see Figure 3.3). Using combinations of these codes, the Daresbury Laboratory researchers have studied radiation damage in nuclear reactors, performed atomistic simulations of resistance to amorphization by radiation damage, and simulated radiation damage in gadolinium pyrochlores.

SUMMARY OF KEY FINDINGS

SBE&S has long played a critical role in materials simulation. As the experimental methodologies to synthesize new materials and to characterize them become increasingly sophisticated, the demand for theoretical understanding is driving an ever greater need for materials simulation. This is particularly so in nanoscience, where it is now possible to synthesize an infinite variety of nanostructured materials and combine them into new devices or systems with complex nanoscale interfaces. The experimental tools of nanoscience—such as scanning probes, neutron scattering, and various electron microscopies—all require modeling to understand what is being measured. Hence, the demand for high-fidelity materials simulation is escalating rapidly, and the tools for verification of such simulations are becoming increasingly available.

The WTEC panel's key findings are summarized below:

- Computational materials science and engineering is changing how new materials are discovered, developed, and applied. We are in an extraordinary period in which the convergence of simulation and experiment at the

nanoscale is creating new opportunities for materials simulation, both in terms of new targets for study and in terms of opportunities for validation.

- World-class research in all areas of materials simulation areas is to be found in the United States, Europe, and Asia.
 - Algorithm innovation takes place primarily in the United States and Europe, some in Japan, with little activity to date in this area in China.
 - China has been mostly concerned with applications in the past; however, as part of the Chinese government current emphasis on supporting creativity in the both the arts and sciences, one can anticipate increased activity in this area in China.
- There is a rapid ramping-up of materials simulation activities in some countries, particularly China and Germany.
- There is an extraordinary revival of science in Germany, tied to recent increases in the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, the equivalent of the U.S. National Science Foundation) budget. In the most recent year, DFG funding jumped over 20%, following on three years of increases in the 5–10% range.
- There is much greater collaboration among groups in code development in Europe compared to the United States. There appear to be several reasons for this:
 - The U.S. tenure process and academic rewards systems mitigate against collaboration.
 - Funding, promotion, and awards favor high-impact science (publications in *Nature* and *Science*, for example), while the development of simulation tools is not considered to be high-impact science. There are numerous counter-examples, in which simulation software has been critical in scientific breakthroughs, but typically the breakthrough science comes a long time after a software tool has been developed, adopted widely, and validated. In other words, the payback on the development of a new simulation tool can be a decade or more after it is first developed.
- Training is a critical issue worldwide, as undergraduate students in the physical and chemical sciences and engineering are increasingly illiterate in computer programming. The situation is perhaps most acute of all in the United States.
- Funding of long-term (5 years at the very minimum), multidisciplinary teams is crucial to enable the United States to develop key materials simulation codes that will run on the largest U.S. computing platforms and on future many-core processors. Such codes will be needed to explore the next generation of materials and to interpret experiments on the next generation of scientific instruments.
- The United States is at an increasingly strategic disadvantage with respect to crucial, foundational codes, as it has become increasingly reliant on codes developed by foreign research groups. Political considerations can make those codes effectively unavailable to U.S. researchers, particularly those in Department of Defense laboratories.
- The utility of materials simulation codes for practical application would be enhanced dramatically by the development of standards for interoperability of codes. This would be similar to the CAPE-OPEN effort undertaken in the chemical process simulation (computer-aided process engineering, or CAPE) field to develop interoperability standards in that field (<http://www.colan.org>).

REFERENCES

- Accelrys Online. <http://accelrys.com/products/materials-studio/modules/CASTEP.html>.
- Allan, R.J., ed. 1999. *High performance computing*. New York: Kluwer Academic/Plenum Press.
- Allison, J., M. Li, C. Wolverton, and X.M. Su. 2006. Virtual aluminum castings: An industrial application of ICME. *JOM* 58(11):28–35.
- Amber Online. <http://ambermd.org/>.

- Carter, E.A. 2007. *Status and challenges in quantum mechanics based simulations of materials behavior*. Presentation at the 1–2 Nov. 2007 WTEC Workshop in Arlington, VA, Review of Simulation-Based Engineering and Science Research in the United States.
- Charmm Online. <http://www.charmm.org>.
- Chatterjee, A., and D.G. Vlachos. 2007. An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *Journal of Computer-Aided Materials Design* 14(2):253–308.
- Cpmd Consortium Online. <http://www.cpmd.org/>.
- Gao, Q., S. Yokojima, T. Kohno, T. Ishida, D.G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura. 2007. Ab initio NMR chemical shift calculations on proteins using fragment molecular orbitals with electrostatic environment. *Chemical Physics Letters* 445(4–6):331–339.
- Glotzer, S.C. and Paul. W, 2002. Molecular and mesoscale simulation of polymers. *Annual Reviews Mater. Res.*, 32:401–436.
- Gonze, X., J.M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, and D.C. Allan. 2002. First-principles computation of material properties: the ABINIT software project. *Computational Materials Science* 25(3):478–492.
- Hattel, J.H. 2008. Integrated modelling in materials and process technology. *Materials Science and Technology* 24(2):137–148.
- Haynes, P.D., C.-K. Skylaris, A.A. Mostofi, and M.C. Payne. 2006. ONETEP: Linear-scaling density-functional theory with local orbitals and plane waves. *Physica Status Solidi B* 243:2489–2499.
- Huang, P., and E.A. Carter 2008. Advances in correlated electronic structure methods for solids, surfaces, and nanostructures. *Annual Review of Physical Chemistry* 59:261–290.
- Kendall, R. A., E. Apra, D.E. Bernholdt, E.J. Bylaska, M. Dupuis, G.I. Fann, R.J. Harrison, J. Ju, J.A. Nichols, J. Nieplocha, T.P. Straatsma, T.L. Windus, and A.T. Wong. 2000. High performance computational chemistry: An overview of NWChem a distributed parallel application. *Computer Physics Communications* 128(1–2):260–283.
- Kobayashi, T., M. Shiga, A. Murakami, and S. Nakamura. 2007. Ab initio study of ultrafast photochemical reaction dynamics of phenol blue. *Journal of the American Chemical Society* 129(20):6405–6424.
- Liao, Q., A.V. Dobrynin, and M. Rubinstein. 2003a. Molecular dynamics simulations of polyelectrolyte solutions: Nonuniform stretching of chains and scaling behavior. *Macromolecules* 36(9):3386–3398.
- . 2003b. Molecular dynamics simulations of polyelectrolyte solutions: Osmotic coefficient and counterion condensation. *Macromolecules* 36(9):3399–3410.
- Likos, C.N. 2006. Soft matter with soft particles. *Soft Matter* 2(6):478–498.
- Mikami, M., S. Nakamura, M. Itoh, K. Nakajima, and T. Shishido. 2002. Lattice dynamics and dielectric properties of yttrium oxysulfide. *Physical Review B* 65(9): 094302.1–094302.4.
- Mikami, M., and A. Oshiyama. 1998. First-principles band-structure calculation of yttrium oxysulfide. *Physical Review B* 57(15):8939–8944.
- . 1999. First-principles study of intrinsic defects in yttrium oxysulfide. *Physical Review B* 60(3):1707–1715.
- . 2000. First-principles study of yttrium oxysulfide: Bulk and its defects. *Journal of Luminescence* 87–9:1206–1209.
- Namd Online. <http://www.ks.uiuc.edu/Research/namd>.
- NRC: Committee on Integrated Computational Materials Engineering of the National Research Council. 2008. *Integrated computational materials engineering: A transformational discipline for improved competitiveness and national security*. Washington, DC: The National Academies Press.
- Peng, Q., Y. Yi, and Z.G. Shuai. 2007. Excited state radiationless decay process with Duschinsky rotation effect: Formalism and implementation. *The Journal of Chemical Physics* 126(11):114302–8.
- Phillpot, S.R., S.B. Sinnott, and A. Asthagiri. 2007. Atomic-level simulation of ferroelectricity in oxides: Current status and opportunities. *Annual Review of Materials Research* 37: 239–270.
- Plimpton, S. 1995. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics* 117(1):1–19.
- Quantum Espresso Online. <http://www.quantum-espresso.org/>.

Siesta Online. <http://www.icmab.es/siesta/>.

Skylaris, C.-K., P.D. Haynes, A.A. Mostofi, and M.C. Payne. 2008. Recent progress in linear-scaling density functional calculations with plane waves and pseudopotentials: The ONETEP code. *Journal of Physics-Condensed Matter* 20:064209.

Theodorou, D.N. 2007. Hierarchical modelling of polymeric materials. *Chemical Engineering Science* 62(21):5697–5714.

Tomić, S. 2006. Electronic structure of InGaAsN/GaAs(N) quantum dots by 10-band kp theory. *Phys. Rev. B*. 73:125348.

Vasp Website. <http://cms.mpi.univie.ac.at/vasp>.

Wang, L.J., Q. Peng, Q.K. Li, and Z.G. Shuai. 2007. Roles of inter- and intramolecular vibrations and band-hopping crossover in the charge transport in naphthalene crystal. *The Journal of Chemical Physics* 127(4): 044506-9.

Wang, Y.X., S. Nakamura, K. Tasaki, and P.B. Balbuena. 2002. Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: How does vinylene carbonate play its role as an electrolyte additive? *Journal of the American Chemical Society* 124(16): 4408–4421.

Wang, Y.X., S. Nakamura, M. Ue, and P.B. Balbuena. 2001. Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: Reduction mechanisms of ethylene carbonate. *Journal of the American Chemical Society* 123(47):11708–11718.

Zeng, Q.H., A.B. Yu, and G.Q. Lu. 2008. Multiscale modeling and simulation of polymer nanocomposites. *Progress in Polymer Science* 33(2): 191–269.

CHAPTER 4

ENERGY AND SUSTAINABILITY

Masanobu Shinozuka

INTRODUCTION

The major energy problem the world faces at this time is the finite availability of fossil fuel. Therefore, energy conservation and the development of reliable clean energy sources should be the main focus of the future research effort. This observation is prompted by the statistics represented in Figure 4.1, which demonstrate that combined energy consumption related to buildings and transportation systems amounts to an overwhelming 75% of the total U.S. consumption. These systems represent backbones of infrastructure systems busily supporting the modern built environment, and yet they can be made substantially more energy-efficient by making use of new materials and by means of new methods of construction, retrofit, operation, and maintenance.

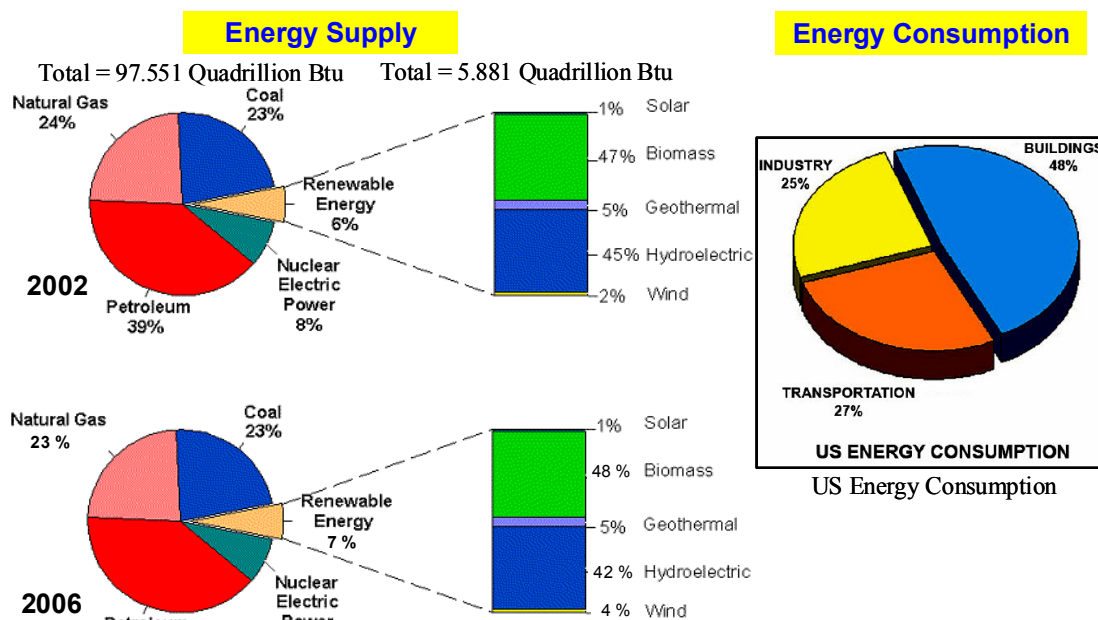


Figure 4.1. Energy supply and consumption in the United States 2006.

Source: http://www.eia.doe.gov/cneaf/solar.renewables/page/prelim_trends/prerends.pdf

Less use of fossil fuel diminishes emission of CO₂ and thus lowers its impact on global warming, assuming that acute correlation exists between them as proposed by some researchers. What is even more important, however, is the fact that the less we consume fossil fuel now, the more time we gain for the development of new, reliable, and substantial sources of clean energy before the arrival of Doom's Day when we deplete all economically affordable fossil fuel on the Earth.

Nuclear power is such a “clean energy” source, although its operational risk and reliability and the treatment of nuclear waste remain the subject of controversy. Biomass energy also appears to serve as a robust clean energy source, as demonstrated by Brazil producing large amounts of ethanol from sugar cane. Other sources such as solar, wind, and geothermal should also be pursued at a more accelerated pace. As also shown in Figure 4.1, the development of renewable energy is too slow and provides too small a percentage of the energy supply and clean air, although some future estimates are more optimistic. Also, we certainly hope that an economically practical hydrogen fuel cell will be delivered in more opportune time.

Table 4.1 shows an interesting comparison of what it takes to achieve annual reduction of one gigaton of CO₂ gas emission (about 3% of total global annual emission) by utilizing alternative energy sources or by means of some emission control technologies. Clearly, some of these technologies entail various socio-economic and/or geo-political repercussions, and others potentially interfere with the balance of the very natural environment that is to be protected. This shows that optimal selection of alternative energy sources requires complex socio-economic balancing act.

Table 4.1. Alternative Energy Sources & Efficiencies: Actions to Cut CO₂ Emissions by 1 Gt/Year*

Today's Technology	Actions that Provide 1 Gigaton / Year of Mitigation
Coal-Fired Power Plants	1,000 “zero-emission” 500-MW coal-fired power plants†
Nuclear	500 new nuclear power plants, each 1 GW in size†
Car Efficiency	1 billion new cars at 40 miles per gallon (mpg) instead of 20 mpg
Wind Energy	50 times the current global wind generation†
Solar Photovoltaics (PV)	1,000 times the current global solar PV generation†
Biomass Fuels from Plantations	15 times the size of Iowa's farmland
CO₂ Storage in New Forests	30 times the size of Iowa's farmland to new forest

* Source: <http://genomics.energy.gov>. Emission of CO₂ in 2010 is estimated to be 27.7 Gigaton (Almanacs)

† in lieu of coal-fired plants without CO₂ capture and storage

In this chapter, the subject matter of “Energy and sustainability” is considered from two points of view. In the first, the issue of the energy and its sustainability is considered as a set, and hence the development and utilization of alternative energy sources play an important role in this consideration. The second focuses on the issue of sustainability of the built environment, particularly infrastructure systems that must remain resilient under natural and man-made hazards such as earthquakes and malicious attacks, respectively. This observation is summarized as follows.

- *Energy and Sustainability*: Timely development of renewable and alternative energy is critical to America's energy security and sustainable adaptation to global climate change. (U.S. Department of Energy)
- *Infrastructure Development and Sustainability*: “Sustainable development is the challenge of meeting human needs for natural resources ... while conserving and protecting environmental quality and the natural resource base essential for future development” (ASCE Policy 418, as first defined in November 1996; http://pubs.asce.org/magazines/asce/news/2008/Issue_01-08/article3.htm).
- *System-Specific Sustainability (a necessary condition)*: A system is sustainable under natural, accidental, and manmade hazards if it is designed sufficiently resilient relative to the return period of the extreme events arising from these hazards (Shinozuka 2009).

The purpose of this chapter is to evaluate the state of the research in the world on energy and sustainability in the context of these definitions from the above-mentioned three points of view, specifically focusing on the research that utilizes simulation-based engineering and science technologies (SBE&S).

SBE&S RESEARCH ACTIVITIES IN NORTH AMERICA

Analysis of Energy and CO₂ Emission

CO₂ emission has been one of the major concerns of the electric power industry; the Electric Power Research Institute (EPRI) carried out a detailed “Prism” analysis that projected by simulation (in a broad sense) the reduction of CO₂ emission that is achievable beyond 2020 and into 2030, as shown in Figure 4.2. The reduction contributions arising from different technologies are color-coded. These technologies include efficiency in generation, use of renewables, nuclear generation, advanced coal generation, carbon capture and storage, plug-in hybrid electronic vehicles, and distributed energy sources; this shows that carbon capture and storage provide the most significant reduction. The “Prism” analysis appears to include evaluation of the effectiveness of suppressing CO₂ emission for each technology. It would be highly useful if the details of this analysis can be made available to the research community.

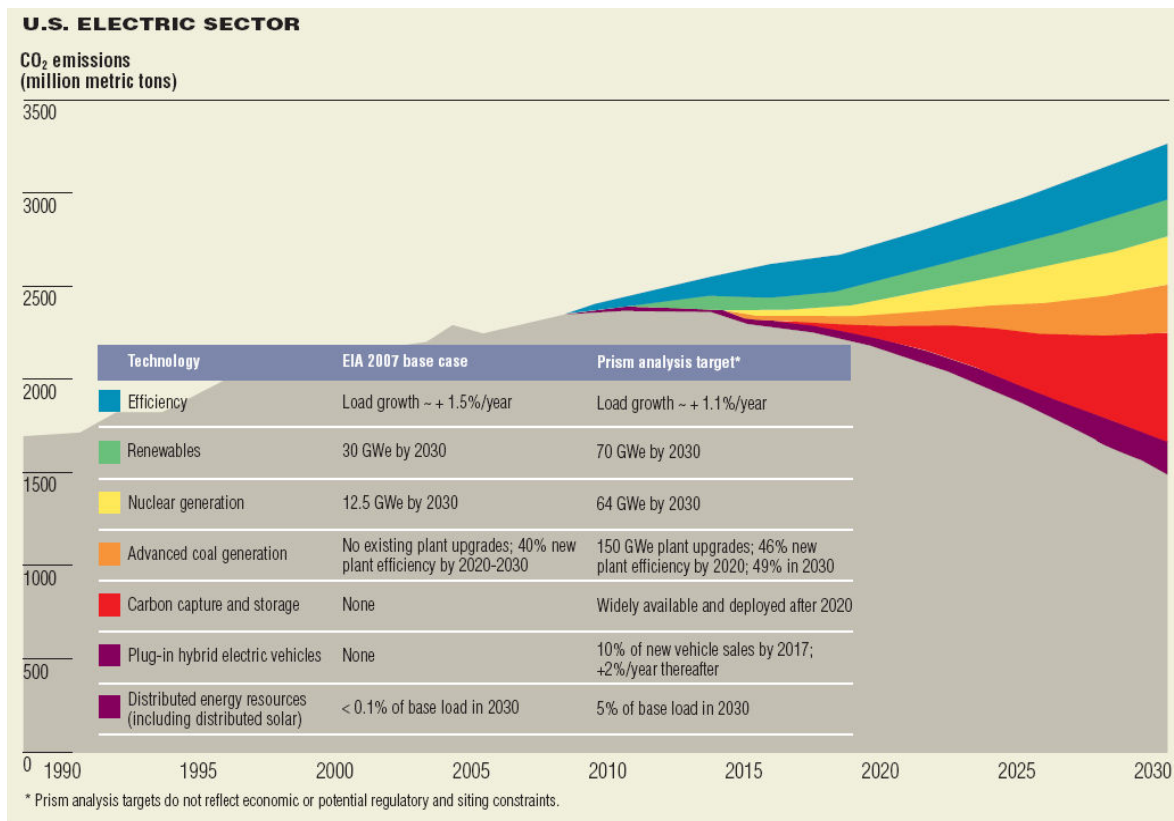


Figure 4.2. EPRI's Prism Analysis (Source: <http://my.epri.com/portal/server.pt?>, “The Portfolio” January/February 2008)

Modeling System Reliability of Electric Power Networks

Analytical modeling and assessment of electric power system reliability have their origins in North America, as represented by a number of books authored by Endrenyi (1978), Billinton and Allan (1988), Billinton and Li

(1994), and Brown (2002). Particularly, the last two books describe how to use Monte Carlo Methods for reliability assessment. In a recent paper by Zerriffi, Dowlatabadi, and Farrell (2007), the authors indicate that they augmented the traditional Monte Carlo reliability modeling framework and compared the performance of centralized to distributed generation (DG) systems under various levels of conflict-induced stresses. They concluded that DG systems are significantly more reliable than centralized systems, and when whole-economy costs are considered, they are also more economical. These simulations are carried out on less complex power network models or IEEE reliability test systems.

An analytical model was developed, improved, and expanded over the last decade by Shinozuka and colleagues for prediction of the seismic performance of power transmission networks under probabilistically defined seismic hazards. The most recent version is summarized in a paper by Shinozuka, Dong, et al. (2007). The model is sophisticated and heavily probabilistic and complex, requiring extensive use of Monte Carlo simulation for the prediction of system performance. The major advantage of using Monte Carlo simulation is that the system can be upgraded and retrofitted “virtually” and then subjected to various probabilistic hazard scenarios. This allowed the researchers to identify the most effective method of retrofit under probabilistic hazard scenarios in terms of shifting of risk curves. The simulation method is applied for transmission network systems using the networks of the Los Angeles Department of Water and Power (LADWP) and of Memphis Light, Gas, and Water (MLGW) as testbeds. The seismic performance in these examples is measured in terms of the remaining power supply capability as a percentage of the intact system capability. The seismic hazard is given by an ensemble of scenario earthquakes consistent with hazard definition of the U.S. Geological Survey. For this reason, the performance is given by a risk curve for LADWP’s power system, in which only transmission-sensitive equipment (transformers, circuit breakers, disconnect switches, and buses) at receiving stations are considered for the failure analysis under seismic conditions. The risk curve indicates the annual exceedance probability that the network will suffer from more than a certain level of loss in power supply capability as a function of that level. The same model also demonstrated its ability to reproduce the actual recovery process of the damaged transmission network. Indeed, a good agreement was observed between actual and simulated states of recovery of the LADWP power transmission network after the 1994 Northridge earthquake (Shinozuka, Dong, et al. 2007) showing the model’s ultimate robustness.

Modeling Civil Infrastructure Systems Resilience and Sustainability

For the purpose of improving resilience and sustainability of civil infrastructure systems under natural and manmade hazards, retrofitting of the systems is often carried out at the expense of significant retrofit cost. Seismic retrofitting is a typical example. In the lack of well-established quantitative definitions for resilience and sustainability at present, a reasonable question that can be raised is if the retrofit is cost-effective, instead of asking if the structural systems increased their resilience and/or sustainability. In recent years, the California Department of Transportation (Caltrans’) engaged in and completed seismic retrofit of its bridges by means of steel-jacketing the bridge columns. The same question of cost-effectiveness was raised here, and indeed, a benefit-cost analysis was performed (Zhou et al. 2007). The tasks required in this analysis consist of the sequential modules, which are in essence the same as the task modules used for the development of risk curves associated with the power transmission network. The difference primarily arises from the fact that in this case the risk curve is developed as a function of the social cost associated with drivers’ delay plus daily loss of opportunity cost, all measured in hour units. This definition of social cost places an emphasis on the Caltrans’ responsibility as the major stakeholder for the degradation of the highway network due to seismically induced bridge damage. The social cost thus given in terms of hours of delay and opportunity loss is converted to equivalent monetary value for the benefit-cost analysis (average hourly wages of \$21.77 for Los Angeles is used, per 2005 U.S. Department of Labor data). In this work, the authors introduced a network-wide probabilistic bridge repair model and simulated the resulting recovery of the network capability. This made it possible to trace the daily social cost, as a function of the elapsed time, taking into account the corresponding progress in the recovery effort. The total expected annual social cost of the network under the given seismic hazard could then be estimated.

The difference between the total annual social cost when no bridges are retrofitted and when all the bridges are retrofitted represents the expected annual cost avoided due to the retrofit—or benefit resulting from the retrofit.

Similarly, expected annual bridge restoration cost avoided can be computed. The total expected annual cost avoided (sum of the social cost avoided and bridge restoration cost avoided) over the expected life (assumed to be 50 years) of the network is converted under a certain discount rate to the present value of the benefit of the retrofit. This value divided by the retrofit cost indicates the benefit-cost ratio. This ratio is computed in this example under the discount rate of 3%, 5%, and 7% for the network life horizon of 50 years, as shown in Table 4.2, which shows that the cost-effectiveness of total bridge retrofitting can be rated as yes, yes, and moderate under the discount rates of 3%, 5% and 7% respectively.

Table 4.2. Cost-Effectiveness Evaluation Summary

Discount Rate	Benefit/Cost Ratio	Cost-Effectiveness*
3%	4.39	Yes
5%	3.23	Yes
7%	2.36	Moderate

* Where $R = \text{Benefit/Cost ratio}$, and the level of cost effectiveness is characterized as
No: $R < 1.5$; **Moderate:** $1.5 \leq R < 2.5$; **Yes:** $R \geq 2.5$

The meteorological conditions under global warming could increase the occurrence frequency and intensity of each adverse event and, therefore, the increased risk to the built environment. Developing new retrofit strategies and design guidelines should reduce or at least contain the damage risk of the built environment at an acceptable level. Typically, retrofitting of existing levy systems to reduce the damage risk is only one of many engineering problems in this case. An example of infrastructure failure due to a weather-related extreme event is studied by Sanders and Salcedo, (2006). In this case, a scenario is considered such that an aging earth dam is breeched and fails under heavy rain. This floods the down-stream areas and inundates residential areas. Thus, the heavy rain causes a multiple hazard: (1) possible human casualty, (2) loss of property, (3) disruption of surface transportation by inundation, and (4) blackout of utility networks due to the malfunction of their control equipment being installed in the lower areas or in basements within utility facilities.

SBE&S RESEARCH ACTIVITIES IN ASIA

Modeling System Reliability of the Electric Power Network

National Center for Research on Earthquake Engineering (NCREE), Taipei Taiwan

The Taiwan Power Company (Taipower) system suffered from devastating damage from the 1999 Chi-Chi earthquake ($M_w=7.2$). The damage was widespread throughout the system, including high-voltage transmission lines and towers, power generating stations, substations, and foundation structures. Drs. Chin-Hsiung Loh and Karl Liu of NCREE in collaboration with Taiwan Power co. were instrumental in developing a computational system not only for modeling high-voltage transmission systems including transmission lines and towers, but also for modeling the entire system, including power-generating stations. The LADWP study reviewed earlier, considered primarily receiving station equipment (transformers, disconnect switches, and circuit breakers) to be seismically vulnerable on the basis of past earthquake experience in California. In this context, NCREE's computer code is able to simulate a broader variety of seismic scenarios. For example, their code can demonstrate by simulation that higher level of seismic reliability can be achieved for a network with a distributed rather than centralized power generation configuration under the same maximum power output.

Central Research Institute for Electric Power Industry (CRIEPI), Tokyo, Japan, and Chugoku Electric Power, Hiroshima, Japan

Dr. Yoshiharu Shumuta of Japan's CRIEPI (Central Research Institute for Electric Power Industry) has developed a support system for Chugoku Electric Power to be able to implement efficient disaster response and

rapid system restoration during and after the extreme wind conditions of Japan's ubiquitous typhoons (see site report Appendix B). While, as in the case of FEMA HAZUS (the U.S. Federal Emergency Management Agency's software program for estimating potential losses from disasters), the wind-induced damage can be correlated with wind speed, direction, and duration, the CRIEPI's tool is more advanced in that it uses early prediction by the weather bureau for these wind characteristics. This will provide critically needed lead time for emergency response and significantly enhance the effectiveness of response and restoration. Figure 4.3 schematically outlines the tool. Dr. Shumuta was also successful in instrumenting electric poles with accelerometers. This allows assessing the integrity of network of electric wires serving as part of the distribution system (as opposed to transmission system). This is the first time the network of electric wires is instrumented for the purpose of system integrity under earthquakes. Dr. Shumuta contends that the same instrumentation is useful for assessment of integrity of the system under Typhoons. Simulation models of distribution network integrity, when developed, can be verified with the observations obtained from this type of measurements.

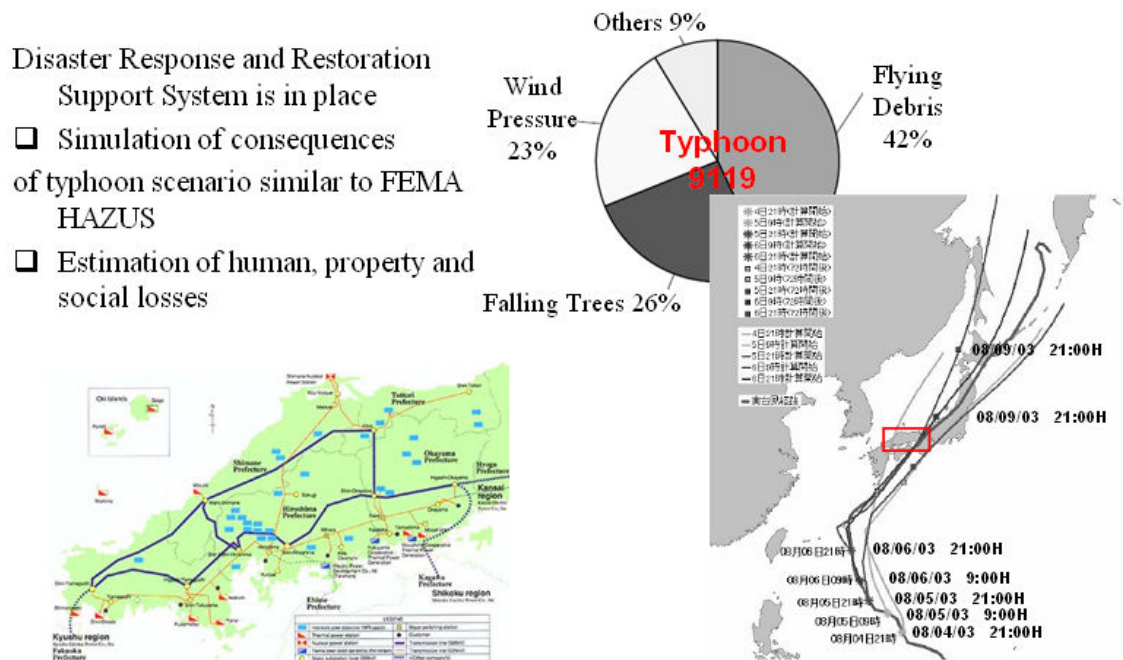


Figure 4.3. Disaster response and restoration support system, Japan.

Modeling Civil Infrastructure Systems Resilience and Sustainability

Pusan Port Authority, Pusan, Korea

The 1995 Hyogoken-Nanbu (Kobe) earthquake (M=6.8) caused serious damage to Kobe, Japan, port facilities, proving that the facilities were not quite resilient to an earthquake of this magnitude, and even worse, that they do not appear to be very sustainable more than one decade after the event. This observation prompted Dr. Ung Jin Na at Pusan Port Authority (Korea) to develop a computer simulation tool that will enable the authority to carry out a probabilistic assessment of seismic risk (and the risk from other natural hazards such as typhoons) that the facilities are likely to be subjected to. Major ports, particularly sea ports, are important nodes of transportation networks providing shipping and distribution of cargos via water. Handling 97% of international cargo through seaports, port's down time results in severe regional and national economic losses. Unfortunately, port facilities are very vulnerable to earthquakes, being often located near to seismic faults and built on liquefaction-sensitive fill or soft natural materials. The most prominent highlight in Dr. Na's simulation work is the use of the spectral representation method (Shinozuka and Deodatis 1996) of random multidimensional simulation for liquefaction analysis. The method is applied to the finite difference liquefaction analysis of saturated backfill soil whose key

mechanical properties are modeled as a two-dimensional random field. A rather dramatic effect of the random field assumption is quite clearly observed when compared with the more subdued liquefaction behavior under the assumption of uniform material properties. The excess pore pressure distributions in the different 2D samples of this random field are obviously quite different from each other. Only with this irregular distribution of pressure, as simulated, was it possible to explain actual observation that some caissons displaced under the earthquake an unusually large distance of 5.5 m displacement, as opposed to the 2.5 m experienced by most of them, and that two adjacent caissons have significantly different values of outward displacements (Na, Ray-Chaudhuri, and Shinozuka 2008). Figure 4.4 depicts a case in which critical infrastructure system, in this case the Kobe port system, demonstrated a reasonable resilience by restoring the cargo handling capability within about a year and half and yet its world ranking in cargo handling degraded from the 5th place before the event to the 35th place in 2004 because of the shifting of the cargo business to other ports while the Kobe port is struggling with the restoration. This represents an example which shows that an apparently reasonable resilience capability does not guarantee the system sustainability (see the description of System-Specific Sustainability in the introduction of this chapter).

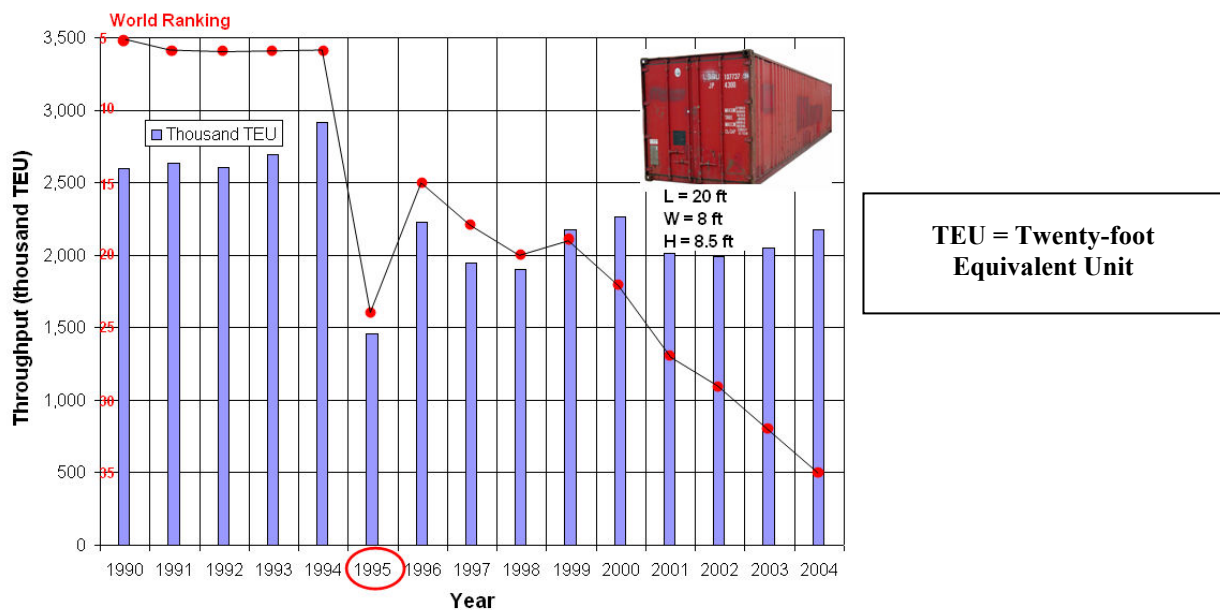


Figure 4.4. Throughput and world ranking of Kobe port before and after the 1995 Kobe earthquake. (Source: Containerization International Yearbook)

Disaster Control Research Center, School of Engineering, Tohoku University, Japan

Professor Shunichi Koshimura developed a numerical model of a tsunami inundation and utilized remote sensing technologies to determine the detailed features of tsunami disaster that killed more than 1,25,000 people (see site report Appendix B). The numerical model predicts the local hydrodynamic features of the tsunami, such as inundation depth and current velocity, based on high-accuracy and high-resolution bathymetry/topography grid. The numerical simulations of the model are validated by the observed records and measured sea levels, e.g., Figure 4.5. Also, high-resolution optical satellite imagery provides detailed information of tsunami-affected area. In recent years, Koshimura is expanding the model's capabilities to comprehend the impact of major tsunami disaster by integration of numerical modeling and remote sensing technologies. Along this line, he integrated his numerical model of tsunami inundation and post-tsunami disaster information obtained from satellite imagery in order to determine the relationship between hydrodynamic features of tsunami inundation flow and damage levels or damage probabilities in the form of fragility functions; development of fragility functions are focused on two

cases. One is the fragility function of house damage, which is useful in estimating quantitatively an important aspect of the tsunami's societal impact. The other is the fragility of the mangrove forest (Koshimura and Yanagisawa 2007). This is important in determining quantitatively the usefulness of the coastal forest to mitigate tsunami disasters.

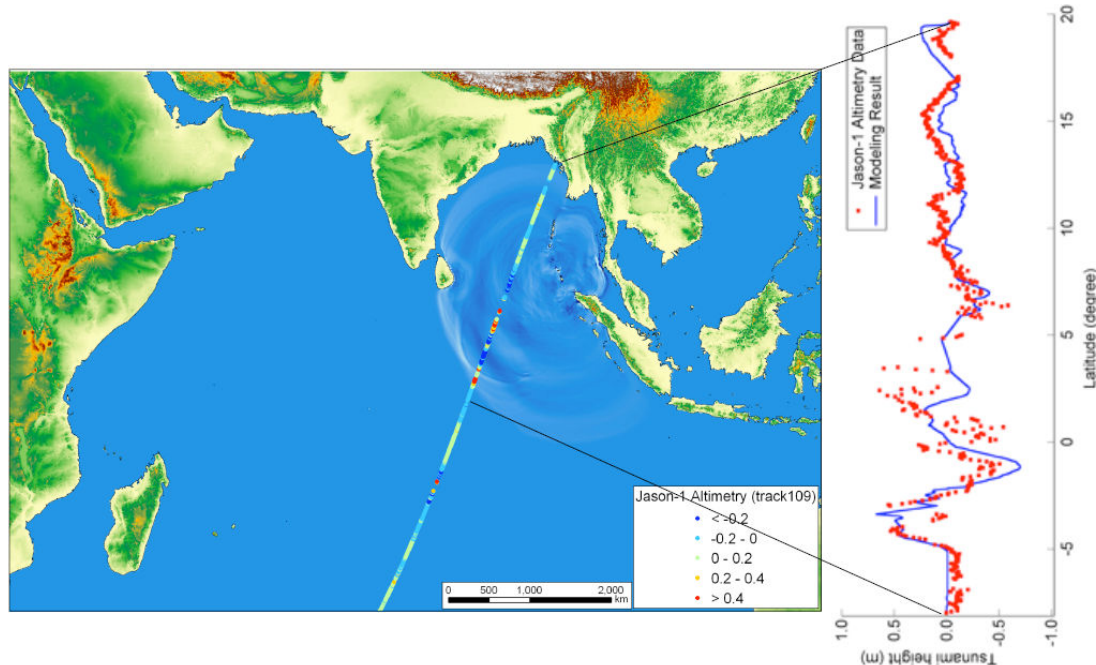


Figure 4.5. Validation of the 2004 Indian Ocean Tsunami model, with Jason-1 altimetry data. Jason-1 is the mission satellite launched by NASA and CNEAS and have captured the sea surface of traveling tsunami along the track approximately two hours after the 2004 Sumatra-Andaman earthquake.

Osaka University Department of Management of Industry and Technology

Professor Masaru Zako made a presentation to WTEC team at Tokyo University focusing on development of new simulation capabilities in energy systems, disaster modeling and damage evaluation, life sciences, nanotechnology, and urban environments is impressive (see site report Appendix B). Professor Zako and his group carried out an extensive study where he uses simulation techniques to deal with hazardous gas dispersion. The aim of their research team is the development of a simulation capability for disaster propagation in chemical plants considering the effects of tank fire and wind speed and direction. Another key aspect is how to estimate associated damages using the results of simulations in conjunction with information from the Geographical Information System (GIS). Monte Carlo methods of estimating radiation heat transfer were studied, and examples from tank fires and gas explosions in realistic scenarios were discussed. The gas originates from leakage from an LNG tank, evaporates and disperses into the air. The dispersion is analytically simulated and traced on a GIS map to predict the regions of contamination. Wireless transmission of real-time wind speed and direction data will make prediction more accurate (see Figure 4.6). (Shinozuka, Karmakar, et al. 2007). It is clear that long-term prediction of such events remains a very difficult task. While this study is in nature the same as the study of vehicular emissions of CO₂ gas, the problem simulated here, which can occur following a severe earthquake as the initiating event, can be even more devastating.

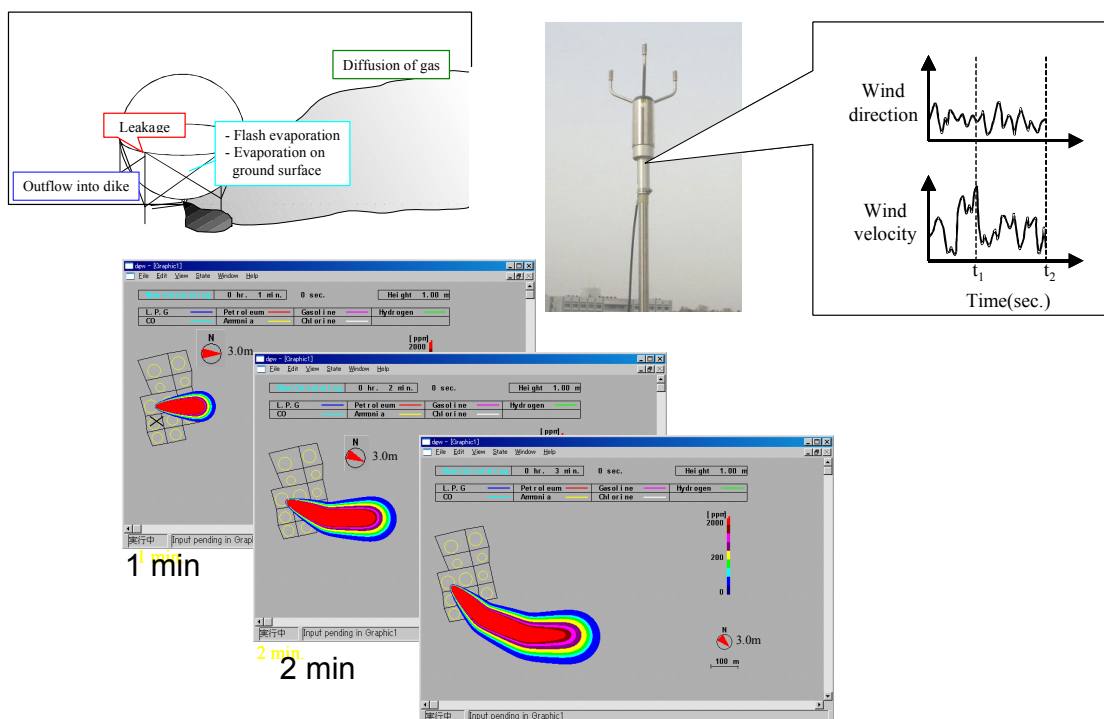


Figure 4.6. Simulation of pollutant dispersion under wind (Osaka University, Japan).

Energy Related Modeling

Toyota Central R&D Labs Inc. (TCRDL), Japan

The depth and breadth of SBES activities at TCRDL are impressive (see site report Appendix B). The role of SBES at the laboratory is in the development and application of simulation tools to aid in understanding phenomena that are observed by experimentalists at the facility. The researchers at TCRDL identified several grand challenge problems for future SBES activities. The most important one is to perform first-principles molecular dynamics simulations of fuel cells to understand the mechanisms of hydrogen atom transport within those systems. The interaction between government and corporate facilities in SBES appears to be beneficial to both parties in Japan.

Fuel Cell Laboratory, Nissan Research Center 1, Japan

The Nissan fuel-cell research group is mostly doing small simulations on workstations (see site report Appendix B). Their main target is the membrane-electrode assembly. They are planning a substantial future project: cost reduction and lifetime extension of the Pt catalyst, and substitution of Pt by a cheaper catalyst. They are considering an application of computational chemistry. Models would include rate equations, molecular dynamics, Monte Carlo transport, and dissipative particle dynamics. They plan to collaborate with an outside party to tackle a series of problems: (1) calculate performance of catalysts in a simple electrode reaction and compare results with experimental results, (2) apply computational chemistry to the more complex cathode reaction, and (3) try to predict performance of alternative catalysts, including non-Pt.

Central Research Institute of Electric Power Industry (CRIEPI), Japan

CREIPI has an impressive array of simulation and experimental capabilities to address the most challenging problems facing the Japanese power industry (see site report Appendix B). Its high-end simulations capabilities in materials science and global climate change are on par with the premier simulation institutions around the world. The institution's mission focus enables its researchers to build strongly coupled simulation and experimental campaigns enabling them to validate their simulation results and build confidence in predictions. The close coordination of simulation and experiment should serve as an example to the rest of the world. The CRIEPI researchers, in collaboration with NCAR (National Center for Atmospheric Research) and Earth Simulator staff, led an international team in performing unprecedented global climate simulations on the Earth Simulator that has influenced international policy on global climate change. The result of the effort was the ability to perform multiple 400-year predictions of future temperature and sea level changes with different CO₂ emissions profiles. The simulations required about six months of wall clock time on the Earth Simulator for four different scenarios to be completed. A challenging high-resolution (a grid spacing of 10 km) ocean simulation was also conducted on the Earth Simulator, and eddy-resolved ocean currents were successfully simulated. Since then, the team at CRIEPI has begun performing regional climate simulations of East Asia to assess the regional impacts of global warming, high-resolution weather forecasting simulations, and global models of ocean currents. Several highlights included simulations of a typhoon making landfall in Japan, and predictions of local flooding in areas surrounding Japan's energy infrastructure that required higher resolution than the Japanese Meteorological Agency typically performs.

CRIEPI researchers provide simulation analysis and tools to the Japanese electric power companies in other areas as well. Examples include CRIEPI's Power system Analysis Tool (CPAT) for electric grid stability, CFD analyses of combustion in coal-fired power plants, CFD analysis of transmission line vibration, and seismic analysis of oil tanks. For the most part, CRIEPI employs commercial software in its routine simulations and outsources code development to tailor the commercial tools for problems of interest. However, the simulation tools for the high-end materials science and climate simulations are developed in-house or with a small set of international collaborators. CRIEPI management considers the codes developed in-house to be proprietary; after establishing the proprietary rights the institution shares its codes with the scientific community.

RESEARCH ACTIVITIES IN EUROPE

System Reliability of Electric Power Network

Union for the Coordination of Transmission of Electricity, Brussels, Belgium

While this subject matter is of vital importance in Europe as in North America and Asia, it is not strongly addressed in Europe from the view point of system failures due to natural and manmade hazards. In North America and Asia, severe earthquakes, hurricanes and Typhoons occur frequently. As a result, it appears, the general public, utility companies and regulatory agencies are much more risk conscious and promote the research on this subject. However, the "Union for the Co-ordination of Transmission of Electricity" (UCTE), the association of transmission system operators in continental Europe provides a reliable market base by efficient and secure electric "power highways". 50 years of joint activities laid the basis for a leading position in the world which the UCTE holds in terms of the quality of synchronous operation of interconnected power systems. Their common objective is to guarantee the security of operation of the interconnected power system. Through the networks of the UCTE, about 450 million people are supplied with electric energy; annual electricity consumption totals approx. 2300 TWh (www.ucte.org). UCTE collected past sequences of black-out events that can be archived to develop scenario of black-out events for future system risk assessment. One example of this is given in Figure 4.7.

The blackout in Southern Sweden and Eastern Denmark, September 2003

A coincidence of technical failures leads to a widespread blackout — sequence of events

Around noon on Tuesday, 23 September 2003, the tripping of a large nuclear power plant (NPP) and an almost concurrent short-circuit in a substation in Western Sweden caused a large-area and long-lasting power outage in Southern Sweden and Eastern Denmark, including the capital city of Copenhagen. **The most severe outage in 20 years within the Nordic power system affected 4 Million inhabitants. About 6350 MW of load were lost.** The full restoration of the system took more than six hours

Stage	Time	Location	Events
Preconditions	Sept. 23, 2003		<ul style="list-style-type: none"> Stable operation with big security margins. Moderate demand: 15 GW in Sweden and 1.9 GW in Eastern Denmark. Several generation units and three 400 kV lines out of service in Southern Sweden and Eastern Denmark due to planned maintenance; similarly, the three HVDC circuits from Zealand (Eastern Denmark) and Southern Sweden to Germany and Poland are not available.
	Before 12:30		
Coincidence of grid faults	12:30		<ul style="list-style-type: none"> Unit 3 of NPP Oskarshamn trips due to internal valve problems in the feedwater circuits; 1.2 GW are lost. The applied grid security standards allow the system to cope with this "standard contingency" by automatically increasing production at other power stations. More power is flowing on the western side to supply the demand in the south.
	12:35		<ul style="list-style-type: none"> Double busbar fault in the 400 kV substation in Horred: below its maximum loading the mechanical structure of a switch breaks apart and causes a short circuit between two adjacent busbars, which are immediately separated by automatic protection devices. As a consequence, two units of the NPP Ringhals (totalling 1.8 GW) and two important 400 kV lines connecting Central and Southern Sweden, all connected to these busbars, are lost.

Restored at appr. 19:00 hrs

Figure 4.7. Sequence of events in the course of the 2003 Nordic blackout (Gheorghe et al. 2006, 168–169).

Energy-Related Simulation Research

There are many interesting studies on harvesting wind and solar energy and integrating the harvested energy into utility power grids. Some of them, which utilize simulation technologies, are selected here from Renewable Energy (Elsevier). These studies are carried out by individual researchers throughout the world demonstrating a strong interest on the subject matter by grass root of the research community. Typical examples of the studies related to wind power are listed below in 1-4 and 5-8 for solar power.

1. Identification of the optimum locations of wind towers within a wind park using Monte Carlo Simulation. (Greece, Marmidis et al. 2008, Renewable Energy, V. 33, 1455 - 1460)
2. Development of the analytical model to simulate the real-world wind power generating system (France, Ph. Delarue et al. 2003, Renewable Energy, V. 28, 1169 - 1185)
3. Identification of the most efficient wind power generator by altering its configuration (such as blade type, radius, # of blades, rotational speed of rotor) (Colombia, Mejia et al. 2006, Renewable Energy, V. 31, 383 - 399)
4. Calculation of efficiency of wind power systems connected to utility grids (Egypt, Tamaly and Mohammed, 2004, IEEE, <http://ieeexplore.ieee.org/iel5/9457/30014/01374624.pdf?arnumber=1374624>)
5. Simulation of off-grid generation options for remote villages in Cameroon (Nfah et al. 2008, Renewable Energy, V. 33, 1064 - 1072)

6. Numerical simulation of the solar chimney power plant systems coupled with turbine in China (Tingzhen et al. 2008, *Renewable Energy*, V. 33, 897 - 905)
7. Use of TRNSYS for modeling and simulation of a hybrid PV–thermal solar system for Cyprus (Soteris A. Kalogirou, 2001, *Renewable Energy*, V. 23, 247 - 260)
8. Two-Dimensional Numerical Simulations of High Efficiency Silicon Solar Cells in Australia (Heiser et al. 1993, *Simulation of Semiconductor Devices and Processes*, V. 5, 389 - 392)

Technical University of Denmark (DTU), Wind Engineering, Department of Mechanical Engineering (MEK), Denmark

The number of kW of wind power generated in Denmark per 1000 inhabitants exceeds 570 and is by far the highest in Europe and indeed the world. Spain and Germany follow with 340 and 270 (kW/1000 inhabitants), respectively. By the end of January 2007, there were 5,267 turbines installed in Denmark, with a total power of 3,135 MW; the total wind power in European Union countries during that period was 56,535 MW with Germany leading at 22,247 MW. The total installed power and the numbers of turbines in Denmark had a continually increase until 2002.

Research on wind energy in Denmark has been taking place for over 30 years, and DTU researchers in collaboration with those of its partner institution Risø have been the leaders in this field (see site report Appendix C). The research approaches employed at DTU are both of the fundamental type (e.g., large-eddy simulations and sophisticated grid techniques for moving meshes) but also of practical use, e.g., the popular aeroelastic code FLEX that uses lumped modeling for wind energy prediction. DTU also maintains a unique data bank of international meteorological and topographical data. In addition, the education activities at DTU are unique and impressive. It is one of the very few places in the world to offer both MSc and PhD degrees on wind engineering, and its international MSc program attracts students from around the world, including the United States, where there are no such programs currently.

Institut Français du Pétrole (IFP), France

IFP's research activities in SBES demonstrate that the institute is deeply grounded in the fundamental mathematical, computational, science, and engineering concepts that form the foundation of SBES projects in the service of technological advances (see site report Appendix C). The coupling of this sound foundation with the institute's significant and ongoing contributions to the solution of "real world" problems and challenges of energy, environment, and transportation form an excellent showcase for the present capabilities of SBES and its even greater future potential. The IFP research is more applied than that typically found in universities, and its time horizon starting from fundamental science is of longer duration than that typically found in industry development projects. Use of SBES to develop thermophysical properties tables of complex fuel mixtures (biodiesels) at high pressures is a good example of precompetitive research and a platform that benefits the entire automotive industry and all bio-fuel industry participants. The strategy of developing an extensive intellectual property (IP) portfolio for the refining of heavy crude with the view of a refinery feedstock mix shifting in that direction (e.g., increasing share of transportation fuels originating from tar sands) is another illustration of the long time horizon of the IFP. IFP's scientific R&D activities are driven by the following five complementary strategic priorities: Extended Reserves; Clean Refining; Fuel-Efficient Vehicles; Diversified Fuels; Controlled CO₂.

Science and Technology Facilities Council (STFC) Daresbury Laboratory (DL), Warrington, United Kingdom

The research group at STFC focuses on the impact of SBES materials research in the energy field, e.g., fuel cells, hydrogen storage, nuclear containment, photovoltaics, novel electronics/spintronics, and solid state battery electrodes (see site report Appendix C). The overarching and unifying theme is that the design space for the new materials in each of these applications is so large that experiments alone are unlikely to lead to the optimal (or even adequate) materials design, and that with recent and expected advances in algorithms and HPC hardware, the SBES approach is entering the "sweet spot" with respect to applicability and relevance for bridging the length and time scales encountered in foundational scientific predictions for structure & composition, thermodynamics of phases, reaction kinetics & dynamics, and electronic structure/correlation. Research at STFC is Timely in the

context of solving the looming energy challenges in the post-fossil-fuel era. Example of SBES material research involving fuel cells is shown in Figure 4.8.

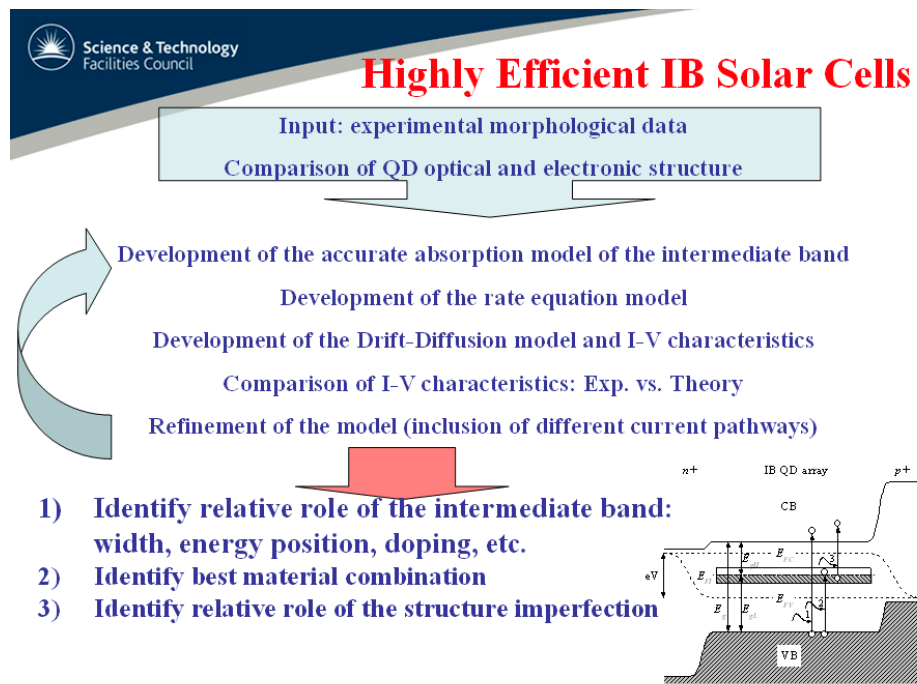


Figure 4.8 Development of a highly improved solar cell.

Modeling Civil Infrastructure Systems Resilience and Sustainability

University of Oxford, Oxford, United Kingdom

Led by Prof. Allstair Borthwick, the research group at University of Oxford carries out SBE&S research on areas such as river co-sustainability, simulation of dyke break, solitary waves, urban flooding, Tsunami simulation, waves up the beach (see site report Appendix C).

CONCLUSIONS

It is quite clear that further research is urgently pursued to discover and develop new generation of fuels, alternative to fossil fuel. At the same time, next generation engines must also be developed that can function efficiently with these next generation fuels.

In the United States, energy consumption for buildings and transportation related activities, consumes almost 75% of the energy supply. It is more than prudent then to pursue the search to conserve the energy for construction and maintenance of these and other systems so that the built environment including such systems must be made sustainable. New materials, method of construction or maintenance, novel design guidelines, and the like must all be examined from energy conservation point of view.

Nuclear energy is recognized as the major source for clean energy. Large investments and rapid progress in SBE&S-driven materials design and safety analysis for nuclear waste containment are needed. In relation to alternative source of energy, the energy density is lower but tax incentives for green sources are driving extensive efforts in wind, solar, tidal and geothermal energy. However, better connection technology of these alternative sources of power to the existing power grid remains to be further improved.

In the field of power networks, significant research efforts in simulation techniques to evaluate performance of large-scale and spatially distributed systems such as power transmission systems that are subject to highly uncertain natural hazards (earthquake, hurricanes, and flood) have been given.

Simulation-based engineering is vital for advancing the physical infrastructure of Energy, Energy Production and Energy for Transportation. It is also an opportunity for creating smart infrastructure. The scarcity of appropriately trained students in the U.S. (relative to Europe and Asia) is a bottleneck to progress.

Overall, U.S. is ahead in modeling of large-scale infrastructure systems, but the gap is closing with Asia in particular. France is ahead in nuclear energy and related research (e.g. containment). Research related to oil production and fossil-fuel supply chain is traditionally led by U.S. Significant research activity is currently taking place in Europe, including new funding models and leveraging strengths in development of community codes. In sustainability front, currently, all regions recognize the need to increase emphasis in this area and research activities are on going.

REFERENCES

- Billinton, R., and R.N. Allan. 1988. *Reliability assessment of large electric power systems*. Boston/Dordrecht/Lancaster: Kluwer Academic Publishers.
- Billinton, R., and W. Li. 1994. *Reliability assessment of electrical power systems using Monte Carlo methods*. New York: Plenum Press.
- Brown, R.E. 2002. *Electric power distribution reliability*. New York: Marcel Dekker.
- Endrenyi, J. 1978. *Reliability modeling in electric power systems*. Chichester (UK); New York: Wiley.
- Gheorghe, A.V., M. Masera, M. Weijnen, and L. De Vries. 2006. *Critical infrastructures at risk: Securing the European electric power system*. Dordrecht: Springer.
- Koshimura, S., T. Oie, H. Yanagisawa, and F. Imamura. 2006. Vulnerability estimation in Banda Aceh using the tsunami numerical model and the post-tsunami survey data. In *Proceedings of the 4th International Workshop on Remote Sensing for Disaster Response*, 25-6 September. 2006, Magdalene College, Cambridge, UK. New York: MCEER. Available online at http://www.arct.cam.ac.uk/curbe/4thInt_workshop.html.
- Koshimura, S., and H. Yanagisawa, 2007. Developing fragility functions for tsunami damage estimation using the numerical model and satellite imagery. In *Proceedings of the 5th International Workshop on Remote Sensing Applications to Natural Hazards*, 10-11 Sept. 2007, Washington, DC. New York: MCEER. Available online at <http://www.gwu.edu/~spi/remotesensing.html>.
- Na, U.J., S. Ray-Chaudhuri, and M. Shinozuka. 2008. Probabilistic assessment for seismic performance of port structures. *Soil Dynamics and Earthquake Engineering* 28(2):147-158.
- Sanders, B., and F. Salcedo, 2006. Private communication.
- Shinozuka, M. 2009. Chapter 3 of the current report.
- Shinozuka, M., and G. Deodatis. 1996. Simulation of multi-dimensional Gaussian stochastic fields by spectral representation. *Applied Mechanics Review* 49(1):29-53.
- Shinozuka, M., X. Dong, T.C. Chen, and X. Jin. 2007. Seismic performance of electric transmission network under component failures. *Journal of Earthquake Engineering and Structural Dynamics* 36(2):227-244.
- Shinozuka, M., D. Karmakar, M. Zako, T. Kurashiki, and M. Fumita. 2007. GIS-based hazardous gas dispersion, simulations and analysis. In *Proceedings of 2nd International Conference on Disaster Reduction*, November 27-29.
- Yassin, M.F., S. Kato, R. Ooka, T. Takahashi, and R. Kouno. 2005. Field and wind-tunnel study of pollutant dispersion in a built-up area under various meteorological conditions. *Journal of Wind Engineering and Industrial Aerodynamics* 93:361-82.
- Zerrieffi, H., H. Dowlatabadi, and A. Farrell. 2007. Incorporating stress in electric power systems reliability models. *Energy Policy* 35(1):61-75.
- Zhou, Y., S. Banerjee, and M. Shinozuka. 2007. Socio-economic effect of seismic retrofit of bridges for highway transportation networks: A pilot study. Accepted for publication in the *Journal of Structure and Infrastructure Engineering*.

CHAPTER 5

NEXT-GENERATION ARCHITECTURES AND ALGORITHMS

George Em Karniadakis

“Gordon Bell’s Prize has outpaced Gordon Moore’s Law...” –David Keyes

INTRODUCTION

One of the main recommendations of the NSF SBES report (Oden et al. 2006) is that “investment in research in the core disciplines of science and engineering at the heart of SBES applications should be balanced with investment in the development of algorithms and computational procedures for dynamic multiscale, multiphysical applications.” This is particularly important at the present time when rapid advances in computer hardware are taking place and new mathematical algorithms and software are required. This was also stressed in the 1999 report of the President’s Information Technology Advisory Committee (PITAC) to the President, which stated, “...information technology will transform science, medicine, commerce, government, and education. These transformations will require new algorithms, new tools, and new ways of using computers...” (Joy and Kennedy 1999, 32), and also, “...Work on system software and algorithms will permit the effective exploitation of parallelism and will contribute to the efficient use of memory hierarchies... Without effective building blocks for software, algorithms, and libraries, high-end hardware is an expensive, underutilized technology...” (Joy and Kennedy 1999, 51).

Recently, in June 2008 and two years earlier than anticipated in the PITAC report, the first ever petaflop benchmark run of Linpack was demonstrated on the Los Alamos parallel computer *Roadrunner*; many more petaflop systems are expected to be operational in the next three years. At the same time, computer designers are developing the first concepts of an exascale (10^{18} ops/sec) computer to be available by 2017 based on more than 10 million processing elements (cores)! While the proponents of these systems that will deliver unique *capability computing* envision new “transformational” and “disruptive” science, it is appreciated by all that the mathematical and system software will play a crucial role for exploiting this capability. The algorithms and computational procedures that the SBES community develops have to take into account these exciting hardware developments, and hence, investments in algorithmic developments and mathematical software should be analogous to investments in hardware.

In 1995 before the teraflop era, a similar argument was raised, namely that “The software for the current generation of 100 gigaflop machines is not adequate to be scaled to a teraflop...” and moreover, “...To address the inadequate state of software productivity, there is a need to develop language systems able to integrate software components that use different paradigms and language dialects” (Sterling, Messina, and Smith 1995). There is, however, a big difference in the technology jump from the gigaflop to teraflop machines compared to what is taking place today in the leap to petaflop and exaflop machines as we approach the limits of transistor technology. While teraflop technology relied primarily on improvements in computer chip speed and a total of less than 1,000 processors, the petaflop and exaflop computers will be based mainly on *concurrency* involving, respectively, hundreds of thousands and millions of processing elements with the number of cores per processor

doubling every 18-24 months. Moreover, petaflop and exaflop systems are much more expensive to acquire and operate, due to high demand on power. For example, a Linpack benchmark run of two hours on the Roadrunner consumes 2.3 MW, which translates into about two barrels of oil. At this cost, even 1% of petaflop or exaflop computing capability is a terrible thing to waste!

Historically, fast algorithms have played a key role in advancing capability computing in SBE&S at the same or faster rate than Moore's law (i.e., the doubling of computer speed every 18 months). An example is the solution of a three-dimensional Poisson equation, required in many different disciplines from computational mechanics to astrophysics: the Gaussian elimination, used primarily in the 1950s, has computational complexity N^7 , which was reduced to N^3 using a multigrid method in the 1980s. This reduction in computational complexity translates into a gain of about 100 million in "effective" flops even for the modest problem size $N=100$.

Improved algorithms have resulted in performance gains of several orders of magnitude in many application domains from magnetic fusion to turbulent combustion to molecular dynamics (MD) (Keyes et al. 2003 and 2004). Figure 5.1 shows the "effective" Moore's law for a first-principles MD code (FPMD), the QBOX code described by Gygi and colleagues (2005). The slope of the Gflops-over-time curve indicates a doubling of performance every eight months, which is much faster than the aforementioned Moore's law on computer speedup. The main software building blocks of QBOX are basic linear algebra routines (BLAS, ScaLAPACK), and the parallel program is based on MPI (the Message-Passing Interface parallel programming standard). By improving these two areas and using clever decomposition techniques, the QBOX developers have achieved better than 50% of the peak speed of the IBM Blue Gene/L.

The same argument can be made by examining the various application codes that have won the prestigious Bell prize for supercomputing performance. It is clear, then, that Gordon Bell's prize outpaces Gordon Moore's law, due to clever algorithmic and software advances. In particular, despite the diverse pool of applications of the Gordon Bell winners—varying from PDE solvers, to integral equations, to MD simulations—the core solvers are common and are based on the fundamental algorithms of solvers for linear systems, graph partitioning, domain decomposition, high-order discretization, and so forth. Improving these commonly used algorithms and solvers would benefit the broader SBES community.

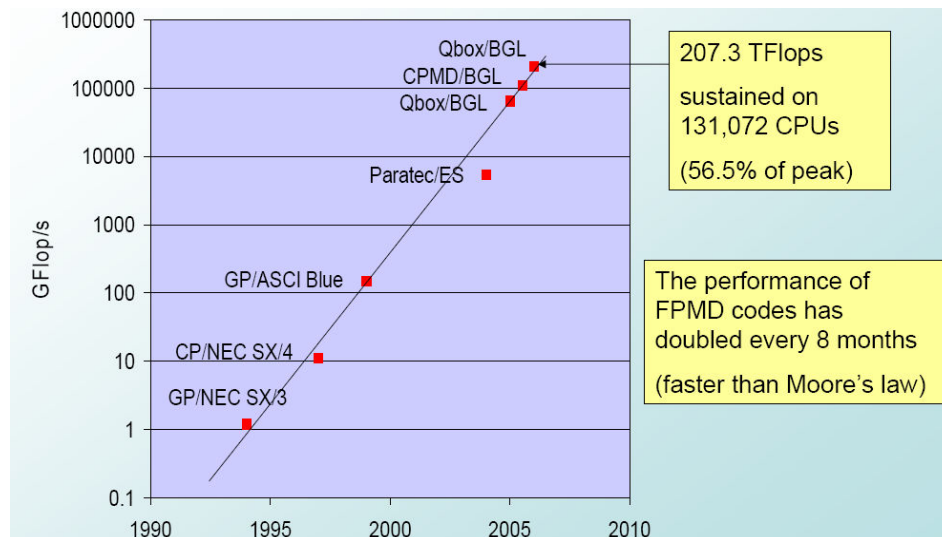


Figure 5.1. History of the performance of FPMD codes on different computer platforms (courtesy of Francois Gygi, University of California, Davis).

HIGH-PERFORMANCE COMPUTING AROUND THE WORLD

Examining the history of the Top500 list (<http://www.top500.org>) of supercomputer sites over the last two decades reveals that the United States has been dominant in the world of high-end computing, having about 58% of the “Top500” supercomputers, as shown in Figure 5.2. In the 31th edition of the Top 500 list released in November 2008, the United States had 290 computers in this list while the U.K. had 46, Germany had 25, and Japan had 17 supercomputers. Switzerland had only 4 supercomputers in the Top500 list, but if we normalize this number by millions of inhabitants the ratio is 0.53, which is more than half the normalized U.S. ratio of 0.94! It is notable that as of November 2008, three out of the top five systems in the Top500 list were non-U.S., and Europe has doubled its number of supercomputing systems in the list in the last few years.

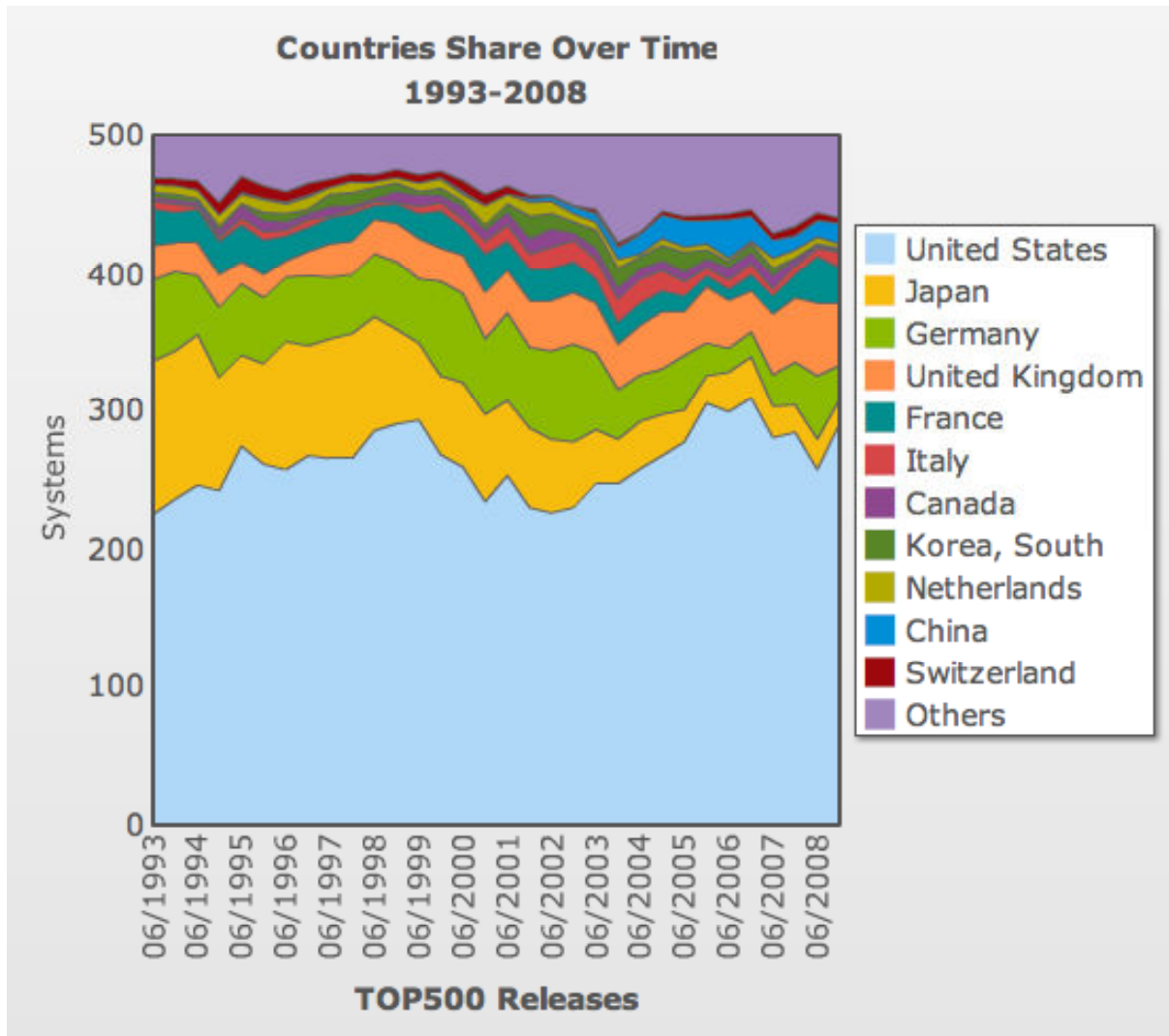


Figure 5.2. History of Top500 supercomputers and distribution across different countries. In November 2008, the United States had 58% of the total share but the same number of systems per million inhabitants as Switzerland (source: <http://www.top500.org/overtime/list/31/countries>).

High-Performance Computing in the United States

In early June of 2008, the world’s fastest supercomputer, Roadrunner, which performed the first sustained *one petaflop* speed on the Linpack benchmark, was unveiled by IBM for use at the Los Alamos National Laboratory.

Its performance is twice as fast as that of the previous number one supercomputer, IBM's Blue Gene system at Lawrence Livermore National Laboratory. Roadrunner cost about \$100 million and consists of 6,948 dual-core Opteron computer chips and 12,960 cell engines; its InfiniBand interconnecting system uses 57 miles of fiber optics.

There are many more systems of the petaflop scale on the horizon both in the United States. For example, a Cray XT (1.38 petaflop/s; more than 150,000 cores) was installed in late 2008 at the Oak Ridge National Laboratory, and an IBM Blue Gene/P (1 petaflop; up to 884,736 cores) will be installed at Argonne National Laboratory in early 2009. Several one-petaflop systems will be installed at the supercomputing centers supported by the National Science Foundation (NSF) and connected through the TeraGrid (<http://www.teragrid.org>). The most impressive NSF project of all, due to be completed in 2011, is the Blue Waters system, an IBM system based on Power7 multicores, with a total of more than 200,000 cores. Its peak speed will exceed 10 petaflops, but the objective is to sustain one petaflop on real applications.

High-Performance Computing in Japan

In Japan, there are three systems of about 100 teraflops peak in the universities, and a 10-petaflop system is on the horizon. Specifically, the University of Tokyo's T2K system is number one in Japan with a peak of 113 Tflop/s. The TSUBAME at the Tokyo Institute of Technology is number two and beating the Earth Simulator. It has a peak speed of 103 teraflops and is a Sun Galaxy 4 system with Opteron dual cores. A 92 teraflops system from Cray was installed at the University of Tsukuba, and a 61 teraflops system from Fujitsu at Kyoto University. The vector machine NEC SX-9 with peak speed of 839 teraflops will be installed at Tohoku University.⁵

The Japanese government puts forward every five to ten years a basic plan for science and technology (S&T) that also targets supercomputing infrastructure. The second S&T plan resulted in the installation of the successful Earth Simulator that made Japan number one in the supercomputing world for a while. Currently, the third S&T plan is being implemented, targeting the construction and operation of the Next Generation Supercomputer at the RIKEN labs (see site report in Appendix B). A 10-petaflop system at a cost of \$1 billion, due to be ready by 2011, this aims to be the world's fastest supercomputer, targeting grand-challenge problems in nanoscale sciences and life sciences. The system's architecture is a heterogeneous computing system with scalar and vector units connected through a front-end unit; all three major Japanese computer companies (NEC, Hitachi, and Fujitsu) are collaborating on this project. The hope is that parallel processing on such a mix of scalar and vector units will endow the programmer with greater flexibility so that large and complex simulations are efficiently performed.

High-Performance Computing in Europe

In Europe, Germany leads the supercomputing activities with three systems in the top 50, including the number eleven spot, a Blue Gene/P system located at the Forschungszentrum (FZ) Jülich. This is the largest national research center in Germany and Europe with an annual budget of €360 million; simulation science is targeted as its key technology. Petaflop systems are planned to be installed at the FZ Jülich and at the University of Stuttgart in the time frame 2009–2011.

France has six systems in the top 50. Of particular interest is the Tera-10 project at the French Atomic Authority (CEA) that resulted in its obtaining the number 48 spot in the top 500 list; this is the first-ever supercomputer built and designed in Europe, by the company Bull in collaboration with Quadrics. The Tera-100 project, also at CEA, aims to produce a petaflop machine by mid-2010.

The UK has two systems in the top 50. The European Centre for Medium-Range Weather Forecasts, based at the University of Reading, has an 8320-core IBM P6 system at number 21. At number 46, the 11328-core Cray XT4

⁵ Several supercomputing centers in Europe that run climate codes, such as the German Weather Service (DWD) and Météo France, are also expected to install this system.

system (HECToR), based at the University of Edinburgh, provides a general-purpose high-end platform for use by UK academic researchers. In addition, the university has a 2560-core IBM P5 system, which is also available to the UK research community and intended to complement the service provided by HECToR. Both of these machines are also made available through the UK National Grid Service.

A new European initiative called Partnership for Advanced Computing in Europe (PRACE) has been formed based on the infrastructure roadmap outlined in the 2006 report of the European Strategy Forum for Research Infrastructures (ESFRI 2006). This roadmap involves 15 different countries and aims to install five petascale systems around Europe beginning in 2009 (Tier-0), in addition to national high-performance computing (HPC) facilities and regional centers (Tiers 1 and 2, respectively). The estimated construction cost is €400 million, with running costs estimated at about €100–200 million per year. The overall goal of the PRACE initiative is to prepare a European structure to fund and operate a permanent Tier-0 infrastructure and to promote European presence and competitiveness in HPC. Germany and France appear to be the leading countries.

Recently, several organizations and companies, including Bull, CEA, the German National High Performance Computing Center (HLRS), Intel, and Quadrics, announced the creation of the TALOS alliance (<http://www.talos.org/>) to accelerate the development in Europe of new-generation HPC solutions for large-scale computing systems. In addition, in 2004 eleven leading European national supercomputing centers formed a consortium, DEISA, to operate a continent-wide distributed supercomputing network. Similar to TeraGrid in the United States, the DEISA grid (<http://www.deisa.eu>) in Europe connects most of Europe's supercomputing centers with a mix of 1-gigabit and 10-gigabit lines.

High-Performance Computing in China

In China, The number 10 system in the world is located in Shanghai at the Shanghai Supercomputer Center with a Chinese Dawning Supercomputer based on AMD processors with a peak speed of 233 Tflop/s. A few HPC centers are located at the major universities, each a few teraflops, connected by gigabit-level network. The top three supercomputers are an IBM/Intel cluster at China Petroleum & Chemical Co., an IBM SP Power4 at the China Meteorological Administration, and the DAWNING Opteron Cluster at the Shanghai Supercomputing Center (see site report in Appendix B). There is a strong government commitment to install by the end of 2009 at least one petaflop in *total* performance and to install in 2010–2011 the first single one-petaflop system.

High-Performance Computing in India

India has a system at position 13 on the Top500 list. That computer installed at the Computational Research Labs (<http://crlindia.com/index.htm>) and funded by Tata & Sons, Inc., delivers 117 Tflops on the Linpack benchmark. It is based on Hewlett-Packard technology and a proprietary interconnect. Overall, the HPC presence in India is currently weak, and researchers are not driven to push their problems to large HPC environments.

Future Directions for High-Performance Computing

Looking to Exascale

During the spring of 2007, researchers at Argonne National Laboratory, Lawrence Berkeley Laboratory, and Oak Ridge National Laboratory held three town hall meetings to chart future directions on exascale computing systems, their hardware technology, software, and algorithms, as well as appropriate applications for such systems. The goal of the exascale initiative is to achieve exascale performance for a range of SBE&S applications within the next ten years; this is likely feasible by 2017–2019 (see Figure 5.3). These higher-performance systems will consist of 10–100 million minicores with chips as dense as 1,000 cores per socket; it is expected that clock rates will accelerate also, but slowly. New 3D chip packaging will be required, and the interconnects will be based on large-scale optics. There are many technical challenges to be overcome, including power consumption and leakage currents—large-scale integration and reliability issues related to both hardware and software faults and their detection. A three-step path approach was discussed in the town hall meetings, targeting a 25-petaflops system by 2012, a 300-petaflops system by 2015, and a 1.2-exaflops system by 2019.

Special-Purpose Processors

The multicores used in the aforementioned petaflop systems are general-purpose processors, but there has been a lot of activity recently to introduce new ways of computing in SBE&S based on special-purpose processors originally built for other applications. For example, Graphic Processing Units (GPUs, e.g., NVIDIA G80) are used in desktop graphics accelerators, Field-Programmable Gate Arrays (FPGAs) are used for digital signal processing, and Sony and IBM cell broadband engines are used for game console and digital content delivery systems. The first petaflop computer, the Roadrunner, used 12,960 IBM Cell chips to accelerate the computation in what is typically termed a *hybrid computing* mode, i.e., mixing standard general-purpose processors with special purpose processors. GPUs, which are ideally suited for applications with fine-grained parallelism, are already yielding speed-ups of factors of 2 – 200 depending upon the application (source: www.nvidia.com/object/cuda_home.html).

FPGAs, on the other hand, allow for high-performance *reconfigurable computing* (HPRC), i.e., that enables programmers to make substantial changes to the data path itself in addition to being able to control flow. Reconfigurable hardware is tailored to perform a specific task; once the task is completed, the hardware is adjusted to perform other tasks. The main processor controls the behavior of the reconfigurable hardware, as depicted in Figure 5.4. HPRC can provide higher sustained performance, and it requires less power compared to microprocessor based systems—a critical issue in HPC.

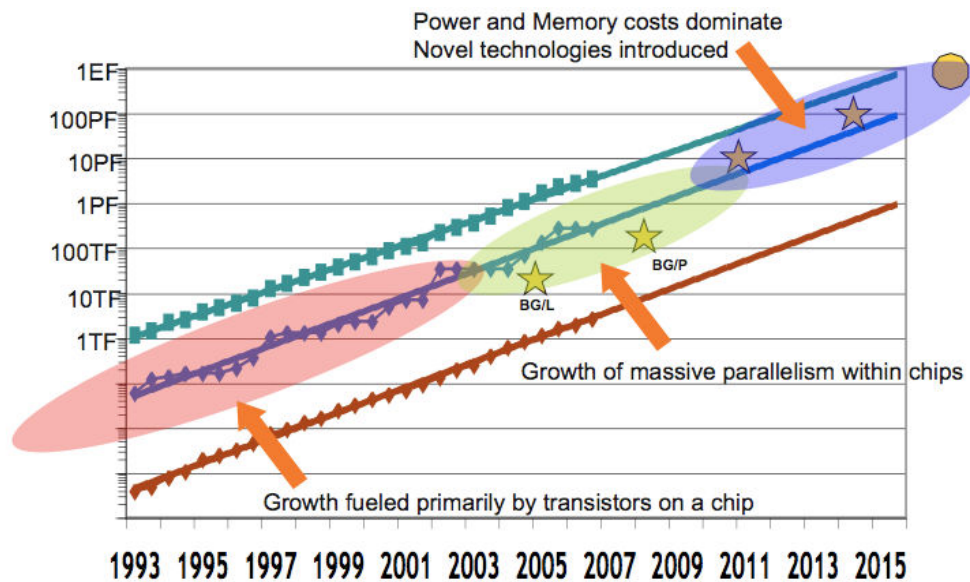


Figure 5.3. The anticipated path from teraflop to exaflop computing is based on three different types of concurrent technology (courtesy of Rick Stevens, Argonne National Laboratory).

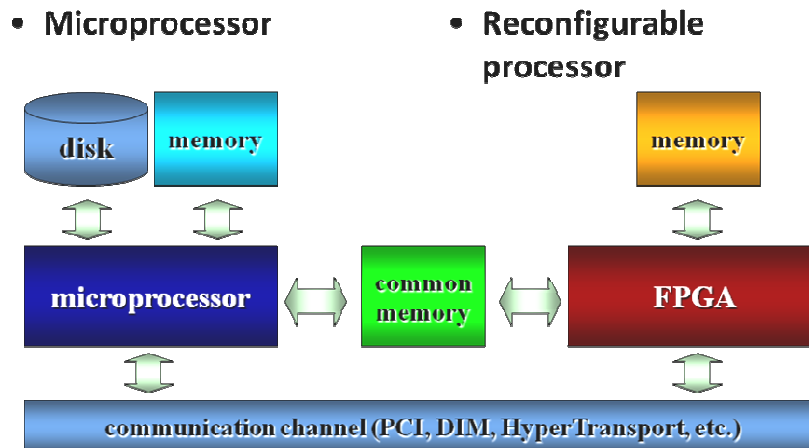


Figure 5.4. Overview of a high-performance reconfigurable system (courtesy of Volodymyr Kindratenko, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign).

A research group at the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign has been evaluating the porting of application codes (cosmology, molecular dynamics, etc) on FPGA-based systems as well as on other special-purpose processors, obtaining significant speedups. Similarly, several U.S. research groups have been performing large-scale simulations on GPU clusters, including full 3D Lattice Boltzmann simulations of urban dispersion modeling in New York City by a group in Stony Brook University and bioinformatics and protein folding by a group at Stanford University. GPUs can be used as a single cluster or for hybrid computing, i.e., as accelerators in conjunction with general-purpose processors. GPUs allow the *data parallel* programming model, because they include more transistors than CPUs devoted to data processing rather than to data caching and flow control. The NVIDIA Tesla GPU was designed specifically for HPC applications.

NEW PROGRAMMING LANGUAGES

Most of the new petascale systems are built on multicores with complex memory hierarchies. From the programming standpoint, these new large multicore systems present another major disruptive technology—a challenge even greater than cluster computing and message passing. According to John Hennessy of Stanford University, this is “the biggest problem Computer Science has ever faced”. Simple porting of highly scalable codes like the QBOX or any computational mechanics codes will not lead to any reasonable efficiency, and it carries the danger of largely underutilizing these very expensive petaflop systems, as the PITAC report suggested. Researchers will need to rewrite their application codes and rethink carefully their corresponding algorithms and software. Numerical libraries such as ScaLAPACK will have to undergo major changes to accommodate multicore and multithread computing. The complexity that this involves can be illustrated with a simple example: Consider a system with 128 cores per socket, with 32 sockets per node, and with 128 nodes per system for a total of 524,288 cores. Let us also assume that the system has four threads of execution per core. Hence, the total number of threads that a researcher has to handle is two million, a large number of threads indeed!

Management of so many threads is an almost impossible task, even for very experienced programmers, and therefore new programming languages are required to deal with this enormous multithreading complexity. New languages from the High Productivity Computing Systems (HPCS) program of the U.S. Defense Advanced Research Projects Agency (DARPA) point the way toward the next-generation programming environment on petaflop computers. Specifically, Partitioned Global Address Space (PGAS) languages for the SPMD (single program multiple data) model, such as Unified Parallel C (UPC), co-Array Fortran, and Titanium, and dynamic languages (X10/IBM, Fortress/Sun, Chapel/Cray), offer many advantages for programming in the new environments. They support both private and shared data, distributed data structures, one-sided memory

communications, synchronization, and collective communications. PGAS languages are a good fit to shared memory computers but also to hybrid shared/distributed architectures, and they have locality that may be important when dealing with more than 100 cores per chip.

UPC, in particular, is expressive enough to allow programmers to hand-optimize. An example is shown in Figure 5.5, where UPC is compared against MPI for gather/scatter operations used in finite element codes; there is a clear factor-of-two speedup by using optimized UPC instead of MPI. In terms of the dynamic languages, X10 emphasizes parallel safety, whereas Fortress and Chapel emphasize expressivity. Fortress uses a math-like representation; X10 and Chapel employ a more traditional programming language front-end. Before the HPC community is likely to adopt these new languages, however, there are many challenges to be addressed, including interoperability with the existing languages, scalable memory models, and parallel I/O procedures.

THE SCALABILITY BOTTLENECK

The main obstacle to realizing petaflop speeds on multi-petaflop-scale systems remains the scalability of algorithms for large-scale problems. For example, even the scalable code QBOX (see Figure 5.1) will have a very poor efficiency on a petaflop system when the size of the problem N is large. The reason is very simple: QBOX and many other FPMD codes employ $O(N^3)$ complexity algorithms, which will not scale well for large values of N ; hence, we need to develop $O(N)$ algorithms for petaflop computing. Other research teams (e.g., see the site report for the University of Zurich in Appendix C) attempt to literally “split” the atom, that is, to develop new domain composition techniques where parts of the calculation for the same atom are distributed on different processors. For codes that involve solution of a linear system, e.g., solutions of the Poisson equation, the bottleneck is the linear system solver. For example, a solver with scaling $O(N^{3/2})$ takes more than 90% of the time, whereas the remaining 10% is spent on the $O(N)$ complexity part of the code. Hence, the entire computation is almost “all-solver” when the size of the problem increases. It is true that most applications groups have not felt this bottleneck—yet, but they will as they start utilizing more than 1,000 processors and scaling up their problems.

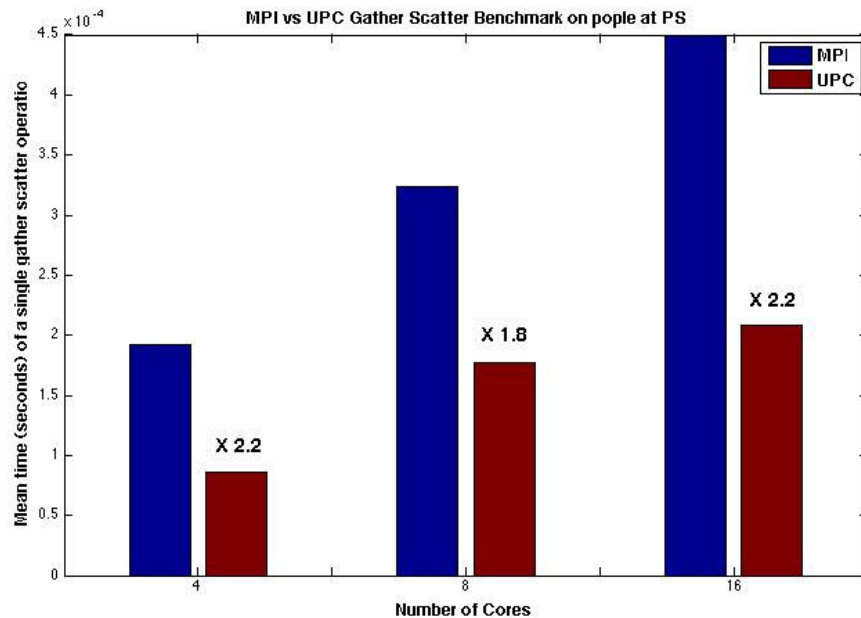


Figure 5.5. Performance (in seconds) of a gather/scatter routine on the SGI® Altix® using MPI and UPC for different number of cores (courtesy of James Noble, Brown University).

Scalable usually implies optimal, which for a numerical analyst implies that the floating point operations grow approximately linearly with the size of the problem N . This, in turn, means that for iterative solvers the number of iterations should be constant or almost constant, e.g., they should be bounded “polylogarithmically.” The number of iterations is strongly dependent on the condition number of the matrix in the linear system unless effective preconditioners are used to weaken this dependency. Unfortunately, some of the best preconditioners are not parallelizable; for example, the often-used geometric multigrid is fundamentally a serial algorithm, unlike the ineffective but embarrassingly parallel Jacobi preconditioner. However, progress can be made by reformulating existing algorithms to be aware of the topology of the computer architecture.

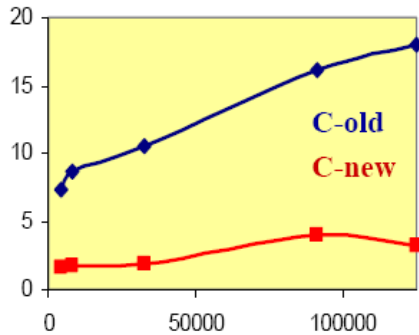


Figure 5.6. Weak scaling (time in seconds versus number of processors) of algebraic multigrid on the Blue Gene/L for up to 125,000 processors, using two different coarsening strategies. The problem size is 15,600 unknowns per processor, with the smallest problem running on 4,096 processors (courtesy of Ulrike M. Yang, Lawrence Livermore National Laboratory).

An example is shown in Figure 5.6, where the algebraic multigrid is employed to solve a linear system on the Blue Gene/L for a problem that increases with the number of processors employed in the computation (weak scaling). In algebraic multigrid, one is not limited by the natural coarse-graining of the grid, and hence aggressive strategies can be employed to achieve high scalability, as shown in the plot, albeit at the price of some loss of the convergence performance. Note that the cost per iteration should also scale proportional to N , and that the Blue Gene architecture allows this favorite scaling with its log-diameter global reduction operation. This example is illustrative of the conflict between fast convergence and scalability and suggests how proper mathematical or computational modifications can be introduced to obtain the most effective parallel solver. To this end, researchers at Columbia University are developing appropriate data transfer strategies to alleviate the communication bottlenecks in the geometric multigrid method as well.

There are two main challenges for scalable computing as we examine *asymptotic* algorithmic scalability. The first one is expressed by *Amdahl's law*, limiting the maximum speedup by the fraction of the purely sequential work, especially for applications

with fixed problem size (strong scaling), e.g., in doubling the number of processors for a fixed problem. The second one is *Little's law* from queuing theory, which, translated into memory bottlenecks, implies that the number of concurrent operations required to avoid waiting on memory is equal to the memory latency in cycles. To deal with these limitations, the simulation scientist has to select carefully the algorithms to be employed, depending on the system architecture. For example, the use of high-order methods increases the memory efficiency and the volume-to-area ratio, hence limiting the required communications. The latter can be overlapped with useful work by splitting operations that use “far away” resources and by employing cross-iteration transformations.

Other approaches may involve development of new algorithms that adapt to memory and concurrency, especially for special-purpose computers, as discussed above. It is also understood from the multigrid example that simple decompositions may lead to large inefficiencies, and similarly, the standard graph partitioning techniques for highly unstructured meshes may be inadequate.

An example of a new type of domain decomposition is shown in Figure 5.7 for simulating blood flow in the cerebrovasculature, including the Circle of Willis (CoW),⁶ the carotids, and other arteries around CoW. The overall domain is split into several subdomains that are colored differently and assigned to different groups of processors; within each subdomain, the Navier-Stokes equations are solved using the high-order spectral element method, while at the interfaces, proper boundary conditions are employed to ensure continuity of the solution and satisfy the conservation principles. Such a problem-decomposition into subproblems is necessary to avoid the

⁶ The Circle of Willis (gray structure at the center of Figure 5.7) is a ring-like arterial network sitting at the base of the brain that functions to evenly distribute oxygen-rich arterial blood throughout the brain.

very large condition number of the linear system if the entire domain is discretized directly, but it is also necessary to provide flexibility in the mapping and optimization of work onto groups of processors for each individual subdomain.

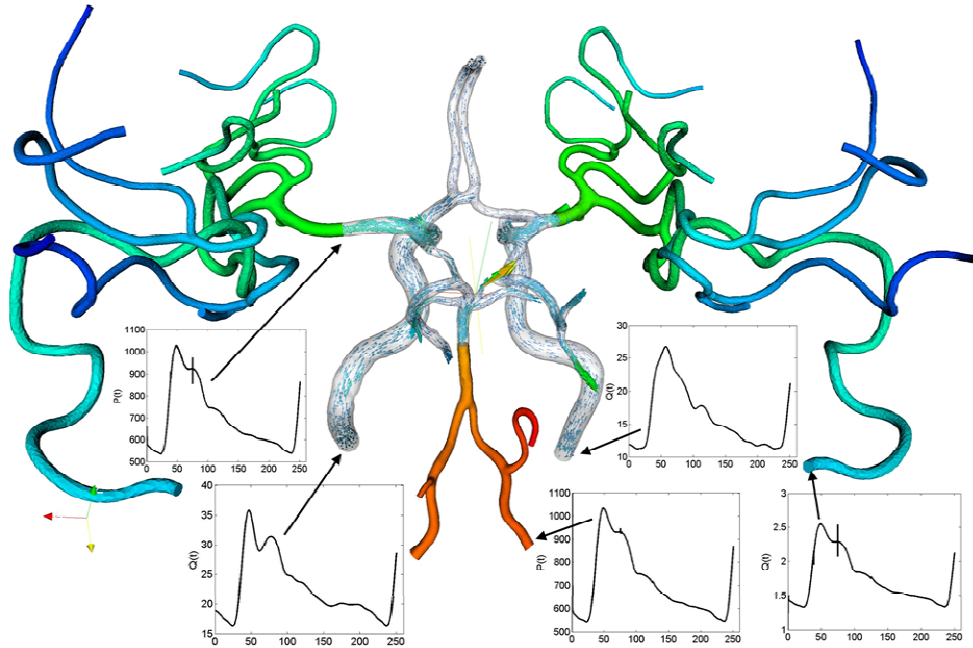


Figure 5.7. Simulation of blood flow in the human cranial arteries using a new multilayer decomposition technique. Domains of different color are loosely coupled, whereas within the domain, tighter coupling is maintained. Computations were performed on the Ranger at the Texas Advanced Computing Center (courtesy of Leopold Grinberg, Brown University).

SUMMARY OF FINDINGS

In the last decade, large investments have been made for developing and installing new computer architectures that are now breaking the petaflop barrier, and we are looking to an exascale era within the next ten years. However, recent investments in mathematical algorithms and high-quality mathematical software are lagging behind. It is appreciated by all SBE&S scientists that the many orders-of-magnitude in speedup required to make significant progress in many disciplines will come from a combination of synergistic advances in hardware, algorithms, and software. Specifically, on the algorithmic front, advances in linear solvers, high-order spatio-temporal discretization, domain decomposition, and adaptivity are required to match the new multicore and special-purpose computer architectures and increase the “effective” performance of these expensive supercomputers.

An example of this synergistic “effective” speedup is offered by a roadmap envisioned by Stephen Jardin of the Princeton Plasma Physics Laboratory and David Keyes of Columbia University on the ITER project, an international research endeavor for fusion energy (Sipics 2006). To simulate the dynamics of ITER for a typical experimental “shot” over scales of interest would require computers with 10^{24} floating point operations, well beyond the terascale platforms available for production computing today. The Jardin-Keyes roadmap outlines a plan on how to achieve the 12 orders of magnitude required to bridge this gap: three orders of magnitude will be gained through hardware, with both processor speed and increased parallelism contributing equally; the other nine orders of magnitude will come from algorithms (software), namely four orders due to adaptive gridding, three orders due to implicit time-stepping, one order due to high-order elements, and one order due to the reformulation of the governing equations in field-line coordinates. Hence, the main speedup (seven orders of magnitude) will come from better spatio-temporal discretization algorithms that will potentially benefit all

SBE&S communities and not just the fusion computational community. Some of these algorithms have already been incorporated in the more advanced parallel codes in the U.S. national laboratories.

Overall, the United States leads both in computer architectures (multicores, special-purpose processors, interconnects) and applied algorithms (e.g., ScaLAPACK, PETSC), but aggressive new initiatives around the world may undermine this position. In particular, the activities of the Department of Energy laboratories have helped to maintain this historic U.S. lead in both hardware and software. However, the picture is not as clear on the development of *theoretical* algorithms; several research groups across Europe (e.g., in Germany, France, and Switzerland) are capitalizing on new “priority programs” and are leading the way in developing fundamentally new work for problems on high dimensions, on $O(N)$ algorithms, and specifically on fast linear solvers related to hierarchical matrices, similar to the pioneering work in Europe on multigrid in the 1980s. Initiatives such as the “Blue Brain” at EPFL and “Deep Computing” at IBM-Zurich (see site reports in Appendix C) have set ambitious goals far beyond what the U.S. computational science community can target at present.

The Japanese government (through the Ministry of Education, Culture, Sports, Science, and Technology, MEXT) is committed to sustaining supercomputing funding and software activities, as indicated by Figure 5.8, and is facilitating partnerships and collaborative projects between national labs, academia, and industry, similar to the model of building and operating the Next Generation Supercomputer at RIKEN (see site report in Appendix B). MEXT is, however, emphasizing hardware and application software, and it is less willing to invest on fundamental work in scientific computing or middleware, unlike the European efforts.

It is clear that barriers to progress in developing next-generation architectures and algorithms are increasingly on the theoretical and software side and that the return onto investment is more favorable on the software side. This is manifested by the fact that the “half-time” of hardware is measured in years; the “half-time” of software is measured in decades; whereas the theoretical algorithms transcend time.

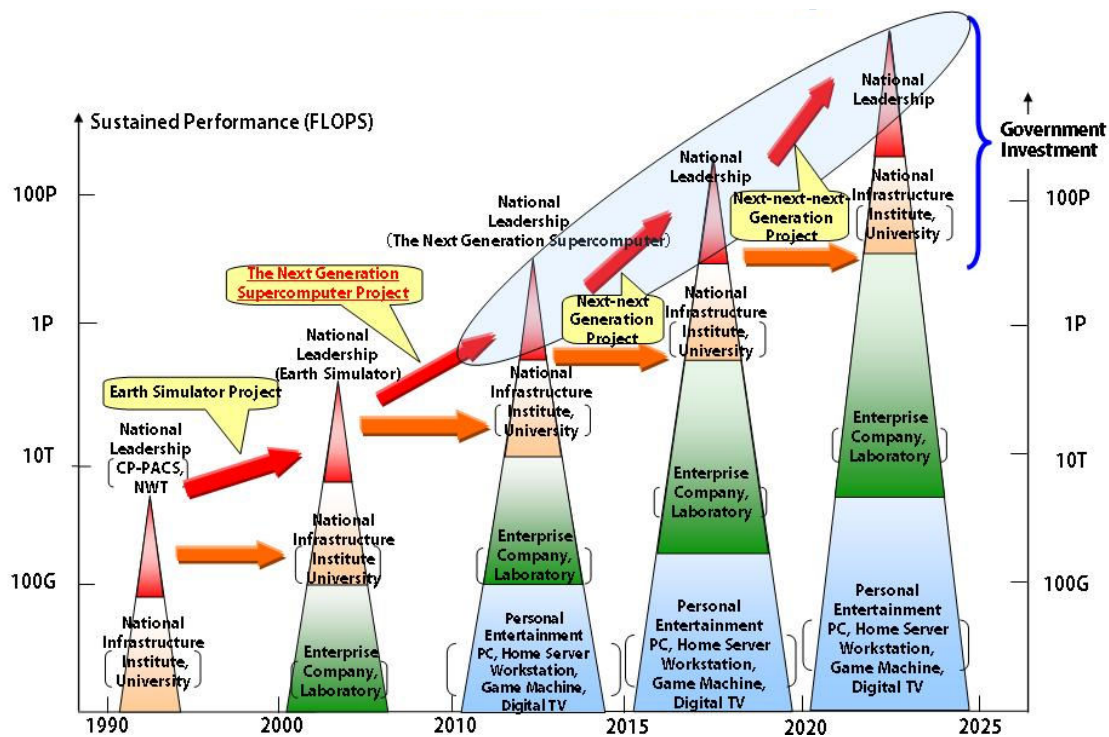


Figure 5.8. The Japanese government has a long-term plan for sustaining leadership in supercomputing (courtesy of Ryutaro Himeno, RIKEN).

ACKNOWLEDGEMENTS

I would like to acknowledge the help and material provided by our hosts (too many to mention separately) in Japan and Europe. The presentations of Jack Dongarra, Bill Gropp, and David Keyes at the November 2007 WTEC workshop reviewing SBE&S research in the United States have formed the basis for this chapter. I am grateful to Jack Dongarra and David Keyes for correcting the original manuscript. I would also like to acknowledge Rick Stevens for providing material related to the exascale project.

REFERENCES

- Brown, D.L., chair. 2008 (June). *Applied mathematics at the U.S. Department of Energy: Past, present and a view to the future*. Panel report prepared by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Available online: <http://siam.org/about/news-siam.php?id=1345>.
- European Strategy Forum for Research Infrastructures (ESFRI). 2006. *European roadmap for research infrastructures: Report 2006*. Luxembourg: Office for Official Publications of the European Communities. Available online: ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf
- Gygi, F., R.K. Yates, J. Lorenz, E.W. Draeger, F. Franchetti, C.W. Ueberhuber, B.R. de Supinski, S. Kral, J.A. Gunnels, and J.C. Sexton. Large-scale first principles molecular dynamics simulations on the BlueGene/L platform using the QBOX code. In *Proceedings of the 2005 ACM/IEEE conference on supercomputing*. Washington, DC: IEEE Computer Society.
- Joy, W., and K. Kennedy, eds. 1999. *Information technology research: Investing in our future*. President's Information Technology Advisory Committee report. Washington DC: PITAC. Available online: <http://www.nitrd.gov/pitac/report/>.
- Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science Foundation blue ribbon panel on simulation-based engineering science*. Arlington, VA: National Science Foundation. Available online: http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.
- Keyes, D., P. Colella, T.H. Dunning, Jr., and W.D. Gropp, eds. 2003. *A science-based case for large-scale simulation*, vol. 1. Washington, DC: Office of Science, U.S. Department of Energy (DOE). Available online: <http://www.pnl.gov/scales/>.
- . 2004. *A science-based case for large-scale simulation*, vol. 2. Washington, DC: DOE Office of Science. Available online: <http://www.pnl.gov/scales/>.
- Sipics, M. 2006. Taking on the ITER challenge: Scientists look to innovative algorithms, petascale computers. *SIAM News* 39(7), September.
- Sterling T., P. Messina, and P.H. Smith. 1995. *Enabling technologies for petaflops computing*. Cambridge, MA: MIT Press.

CHAPTER 6

SOFTWARE DEVELOPMENT

Martin Head-Gordon

INTRODUCTION

The practical applications of computational engineering and science are built around effective utilization of the products resulting from *software development*. These products span a huge range of uses that impact research and development in industries ranging from pharmaceutical and chemical, to semiconductor microelectronics, to advanced materials, to engineering design in automotive and aerospace sectors and many more. The economic significance of these industries is very large – vastly greater than the relatively small companies that typically produce the software itself. Their task is to create integrative tools that bridge basic advances in computer science and the physical and biological sciences from universities with the need for turnkey solutions in engineering design, research and development.

There are fundamental differences between software for computational engineering and science and the software we buy for home or home office. The number of customers is small – for instance there may only be on the order of 100 significant companies in a particular target industry – rather than enormous. At the same time, the simulation problem to be solved – such as perhaps the modeling a catalyst for fuel conversion – involves such complexity and physical realism that the resulting software is often on the order of millions of lines of code, and its development represents research much more than development. Therefore there is a crucial role for government research support, as part of the overall question of research in simulation-based science and engineering.

There are three broad levels of software relevant to simulations. At the topmost level, discipline-specific applications software is what is employed by a user to solve practical problems, ranging from engineering to physical science to biology. That is what a majority of this chapter will focus on. However it is important to realize that most applications software also depends upon two further layers of lower-level tools. At the bottom lies the programming environment (what computer language is employed, and what model is used to direct multiple processors), whose development is an applied branch of computer science. Advances in programming paradigms represent an investment in applied computer science that is essentially a mandatory prerequisite for exploiting changes in computer hardware such as massive parallelism or parallelism on a chip. Programming environments tend to lag behind new developments in hardware, and applications software in turn tends to trail behind the provision of appropriate development environments. Finally in between the programming environment and the applications software often lies a crucial middle layer of software – indeed often called middleware – that might include numerical libraries, collections of objects for building applications or interfaces, etc.

Advances in software capabilities have revolutionized many aspects of engineering design and scientific research already – ranging from computer-aided design in the automotive industry to the development of the

chlorofluorocarbon replacements that have allowed us to start reversing stratospheric ozone depletion. The significance of these end-use examples is why government support for the development of the next generation capabilities is a good strategic investment – particularly when pursued consistently over the long term. A famous example of simulation software at the applications level that originated with basic research is the Aspen program for process design and optimization in chemical engineering. Now used by virtually all the world's leading chemical and oil companies (as well as many others), AspenTech originated as a spin-off from basic research sponsored by the U.S. government starting in the late 1970's. Similarly, one of the most successful pieces of middleware – the dense linear algebra library called LAPack – is an outgrowth of basic U.S.-government-sponsored research in applied mathematics, and now comprises a package that is ported to and then carefully optimized for essentially all significant computer architectures by the vendors themselves, to encourage subsequent porting of applications software that depends on this middleware.

Those are established historical success stories. The purpose of this chapter is to provide a comparative assessment and discussion of the present American position relative to the rest of the world regarding the software development for simulation-based engineering and science. The discussion and analysis will attempt to address questions such as the following. What is the role of government, if any, in the facilitation of software development for simulation-based engineering and science? If such government role is appropriate, then is the 2008 U.S. level of government investment sufficient to ensure leadership capabilities that will permit us to continue to reap the downstream technological and economic benefits? Is any lack of investment something that is particular to simulation-based software, or is it simply part of larger trends? What are some specific examples of areas where lack of investment has had consequences for software leadership, and what, if any, are the kick-on implications? After a series of sections that attempt to address these issues in the context of activities occurring around the world, and particularly in Europe and Asia, the chapter ends with a short summary of main findings.

ROLE OF UNIVERSITIES, NATIONAL LABORATORIES, AND GOVERNMENT

There is worldwide recognition that state-of-the-art simulation software development relies on vibrant activity at the university and national laboratory level from which to obtain new advances – for instance more accurate and realistic simulation models, or new simulation paradigms, or new software engineering advances to enable code development. With globalization, America's basic industrially-sponsored research in computational science and engineering as previously represented by flagships such as Bell Laboratories and IBM's Research Division has declined. While there is still tremendous investment in product-level development, the gap in basic research has had to be made up through increased activities at universities and government laboratories. Furthermore, in science and engineering, software development activities are tightly integrated with innovations in simulation models and algorithms, as well as training. Therefore a strong university software development community working on simulation methods trains the scientists and engineers who will then deploy simulation methods in industry or further develop them in their own research careers. It also produces software that might not otherwise be developed by industry, as shall be discussed later in this section.

Fundamentally, we have found the situation to be broadly similar in America's main Asian and European rivals. The leading industrial sites we were able to visit described their main simulation-based activities as involving the application of tools that were largely developed elsewhere (ie. in universities and national laboratories). For instance, at the research laboratories of Nissan in Japan, simulation-based modeling is being used as a way of exploring new ideas for improvements in fuel cells, for future vehicles (see site report). The second example was Mitsubishi Chemical (see site report), where simulation software is extensively employed for process simulation, polymer modeling, and materials development problems, as well as others. The diversity of simulation needs relative to the number of staff directly involved makes software development prohibitive. However Mitsubishi researchers have developed software tools to allow information passing from one program to another (for instance from chemical reaction modeling to flow modeling software), and additionally have contributed to community codes such as AbInit (Gonze et al. 2002). At Toyota Corporation Central Research and Development Laboratories, not only were simulation methods applied to relevant problems, including fuel cell development, aspects of efficient combustion, crash modeling, reduction of engine vibrations, optimizing aerodynamics, but a number of the key pieces of software were developed in-house.

In Europe, the WTEC panel visited BASF, probably the world's leading chemical company, which has extensive applications-based usage of simulation methods. On the whole, these efforts use codes that are developed elsewhere, although in areas where BASF feels there are critical unmet development needs, or where they feel the company could benefit significantly from advances in simulation capabilities, they follow the strategy of funding extramural research at university groups who can supply such tools (for example in process scheduling), or they participate actively in European Union research networks (for example BASF coordinates one initiative entitled NanoModel, and also participates in a consortium for the development of theoretical chemistry software). This model of tracking developments in academic research that are of relevance to the company in turn couples to the extensive uses of simulation within BASF. Within China, simulation activity in industrial research is still quite limited with the exception of the oil industry (Sinopec) which we could not visit. Academic researchers whom we talked to felt that there would be greater adoption of simulation methods within industry in the near future, as Chinese industry aims to move up the so-called value chain towards design as well as manufacturing. It is nearly certain that universities will play a crucial enabling role in software development as this transition develops.

Of course the reason that simulation codes are generally not directly developed in industry is because they are only a means towards a vastly more valuable end – the target drug, or fuel cell or aerospace product, which is the real goal of corporate activity. In fact, because the market for simulation codes is so small (perhaps 100 leading chemical companies, perhaps 100 oil companies, perhaps 100 drug companies, etc.), there is only marginal value in a software industry to produce the simulation software directly. The value equation is further tilted by the fact that the simulation software is more often than not directly based on new results from research into simulation models and algorithms. Thus the production of software is inseparable from long-range basic research into models and algorithms. The third (and strongly related) factor disfavoring direct industrial development of simulation software is discussed in more detail in the section on the complexity inherent in state-of-the-art simulation software. This means that the development time can be very long.

For all these reasons, in fields such as materials design, drug design, and chemical processing, where these software factors are at play, software development at universities and national laboratories plays a crucial role. In effect they serve as engines to produce vastly subsidized products for use in industrial research and development – products that for the reasons discussed above would probably not otherwise exist. This transfers direct costs and risk away from the industries that benefit from the use of the resulting software, but would likely not be willing to directly support the real cost of the development of the tools. Of course, in some other fields, such as computer-aided design and computer-aided manufacturing, where the market is larger and the tools are more standardized, this is not so much the case in 2008. However, the closer we are to the frontier applications areas of simulation, the greater the role of universities and national laboratories becomes in pioneering the new simulation capabilities and directly providing the software necessary to deploy such capabilities.

In turn, this leads us directly to the need for vigorous government funding of software development for simulation-based research in engineering and science. The situation in this regard will be discussed in more detail in the later section on funding trends. It also leads us directly to the need for effective interactions between universities and industries – such as software companies to commercialize university developments. Going beyond funding, this leads to the question of how efficiently and fairly intellectual property (IP) rights can be obtained for software developments performed at universities to enable commercialization. We heard many comments from both industry and European academics that American universities are at present much more difficult to deal with regarding IP issues. In turn this is leading many companies, even American companies, to deal with foreign universities as more attractive partners for advances in research and development. Of course this situation is not restricted to software development by any means, but in fact applies quite broadly, and is one that American universities should urgently respond to. Whether the future will see foreign universities more aggressively asserting IP rights similar to their American counterparts or whether they will continue to exploit their present competitive advantage is of course an open question.

SOFTWARE LIFE CYCLE – MANAGING COMPLEXITY

Nontrivial simulation software for production calculations blends algorithmic and computing advances through typically tens to hundreds of man-years of development and often up to millions of lines of source code. *The*

resulting software life-cycle usually exceeds the life-span of computing paradigms, and certainly exceeds the length of many funding initiatives. It is important to recognize this reality, and the associated implication that the greatest rewards are the result of sustained investment in high-quality work. Of course there are exceptions, where well-defined technological developments can be greatly boosted by relatively short-term strategic investments (for instance, the development of new programming models and middleware for the coming massively multicore hardware – see the section below on Emerging Opportunities in Software Development), but these become relatively scarce in the applications disciplines. In this section we briefly discuss the reasons for the long life-cycle, the consequences for software development in universities around the world, and how trends in funding affect software development activity.

Typically a sophisticated simulation software package that is built from an initially blank slate will take between 5 and 10 years to be mature enough to be released. While there are exceptions that we shall mention later, many such efforts are built with a small nucleus of highly committed scientists and engineers as the initial developers. They may typically be led by a professor and involve a small number of graduate students or postdocs who essentially launch their research careers in the process of developing efficient implementations of novel simulation algorithms (and indeed sometimes developing the algorithms themselves). The fact that new simulation codes are most worth developing when there are new ideas to base them on means that significant effort is involved, including perhaps several false starts and wrong turns. Furthermore the scale of a working production code and the need to cover all bases --- from new discipline-specific algorithms to producing appropriate tools for building inputs and analyzing outputs – means that a large code base is required. If the effort is aiming to produce a commercializable product in the end, then all components above middleware may well need to be constructed from scratch. Even new public domain codes tend to build all pieces as new simply because of advances in software languages, and possibly changes in target computer architectures.

In the section on World Trends in Simulation Software Development, we discuss electronic structure software as a case study. Within this area, two relatively recent “clean sheet” software designs coming from Europe are the ONETEP (<http://www.onetep.soton.ac.uk>) and Conquest codes (<http://www.conquest.ucl.ac.uk>) (as new codes, they do not yet qualify as widely used codes, documented there), and serve as examples of this small group paradigm. They were efforts initiated to try to achieve simulation costs that scale only linearly with the number of atoms in the calculation (in contrast to the conventional cubic scaling with particle number for conventional codes), and, as a second goal, be capable of running efficiently on parallel computers. Development of these codes began around 2000, and both are now just reaching the stage where they are usable by simulators beyond the immediate developers. The students and postdocs who have done the hard work of algorithm and code development were not able to publish as many papers as their contemporaries who were working on applications of simulation codes to systems of topical interest in materials science or nanoscience. However, they have contributed significantly to the development of linear scaling methodology, as well as addressing important software engineering aspects such as support for sparse matrices that exploits the particular aspects of electronic structure for efficiency. The key junior figures in both codes have all obtained academic positions in the United Kingdom, which indicates an encouraging degree of recognition of the significance of their work by the U.K. materials science community despite relatively limited ability to publish many scientific papers during the slow process of constructing their programs.

Given the long lead times for development of simulation codes, it is natural that the resulting packages have lifetimes that, in healthy, successful cases, are typically longer than 10 years, and in many cases are measured in decades. The lifetime is related to the complexity of the application, and the variety of functionality that must be supported. New codes that support only limited functionality may find initial success through offering greater performance or a particular capability that is unique. By contrast, mature and established codes continue to find success by providing a greater range of functionality, or even simply a familiar user interface. Furthermore, they are often supported by significant developer communities which provide ongoing enhancements. It is only when their performance becomes uncompetitive and their range of functionality is matched by newer rivals (and ceases to be expanded) that they become eclipsed. This situation is certainly very much the case in many fields. For example, in computational quantum chemistry, the Gaussian and GAMESS programs both have roots stretching back to the early 1970s, so that as of 2008, they are approaching 40 years old. In biomolecular simulations, the

first programs to emerge in the early 1980s, CHARMM (see Charrmm Online) and Amber (see Amber Online), are still among the leading programs employed as of late 2008.

The consequences for support of software development and for the rate of adoption of new computing paradigms and technologies are interesting to consider. A software development project starting from a clean new design will not be completed during the lifecycle of a typical funding initiative. An example of this type is the NWChem program for computational chemistry that was initiated as a public domain effort as part of the construction project for the Environmental Molecular Science Laboratory at PNNL, running from 1990 to 1995. Shifts in scientific computing paradigms, such as vector processors in the 1980s, and massively parallel computers more recently, will be adopted more gradually than the hardware permits, as software trails in its wake. Codes that are relatively simple will be the first to be ported or rewritten, while complex multimillion line codes will migrate more gradually. The bulk of scientific computing is accordingly performed on established architectures where there is a suitable balance between hardware cost and ease of installation/maintenance, and software availability, as dictated by the factors discussed here that currently determine it.

From the applications software perspective, it is desirable to see improved developer tools that can abstract as many details of the architecture of both present and coming computers into natural constructs of programming languages. This is a clear area of challenge for applied computer scientists, in collaboration with hardware vendors, and applications developers. Looking ahead to programming paradigms that improve present languages such as C++ and Fortran 95 by permitting easier abstraction of data and functionality, without unduly compromising performance is one such goal. It may also dovetail with the need to go beyond current tools for controlling parallelism such as OpenMP (primarily for shared memory small scale parallelism) and MPI (for distributed memory parallel machines). The ready availability of robust and usable development tools that support the coming generations of new hardware will in large measure determine how quickly and broadly it is adopted – either at the desktop level or at the supercomputer level. One may also hope that improved software development tools will help to shorten the software development cycle, and therefore more effectively enable the development of next generation applications. This can in part be true, although it must again be stressed that the development of truly new applications is equally bottlenecked by the intellectual barriers to developing the new simulation models and algorithms that define them. It is not clear, for instance, that multimillion-line codes based on dated languages such as Fortran 77, are appreciably handicapped in practice against codes that are developed in more up-to-date languages, at least as measured by their continued usage and evolution. For example, several of the codes being ported for use with the Japanese next generation supercomputer are indeed Fortran 77 programs.

Finally, another aspect of complexity is that with increasing focus on multiscale modeling (see Chapter 9), there is a need for improved interoperability of different software programs. This was a need expressed by practicing simulation scientists and engineers at a number of sites visited by the WTEC panel, such as Mitsubishi in Japan and BASF in Germany. For example, software that models chemistry at the atomistic level should be able to provide inputs to software working on longer-length scales that describes fluid dynamics. Or programs with complementary capabilities should be able to exchange data so that one provides input for the other. Neither of these needs is adequately met in the major applications disciplines at present. This could be met by a targeted initiative in this particular form of middleware.

SUPERCOMPUTING SOFTWARE VERSUS SOFTWARE FOR MIDRANGE COMPUTING

Supercomputers represent the leading edge of hardware development, and correspondingly offer the ultimate potential for applications performance at any given time. With the further passage of time, advances made at this level can, when appropriate, trickle down to benefit future developments in midrange computing software. As discussed in detail in Chapter 5, we are in the petaflop era in 2008, and we expect that by early in the next decade, the world's leading computers will be in the 10 petaflop class. The development of software for these machines poses challenges that are at least equal to those of previous generations of supercomputers. The reason is that the new hardware is dramatically different from those where software development is concentrated today – in particular the extent of parallelism is vastly greater. Indeed the degree of parallelism will be the primary differentiator between midrange computing and supercomputing. Most nonclassified supercomputer software

development is directed towards long-term grand-challenge scientific applications of the type described in previous chapters – spanning materials, energy, nanoscience, health and medicine. Some applications with a more immediate focus include oil industry simulations of reservoirs, and turbulent combustion for engine or turbine design. However given the tremendous cost of supercomputing and the need for customized software, it is only in industries where the results of simulations are most mission-critical where it can be justified.

Japan is investing heavily in software development for the next-generation supercomputer project, an effort running from 2006 to 2012 to replace the Earth Simulator. This software investment is because it would be a tremendous waste of resources to produce the hardware, without a corresponding effort to have software in place for this machine, as soon as possible after it is commissioned. Therefore a set of 6 scientific software challenges have been identified, three originating primarily from the physics community and three from the chemistry community. The physics-derived software will comprise codes for real-space density functional theory (DFT), dynamical density matrix renormalization group (DMRG), and quantum Monte Carlo (QMC). The chemistry-derived applications will comprise software for molecular dynamics (MD), the reference interaction site model (RISM) approach to solvation, and the fragment molecular orbital (FMO) approach to treating the electronic structure of very large molecules. One particularly interesting aspect of the software development projects is that they are underway now before the exact form of the next generation supercomputer is known. This is to allow time for innovation in the algorithms and models, rather than simply being forced to adapt existing codes to the machine when it is available. Since it is widely recognized (see Chapter 5) that advances in applications performance derive at least as much from improvements in models and algorithms as hardware advances, this is a very reasonable attempt to blend the realities of supercomputer hardware development with those of software development.

A second example of very active supercomputer software development was seen at Karlsruhe Institute of Technology (KIT) in Germany. KIT is home to the Steinbuch Center for Computing (SCC), which, in terms of personnel, is Europe's largest with more than 200 staff members. In contrast to supercomputer centers in the United States, the SCC has a number of professors whose positions are explicitly within the center. One of these positions is in applications sciences, and involves the development of a variety of simulation software, with capabilities ranging from modeling hull design for an entry in the America's cup, to the modeling of air flow through the nose. All this software development is within an open source model. Other professors have specializations in numerical methods and aspects of high performance computing research. This embedding of long-term research projects in high-performance computing into a supercomputer center is an interesting alternative to the U.S. system where most activity at supercomputer centers is more closely tied to supporting users.

A third example of supercomputing software development was at the Institut Français du Pétrole in Paris, which performs a wide and deep variety of simulations in support of basic and applied petroleum research (see site report). In areas such as molecular modeling, they employ third-party codes. However they have an extensive effort in high-performance computing software development, such as computational fluid dynamics codes to enable the design of improved internal combustion engines. New-generation large-eddy simulation (LES) codes specifically enable modeling of fast transients, cold starts, etc. Previous-generation codes have been commercialized and then more broadly distributed. Parallel scaling of research codes was demonstrated to the WTEC panel on up to 4000 processors, consistent with the IFP goal of keeping a high-performance computing capability that is roughly a factor of ten behind the state of the art.

While supercomputer applications are valuable as discussed above, support for software development targeted specifically at supercomputers must be tempered by the fact that its impact is relatively narrow in the sense that the vast majority of useful simulations in scientific and engineering research is accomplished using midrange (or "capacity") computing, usually on clusters that are controlled by small groups of researchers. Relative to fifteen years ago, the most remarkable development in scientific computing is the merging of the standard office personal computer with the scientific workstation – there has been no meaningful distinction for a decade or so. Midrange computing is accomplished with clusters that are essentially collections of personal computers, with an extraordinary reduction in the cost relative to the situation that previously existed when such machines were specialized research tools.

This has contributed to two important world trends, both of which were evident in our study trip. First, due in part to the tremendously decreased cost of the computing resources is an enormous broadening of the user base in many fields. This can be illustrated for computational quantum chemistry in the academic literature by the data shown in Figure 6.1. It is clear that there has indeed been a tremendous increase in adoption of these simulation methods during the period 1999-2007. This is an example of the usage penetrating far beyond the (relatively small) community of simulation experts and into a much larger community of nonspecialist users. This reflects both the diminished cost of entry and the increasing capability of the methods themselves. While it is not possible to survey many other fields in the same way, we strongly suspect that similar considerations apply.

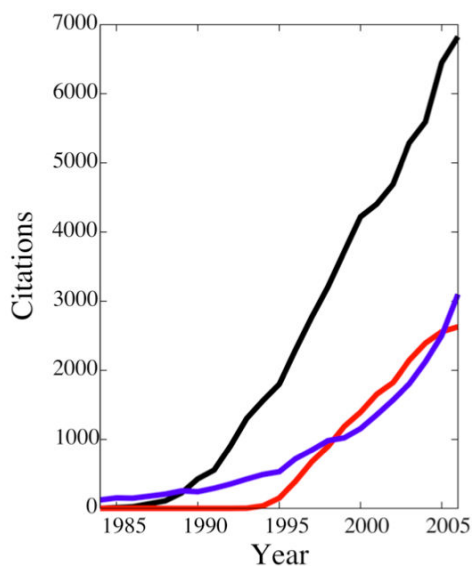


Figure 6.1. Development of the usage of computational quantum chemistry, as measured by two different metrics. The top curve gives the number of citations to software packages per year, while the lower curves provide the number of citations of particular electronic structure methods (specifically pure and hybrid density functionals). Data are from the Web Science (courtesy of Emilio Artacho). Note it would be interesting to illustrate a modified version that tracks U.S. and worldwide usage separately.

The second trend concerns the expert community itself. Groups that develop and apply simulation methods in the United States traditionally had a large advantage in their computational facilities relative to rivals in either Europe or Asia. This is no longer clearly the case now when the dominant cost in the United States at the midrange level is labor rather than equipment. Indeed the panel found that development resources were broadening across the world and leveling out.

We conclude this section by observing that strategic investments in simulation software development must in the end balance support for advancing the state of the art in supercomputing, with support for innovative developments that affect the far larger communities that rely on midrange computing. As a senior simulation scientist at Karlsruhe, site of Europe's largest supercomputer center (by personnel) remarked: "supercomputing is fine but most science gets along without it."

WORLD TRENDS IN SIMULATION SOFTWARE DEVELOPMENT

It is not possible to cover all areas of software applications development, or even a significant fraction, in any meaningful detail in a document of reasonable length. The approach that will be taken here is to first discuss some generalities about the present world situation regarding simulation software development. We shall consider middleware briefly, and then focus primarily on scientific and engineering applications software. We then consider software development in a representative applications area in some detail (electronic structure

codes in condensed matter physics and chemistry). This permits some specific illustrations of general trends, which at least in part, show a strong shift away from American leadership in software development, and some of the consequences.

We first discuss some broad generalities. In 2008, the US continues to dominate in the development of system-level tools and middleware, as has been the case virtually since the large-scale computing first became viable in the 1950's and 1960's. In large part this is because of the strength of American leadership in hardware development, and the resulting impetus from hardware manufacturers to spur the development of appropriate software tools that enable effective usage of their platforms. At the supercomputing level, this is often supported by developments at national laboratories and universities because of the relatively small scale of the market. There has been a trend towards moving these tools into the public domain, driven by the success of the open-source development of the Linux operating system and the closely related Gnu software tools. Thus obtaining access to existing system-level software and middleware is now easier and less expensive than it has ever been. Without a doubt this has contributed to a leveling of the playing field for applications development, since there is no clear-cut technical advantage in the tools of software development (or, to a considerable extent in the hardware, as discussed in the midrange computing above).

Leadership in applications software development, therefore, is not at present determined by access to either hardware or development software, and will instead be determined by other factors. Those factors are diverse, but clearly include the strength of the relevant scientific and engineering research communities in the particular discipline, and the incentives or disincentives for software development activity relative to applications of existing software to solve problems of interest. Those incentives and disincentives revolve around what type of work is considered of highest priority for funding, the level of professional recognition and rewards in software development activity relative to targeted applications, and the appropriateness of either activity for graduate student education and training. Indeed, it is possible to even adopt the strategic view that because of the level playing field for hardware and standard middleware, new investments can be focused on trying to drive the applications. In software development for Japan's next generation supercomputer, discussed above, this point of view has been adopted – there will be no concerted development of middleware because of their ability to obtain American-developed middleware, and their desire to compete for software development leadership in applications that impact energy, new materials, and biology.

Our impression is that on the whole, American leadership in applications software development for simulation-based science and engineering has diminished quite noticeably relative to a decade ago (2008 versus 1998). In some areas, U.S.-based applications software development is now clearly trailing competitive efforts in Europe particularly. Japan also has some examples of flagship software development (such as supercomputing applications), but their impact in terms of widespread distribution for mid-range capacity computing has been more limited. One examples of such an area is computational condensed matter physics, discussed below. The related area of computational quantum chemistry codes has increasingly strong world-wide competition, but the most widely used commodity codes are still American-developed. In other areas, Europe has also made large strategic investments in software development, such as preparing for the Large Hadron Collider (LHC) experiment, and thus the strategic balance is changing. The panel found that China at present (2008) is not a major player in software development for simulation-based science and engineering. However the strong push to raise the level of excellence of the leading Chinese universities is leading to a new generation of talented academic scientists who are training talented students in increasingly large numbers, and we should expect a greatly increased Chinese presence in simulation software development in the future. India is also increasingly laying the groundwork for a future presence in simulation software development, due to the combination of a strong presence in commercial software development, a strong academic tradition at its leading universities, and now increasing government investment.

Let us now turn to discuss software for electronic structure calculations as a specific example. Such software solves sophisticated approximations to formally exact but practically intractable quantum mechanics to describe the behavior of electrons in bulk solids (including new materials), surfaces and interfaces (of relevance for catalysis – see the WTEC report on catalysis research by Davis et al. 2009), and molecules and nanoscale materials (relevant for chemical reactions generally). Software development in this area was pioneered in the

United States starting in the late 1960's, and, with the passage of time, reached a level of sophistication that began to have a substantial impact on not only academic but also industrial applications. One such example was its use in the 1990s to assist in the design of replacements for chlorofluorocarbons (CFCs) which were by then recognized as contributing to the destruction of the ozone layer. The American physicist, Walter Kohn, who developed the theorems that underlie modern electronic structure software, and the chemist, John Pople, who led the development of the software (the Gaussian series of programs) that first made a major impact on diverse areas of chemical research, in fact shared the Nobel Prize in Chemistry in 1998.

To assess the software development situation in 2008, we shall first discuss the codes that are used to treat extended materials, which are primarily developed in the condensed matter physics community. The first codes in this area were developed by leading American researchers at institutions including Berkeley, Northwestern, Lawrence Livermore, Los Alamos, and many others. Today, however, the codes that are most widely used for practical applications come almost entirely from Europe. They are summarized in Table 6.1 below. These software programs include both commercial programs and public domain programs. They are dominant in applications calculations around the world, including the United States, where it is common for leading research groups to employ them. This is despite the fact that scientific leadership in the research that drives new developments in the programs is more evenly balanced than the lopsided situation represented by the national origin of the software packages shown in Table 6.1. Indeed, there are still numerous “in-house” programs in the United States that exist within one or a handful of groups. These codes, it appears, are mostly used to explore new ideas, though in some cases they are also used for “in-house” production calculations. However, as nondistributed codes, they are often lacking either in performance or standard features, and therefore even in groups that have such programs, it is common for them to employ the standard European codes for production calculations.

Table 6.1. Widely Distributed Physics Codes for Electronic Structure Calculations on Extended Systems

Program name	Function	Type	Origin
VASP	Plane wave DFT	Commercial	Vienna, Austria
CASTEP	Plane wave DFT	Commercial	Cambridge, United Kingdom
Siesta	Local orbital DFT	Public domain	Barcelona, Spain
CP2K	Plane wave DFT	Public domain	Zurich, Switzerland
Quickstep	Local orbital DFT	Public domain	Zurich, Switzerland
Quantum Espresso	Plane wave DFT	Public domain	Princeton, United States
CRYSTAL	Gaussian orbital DFT/HF	Commercial	Pisa, Italy

Source: WTEC panel.

In Table 6.2, the codes that are widely used to treat molecular systems are shown. They are developed primarily in the computational quantum chemistry community. The situation here is far better balanced between Europe and the United States than is the case for physics, but this author feels that there has been a relative change in the amount of activity in the United States relative to the rest of the world. In straightforward terms, the rest of the world is catching up, and even the U.S.-developed codes often involve substantial and important contributions from groups based in Europe or Japan. The U.S.-based Gaussian program, for example, includes large amounts of software developed in Japan (Kyoto University) as well as Europe (groups in Italy and England). Likewise, the U.S.-originated Q-Chem program includes key contributions from groups in Australia and Germany. The public domain GAMESS program is not significantly different, with important contributions from Japan (IMS). Therefore some of the slippage in the U.S. position is disguised by the historical advantage that major codes originated in the United States, but it is occurring nonetheless.

Table 6.2. Widely Distributed Chemistry Codes for Electronic Structure Calculations on Molecular Systems

Program name	Function	Type	Origin
Gaussian	Gaussian quantum chem.	Commercial	Connecticut, United States
GAMESS	Gaussian quantum chem.	Public domain	Iowa, United States
Q-Chem/Spartan	Gaussian quantum chem.	Commercial	Pittsburgh, United States
NWChem	Gaussian quantum chem.	Public domain	Washington, United States
ADF	Local orbital DFT	Commercial	Amsterdam, Holland
TurboMole	Gaussian quantum chem.	Commercial	Karlsruhe, Germany
MolPro	Gaussian quantum chem.	Commercial	Stuttgart, Germany
MolCAS	Gaussian quantum chem.	Commercial	Lund, Sweden
Dalton	Gaussian quantum chem.	Public domain	Oslo, Norway

Source: WTEC panel.

What are the consequences of a loss of U.S. leadership in software development for an applications area such as condensed matter physics electronic structure programs? They begin with issues that may seem essentially benign such as U.S. researchers acquiring the European-sourced codes to enhance their often-leading research on applications problems using simulations. This is arguably not so different from the world reliance on U.S.-developed middleware discussed above. However, it is important to remember that the United States restricts the export of high technology items to organizations that may use them to compete militarily against the United States. With software applications dominated by overseas developers, the same effect can and does start to occur in reverse. The WTEC panel was made aware of one specific recent example of this type in the computational condensed matter physics area. The Siesta program, noted in Table 6.1, has an add-on module for calculating transport properties, TranSiesta, with development headquartered in Spain. The request for a license for this program for the AFOSR research base was refused by the TranSiesta developers based on the principle of not contributing to research with potential military applications. Therefore when simulation software reaches the stage of having the ability to contribute to research on issues regarding national security, the consequences of a loss of national leadership in software development can, and in this case, did, translate to inability to access that software.

There are also consequences for leadership in developing the new ideas for improved algorithms or new simulation models in the future. Since there is a close relationship between producing new ideas and turning them into software (without the latter, the former is not proven useful), there is a real possibility that leadership in software development will synergize with leadership in developing the new simulation paradigms. There is also the question of student training (see Chapter 11 for much more detail on this problem), which can only be effective for software development if research is performed on a state-of-the-art program. Furthermore, students trained primarily to run simulation software are less flexible simulation scientists and engineers than those who are also trained to be able to handle large-scale software development projects.

COMPARATIVE ASPECTS OF FUNDING FOR APPLICATIONS SOFTWARE DEVELOPMENT

We have presented evidence that concerted European and Japanese efforts are leading to a tipping of the competitive balance in some fields, while China is fast developing. Generally, this is a reflection of larger trends – U.S. investment in science and engineering research of most types has been gradually declining in real terms during this decade, while overseas rivals have been enhancing their own spending. In the case of China, the doubling time for the overall science budget has been on the order of every 5 years since around 1990. The most important single thing that could be done is to reinvigorate U.S. research in science and engineering at the single investigator level, based simply on the criterion of excellence. Within the United States numerous researchers have complained to the panel that in addition to gradual budget cuts, agencies such as the National Science Foundation are overemphasizing so-called broader impacts (communicating research results and educating

society at large) relative to scientific excellence, as well as overemphasizing short-term initiatives relative to the long-term health of basic research.

However, in light of the long software development process, there are also specific challenges in obtaining funding for simulation software development within the United States. A typical funding initiative will involve a high level of research support for a period such as 3 to 5 years, sometimes with the option for a second renewal period subject to achieving satisfactory performance in the first grant. Such a timescale is only marginally suitable for new software construction – though it can be very valuable for enhancing software that is already established. In general, however, short-term U.S. (and overseas) initiatives do not promote the long-term health of innovation-driven software development – rather they tend to encourage novel applications of the tools that already exist. In turn this reduction in applications software development in the United States is leading to reduced production of appropriately trained students which will threaten our long-run competitive position.

Indeed several European investigators commented on the fact that there have been numerous short-term initiatives to build new U.S.-based codes that have then ceased or stagnated when funding ended. They also commented that this situation tended to make U.S.-based scientists somewhat unreliable partners for long-term software development projects. While the academic simulation software development situation is not entirely satisfactory in either Europe or Asia either, it is nonetheless striking that Europe in particular has been able to better sustain a strong tradition of both community and commercial code development. European researchers complained that they were unable to hire personnel specifically for programming tasks, and that obtaining genuinely long-term support for software development was very hard. One exception is that long-term software maintenance of community chemistry codes is supported to a modest degree in the United Kingdom by several long-term Computational Chemistry Projects (CCP's).

One may argue that this state of affairs has at least the merit of encouraging innovation in any new software development project, as that is the primary mechanism by which funding can be obtained. Longer term, the commercialization of software is the most widely used mechanism to obtain funding for code maintenance and development. This approach applies to virtually all of the codes that are commercial which were listed in Tables 6.1 and 6.2, and this is a real incentive to make a code commercial rather than public domain. Users thereby support part of the cost of ongoing development, even though government provides essentially all funding for the addition of new algorithms and models which can be classified as novel research. If government wishes to specifically encourage the development of open source and/or public domain codes, then it is essential that some support for programmers be provided, independent of research funding.

A few words should be added on the type of teams that develop applications-based simulation software. Due to the size and longevity of software codes, they are usually developed by large teams. Generally such teams are based in multiple locations and are quite loosely coupled. Different team members typically work on different pieces of functionality. It is important to note that such teams are often based around a very small core of people at a central site who carry primary responsibility for managing the collaboration and the code. There are countless such examples, such as documented in Tables 6.1 and 6.2. Therefore funding the development of codes does not necessarily require funding the entire team, but may naturally separate into smaller grants that can be judged based on traditional requirements of innovation and excellence.

Of course these considerations apply to software development as the WTEC panel has seen it in action in 2008. It is possible to imagine that higher levels of funding could support an entire team, including team members such as applied computer scientists or professional programmers who are not often incorporated in typical simulation code teams today. However, this type of support will only be effective if provided for a sufficient period of time to be commensurate with the decade time scale that the WTEC panel has typically observed for the development of state-of-the-art simulation software.

EMERGING OPPORTUNITIES IN SOFTWARE DEVELOPMENT

The specifics of innovation in software development across the vast diversity of active disciplines in simulation-based science and engineering are not predictable. However, the synergistic effects of continuing exponential

improvement in hardware performance (including that which can be obtained for an individual researcher, as well as supercomputers) together with improvements and sometimes revolutions in simulation models and algorithms guarantee new opportunities. We could not know that investments in computational chemistry in the 1970s and 1980s would help to save the ozone layer in the 1990s. The energy, materials and health challenges discussed in earlier chapters set the broad context where strategically important benefits are likely to be gained in the future as the result of new simulation capabilities. Largely it will be ensuring dynamism within the applications fields in general that will be crucial rather than large centrally organized investments in computational science and engineering. The latter can only serve to partially and unevenly compensate for a general lack of investment in research in the relevant disciplines.

An important theme regarding software development for simulation-based engineering and science is that it should be seen as an opportunity whose time has come to different extents in different fields. Within a discipline it depends upon whether or not we have crossed the point where computational resources permit predictions with useful accuracy. In the example of computational quantum chemistry discussed in the context of Figure 6.1, it is evident that we have, and therefore further advances in models, algorithms and computer hardware all provide a synergistic effect that enables users to treat larger and more complex molecules. In other fields, such as evolutionary biology, it is reasonably clear that we do not yet have useful computational models at all, and therefore advances in computer capabilities are for the time being entirely irrelevant. Finally in some other fields, the computational requirements are so low that advances in computing do not materially change their ability to do the relevant simulations which were already possible. Thus the automobile industry used to be a major purchaser of supercomputers but this is no longer necessary for CAD/CAM purposes, which can be met with high-performance workstations and midrange computing. By contrast, the oil industry continues to be major users and thus purchasers of high-end computer hardware.

For those fields in which useful simulations are currently computationally limited, new government investments in software development will likely reap future rewards. Many of the opportunities that are highlighted in the chapters on materials, life sciences, energy, multiscale modeling, and algorithms constitute appropriate examples. The algorithms that are contained in software developed for recent generations of hardware contains new capabilities that are only feasible as a result of hardware advances. A good example in the context of electronic structure calculations discussed above is the emerging ability to exploit separations of length scales to obtain algorithms with complexity that scales only linearly with the size of the molecule. In turn, new types of applications calculations are enabled by the synergism between improvements in hardware performance and algorithms, together embodied in functional software. The transformational power of new simulation capabilities is greatly magnified when the computing requirements are as inexpensive as commodity personal computers are in 2008.

Therefore it is worth highlighting the possibilities of the next generations of commodity computers. Advances in graphical processing units (GPUs) have been outpacing advances in general purpose central processing units (CPUs). Indeed the hardware associated with the current generation of video games such as the Microsoft Xbox and the Sony Playstation III, has computational capability that is very similar to state-of-the-art PCs (if one disregards other aspects of system performance such as disk-based IO and memory bandwidth). It is very likely that future commodity processors from companies such as Intel and AMD will integrate general purpose GPUs (GPGPUs) with CPUs. GPUs demand very high degrees of parallelism in order to achieve a reasonable fraction of their peak performance. Accordingly it is likely that programming tools, which are currently a relatively rudimentary stage for GPUs, should be much further developed as a prerequisite for widespread software development activity to exploit the capabilities of these devices for scientific and engineering applications. It is also very likely that other advances in general-purpose commodity computers will be in the direction of far greater parallelism, because as discussed in the Chapter 5 advances in computer speed based on increasing clock speed are no longer possible due to physical limitations associated with heat generation.

SUMMARY OF FINDINGS

Modern software for simulation-based engineering and science is sophisticated, tightly coupled to research in simulation models and algorithms, and frequently runs to millions of lines of source code. As a result, the

lifespan of a successful program is usually measured in decades, and far surpasses the lifetime of a typical funding initiative or a typical generation of computer hardware. Leadership in many disciplines remains largely in U.S. hands, but in an increasing number of areas it has passed to foreign rivals, with Europe being particularly resurgent in software for midrange computing, and Japan particularly strong on high-end supercomputer applications. Moreover, defense research in the United States has been denied access to foreign-developed software as a matter of principle.

Simulation software is too rich and too diverse to suit a single paradigm for progress. Choices as disparate as software targeting cost-effective computer clusters versus leading edge supercomputers, or public domain software versus commercial software, choices of development tools, etc, are mapped across the vast matrix of applications disciplines. Best outcomes seem to arise from encouraging viable alternatives to competitively co-exist, because progress driven by innovation occurs in a bottom-up fashion. Thus strategic investments should balance the value of supporting the leading edge (supercomputer class) applications against the trailing vortex (midrange computing used by most engineers and scientists).

Fundamentally, the health of simulation software development is inseparable from the health of the applications discipline it is associated with. Therefore the principal threat to U.S. leadership comes from the steady erosion of support for first-rate, excellence-based single-investigator or small-group research in the United States. A secondary effect that is specific to software development is the distorting effect that long development times and modest numbers of publications have on grant success rates. Within applications disciplines, it is important to recognize and reward the value of software development appropriately, in balance with the direct exploration of phenomena. Software development benefits industry and society through providing useful tools too expensive, long-term, and risky to be done as industrial R&D, trains future scientists and engineers to be builders and not just consumers of tools, and helps to advance the simulation state of the art.

Future investments in software development at the applications level are best accomplished as part of re-invigorating U.S. physical and biological sciences generally. More specific investments in simulation software can be justified on the basis of seeking targeted leadership in areas of particular technological significance, as discussed in more detail in chapters on opportunities in new energy, new materials, and the biological sciences.

REFERENCES

Amber Online. <http://ambermd.org/>.

Anjyo, K. 1991. Semiglobalization of stochastic spectral synthesis. *Visual Computer* 7:1-12.

Anjyo, K., Y. Usami, and T. Kurihara. 1992. A simple method for extracting the natural beauty of hair. *Computer Graphics* 26(2):111-120.

Charmm Online. <http://www.charmm.org>.

Davis, R.J., V.V. Gulians, G. Huber, R.F. Lobo, J.T. Miller, M. Neurock, R. Sharma, and L. Thompson. 2009. *WTEC panel report on international assessment of research and development in catalysis by nanostructured materials*. Baltimore: World Technology Evaluation Center, Inc.

Gonze, X., J.M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.Y. Raty, and D.C. Allan. 2002. First-principles computation of material properties: The ABINIT software project. *Computational Materials Science* 25:478-492.

CHAPTER 7

ENGINEERING SIMULATIONS

Abhijit Deshmukh

INTRODUCTION

The use of simulation and computational models is pervasive in almost every engineering discipline and at every stage in the life cycle of an engineered system. Typical examples of the use of simulation in engineering include manufacturing process modeling; continuum models of bulk transformation processes; structural analysis; finite element models of deformation and failure modes; computational fluid dynamics for turbulence modeling; multiphysics models for engineered products; system dynamics models for kinematics and vibration analysis; modeling and analysis of civil infrastructures; network models for communication and transportation systems; enterprise and supply chain models; and simulation and gaming models for training, situation assessment, and education (PITAC 2005; Oden et al. 2006).

Simulation and modeling approaches used in engineering can be classified along various dimensions, other than application domains. One classification is based on the use of the simulation models, whether they are used for analysis or synthesis of the system. In most cases, solving the analysis or forward problem is significantly easier than solving the synthesis or inverse problem. Another classification of engineering simulation models is based on whether the models are deterministic or stochastic. A key issue in using simulation models to design engineered systems revolves around making predictions about the behavior and performance of these systems. While deterministic simulations may be useful in recreating past events or testing a hypothesis in the presence of perfect knowledge, any simulation model used for prediction of the future needs to consider the uncertainty involved in system performance in order to be meaningful. Another classification of simulation models is based on the underlying mathematical and computational constructs used to model the engineered system. The constructs used to model engineered systems range from partial differential equations used to model continuous processes, such as material removal and bulk reactions, to discrete event models used to capture the event-driven nature of shop-floor and enterprise-level decisions. Finally, engineering simulations differ significantly based on how the results from the models are used by the decision-makers. Offline models are not as conservative about the use of computational resources and time requirements as online or real-time models, where latency in obtaining results is a critical factor and data from the real-world observations can alter the functioning of the models during execution (Douglas and Deshmukh 2000).

In the past, the unavailability of computational resources to model most real-world engineered systems at a meaningful level of granularity has been a major constraint. The computational capabilities, in terms of compute cycles, data storage, and network bandwidth, offered by the next-generation cyberinfrastructure are quickly approaching the exascale range. This provides unprecedented opportunities to model engineered systems at a level of fidelity where meaningful predictions can be made in actionable time frames (Atkins et al. 2003).

ENGINEERING SIMULATION HIGHLIGHTS

The sites visited by the WTEC panel presented examples of the use of simulation and modeling in a variety of engineering domains and application areas. (Details on the site visits are available in Appendixes B and C.) The examples highlighted below represent only a subset of the research conducted at the sites visited by the WTEC panel, which in turn represent only a subset of all modeling and simulation activities across engineering. These examples were selected to emphasize the diversity of application areas at the sites visited; the reader should not infer that these sites were conducting higher-quality research than those *not* included in this report. It is also important to note that given the scope of the study, most of the sites visited focused on simulation and modeling of physical processes and devices. Modeling of systems-level issues, involving logical and discrete event interactions, and the use of simulation output in decision-making, prediction, and design were not investigated in this WTEC study.

Energy Systems

The use of simulation and modeling related to energy systems revolves around several themes, such as combustion, wind turbines, fuel cells, hydrogen storage, photovoltaic interfaces, biofuel processes, nuclear reaction, and waste containment (ESF 2007).

The Central Research Institute of Electric Power Industry (CRIEPI), a Japanese nonprofit corporation founded in 1951 with a broad mission of “solving global environmental problems while ensuring energy security,” is conducting research on simulation analysis and tools for the Japanese electric power companies. Several examples of CRIEPI’s activities are its Power system Analysis Tool (CPAT) for electric grid stability, computational fluid dynamics (CFD) analyses of combustion in coal-fired power plants, CFD analysis of transmission line vibration, and seismic analysis of oil tanks. For the most part, CRIEPI employs commercial software in its routine simulations, and it outsources code development to tailor the commercial tools for problems of interest.

Professor Nobuhide Kasagi in the Department of Mechanical Engineering at the University of Tokyo is heading a project on simulation of microturbines as mobile energy systems. His work focuses on developing a microturbine 30 kW energy generation system combined with a solid oxide fuel cell that would address the power supply needs for mobile systems.

Significant research is being conducted in Denmark on modeling and simulation of wind turbines. The Danish focus on this research area is quite understandable since the amount of wind power generated in Denmark per 1000 inhabitants exceeds 570 kW, which is by far the highest in Europe and, indeed, the world. Many advances in simulation-based design of wind turbines are associated with researchers at the Technical University of Denmark (DTU) and at Risø DTU National Laboratory for Sustainable Energy. The research at DTU is focused in the following areas: design of optimum airfoils; dynamic stall, especially 3-D stall; tip flows and yaw; heavily loaded motors; and interference (wake and park) effects (see Figure 7.1). A few examples of successful application of basic research at DTU are the development of the popular aeroelastic code FLEX and the design of parts of the Nible and Tjaereborg turbines. In particular, the FLEX code has undergone developments over many generations (currently at FLEX5) and is used for designing wind turbines and analyzing loadings.



Figure 7.1. Structural analysis of a complete wind turbine installation (courtesy of DTU).

The Institut Français du Pétrole (IFP), a state-owned industrial and commercial establishment in France, is focusing on advancing research in energy, transportation, and the environment, and catalyzing the transfer of technology from fundamental research to industrial development. IFP's research activities in the energy systems area are targeted towards extended reserves, clean refining, and fuel-efficient vehicles. Extended reserves are based on the assumption that oil and other fossil fuels will remain the dominant source of transportation fuels and chemical feedstock; the research themes in this area target increasing the success rate in exploration, improving the recovery ratio in reservoirs, and developing new fields in extreme environments. Clean refining focuses on obtaining the highest possible yields of transport fuels from a unit basis of raw materials in an environmentally responsible fashion; the research themes in this area are the production of high-quality fuels; the conversion of heavy crudes, residues, and distillates; and the production of petrochemical intermediates. The fuel-efficient vehicles program recognizes the importance of reducing fuel consumption and the development of new powertrain systems for alternative fuels (e.g., biofuels); the research themes in this area are adaptive simulation of combustion and load balancing to develop highly efficient engine technologies, including conventional and hybrid powertrains; development of pollutant after-treatment technologies; development of electronic control strategy and onboard software; and validation and specification of alternative fuels (e.g., biofuels and NGV) with low CO₂ emissions.

Researchers in the United Kingdom are focused on developing a virtual power plant that will enable safe production of high-quality and cost-effective energy while extending the lifespan of power plants to 60 years, guaranteeing optimum fuel use, and better waste management. These challenges demand access to petascale machines to perform advanced simulations, along with development of a new generation of codes and simulation platforms (EPSRC 2006).

Disaster Planning

Simulation and computational models are often used to evaluate civil structures and critical infrastructures in the event of disasters. These models are also needed to devise mitigation and rescue strategies. Examples of emergencies and critical infrastructures that are modeled include environmental phenomena such as earthquakes in a high-population-density urban environment; fires in large buildings and public facilities; chemical, biological, or radiation contamination from hazardous spills (or terrorism); large, near-shore oil spills in environmentally sensitive locations; and major floods and tidal waves (tsunamis). A key requirement in developing successful mitigation and rescue strategies is the ability to “predict” sufficiently accurately and in advance the specific events that may characterize the evolution of an emergency. The only way that this can be achieved is to run realistic simulations of the emergency in real time (EPSRC 2006).

Significant research is being conducted at the University of Tokyo on modeling the impact of various natural and man-made disasters on large-scale infrastructures. Professor Shinobu Yoshimura's group is developing multiscale and multiphysics simulations for predicting quake-proof capability of nuclear power plants. The main objective of this project is full-scale simulation of components, but it also considers buildings subject to earthquakes. Most of the simulations involve one-way coupling on flow-structure interactions. A fully coupled simulation will require access to the next-generation supercomputer system at RIKEN. Professor Masaru Zako is leading a project on simulating disasters in chemical plants. The goal of this project is the development of a simulation capability for disaster propagation in chemical plants considering the effects of tank fire and wind speed and direction (see Figure 7.2). Another key question addressed by this research is the estimation of associated damages using the results of simulations in conjunction with information from the Geographical Information System (GIS).

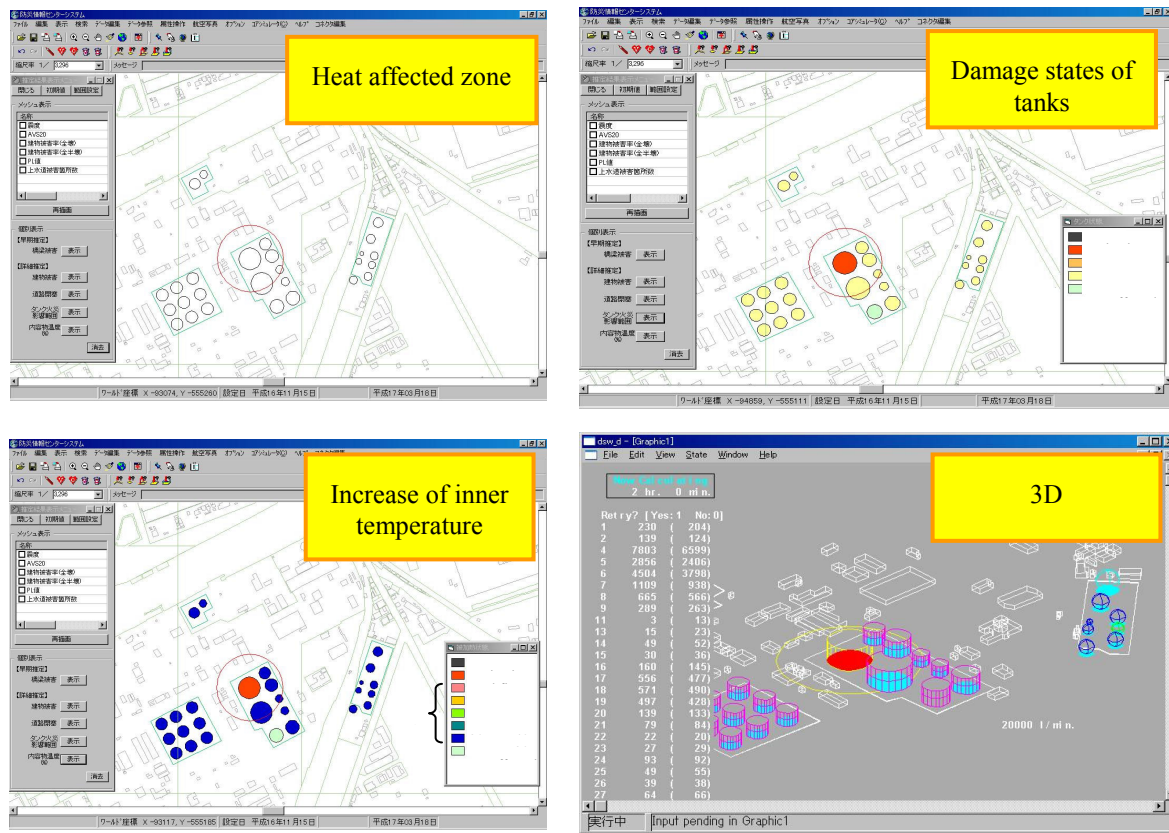


Figure 7.2. Assessing impact of fire in a chemical plant (courtesy of University of Tokyo).

Professor Shunichi Koshimura's group at the University of Tokyo is working on developing fragility functions for tsunami damage estimation using numerical models and post-tsunami survey data from Banda Aceh, Indonesia. The purpose of this project is simulation of tsunamis and also estimation of damage caused based on simulation results and survey data. Advanced methods in solving the shallow water equations were developed for realistic terrains, and a "fragility function" that estimates the probability of damage was constructed for first time for Banda Aceh.

Earth Simulator Center (ESC) of the Japan Agency for Marine-Earth Science and Technology has been a leading research center on global climate simulation. The Earth Simulator Project has achieved resolution of 10 km, that can be further refined to 1 km resolution via adaptive mesh refinement. The ESC researchers validate the model with historical data. The group has also achieved some impressive results for weather prediction, primarily due to the speed at which predictions can be generated by the Earth Simulator. The group can obtain predictions for wind stream between buildings in downtown Tokyo for use in urban planning.

Product and Process Modeling

Development of new engineered product offers several opportunities for the use of simulation and computational modeling. A typical design and manufacturing cycle requires the product first to be modeled using computer-assisted design (CAD) software, either as a solid model or a B-rep model, which can be used for geometric analysis. The same product then needs to be represented in computer-assisted engineering (CAE) software, typically as a finite element mesh, which can be used to conduct distortion analysis during and after manufacturing. The model then is represented in computer-assisted manufacturing (CAM) software, typically as a solid model, to determine the optimal tool path for machining. Finally, the machine tool controller instructions need to be generated using the tool path model to drive a computer numerical control (CNC) machine to complete the manufacturing process. The current state of the art in different areas of design and manufacturing

models also differs significantly. For example, finite element models of metal removal and forming processes (as shown in Figure 7.3), which incorporate high strain rate, large deformation constitutive models, adaptive meshing, generalized contact algorithms, and heat transfer, can handle multiple length scales (10^{-4}m to 10^2m), and analysis of short (10^{-6}s) and long temporal (10^3s) events. In contrast, the modeling tools to support conceptual design are extremely primitive.

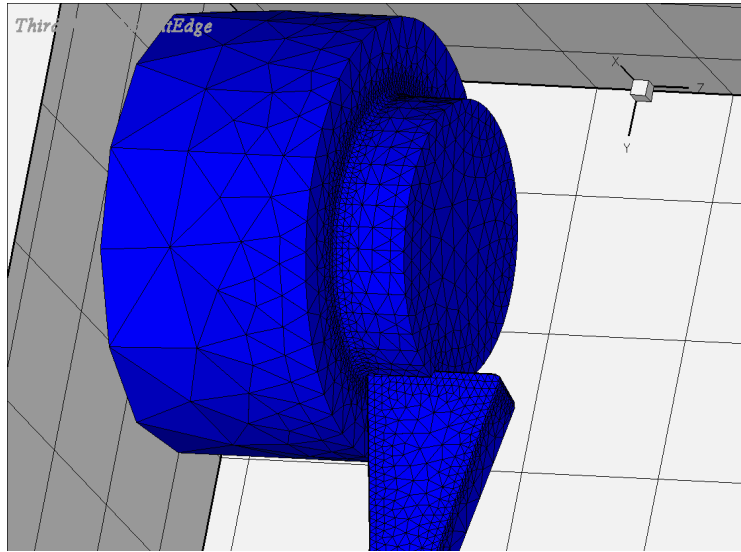


Figure 7.3. FEA model of metal cutting/chip removal (courtesy of Third Wave Systems).

In a typical enterprise, the product development lifecycle consists of a sequence of complex tasks and decisions spread over a period of time. These tasks are either virtual, as in design, or physical, as in manufacturing and maintenance operations. Nevertheless, the success (both economic and operational) of an engineered system hinges on accurate prediction of the product performance during each stage of the life cycle. For a successful product, all the decisions and operations have to be based on sound expectation of the implications of current decisions on downstream tasks. Most, if not all, of these tasks are performed in an *open-loop* fashion, i.e., very little feedback is available to any decision-maker on the implications. For example, consider the issue of level of permissible noise in a car. The level of noise tolerated in a small car in Europe is higher than in the United States. Initial designs of small cars were done for the European market. When the small car was introduced to the United States, the issue of isolation of noise was discovered to be a critical problem. Packaging or other simple redesign attempts were not able to resolve it. This problem could have been addressed in the initial design phase easily if it had been an explicit constraint in the optimal design formulation. However, the lack of information about the U.S. market and uncertainty about the importance of this issue led to expensive product redesign and manufacturing changes.

Toyota Central R&D Labs (TCRDL) has significant research activities focusing on issues related to automobile production, such as resource and energy conservation, environmental preservation, enhancement of comfort and safety, and advanced information processing. The highly publicized simulations of the structural responses of Toyota automobile frames to impacts, and simulations of the wind resistance levels of auto bodies, have reached a level of maturity such that they are no longer conducted at the TCRDL but at the Toyota Motor Company. Toyota has developed EVAS: Engine Vibration Analysis System, which is an engine design tool that is helping to develop lighter and quieter engines by providing highly accurate vibration predictions. The Aerodynamic Noise Simulator COSMOS-V developed at TCRDL is used to calculate temporally fluctuating airflow and to predict noise such as wind noise and wind throb, as well as vibrations caused by pressure fluctuations. Computational resources at Japan's Riken have also been used to conduct multiphysics simulations that combine airflow, heat distribution, structural deformations, and noise analysis for automobiles (see Figure 7.4). Additionally, TCRDL has developed Mill-Plan software to determine the optimal number of machining processes and the tool form and

cutting conditions for each process by closely examining the machining procedures and evaluating their overall efficiency.

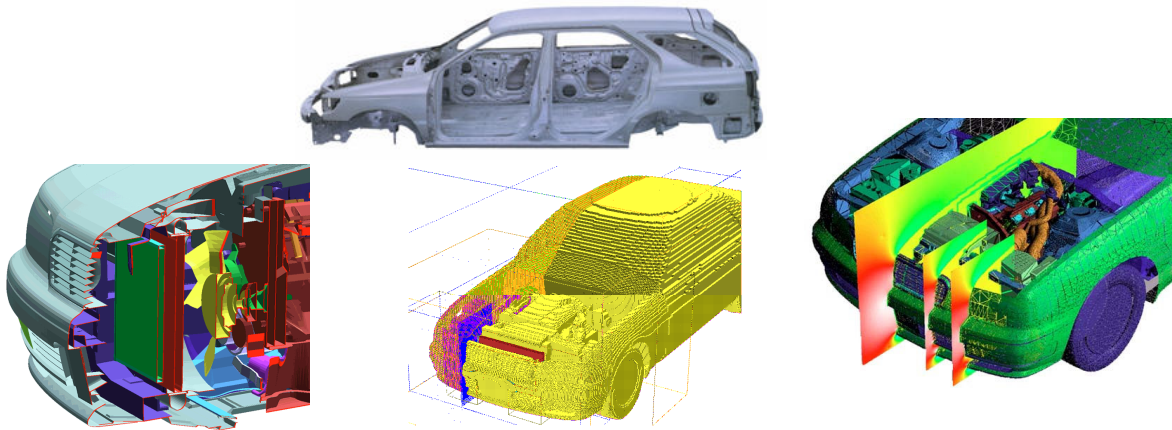


Figure 7.4. Digital automobile production enabled by voxel-based multiphysics: flow/heat/structure/noise (courtesy of Riken, Japan).

In the safety/human engineering area, TCRDL is studying human body dynamics to determine the kind of injuries people are likely to suffer in auto collisions in order to help design safer automobiles. THUMS (Total HUMAN Model for Safety) is a suite of detailed finite element models of the human body that can be subjected to impacts to assess injury thresholds. Models included detailed meshing of musculoskelature and internal organs. It was evident to the WTEC panelists that these and other simulation activities are a result of TCRDL management's growing interest in modeling human behavior and interaction with machines, and social network interactions to understand disruptions to Toyota's production lines and/or to predict the behavior of markets towards disruptive technologies such as the hybrid automobile.

The European Union has several initiatives focused on developing complex engineered systems using computational models and simulation. For example, the European helicopter industry has a goal of developing a complete helicopter simulation that involves CFD-based simulations of aerodynamics, aeroacoustics, and dynamics of rotorcraft. The EC ACARE 2020 Green Aircraft directive provides the politically agreed-on targets for an acceptable maximum impact of air traffic on people and the environment, while at the same time allowing for a constantly increasing amount of air travel. The goals include a considerable reduction of exhaust gas and noise while air traffic increases by a factor of 3; accidents are expected to go down by 80%; passenger expense should drop by 50%; and flights must become largely weather-independent. Significant advances in multiscale and multiphysics modeling are needed to develop products that have a chance of achieving these goals.

Simulation and modeling also play a critical role in engineering one-of-a-kind devices for scientific exploration. CERN, The European Center for Nuclear Research, used simulation in the form of finite element modeling (FEM) at the beginning of construction of the Large Hadron Collidor (LHC) to estimate the deflection expected in the detector itself. Using this approach, the maximum deflection (sag) in the 22-m-diameter detector, caused by its own weight of 7000 tons (~9 kN), was predicted to be 25 mm, which required compensation during the design stage. Accurate estimation of deflection to the cabin was critical to achieve the required alignment of the proton beam of the LHC. This example of civil engineering, in which the mass and volume of the detector was compensated by the removal of bedrock, is critical to the success of the forthcoming experiments. This was the first application of simulation to the engineering of the LHC.

Professor Gumbsch at the Fraunhofer Institute for Mechanics of Materials (Institut Werkstoffmechanik, IWM) has developed microscopic, physics-based models of materials performance and then inserted these subroutines into FEM codes such as LSDYNA, ABAQUS, and PEMCRASH. Specific applications of the IWM approach include modeling of materials (micromechanical models for deformation and failure damage analysis), simulation of manufacturing processes (pressing, sintering, forging, rolling, reshaping, welding, cutting), and simulation of

components (prediction of behavior, upper limits, lifetime, virtual testing). The IWM also has a strong program in metrology at small scales, developing models of failure/fracture, coupling data from nano-indentation with engineering- and physics-based models of same. The IWM has a substantial simulation effort in powder processing/sintering, and it has demonstrated the full process history from filling to densification and subsequent prediction of materials properties. Given its strong affiliation with the automobile industry, the IWM has sophisticated programs in modeling springback in stamped components, rolling, forming, and friction. It also has developed a detailed “concurrent,” multiscale model of diamond-like-carbon thin-film deposition. The model integrates models from Schrödinger’s equation up to classical models of stress-strain (FEM) and captures such properties as surface topography, internal structure, and adhesion.

Computational Fluid Dynamics (CFD)

CFD has been one of the primary engineering application domains for simulation using high-performance computing. The importance of this domain comes from the enormous range of practical applications of CFD, and from the existence of several different approaches to CFD itself. At one end, Direct Numerical Simulation (DNS) can be used to solve the governing Navier-Stokes equations using high-accuracy numerical methods. On the other end, Reynolds-Averaged Navier-Stokes (RANS) simulation is a computationally inexpensive method for complex problems, which removes small scales of turbulence by simple averaging.

In Europe there are several centers of excellence for CFD research. Large investments and effort have been put into CFD research in the European Union over the last 20 to 30 years. As an example, in the aerospace sector, the European Union has funded via the Framework Programmes many Specific Targeted Research Projects (STREP) related to CFD research. The European Union research community has also focused on collaboration within the community through a number of fora. First, ERCOFTAC (European Research Community on Flow, Turbulence and Combustion) is a well-established organization, funded through member subscriptions, in the area of fluid mechanics, turbulence, and combustion. The European Union research community has also focused on developing community codes that would be of value across several projects and reduce duplication of effort. Table 7.1 gives a list of some of the community codes developed by European Union researchers.

Table 7.1. European Community Codes in Engineering Disciplines

Name	Scientific Area	Brief Description	Licensing	# Users
Code_Aster	Engineering	Mechanical and thermal analysis	Free	300 (EDF) + 22000 download since 2001
Code_Saturne	Fluid dynamics	Incompressible + expandable fluids + heat transfer + combustion	Free	80 (EDF) + 20 industrial/academic partners + 5 for teaching
OpenFOAM	Fluid dynamics + structural mechanics	Finite volume on an unstructured grid	Free (fee for support)	~2000
Salome	Framework for multiphysics simulations	Actually used for engineering applications	Free	50 (EDF) + 21 institutions

The project of Ecole Polytechnique Federale de Lausanne Chair of Modeling and Scientific Computing (EPFL-CMCS) “Multiphysics in America’s Cup,” is a tour-de-force in fluid flow modeling. This was a formal collaboration of EPFL researchers with the designers of Italy’s entry into the America’s Cup sailing race, the *Alinghi*, with substantial proprietary information being shared by the designers. Commercial codes were used, initially Fluent, then CFX, because of the perceived superiority of the included turbulence models. The implementation required a supercomputer. (In order to iterate the design with the developers, a turnaround time of 24 hours was required.) The simulations had approximately 20 million elements, with 162 million unknowns in the solution. The work was rewarded with two America’s Cup victories.

Professor Leonhard Kleiser at the Computational Collaboratory at ETH Zurich is conducting research on turbulent flows and laminar-turbulent transition. Turbulent and transitional flows are investigated by direct numerical simulations. Here no turbulence models are employed, but the basic equations of fluid dynamics are numerically integrated on large computers whereby all relevant length and time scales of the flow must be resolved. Such investigations contribute to a better understanding of the fundamental phenomena and mechanisms of transition and turbulence, with respect to the conceptual exploration of methods for flow control and to the development of improved models for practical calculation methods.

The Institute of Fluid Mechanics of Toulouse (IMFT) conducts research in the mechanics of fluids with close experimental validation for applications in the areas of energy, processes, the environment, and health. The group of Dr. Braza, Director of Research of France's Centre National de la Recherche Scientifique (CNRS), has been working for a long time on direct numerical (DNS) and large-eddy simulations (LES) of turbulent prototype flows, and more recently on the detached eddy simulation (DES) of industrial-complexity flow applications. Specifically, DES was the central theme of a European initiative with the name DESider, (Detached Eddy Simulation for Industrial Aerodynamics; <http://dd.mace.manchester.ac.uk/desider>), a €10 million program motivated by the increasing demand of the European aerospace industries to improve their CFD-aided tools for turbulent aerodynamic systems with massive flow separation. DES is a hybrid of LES and statistical turbulence modeling, and it is designed as a multiscale framework for modeling and capturing the inner and outer scales of turbulence. Furthermore, IMFT participates in the UFAST (Unsteady Effects in Shock-Wave-induced separation) project of the European Sixth Framework Programme, which is coordinated by the Polish Academy of Science (Gdansk) Institute of Fluid-Flow Machinery, and which includes about 18 partners. The objectives are to perform "closely coupled experiments and numerical investigations concerning unsteady shock wave boundary layer interaction (SWBLI)" to allow for feeding back numerical results to the experiments, and vice versa, for the sake of physics and modelling of compressibility effects in turbulent aerodynamic flows. Using RANS/URANS and hybrid RANS-LES methods, UFAST aims at assessing new methods for turbulence modelling, in particular for unsteady, shock dominated flow. UFAST investigates the range of applicability between RANS/URANS and LES for transonic and supersonic flows.

SUMMARY OF KEY FINDINGS

Simulation and modeling are integral to every engineering activity. While the use of simulation and modeling has been widespread in engineering, the WTEC study found several major hurdles still present in the effective use of these tools:

- Interoperability of software and data are major hurdles
 - Commercial vendors are defining *de facto* standards
 - Very little effort goes beyond syntactic compatibility
- Use of simulation software by nonsimulation experts is limited
 - Codes are too complicated to permit any user customization
 - Effective workflow methods need to be developed to aid in developing simulations for complex systems
- In most engineering applications, algorithms, software and data are primary bottlenecks
 - Computational resources (flops and bytes) were not limiting factors at most sites
 - Lifecycle of algorithms is in the 10–20-year range, whereas hardware lifecycle is in the 2–3-year range
- Visualization of simulation outputs remains a challenge
 - Use of HPC and high-bandwidth networks has exasperated the problem
- Experimental validation of models remains difficult and costly
 - Models are often constructed with insufficient data or physical measurements, leading to large uncertainty in the input parameters

- The economics of parameter estimation and model refinement need to be considered; Figure 7.5 shows (qualitatively) the tradeoff between the cost of parameter estimation and cost of errors in prediction

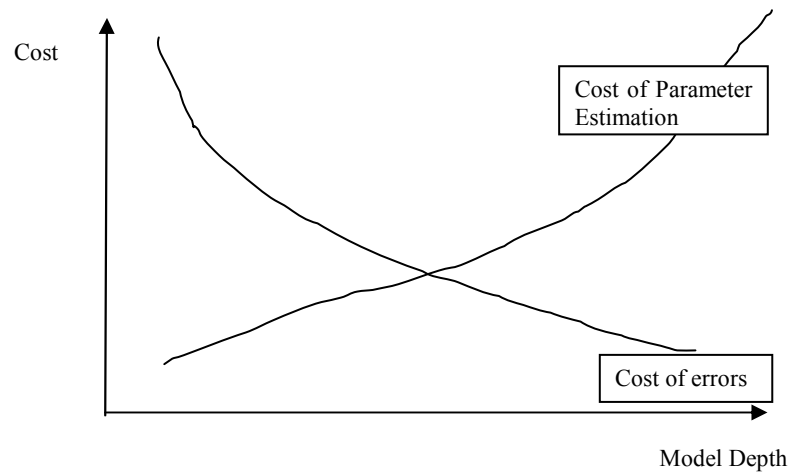


Figure 7.5. Tradeoffs between costs of parameter estimation and costs of errors.

- Uncertainty is not being addressed adequately in many of the applications
 - Most engineering analyses are conducted under deterministic settings
 - Current modeling and simulation methods work well for existing products; however, they are not ideally suited for developing new products that are not derivatives of current ones
 - Existing models are mostly used to understand/explain experimental observations
- Links between physical- and system-level simulations are weak
 - There is very little evidence of atom-to-enterprise models
 - Enterprise-level modeling and decision-making are not coupled tightly with the process- and device-level models
- Engineers are not being trained adequately in academia to address simulation and modeling needs
 - Possession of a combination of domain, modeling, mathematical, computational, and decision-making skills is rare in program graduates

COMPARISON OF U.S. AND WORLDWIDE ENGINEERING SIMULATION ACTIVITIES

- On average, U.S. academia and industry are ahead (marginally) of their European and Asian counterparts
 - Pockets of excellence exist in Europe and Asia that are more advanced than U.S. groups (Toyota, Airbus, University of Stuttgart)
- European and Asian researchers rely on the United States to develop the common middleware tools
 - Their focus is on application-specific software
- In the United States, the evolution from physical systems modeling to social-scale engineered systems lags behind that of Japan
 - Japan is modeling behavioral patterns of 6 billion people using the Life Simulator
- European universities are leading the world in developing curricula to train the next generation of engineering simulation experts (COSEPUP 2007)

REFERENCES

- Atkins, D.E., K.K. Droegemeier, S.I. Feldman, H. Garcia-Molina, M.L. Klein, D.G. Messerschmitt, P. Messina, J.P. Ostriker, and M.H. Wright. 2003. *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Arlington, VA: National Science Foundation. Available online: <http://www.nsf.gov/od/oci/reports/toc.jsp>.
- Committee on Science, Engineering, and Public Policy (COSEPUP; a joint unit of the National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine). 2007. *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology. Washington, DC: National Academies Press. Available online: http://www.nap.edu/catalog.php?record_id=11463#orgs.
- Douglas, C., and A. Deshmukh, eds. 2000 *Dynamic data driven application systems*. Report of the NSF Workshop, March 2000. Arlington, VA: National Science Foundation. Available online: <http://www.nsf.gov/cise/cns/dddas/>.
- Engineering and Physical Sciences Research Council of the UK (EPSRC): High End Computing Strategic Framework Working Group of the High End Computing Strategy Committee. 2006. *A strategic framework for high end computing*. Available online: <http://www.epsrc.ac.uk/ResearchFunding/FacilitiesAndServices/HighPerformanceComputing/HPCStrategy>.
- . 2006. *Challenges in high end computing* (companion document to the Strategic Framework document. Available online: <http://www.epsrc.ac.uk/ResearchFunding/FacilitiesAndServices/HighPerformanceComputing/HPCStrategy>.
- European Computational Science Forum of the European Science Foundation (ESF). 2007. *The Forward Look Initiative. European computational science: The Lincei Initiative: From computers to scientific excellence*. Information available online: <http://www.esf.org/activities/forward-looks/all-current-and-completed-forward-looks.html>.
- Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science Foundation blue ribbon panel on simulation-based engineering science*. Arlington, VA: National Science Foundation. Available online: http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.
- President's Information Technology Advisory Committee (PITAC). 2005. *Computational science: Ensuring America's competitiveness* (Report to the President). Washington, DC: Executive Office of the President, Committee on Technology. Available online: <http://www.nitrd.gov/pitac/reports/>.

CHAPTER 8

VERIFICATION, VALIDATION, AND UNCERTAINTY QUANTIFICATION

George Em Karniadakis

“...Because I had worked in the closest possible ways with physicists and engineers, I knew that our data can never be precise...” –Norbert Wiener

INTRODUCTION

Uncertainty quantification (UQ) cuts across many disciplines, from weather modeling, to human dynamics, to reservoir modeling, systems biology, and materials modeling, to name just a few areas. The sources of uncertainty may be associated with initial and boundary conditions, material properties, equations of state, loads, reaction constants, geometry, and topology, but also constitutive laws. In time-dependent systems, uncertainty increases with time, hence rendering simulation results based on deterministic models erroneous. In engineering systems, uncertainties are present at the component, subsystem, and complete system levels; therefore, they are coupled and are governed by disparate spatial and temporal scales or correlations.

Designing safe and robust engineering systems requires sensitivity UQ and sensitivity analysis at all levels. Despite its importance in all branches of Simulation-Based Engineering and Science (SBE&S), the field of UQ is not well developed. However, recently there has been an intense interest in *verification and validation* (V&V) of large-scale simulations and in modeling and quantifying uncertainty, as it is manifested by the many workshops and special volumes in scientific journals (Karniadakis and Glimm 2006; Schuëller 2005; Ghanem and Wojtkiewicz 2003; Karniadakis 2002) that have been organized to address these issues in the last few years. The U.S. Defense Modeling and Simulation Office (DMSO; <http://www.dmsso.gov>) of the Department of Defense (DOD) has been the leader in developing fundamental concepts as well as terminology for V&V. In 1994, DMSO published definitions of V&V that have been adopted by other professional engineering communities such as the American Institute of Aeronautics and Astronautics (e.g., see AIAA 1998).

The DMSO language is very specific: *Verification* is the process of determining that a model implementation accurately represents the developer’s conceptual description of the model and the solution to the model. Hence, by verification we ensure that the algorithms have been implemented correctly and that the numerical solution approaches the exact solution of the particular mathematical model—typically a partial differential equation (PDE). The exact solution is rarely known for real systems, so “fabricated” solutions for simpler systems are typically employed in the verification process. *Validation*, on the other hand, is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. Hence, validation determines how accurate are the results of a mathematical model when compared to the physical phenomenon simulated, so it involves comparison of simulation results with experimental data. In other words, verification asks “Are the equations solved correctly?” whereas validation asks “Are the right equations solved?” Or as stated in Roache (1998), “verification deals with mathematics; validation deals with physics.”

This V&V framework is not new; Figure 8.1 shows a schematic of the simulation cycle proposed by the Society for Computer Simulation in 1979 (Schlesinger et al. 1979), where two models are identified: a *conceptual* and a *computerized* model. The former is the mathematical model, whereas the latter is what is called today the “code.” It is the output of the computerized model that is validated as shown in the schematic. A third stage, named “qualification” in the schematic, determines the adequacy of the conceptual model in providing an acceptable level of agreement for the domain of intended application. Hence, qualification deals with the broader issues of the correct definition of the system, its interactions with the environment, and so forth.

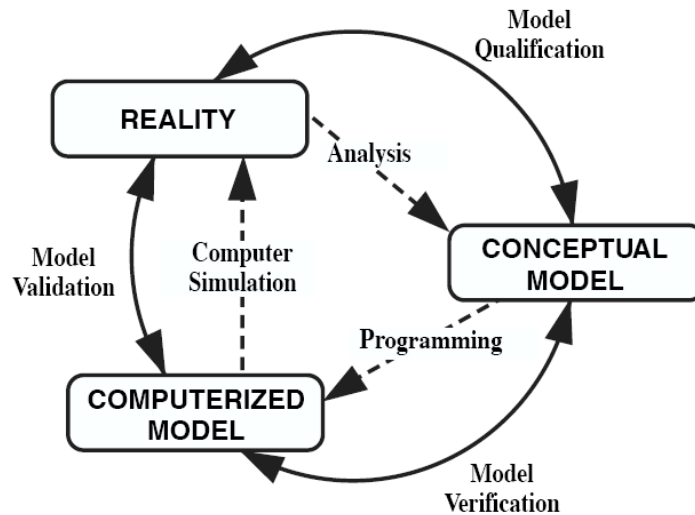


Figure 8.1. Schematic illustrating the V&V stages as proposed in 1979 by the Society for Computer Simulation (Schlesinger et al. 1979).

Validation is not always feasible (e.g., in astronomy or in certain nanotechnology applications), and it is, in general, very costly because it requires data from many carefully conducted experiments. To this end, characterization of experimental inputs in detail is of great importance, but of equal importance are the *metrics* used in the comparison, e.g., the use of low-order moments or of probability density function distributions (PDFs). Validation involves many aspects from physics, experimentation, sensitivity analysis, and even philosophy of science. The comparison of simulation results with physical data represents to some extent a test of accuracy of a specific scientific theory. This interpretation is a topic of current debate, as there are many researchers who adopt the thinking of the 20th-century science philosopher Karl Popper, who asserted that a scientific theory can only be invalidated but never completely validated! As Stephen Hawkins further explains, "No matter how many times the results of experiments agree with some theory, you can never be sure that the next time the result will not contradict the theory" (Hawkins 1988).

It may be pragmatically useful to accept a well-tested theory as true until it is falsified, but this does not solve the philosophical problem of induction. Indeed, invalidation of a theory or a corresponding model can be done with a single point in the parameter space, but validation would require an infinite number of tests to cover the entire parameter space, an impossible task indeed! Nevertheless, experience has shown that in practical situations, comparisons of the simulation results with data from physical experiments is very useful and leads to the general acceptance of the model and its effective use in other similar situations. How accurate is the prediction of a model validated at one set of parameter values to another regime is an open scientific question, and no general guidelines exist. Hereafter, we will use the term “validation” in the conventional sense, and we direct the interested reader to the literature (Kleindorfer, O’Neil, and Ganeshan 1998) for other views on this subject.

Quantifying numerical inaccuracies in continuum or atomistic simulations is a difficult task; however, significant progress has been made, especially for PDE systems in *a posteriori* error estimation (Ainsworth and Oden, 2000). Uncertainty quantification in simulating physical systems is a much more complex subject; it includes the aforementioned *numerical* uncertainty, but often its main component is due to *physical* uncertainty. Numerical

uncertainty includes in addition to spatial and temporal discretization errors, errors in solvers (e.g., incomplete iterations, loss of orthogonality), geometric discretization (e.g., linear segments), artificial boundary conditions (e.g., infinite domains), and others. Physical uncertainty includes errors due to imprecise or unknown material properties (e.g., viscosity, permeability, modulus of elasticity, etc.), boundary and initial conditions, random geometric roughness, equations of state, constitutive laws, statistical potentials, and others. Numerical uncertainty is very important and many scientific journals have established standard guidelines for how to document this type of uncertainty, especially in computational engineering (AIAA 1998).

Physical uncertainty is much more difficult to quantify; it can be broadly characterized as *epistemic*, i.e., reducible, or as *aleatory*, i.e., irreducible. For example, given the current rapid advances in quantitative imaging technologies, we can expect that the rock permeability of an oil reservoir will be measured much more accurately in the future—this is an example of epistemic uncertainty. However, even in this case, and in many simulations of realistic configurations, uncertainty is irreducible beyond some level or scale, e.g., background turbulence—there are no absolutely quiet wind tunnels, and the atmosphere or the ocean are inherently noisy environments. For *atomistic* simulations, it should be appreciated that inter-atomic potentials even for simple media (e.g., water) are not known precisely.

Uncertainty modeling has been of interest to Operations Research for a long time, but there has been recently renewed interest in the role that uncertainty plays in *decision making*, where a distinction is made between situations where the probabilities are totally unknown as opposed to situations under risks with known probabilities. The former type of uncertainty is named Knightian uncertainty, in honor of Frank Knight who made this distinction in the 1920s (Starke and Berkemer 2007). Another characterization is *deep uncertainty*, referring to the serious case where decision makers cannot even agree on the proper model or suitable cost function to be used (Lempert et al. 2004). Moreover, uncertainty can be exogenous, because the behavior of other systems is uncertain, e.g., the effect of weather. This type of uncertainty quantification is of great interest to insurance and financial sectors but also to industry, including the automotive industry, as we discuss further in subsequent sections.

This chapter refers to “uncertainty” somewhat loosely in order to address diverse sources of stochasticity, e.g., from imprecise values of parameters in a physical system to inherent noise in biological systems or even to *variability* of parameters or operation scenarios. The following text gives a few examples of sources of uncertainty and subsequently reviews typical methods, the industry view, and finally summarizes the findings of the WTEC panel.

EFFECTS OF UNCERTAINTY PROPAGATION

A few examples from physical and life sciences illustrate the effects of uncertainty in the boundary conditions, material properties, and equation of state. In addition, inherent noise in the system, for example, in cellular and gene networks, may cause dramatic changes to the response.

Fluid Mechanics

The first example is on direct numerical simulations (DNS) of turbulence—a field established with the emergence of the first Cray supercomputers. Resolving the high wave number regime is typically a computationally challenging task, because it requires millions of grid points. This example, however, is from a comparison of DNS and experimental results in terms of the energy spectrum obtained in the wake of a circular cylinder (Ma et al. 2000). As shown in Figure 8.2, there is agreement between the DNS and experiment in the high wave number regime, whereas there exists a rather large disagreement in the low wave number regime. This is surprising from the numerical discretization standpoint, but it is explained by the fact that for sufficiently high resolution (here more than 100 million degrees of freedom), the small scales are sufficiently resolved and do not depend on the boundary conditions, unlike the large scales which do depend on the boundary conditions. Since the inflow conditions in the experimental setup are not precisely known, simplified boundary conditions are employed in the DNS leading to this large disagreement.

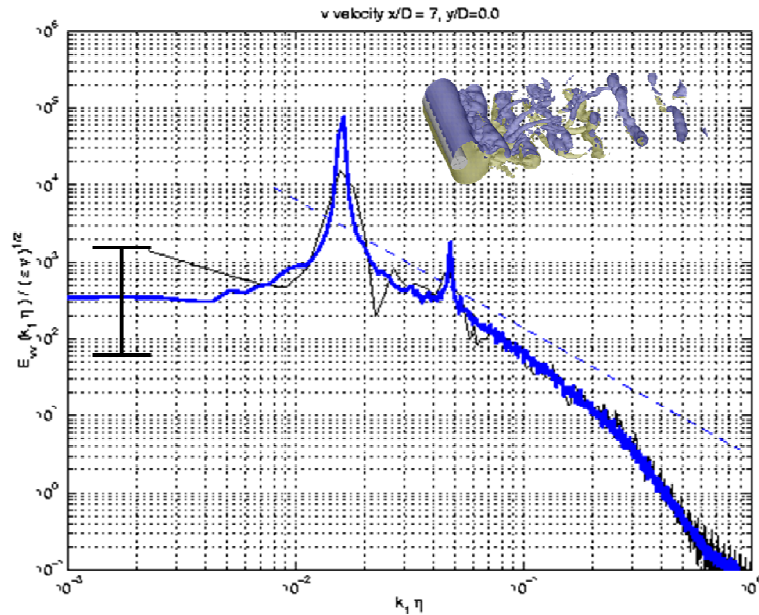


Figure 8.2. Direct numerical simulation (DNS) of turbulent flow past a cylinder. The energy spectrum is plotted against the wave number. The DNS results (black line, decreasing at the start) are in agreement with the experimental results (blue line) at high wave numbers but not at low wave numbers.

Plasma Dynamics

The second example is from simulations of pulsed plasma microthrusters used in maneuvering satellites, and it involves ablation of Teflon®. The uncertainty here comes from the equation of state of Teflon® and its material properties. Figure 8.3 (left) shows results from two simulations using the same plasma code MACH2 (Kamhawi and Turchi 1998) but two different equations of state. We observe that using the Los Alamos library SESAME in the MACH2 simulation, the pressure is almost 20 times greater than using another equation of state described in Kamhawi and Turchi (1998). The same magnitude in discrepancy has been demonstrated in another set of simulations (Gatsonis et al. 2007) for the microthruster problem using Teflon® properties from different literature sources; see Figure 8.3 (right). The ablation here is modeled as a three-stage process that takes into account the phase transformation, two-phase behavior, and the depolymerization kinetics of Teflon®.

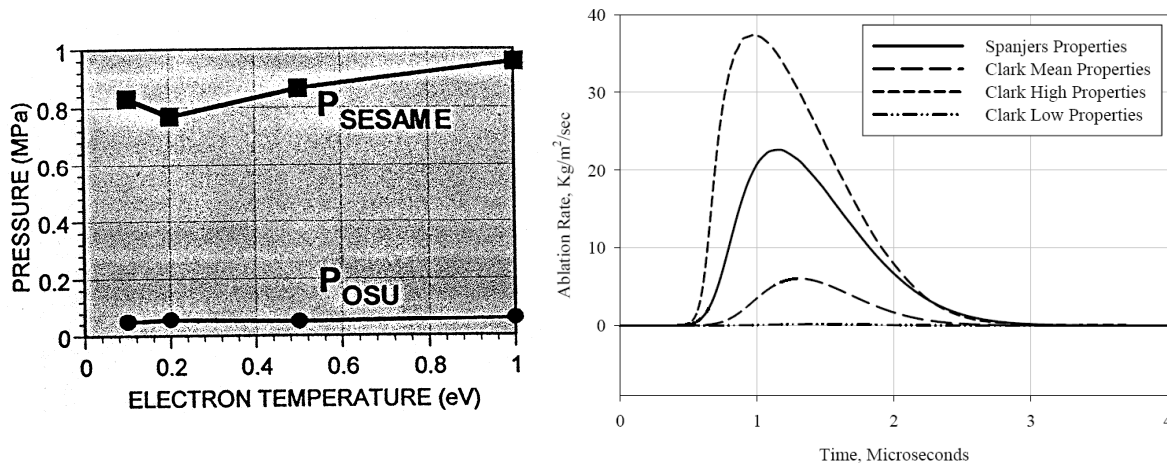


Figure 8.3. Simulation of pulsed plasma micro-thrusters. (Left) Pressure versus electron temperature for two different models of the equation of state (Kamhawi and Turchi 1998). (Right) Ablation rate versus time for different material properties (Gatsonis et al. 2007).

Biomedical Applications

In simulations of *biomechanics*, the values of the elastic properties of the vessel walls are important in order to capture accurately the blood flow–arterial wall interaction. The mean values of some of the required properties (e.g., modulus of elasticity, Poisson ratio) for the main arteries are tabulated in medical handbooks, but there is great variability among patients, even those with the same gender, age, or background. This is shown in Figure 8.4 (left), where the Cauchy stress was measured for five patients expected to have similar response. There is a very large discrepancy in the response and hence large discrepancy in the elastic properties, even under identical testing conditions.

In *systems biology*, uncertainty in initial conditions and reaction constants play a key role in the accuracy of the models, and hence sensitivity studies should be performed to assess this effect (Cassman et al. 2007). Noise is inherent also at the gene level, and stochasticity has a defining role in gene expression because it may determine what type of phenotype is produced (Elowitz et al. 2002; Pedraza and van Oudenaarden 2005). Recent work has shown that even cells grown in the same environment with genetically identical populations can exhibit drastically different behaviors. An example is shown in Figure 8.4 (right), from experiments with bacterial cells expressing two different fluorescent proteins (red and green) from identical promoters. Because of stochasticity (noise) in the process of gene expression, even two nearly identical genes often produce unequal amounts of protein. The resulting color variation shows how noise fundamentally limits the accuracy of gene regulation.

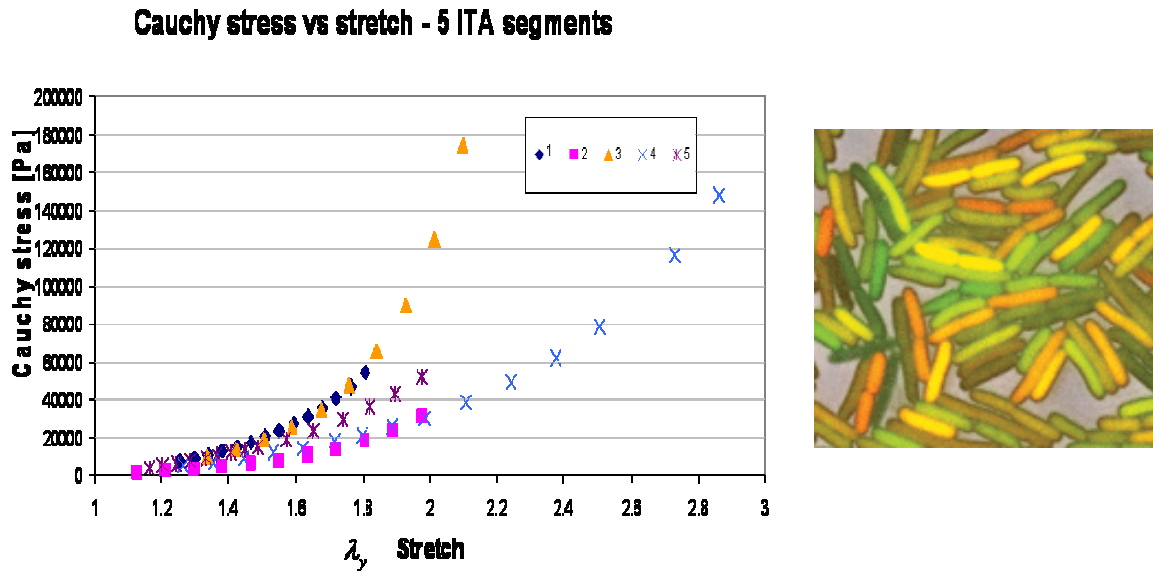


Figure 8.4. Uncertainty in life sciences: (Left) Cauchy stress versus stretch of internal thoracic arteries for five patients (Courtesy of Z. Yosibash and G. Sahar). (Right) Gene expression of bacterial cells (Elowitz et al. 2002).

METHODS

Most of the research effort in scientific computing so far has been in developing efficient algorithms for different applications, assuming an ideal input with precisely defined computational domains. Numerical accuracy and error control via adaptive discretization have been employed in simulations for some time now, but mostly based on both heuristics and in simple cases *a posteriori* error estimation procedures. With the field now reaching some degree of maturity, the interest has shifted to deriving more rigorous error estimators. It is also timely to pose the more general question of *how to model uncertain inputs and how to formulate new algorithms* in order for the simulation output to reflect accurately the propagation of uncertainty. To this end, the standard Monte Carlo (MC) approach can be employed, but it is computationally expensive and it is only used as the last resort. A faster version, the quasi Monte Carlo method (QMC), has been used successfully in computational finance and is particularly effective in applications with a relatively small *effective* dimensionality. The sensitivity method is an alternative, more economical, approach, based on moments of samples, but it is less robust and it depends strongly on the modeling assumptions. There are also other more suitable methods for physical and biological applications. The most popular technique for modeling stochastic engineering systems is the perturbation method, where all stochastic quantities are expanded around their mean value via a Taylor series. This approach, however, is limited to small perturbations and does not readily provide information on high-order statistics of the response. Another approach is based on expanding the inverse of the stochastic operator in a Neumann series, but this too is limited to small fluctuations. Bayesian statistical modeling has been used effectively in several different applications to deal with large uncertainties. At the heart of this framework is the celebrated Bayes' Theorem that provides a formal way of linking the prediction with the observed data.

A nonstatistical method, the polynomial chaos expansion (Ghanem and Spanos 1991) and its extensions, has received considerable attention in the last few years, as it provides a high-order hierarchical representation of stochastic processes, similar to spectral expansions. Its main drawback is that currently it cannot handle high-dimensional problems; however, new algorithms under development aim to overcome this difficulty. One such approach is the use of sparse grids in multidimensional integration (Bungartz and Greibel 2004) instead of full tensor-product forms, which in conjunction with direction-based adaptivity can result in very accurate and computationally efficient algorithms for problems with one hundred parameters or so. The “curse of dimensionality” is a fundamental problem that currently has no effective treatment, and it is at the heart of the major computational bottleneck in stochastic modeling in many different applications. To this end, a unique new

“priority program” was launched in 2008 by the German National Science Foundation (DFG) with focus on *high dimensions*. Specifically, the new priority program is entitled “Extraction of quantifiable information from complex systems,” its duration is six years, and its aim is the development of new mathematical algorithms for modeling and simulation of complex problems related to high-dimensional parameter spaces; uncertainty quantification is one of the main thrusts in this program. Researchers in computational finance have also been making progress in dealing with many dimensions, and ideas from that field may also be applicable to physical and life science problems.

Engineering Systems

Materials

Another class of problems that illustrate the necessity of incorporating the effects of uncertainty (in the lower scales) is the design of components made of heterogeneous random media. Experimental evidence has shown that microstructural variability in such materials (e.g., polycrystalline materials) can have a significant impact on the variability of macroscale properties such as strength or stiffness as well as in the performance of the devices fabricated from such materials. It is well understood that the thermomechanical behavior at the device scale is significantly affected by variations in microstructure features. Texture variations have been found to significantly affect yield and ultimate tensile strength, bend ductility, and total fatigue life to failure. Similarly, grain-size variations have been found to affect ultimate and fracture strength of polycrystalline materials. No predictive modeling or robust design of devices made of heterogeneous media (MEMS, energetic materials, polycrystals, etc.) is practically feasible without accounting for microstructural uncertainty. The recent report by the National Academies of Science on Integrated Computational Materials Engineering (ICME) (NRC 2008) stresses the importance of uncertainty quantification in material properties that may be obtained by either direct measurements or modeling, and also stresses the need for UQ in each stage of the design process (NRC 2008).

Uncertainty-Based Design

One of the most important goals of incorporating uncertainty modeling in large-scale simulations is that it will lead to new *nonsterilized* simulations, where the input parameters and geometric domain have realistic representations. The simulation output will be denoted not by single points but by distributions that express the sensitivities of the system to the uncertainty in the inputs. This is a key element to reliability studies and will provide the first step towards establishing simulation-based certificates of fidelity of new designs. It will also be a valuable tool for experimentalists as it will quantify individual sensitivities to different parameters, thereby suggesting new experiments and instrumentation.

Uncertainty management for engineering *systems planning and design* is also of great significance (de Neufville et al. 2004). In the more traditional approach, probabilistic design is employed to deal with the risk and reliability associated with a specific design. However, a new paradigm where one plans for uncertainty *explicitly* can endow the design with flexibility and robustness and lead to new architectures that are markedly different from designs created to simply meet fixed specifications, e.g., design codes or government regulations. This new paradigm will require investigation of scenarios where changes in economic and political events or shifts in customer preferences may take place in the future. Such a design paradigm change will require incorporation of cost and option analysis, so that system managers responsible for the operation of the final product are able to redirect the enterprise to avoid downside consequences or exploit upside opportunities. This new concept, termed *strategic engineering* by de Neufville and colleagues (2004), is illustrated in Figure 8.5 using the Weibull distribution often employed in bridge engineering.

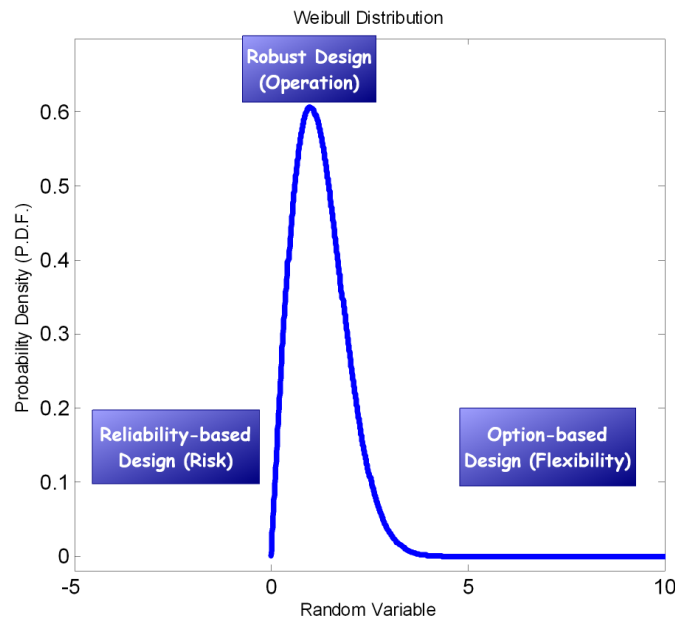


Figure 8.5. Illustrating the concept of uncertainty-based design. Unlike risk-based design, where only the left side of the PDF is important, in the new paradigm the trade-off between flexibility and real options is studied by also examining the right tail of the PDF.

For example, robust design of a bridge would require examining loads around the mean value, whereas risk-based design would require structural integrity of the bridge at extreme loads as specified, e.g., by existing regulations (left tail of the distribution). In the new design paradigm, an extra step is taken to investigate the possible enhancement of the strength of the structure well beyond the existing regulations, in anticipation that a construction of a second deck of the bridge may be needed in a few years. The designer then has to perform a financial analysis to investigate if such enhancement is cost-effective at this juncture. Recent work by financial analysts has developed the theory for *real options*, i.e., options analysis applied to physical systems (Trigeorgis 1996), and this holds the promise of enabling engineers to calculate the value of flexibility even at an early design stage. Uncertainty-based analysis is particularly important for the early stages of design in decision making, including acquisitions of expensive new products, e.g., for the Department of Defense.

Certification/Accreditation

For engineering systems, a *hierarchical* approach to validation is required to identify a set of possible experiments against which the simulation results from the various codes used will be tested. This is a complex and often expensive process, as it requires testing of the complete system. For example, in the aerospace industry two entire aircrafts should be tested for regulatory assessment. To be effective, the validation hierarchy should be *application-driven* rather than code-driven. It is common to construct a *validation pyramid* consisting of several tiers, as shown in Figure 8.6. On the left hand of the validation pyramid are placed the selected experiments; on the right side are placed the simulation results to be compared against the experiments. Both the complexity and the cost of the experiments grow as we move up the pyramid. As we move down the pyramid, the emphasis changes from validating multiphysics codes to singlephysics codes. At the lowest tier, we need to identify simple-geometry experiments that can be tested by general codes, whereas as we move up the pyramid at the subsystem, system, and complete-system tiers, there exist multiscale and multiphysics coupling in complex-geometry domains, and it is much more difficult to select the required experiments. Each face of the pyramid corresponds to specific validation experiments for each computer code for each component of the design. The edges of the pyramid represent coupling between the different codes.

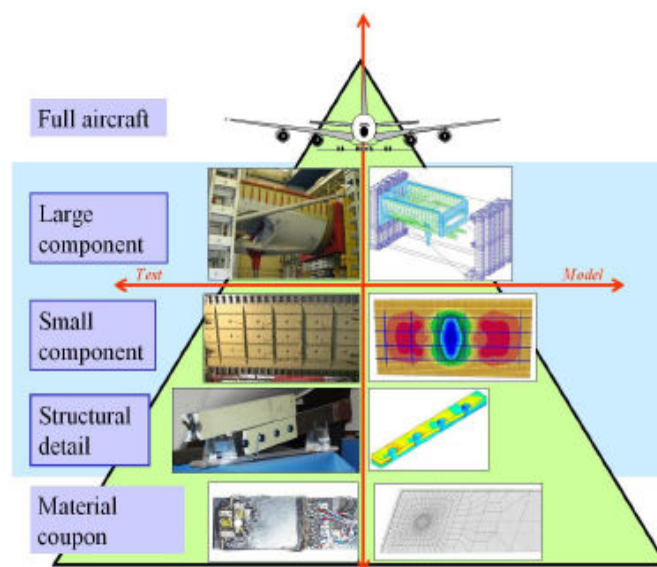


Figure 8.6. Validation hierarchy for a full aircraft: On the left of the pyramid the selected experiments are placed; on the right are the corresponding simulation results. This is an idealized validation pyramid targeting aircraft structural design (Babuska, Nobile, and Tempone 2007).

It is clear that the validation pyramid is not a unique construct and certainly not optimum for every application scenario of the complete system. However, it is in a sense an optimization problem, because the objective is to maximize reliability of the design given the constraint of a specific financial budget (Babuska et al. 2007). The validation pyramid represents the engineering systems viewpoint in modeling and simulation-based design rather than the viewpoint of a specific discipline (Oberkampf et al. 2004).

A case study of how the accreditation tests may lead to the rejection of the model is discussed by Babuska and colleagues (2007) for the Airbus A380 wing. The wing test failed on February 14, 2006. The European Aviation Safety Agency (EASA) specifies that the wing has to endure a load that is 150% of the limit load for 3 seconds, but the wing failed at 147% of the limit load.

Molecular Systems

Verification and validation of atomistic simulations is not well developed. Most researchers perform routine checks of convergence of statistics, especially with respect to the size and number of time steps. This is particularly important in mesoscopic methods where large time steps are employed. For example, in density-functional theory (DFT) a further step in verification could involve benchmarking of DFT codes against standard community codes. UQ activities are rare, but some recent work at the Technical University of Denmark (DTU) (Mortensen et al. 2005) has developed practical schemes based on Bayesian error estimation in DFT. This approach involves the creation of an ensemble of exchange-correlation functionals by comparing with an experimental database of binding energies of molecules and solids. The fluctuations within the ensemble are used to estimate errors relative to the experimental data for the simulation outputs in quantities such as binding energies, vibrational frequencies, etc. This procedure also checks for converged statistics as well as discretization and stationarity errors.

An example of the Bayesian error estimation in DFT is shown in Figure 8.7. The two plots show binding energies versus bond lengths for some solids (left) and molecules (right). The green spots (the topmost spots in the legend) indicate experimental values while the crosses are calculated with a best-fit model from the class of gradient-corrected density-functional-theory (GGA-DFT) models. The “clouds” represent ensemble calculations predicting error bars on the calculated properties (shown as horizontal and vertical bars). The error bars are seen to provide reasonable estimates of the actual errors between best-fit model and experiment.

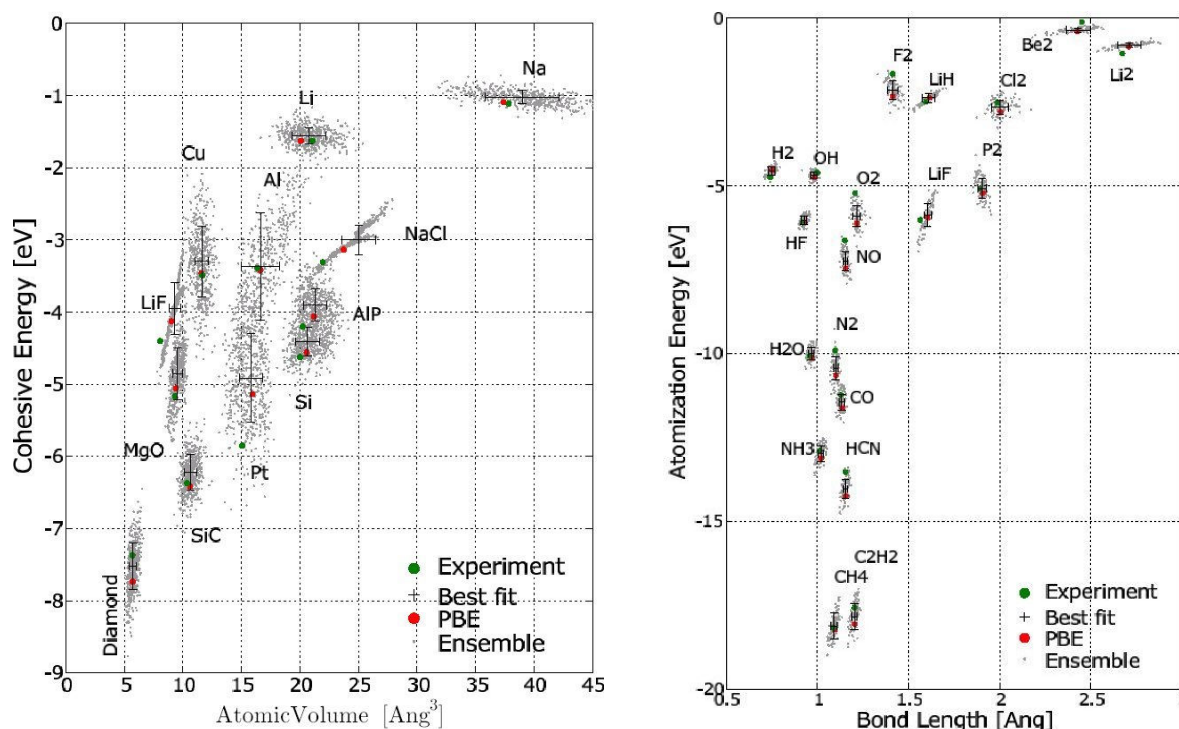


Figure 8.7. GGA-DFT ensemble results for solids and molecules with uncertainty quantification (Mortensen et al. 2005).

INDUSTRIAL VIEW

Industry, in general, is greatly interested in quantifying *parametric* uncertainty, and in major companies like BASF (Germany), ENI (Italy), and Mitsubishi Chemicals (Japan) that the WTEC panel visited (see site reports in Appendixes B and C), this seems to be the main UQ-oriented concern. Researchers often design several experiments in order to obtain the range of the parameters in their models so they can perform their simulations. In electrical power systems, both in land projects or in new projects such as the all-electric ship (see websites of the Electric Ship Research and Development Consortium: <http://www.esrdc.com> and <http://www.esrdc.mit.edu>) or future Boeing airplanes, uncertainty modeling is of paramount importance, for estimating robustness both of the electric components and of the entire system and for evaluating risk and assessing various reconfiguration scenarios. A similar type of modeling is required in the auto industry as researchers there attempt to simulate the noisy RLC circuits in fuel cells. Uncertainties in such problems are due to load models, fault time, frequency, angle and voltage variations, operating conditions, and manufacturing variability of the different components.

An example from the chemical industry of the effect of the uncertainty in the thermodynamic properties on the design process is discussed by Larsen (1986). There are typically three stages in process design. In the first stage, where process screening takes place, only modest accuracy of properties is required. Physical properties are frequently estimated by engineering correlations or measured by simple, low-cost experiments (e.g., infinite dilution activity coefficients by gas chromatography). An accuracy of $\pm 25\%$ is acceptable in cost estimates. The demands for data accuracy vary: for example, a 20% error in density may result in a 16% error in equipment size/cost, whereas a 20% error in diffusivity may result in 4% error in equipment size/cost. The errors in density are usually small for liquids, but the errors in diffusivity are frequently large (factor of two or more). On the other hand, a 10% error in activity coefficient results in negligible error in equipment size/cost for easily separated mixtures, but for close-boiling mixtures (relative volatility < 1.1) a 10% error can result in equipment sizes off by factor of two or more.

In stages II and III of the design process, more precise and additional thermophysical properties data are required, so selective experiments may be commissioned at stages II and III. In particular, the Stage III design is typically performed using interpolated data based on experimental measurements. The cost estimates at the conclusion of Stage III is $\pm 6\%$ in this particular study (or $\pm 3\%$ if at engineering drawing stage). This case study is typical of the effect of parametric uncertainty in the design process, and it has a direct effect on the decision making process.

In addition to parametric uncertainties, other types of uncertainties are of interest to industry. For example, in the Toyota Central R&D Laboratory (CRDL), researchers are interested in quantifying uncertainty primarily in the context of human dynamics and decision making. To this end, they organized a workshop entitled “Prospects and Limitations of Mathematical Methods for Decision Making in Nonlinear Complex Systems” that took place in Denmark (Helsingør) 21–22 November 2006 (Starke and Berkemer 2007). A second workshop on “Complex Systems: Interaction and Emergence of Autonomous Agents” was held in Baden, Austria in 2007. These workshops promoted the development and application of new methods for decision making and strategic planning, in particular with respect to uncertainty and nonlinear effects in complex systems through the discussion of a limited number of invited scientists with different research backgrounds from various fields. Their aim is to quantify uncertainty in both human activities and environmental issues.

The aerospace industry employs sensitivity analysis and hierarchical validation, such as validation pyramids, to quantify uncertainties both at the component and system level. In Europe, the MUNA (Management and Minimization of Uncertainties in Numerical Aerodynamics) initiative aims to quantify uncertainty in flow-structure interactions, e.g., in large aeroelastic motions, for aircraft technology. It is managed by the DLR Institute of Aerodynamics and Flow Technology in Braunschweig, with participation of eight university institutes, Airbus Germany, EADS-MAS, and Eurocopter. In four steps, error sources will first be identified, their effects will be quantified and made accessible to users via uncertainty margins, and finally strategies will be formulated to minimize simulation errors.

SUMMARY OF FINDINGS

The NSF SBES report (Oden et al. 2006) stresses the need for new developments in V&V and UQ in order to increase the reliability and utility of the simulation methods at a profound level in the future. A report on European computational science (ESF 2007) concludes that “without validation, computational data are not credible, and hence, are useless.” The aforementioned National Research Council report (2008) on integrated computational materials engineering (ICME) states that, “Sensitivity studies, understanding of real world uncertainties and experimental validation are key to gaining acceptance for and value from ICME tools that are less than 100 percent accurate.” A clear recommendation was reached by a recent study on Applied Mathematics by the U.S. Department of Energy (Brown 2008) to “significantly advance the theory and tools for quantifying the effects of uncertainty and numerical simulation error on predictions using complex models and when fitting complex models to observations.”

The data and other information the WTEC panel collected in its year-long study suggests that there are a lot of “simulation-meets-experiment” types of projects but no systematic effort to establish the rigor and the requirements on UQ and V&V that the cited reports have suggested are needed. Overall, the panel found that the United States leads the research efforts at this juncture—certainly in terms of volume—in quantifying uncertainty, mostly in computational mechanics; however, there are similar recent initiatives in Europe. Specifically, fundamental work in developing the proper stochastic mathematics for this field, e.g., in addressing the high-dimensionality issue, is currently taking place in Germany, Switzerland, and Australia.

In the United States, the DOD Defense Modeling and Simulation Office has been the leader in developing V&V frameworks, and more recently the Department of Energy has targeted UQ through the ASCI program and its extensions (including its current incarnation, PSAAP). The ASCI/PSAAP program is focused on computational physics and mechanics problems, whereas DMSO has historically focused on high-level systems engineering, e.g., warfare modeling and simulation-based acquisition. One of the most active groups in V&V and UQ is the Uncertainty Estimation Department at Sandia National Labs, which focuses mostly on quantifying uncertainties

in complex systems. However, most of the mathematical developments are taking place in universities by a relatively small number of individual researchers.

There are currently no funded U.S. national initiatives for fostering collaboration between researchers who work on new mathematical algorithms for V&V/UQ frameworks and design guidelines for stochastic systems. In contrast, there are several European initiatives within the Framework Programmes to coordinate research on new algorithms for diverse applications of computational science, from nanotechnology to the aerospace industry. In Germany, in particular, UQ-related activities are of utmost importance at the Centers of Excellence; for example, at the University of Stuttgart there is a Chair Professorship on UQ (sponsored by local industry), and similar initiatives are taking place in the Technical University of Munich and elsewhere.

On the education front, the WTEC panel found that existing worldwide graduate-level curricula do not offer the opportunity for rigorous training in stochastic modeling and simulation in any systematic way. Indeed, very few universities offer regular courses in stochastic partial differential equations (SPDEs), and very few textbooks exist on numerical solution of SPDEs. The typical graduate coursework of an engineer does not include advanced courses in probability or statistics, and important topics such as design or mechanics, for example, are taught with a purely deterministic focus. Statistical mechanics is typically taught in Physics or Chemistry departments that may not fit the background or serve the requirements of students in Engineering. Courses in mechanics and other disciplines that emphasize a statistical description *at all scales* (and not just at the small scales) due to both intrinsic and extrinsic stochasticity would be particularly effective in setting a solid foundation for educating a new cadre of simulation scientists.

REFERENCES

- American Institute of Aeronautics and Astronautics (AIAA). 1998. *AIAA guide for verification and validation of computational fluid dynamics simulations*. Reston, VA: AIAA. AIAA-G-077-1998.
- Ainsworth, M., and J.T. Oden. 2000. *A posteriori error estimation in finite element analysis*. New York: John Wiley.
- Babuska, I., F. Nobile, and R. Tempone. 2007. Reliability of computational science. *Num. Methods for PDEs* 23(4):753–784.
- Brown, D.L. (chair). 2008. *Applied mathematics at the U.S. Department of Energy: Past, present and a view to the future*. May, 2008.
- Bungartz, H.-J., and M. Greibel. 2004. Sparse grids. *Acta Numerica* 13:10123.
- Cassman, M., A. Arkin, F. Doyle, F. Katagire, D. Lauffenburger, and C. Stokes. 2007. *Systems biology: International research and development*. Dordrecht, The Netherlands: Springer.
- de Neufville, R., et al., 2004. *Uncertainty management for engineering systems planning and design*. Monograph of the 1st Engineering Systems Symposium, MIT, March 29-31. Cambridge: MIT Engineering Systems Division.
- Elowitz, M.B., A.J. Levine, E.D. Siggia, and P.S. Swain. 2002. Stochastic gene expression in a single cell. *Science* 297:1183–1196.
- Gatsonis, N.A., D. Juric, D.P. Stachmann, and L. Byrne. 2007. Numerical analysis of Teflon ablation in pulsed plasma thrusters. *Meeting paper of the 43rd AIAA/ASME/SAI/ASEE Joint Propulsion Conference and Exhibit*, 8–11 July 2007, Cincinnati, OH. AIAA: AIAA-2007-5227.
- Ghanem, R.G., and P.D. Spanos. 1991. *Stochastic finite elements: A spectral approach*. New York: Springer-Verlag.
- Ghanem, R.G., and S.F. Wojtkiewicz, eds. 2003. Uncertainty quantification. Special issue, *SIAM Journal on Scientific Computing* 26(2).
- Hawkins, S. 1988. *A brief history of time*. London: Bantam Books.
- Kamhawi, H., and P.J. Turchi. 1998. Design, operation, and investigation of an inductively-driven pulsed plasma thruster, AIAA 98-3804, *34th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, July 13-15, 1998, Cleveland, OH.
- Karniadakis, G.E., guest ed. 2002. Quantifying uncertainty in CFD. *Journal of Fluids Engineering* 124(1):2–3.
- Karniadakis, G.E., and J. Glimm, eds. 2006. Uncertainty quantification in simulation science. *Journal of Computational Physics* 217(1):1–4.

- Kleindorfer, G.B., L. O'Neil, and R. Ganeshan. 1998. Validation in simulation; Various positions in the philosophy of science. *Management Science* 44:1087–1099.
- Larsen, A.H. 1986. Data quality for process design. *Fluid Phase Equilibria* 29:47–58.
- Lembert, R., N. Nakicenovic, D. Sarewitz, and M. Schlesinger. 2004. Characterizing climate-change uncertainties for decision-makers: An editorial essay. *Climatic Change* 65(2):1–9.
- European Computational Science Forum of the European Science Foundation (ESF). 2007. *The Forward Look Initiative. European computational science: The Lincei Initiative: From computers to scientific excellence*. Information available online: <http://www.esf.org/activities/forward-looks/all-current-and-completed-forward-looks.html>.
- Ma, X., G.-S. Karamanos, and G.E. Karniadakis. 2000. Dynamics and low-dimensionality of a turbulent near wake. *J. Fluid Mech.* 410:29–65.
- Mortensen, J.J., K. Kaasbjerg, S.L. Frederiksen, J.K. Nørskov, J.P. Sethna, and K.W. Jacobsen. 2005. Bayesian error estimation in density-functional theory. *Phys. Rev. Lett.* 95:216401.
- National Research Council (NRC) of the National Academies. 2008 (in press). *Integrated computational materials engineering: A transformational discipline for improved competitiveness and national security*. Washington, DC: National Academies Press.
- Oberkampf, W.L., T.G. Trucano, and C. Hirsch. 2004. Verification, validation, and predictive capability in computational engineering and physics. *Appl. Mech. Rev.* 57(5):345–384.
- Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science blue ribbon panel on simulation-based engineering science*. Arlington: National Science Foundation. Available online: http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.
- Pedraza, J.M., and A. van Oudenaarden. 2005. Noise propagation in gene networks. *Science* 307:1965–1969.
- Roache, P.J. 1998. *Verification and validation in computational science and engineering*. Albuquerque,: Hermosa Publishers.
- Schlesinger, S., R.E. Crosbie, R.E. Gagne, G.S. Innis, C.S. Lalwani, and J. Loch. 1979. Terminology for model credibility. *Simulation* 32(3):103–104.
- Schuëller, G.I., ed. 2005. Special issue on computational methods in stochastic mechanics and reliability analysis. *Comput. Methods Appl. Mech., Engrg.* 194(12–16).
- Starke, J., and R. Berkemer. 2007. Prospects and limitations of mathematical methods for decision making in nonlinear complex systems: Synopsis of the Workshop. *Decision making and uncertainty in nonlinear complex systems*.. Conference Hotel Marienlyst, Helsingør, Denmark, 21–22 Nov. 2006, funded by Toyota CRDL. Mat-Report No 2007-05, ISSN #0904-7611, Lyngby, Denmark: Technical University of Denmark Department of Mathematics.
- Trigeorgis, L. 1996. *Real options: Managerial flexibility and strategy in resource allocation*. Cambridge: MIT Press.

CHAPTER 9

MULTISCALE SIMULATION

Peter T. Cummings

INTRODUCTION

The range of temporal and spatial time scales covered by simulation-based engineering and science (SBE&S), as illustrated in Figure 9.1, is truly extraordinary. Even within a single subdomain, such as materials or biological systems modeling, the spatiotemporal range can be many orders of magnitude. In many applications, phenomena at one particular spatiotemporal scale have significance and impact at spatiotemporal scales both above and below. Hence, a major focus of SBE&S research, in all of its subfields, is the development of *multiscale simulation* methodologies—techniques that allow modeling of phenomena across disparate time and length scales. The increasing demands of engineering design to push beyond the limits of past engineering and scientific experience elevate the practical importance of achieving true multiscale simulation.

The term “multiscale simulation” is often loosely used to describe a wide variety of simulation activities in science and engineering. Broadly defined, all simulation could be classified as multiscale, since by definition simulation involves spanning at least one temporal and/or spatial scale. This is because in simulation we take a description of a phenomenon at one scale, derive the equations for the elements that interact at that scale and by solving dynamical equations of motion, predict the collective behavior of these elements as a system. For example, if we define models for how water molecules interact with each other, and solve the dynamical equations of motion for a large group of water molecules at the appropriate density and temperature, we can predict the properties of bulk water. In a second example, if we wish to understand the mechanical behavior of an engineering structure, such as an aircraft wing, we break it down into small components (finite elements) over which the mechanical properties are constant and behave in relation to each other in a well-defined way, and by solving the dynamical equations that specify continuity of the appropriate properties within and at boundaries between the finite elements, we can simulate the mechanical response of the whole structure. In a third example, if we wish to predict how a cancerous tumor will grow and spread, one approach is to model it as a collection of cells, define the equations that govern the motion and state of each cell (single cell motion, cell-cell interactions, cell-environment interactions, cell growth and death), and by solving the dynamical equations for the cells, we can predict the behavior of the tumor. These are examples of simulation in the physical, engineering, and biological sciences, respectively.

More specifically, the foregoing are examples of *upscaling*, in which knowledge at a one level of description is obtained by averaging over dynamics at a level of description that is lower (i.e., more fundamental). Upscaling through one level of description is the workhorse of scientific and engineering simulation. However, we typically reserve the term “multiscale” for those simulations that involve multiple levels of upscaling, and that feature some degree of *downscaling* as well. By downscaling we mean that actions (or changes in boundary conditions) at one level of description of interest propagate down to the more fundamental levels.

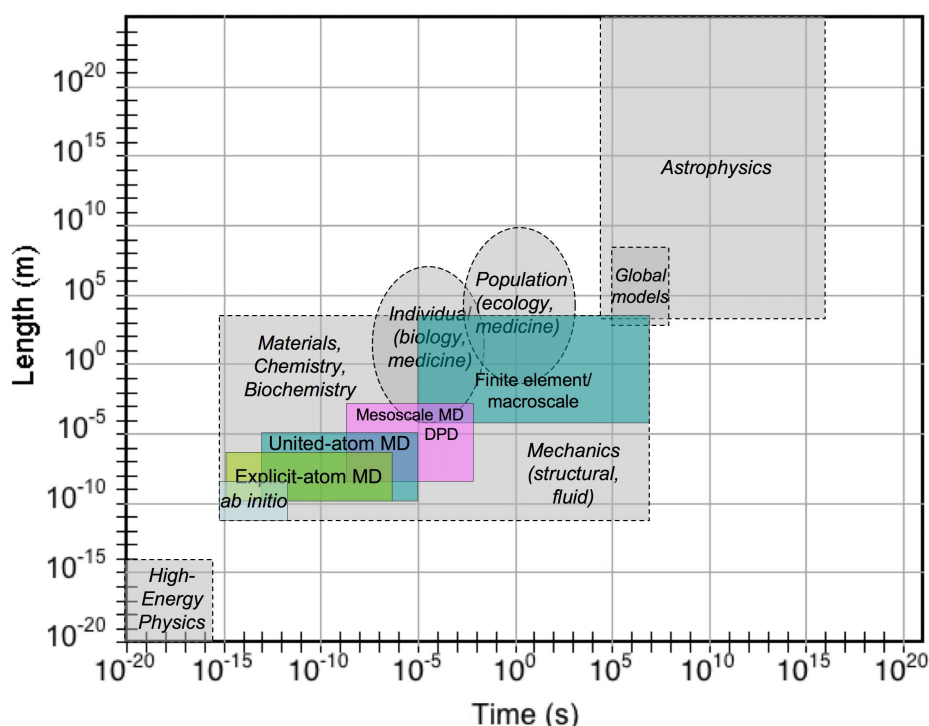


Figure 9.1. Typical temporal (horizontal axis, in seconds) and spatial (vertical axis, in meters) scales for several SBE&S problem domains. High-energy physics is at the low end, with time scales set by subatomic particle collision and decay times, and spatial domains determined by subatomic particle size and relative distance. Material modeling methods, ranging from *ab initio* methods based on quantum mechanics, to explicit-atom molecular dynamics (MD, in which each atom is represented explicitly and its interactions and dynamics solved for), to united-atom MD (in which atoms clustered in functional groups, such as CH_2 and CH_3 , are treated as single centers of force), to mesoscale MD and dissipative particle dynamics (in which whole molecules or clusters of molecules are frequently treated as a single entity, and, in the case of suspensions, the solvent degrees of freedom are integrated over to eliminate explicit solvent), and finally to finite element methods used to solve macroscopic balance equations, including the equations for structural mechanics. Also shown are the scales for modeling the biology of individual organisms and of populations, as well as astrophysics, where the low end is determined by the time scale (10^4 s) for a star spiraling into a million solar mass black hole with an innermost stable circular orbit of 10^8 meters, and the upper end by simulations the age of the universe (6 billion years, or 10^{16} s). The time and spatial scales for global climate modeling of galaxy formation (a billion light years, or 10^{25} m, in width) over time scales of (labeled global models) are included for comparison.

As an example of downscaling, consider the simulation of an aircraft wing using finite elements (one level down). At sufficient levels of stress and fatigue, the possibility arises of a crack forming in part of the structure, near which point traditional macroscopic representations of the stress-strain relationship will fail. In order to model this at a high level of physical rigor, the finite elements subject to the highest strains may require modeling at a mesoscale level in which grain boundaries are taken into account. Finally, in the vicinity of the point at which a crack initiates, an explicit atomistic description is desirable, with the atoms at the point at which the crack initiates and propagates modeled via a first-principles (*ab initio*) method to accurately reflect the bond-breaking processes taking place at the point of rupture. Such a simulation, which involves many scales of description (*ab initio*, atomistic, mesoscale, and finite-element), in which information flows both upwards (upscaling) and downwards (downscaling), is an example of true multiscale materials simulation. Accordingly, an enduring goal in the materials modeling community is true multiscale simulation, in which the most fundamental levels of describing materials (electrons and nuclei, described by Schrödinger's equation) are connected seamlessly with

the macroscopic scale to describe physical and chemical phenomena and devices at whatever level of detail is required for data-free prediction at some prescribed accuracy.

In biology, the range of scales is potentially just as great as or greater than the range of scales for materials. However, organizing principles at various levels (protein structure at the level of molecules, intracellular signaling at the level of single cells, genes, etc.) often mean that the most fundamental level in multiscale modeling of biological systems is at a higher level than the electronic level. Indeed, within biology, the term “*ab initio* protein folding” does not involve electronic structure calculations, as the name might imply to a materials simulator. *Ab initio* protein folding, also termed *de novo* protein folding, involves the prediction of protein structure from its molecular composition (amino acid sequence), as described in, for example, the reviews by Osguthorpe (2000) and Klepeis et al. (2002). The amino acid sequence is typically represented by an atomistically detailed model, or a coarse-grained version of this. The objective is to locate the structure with the global energy minimum. Thus, in principle, one approach to the problem of *ab initio* protein folding is molecular dynamics (MD) simulation. However, a fundamental problem in applying MD is the time scale for protein folding, which for even the simplest proteins is of the order of ms or longer. One of the heroic attempts to fold an atomistically detailed model for a protein in aqueous solution was that of Duan and Kollman (1998), who performed a 1- μ s-long MD simulation. In short, in biology, the most fundamental level

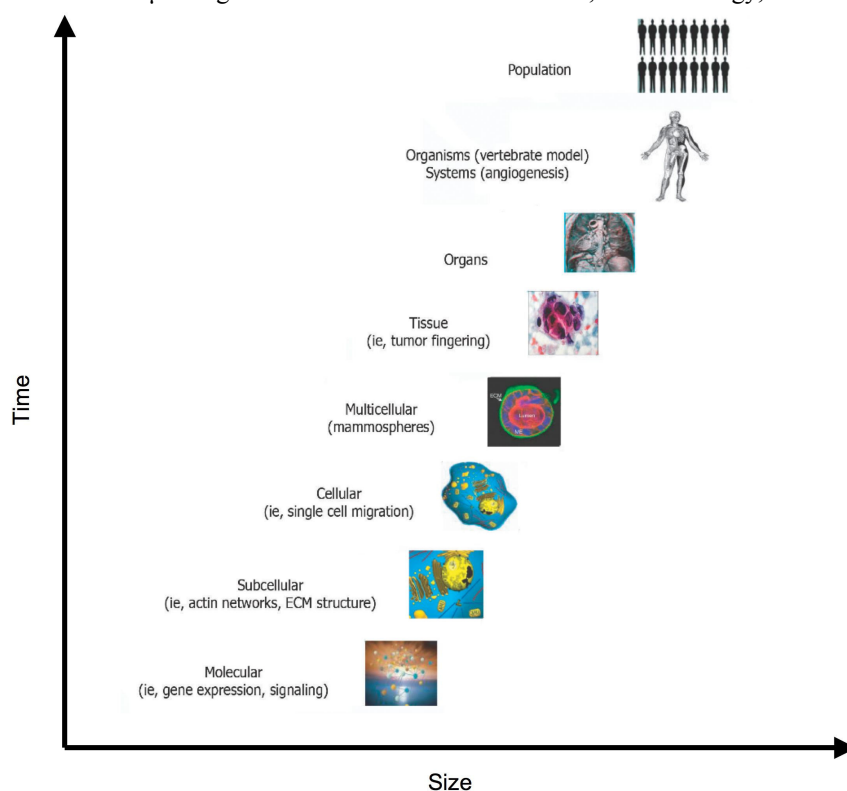


Figure 9.2. A hierarchy of spatial and temporal scales relevant for modeling the initiation of cancer (oncogenesis) and its spread throughout the body (metastasis). At the highest level, bioinformatics is used to identify risk factors in populations, which can then be related through experiment to genetic abnormalities at the molecular level. The challenge in multiscale cancer modeling is to bridge between these two extremes to understand the mechanics of oncogenesis and metastasis, and identify effective patient-specific treatment strategies.

is molecules (proteins, ligands, solvent, counter-ions), but many biological models begin with a cell as the fundamental building block (or, as in ecological modeling, a single organism). A hierarchy of the spatial and temporal scales relevant to modeling the onset and spread of cancer in humans is shown in Figure 9.2.

Energy and environmental sustainability are areas in which multiscale simulation have played, and will continue to play, important roles. One notable and familiar example is global climate simulation (GCS), in which global circulation models (based on mass, energy, and momentum balances) for the oceans and the atmosphere are combined with models for relevant terrestrial processes in order to predict future climate trends. Two prominent examples of GCS codes are the United Kingdom's Meteorological Office Hadley Centre HadCM3 code, and the U.S. Geophysical Fluid Dynamics Laboratory's CM2 code. A large fraction of the computing cycles at several of the world's largest supercomputing center, including the Earth Simulator in Japan (<http://www.jamstec.go.jp>) and the National Center for Computational Sciences at Oak Ridge National Laboratory (<http://nccs.gov>), are devoted to running GCS codes, most often to investigate scientific questions but also to provide insight for policymakers. GCS may well be the application of SBE&S most familiar to the general public. The debate about the reliability of global warming simulation predictions has resulted in the general public becoming aware of predictive computational simulation, and that there can be potential economic implications incurred by relying (or not relying) on the results of simulation. The predictions of global warming from GCS have large political and economic implications, and indeed the importance of GCS modeling is underlined by the 2007 Nobel Peace Prize being shared by the Intergovernmental Panel on Climate Change (<http://www.ipcc.ch>), much of whose work and conclusions are informed by GCS.

This chapter discusses some of the multiscale simulation issues and trends identified in the course of the WTEC international assessment. The following briefly reviews the state of the art, then presents examples of multiscale simulation viewed by the international assessment panel, and concludes with the findings of the panel in relation to multiscale simulation.

CURRENT STATE OF THE ART IN MULTISCALE SIMULATION

True multiscale simulations, with both automated upscaling and adaptive downscaling, are actually quite rare today. This is because there has been no broadly successful general multiscale modeling simulation methodology developed to date. Thus, multiscale simulation research efforts today could be said to fall into two main categories: methods designed to have general applicability in any multiscale simulation application, and multiscale simulation methods designed for a specific application or problem domain.

In specific problem domains, there are examples of successful applications of multiscale simulations. In crack propagation in metals, for example, multiscale simulation, spanning semiempirical *ab initio* methods, atomistic simulation, and finite element methods, was achieved in the last decade (Broughton et al. 1999). Multiscale simulation methods, particularly relevant to nanoscale mechanical systems, are reviewed in a number of recent papers and monographs (Curtin and Miller 2003; Liu et al. 2004; Liu, Karpov et al. 2005; Farrell, Karpov et al. 2007). The OCTA project (<http://octa.jp>) is another example of multiscale simulation modeling within a specific problem domain (polymer structure and dynamics); it is also a notable example of how a large government investment, combined with industry participation, can result in the development of a complex, integrated software environment useful both for fundamental research and for industrial application.

Among general multiscale simulation approaches, the equation-free modeling approach of Kevrekidis and coworkers (see, for example, Bold et al. 2007; Kavousanakis et al. 2007; Papavasiliou and Kevrekidis 2007; Roberts and Kevrekidis 2007) offers a framework for using existing modeling codes, enclosed with appropriate "wrappers," within a multiscale simulation framework. The equation-free modeling approach seems particularly suited to developing a coarse-grained model that computes properties as needed on the fly from simulations performed at a more detailed level. When implemented appropriately, it assumes no details of the model at the more detailed level, and uses standard integration of equations of motion at the coarse-grained level using properties computed at the more detailed level. Specific instances of this approach have been used before. For example, the Gibbs-Duhem method (Kofke 1993) for computing phase equilibrium boundaries integrates a specific equation (the Gibbs-Duhem equation) to compute the saturated vapor pressure of a fluid by using two atomistic simulations (one of the vapor phase, one of the liquid phase) to determine the needed thermodynamic properties (differences in enthalpy and volume between the coexisting phases).

Despite advances in multiscale simulation, there is clearly much to be done in the future. The development of multiscale simulation methodologies, both general and specific, will continue to be the subject of intense effort, both in the United States and internationally. The national and industrial investment that led in Japan to Masao Doi's OCTA integrated simulation system for soft materials may provide insight into how domain-specific multiscale simulation methodologies can be realized.

MULTISCALE SIMULATION HIGHLIGHTS

This section highlights some of the multiscale simulations research being conducted at sites visited by the WTEC study panel. The reports below are not comprehensive, nor are their inclusion in this report meant to suggest that these sites conducted higher-quality research than other sites visited; rather, the intent is to choose a selection of sites that are representative of the range of activities observed during the WTEC study.

Mitsubishi Chemical

The Mitsubishi Chemical Group Science and Technology Research Center (MCRC), headquartered in Yokohama, Japan, is profiled in more detail in Chapter 3, Materials Simulation. It is also mentioned in this chapter because it exemplifies the industrial need for multiscale simulation, and because its efforts relate to specific findings.

MCRC staff use a variety of tools to design, construct, and operate optimal chemical and materials manufacturing processes. In order to achieve its design goals, MCRC couples commercial and in-house codes to create multiscale models of its processes, both existing and planned. Its researchers achieve this by writing their own in-house codes to provide coupling between the commercial codes. MCRC supports an internal multiscale molecular modeling effort on polymers, which is linked to the process modeling efforts. For MCRC staff, the lack of interoperability of commercial codes at different levels of spatio-temporal resolution makes the development of effective multiscale models difficult.

Theory of Condensed Matter Group, Cavendish Laboratory, Cambridge University

The Theory of Condensed Matter (TCM) group at Cambridge, headed by Mike Payne, is well known for the development of several DFT-based codes, notably the Cambridge Sequential Total Energy Package, (CASTEP, <http://www.tcm.phy.cam.ac.uk/castep/>) and the order-N electronic total energy package (ONETEP, <http://www.onetep.soton.ac.uk>). Additional information about these codes is available in the site report in Appendix C and Chapter 3.

In addition to his efforts developing *ab initio* codes, Payne has been developing a multiscale simulation method for modeling of dynamics of materials, specifically crack propagation in graphine sheets, illustrated in Figure 9.3. The technique used is called learn on the fly (LOTF) (Csányi et al. 2004 and 2005). As with many multiscale simulation methods that bridge between scales by constructing a hybrid Hamiltonian that combines different models, the LOTF method uses a unique short-ranged classical potential whose parameters are continuously tuned to reproduce the atomic trajectories at the prescribed level of accuracy throughout the system. Payne and coworkers have developed LOTF simulations with interfaces to a number of *ab initio* codes, in addition to the SIESTA code (Soler et al. 2002) used to generate the configurations in Figure 9.3.

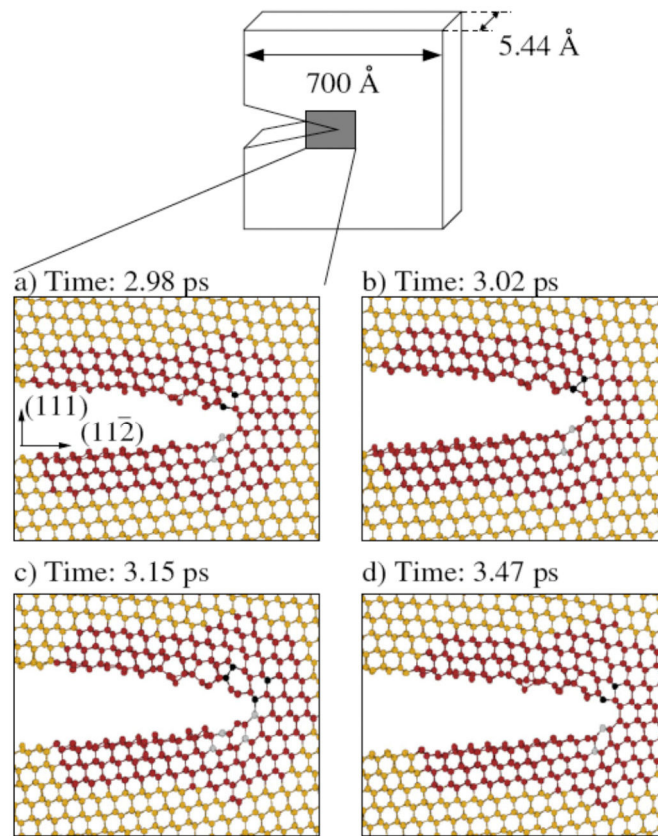


Figure 9.3. Propagation of a crack in silicon using the LOTF scheme. Silicon atoms treated classically are shown in gold (gray), and those described using the SIESTA code are shown in red (black) (Csányi et al. 2005).

Blue Brain Project, École Polytechnique Fédérale de Lausanne

The Blue Brain Project is a joint collaboration between the Brain Mind Institute at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, and IBM, maker of the Blue Gene high-performance computing (HPC) platform. The success of the Blue Gene HPC platform is evident in the fact that four of the top ten computers on the June 2008 top 500 supercomputer list (<http://top500.org>) were IBM Blue Gene architecture machines. In the same list, the IBM Blue Gene computer in the HPC center at EPFL (<http://hpc.epfl.ch>) was ranked 103.

The goal of the Blue Brain computational project, directed by Henry Markram and managed by Felix Schürmann and involving a team of 35 international scientists, is to model in great detail the cellular infrastructure and electrophysiological interactions within the cerebral neocortex, which constitutes about 80% of the brain and is believed to host cognitive functions such as language and conscious thought. Initiated by Henry Markram, an experimental biologist, the project started out with the goal of reverse-engineering the neocortex (Markram 2006). Experimental data has been collected for different types of cells, electrical behaviors, and connectivities. Experimental imaging data is used to construct a faithful *in silico* replica at the cellular level of the neocortical column of a young rat. The detailed multiscale model consists of 10,000 neurons (involving approximately 400 compartments per neuron, using a Hodgkin-Huxley cable model), 340 types of neurons, 200 types of ion channels, and 30 million dynamic connections (synapses) (Kozloski et al. 2008). The whole system is described by millions of ordinary differential equations modeling electrochemical synapses and ion channels. The main objective of the model is to explain the electrical, morphological, synaptic, and plasticity diversity observed in the laboratory, and to understand emergent network-level phenomena. Because of the nature of the model, it can be used both in an upscaling fashion (what network-level behavior emerges from a given cellular/connection

structure?) as well as in a downscaling fashion (how do changes in the neocortical environment such as injury and drug regimens, and changes in physiological parameters such as electrolyte concentration, impact network behavior and individual neurons and synapses?).

Of additional note is another neocortex simulation project being pursued by the Computational Biology and Neurocomputing Group at the Royal Institute of Technology (CBNG-RIT) in Stockholm, Sweden (Djurfeldt et al. 2008). At its largest incarnation, this simulation includes 22 million neurons and 11 billion synapses; it also is implemented on the IBM Blue Gene/L computing platform. The CBNG-RIT model involves a more coarse-grained description than the Blue Brain Project.

Yoshimura Group, Department of Systems Innovation, School of Engineering, University of Tokyo

The 21st Century Center of Excellence Program on Mechanical Systems Innovation in the School of Engineering of the University of Tokyo supports a number of projects that, if successful, will result in multiscale simulations relevant to energy and sustainability (see the site report in Appendix B). As one example, Shinobu Yoshimura, a faculty member in the Department of Systems Innovation, is developing a multiscale/multiphysics simulation for predicting the degree to which nuclear power plants can be made quake-proof. He is building on his expertise and prior experience in large-scale mechanical systems simulation (Yoshimura et al. 2002; Kim et al. 2004; Chang et al. 2005; Yoshimura 2006).

SUMMARY OF KEY FINDINGS

True multiscale simulation has long been a goal of SBE&S. It quite naturally arises as a desirable capability in virtually every application area of SBE&S. It particularly arises when the notion of design and/or control comes into play, since in this case the impact of changes in problem definition at the larger scales needs to be understood in terms of the impact on structures at the smaller scales. For example, in materials, the question might be how changes in the desirable performance characteristics of a lubricant translate into differences in molecular structure of the constituent molecules, so that new lubricants can be synthesized. In biology, the question might be what needs to happen at the cell-cell interaction level in order to control the spread of cancer. In energy and sustainability, the goal might be to understand how setting emissions targets for carbon dioxide at power plants in the United States may affect the growth and biodiversity of South American rain forests. Answering all of these questions requires a combination of downscaling and upscaling, and so fall in the realm of multiscale simulation.

The WTEC panel's key findings on multiscale simulation are summarized below:

- Multiscale modeling is exceptionally important. It holds the key to making SBE&S more broadly applicable in areas such as design, control, optimization, and abnormal situation modeling.
- Successful examples exist within narrow disciplinary boundaries; these include, but are not limited to
 - Crack propagation within materials
 - Computational neuroscience
 - Global climate modeling
- Attempts to develop general strategies have not yet succeeded. Some general mathematical principles exist for upscaling, such as homogenization theory (Cioranescu and Donato 2000). As noted in the state-of-the-art section, the equation-free method of Kevrekidis and coworkers (Bold et al. 2007; Kavousanakis, Erban et al. 2007; Papavasiliou and Kevrekidis 2007; Roberts and Kevrekidis 2007) is another attempt to develop a general approach to multiscale modeling.
- Because of the immediate and unavoidable needs for design and control, industry performs multiscale modeling with varying degrees of success. The lack of standards-based interoperability of codes is major impediment to them in linking together codes at various scales. This problem was cited by a number of the companies the WTEC panel visited. Interoperability of codes does not, in itself, result in a multiscale model.

Code interoperability should be regarded as a necessary but not sufficient requirement for the development of general-purpose multiscale simulation capabilities.

- U.S. research in multiscale simulation is on a par with Japan and Europe. In fact, U.S. leadership in high-performance computing resources has helped the United States to be competitive in this area. This is because in order to validate a multiscale simulation methodology it is necessary to perform the simulation at full detail (i.e., using the most fundamental level of description) enough times to provide a validation data set. For example, to validate a crack propagation simulation, the whole problem should be simulated multiple times at the first-principles level to provide the set of data to which the multiscale simulation should be compared. However, one concern is that access to the largest computational facilities (such as the emerging petascale platforms) may be difficult for a project whose stated aim is to run a series of petaflop-level simulations in order to develop and validate a multiscale modeling methodology. Access to petaflop-level resources today is generally focused on solving a small set of problems, as opposed to developing the methodologies that will enable the solution of a broader range of problems.
- For the most part, U.S. research in multiscale modeling is diffuse, lacking focus and integration, and federal agencies have not traditionally supported the development of codes that can be distributed, supported, and successfully used by others. By contrast, the Japanese and European approach is to fund large interdisciplinary teams, such as the OCTA project in Japan and the Blue Brain project in Switzerland, often with the goal of distributing the product codes either in open-source or commercial form.
- Solving the multiscale simulation challenge is without doubt an interdisciplinary endeavor; however, the tradition of interdisciplinary collaboration leading to community solutions is much stronger in Europe and Japan than in the United States. Some of the issues related to this are discussed in more detail in Chapter 3.

REFERENCES

- Bold, K.A., Y. Zou, I. Kevrekidis, and M. Henson. 2007. An equation-free approach to analyzing heterogeneous cell population dynamics. *Journal of Mathematical Biology* 55(3):331–352.
- Broughton, J.Q., F.F. Abraham, N. Bernstein, and E. Kaxiras. 1999. Concurrent coupling of length scales: Methodology and application. *Physical Review B* 60(4):2391–2403.
- Chang, Y.S., H.O. Ko, J.-B. Choi, Y.-J. Kim, and S. Yoshimura. 2005. Parallel process system and its application to steam generator structural analysis. *Journal of Mechanical Science and Technology* 19(11):2007–2015.
- Cioranescu, D., and P. Donato. 2000. *An introduction to homogenization*. Oxford: Oxford University Press.
- Csányi, G., T. Albaret, G. Moras, M.C. Payne, and A. De Vita. 2005. Multiscale hybrid simulation methods for material systems. *Journal of Physics-Condensed Matter* 17:R691–R703.
- Csányi, G., T. Albaret, M.C. Payne, and A. De Vita. 2004. “Learn on the fly”: A hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys. Rev. Lett.* 93(17):Art. No. 175503.
- Curtin, W.A., and R.E. Miller. 2003. Atomistic/continuum coupling in computational materials science. *Modelling and simulation in materials science and engineering* 11(3):R33–R68.
- Djurfeldt, M., M. Lundqvist, C. Johansson, M. Rehn, Ö. Ekeberg, and A. Lansner. 2008. Brain-scale simulation of the neocortex on the IBM Blue Gene/L supercomputer. *IBM Journal of Research and Development* 52(1-2):30–41.
- Duan, Y., and P.A. Kollman. 1998. Pathways to a protein folding intermediate observed in a 1- microsecond simulation in aqueous solution. *Science* 282(5389):740–744.
- Farrell, D., E. Karpov, and W.K. Liu. 2007. Algorithms for bridging scale method parameters. *Computational Mechanics* 40(6):965–978.
- Kavousanakis, M.E., R. Erban, A. G. Boudouvis, C.W. Gear, and I.G. Kevrekidis. 2007. Projective and coarse projective integration for problems with continuous symmetries. *Journal of Computational Physics* 225(1):382–407.
- Kim, J.C., J.B. Choi, Y.-J. Kim, Y.-H. Choi, Y.-W. Park, and S. Yoshimura. 2004. Development of an integrity evaluation system on the basis of cooperative virtual reality environment for reactor pressure vessel. *Advances in nondestructive evaluation*, Pt 1-3. Zurich-Uetikon: Trans Tech Publications Ltd. 270-273:2244–2249.

- Klepeis, J.L., H.D. Schafroth, K.M. Westerberg, and C.A. Floudas. 2002. Deterministic global optimization and ab initio approaches for the structure prediction of polypeptides, dynamics of protein folding, and protein-protein interactions. *Computational Methods for Protein Folding*. New York: John Wiley & Sons, Inc. 120:265–457.
- Kofke, D.A. 1993. Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line. *Journal of Chemical Physics* 98(5):4149–4162.
- Kozloski, J., K. Sfyrakis, S. Hill, F. Schuermann, C. Peck, and H. Markram. 2008. Identifying, tabulating, and analyzing contacts between branched neuron morphologies. *IBM Journal of Research and Development* 52(1-2):43–55.
- Liu, W.K., E.G. Karpov, and H.S. Park. 2006. *Nano mechanics and materials: Theory, multiscale analysis and applications*. New York: John Wiley and Sons.
- Liu, W.K., E.G. Karpov, S. Zhang, and H.S. Park. 2004. An introduction to computational nanomechanics and materials. *Computer Methods in Applied Mechanics and Engineering* 193(17-20):1529–1578.
- Markram, H. 2006. The Blue Brain Project. *Nature Reviews Neuroscience* 7(2):153–160.
- Osguthorpe, D.J. 2000. Ab initio protein folding. *Current Opinion in Structural Biology* 10(2):146–152.
- Papavasiliou, A., and I.G. Kevrekidis. 2007. Variance reduction for the equation-free simulation of multiscale stochastic systems. *Multiscale Modeling & Simulation* 6(1):70–89.
- Roberts, A.J., and I.G. Kevrekidis. 2007. General tooth boundary conditions for equation free modeling. *Siam Journal on Scientific Computing* 29(4):1495–1510.
- Soler, J.M., E. Artacho, J.D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal. 2002. The SIESTA method for ab initio order-N materials simulation. *Journal of Physics-Condensed Matter* 14(11):2745–2779.
- Yoshimura, S. 2006. Virtual demonstration tests of large-scale and complex artifacts using an open source parallel CAE system, ADVENTURE. *Advances in safety and structural integrity 2005*. Zurich-Uetikon: Trans Tech Publications, Ltd. 110:133–142.
- Yoshimura, S., R. Shioya, H. Noguchi, and T. Miyamura. 2002. Advanced general-purpose computational mechanics system for large-scale analysis and design. *Journal of Computational and Applied Mathematics* 149(1):279–296.

CHAPTER 10

BIG DATA, VISUALIZATION, AND DATA-DRIVEN SIMULATIONS

Sangtae Kim

INTRODUCTION

One hundred meters under the Franco-Swiss border, along a circular path 17 miles in circumference, the collision of opposing beams of protons streaming at 99.999999% of the speed of light generates 1 petabyte per second of raw data. (See Figure 10.1.) The gold standard in data-intensive scientific cyberinfrastructure distills this massive stream of data to “merely” petabytes per year of interesting archival events for further scientific analysis in the quest for a fundamental understanding of the nature of our material universe. The smooth transition and acceptance of “petascale” in our scientific lexicon belies the staggering magnitudes – peta, or 10^{15} , is the scale that bridges the macroscopic and molecular worlds; consider that each byte in a 30-petabyte storage system would map one to one to each molecule in a 1 mm by 1 mm by 1 mm cubic air sample. (Note: .224 cc = 6×10^3 peta and .00112 cc = 3×10 peta). More than two thousand miles away, in the massive Khurais oil fields of Saudi Arabia, 100 injection wells ring an area of 2700 square miles. (See Figure 10.2.) Sophisticated SBE&S reservoir models use over 2.8 million 3D images of the Khurais underground strata to dictate dynamically the real-time operation of injection and production wells (King 2008). Faced with declining output from aging fields, the optimal development of Khurais is key to maintaining the “growth option” for output in the world’s largest exporter of oil. These two examples, at opposite ends of the spectrum of fundamental research/applied technology, and the extremum between the quest for knowledge of the nature of the universe and the pragmatic concerns for energy to fuel global economic development, underscore the pivotal role of “big data” and data-driven SBE&S. Furthermore, consider that the flood of data in the post-genomic era is the daunting barrier to progress in life sciences research and the search for cures. Perhaps more urgently, consider the tantalizing prospect of data-intensive SBE&S for fresh approaches to understanding the complex system dynamics in the global financial marketplace and creation of new and reliable tools for governmental financial policymakers. It is no wonder then that the popular science and technology media prescribe the data deluge of the petabyte era as the end of the scientific method (Anderson 2008).

In our global travels for the SBE&S project and the search for leadership in big data issues, an overarching finding is that aside from notable exceptions as mentioned in the following pages, there is a significant gap in the level of commitment to data vs. computational infrastructure. We use the term “infrastructure” broadly, to denote not just the hardware and software, but human infrastructure (training and education) as well. The gap is narrower in the life sciences disciplines and in industrial R&D laboratories, but these efforts are in turn limited by the overall gap across the SBE&S educational landscape. Therefore, this motivated the organizational scheme for this chapter on “big data:” we lead with the uniformly strong cyberinfrastructure organization of the particle physics research community, then describe the life sciences research examples across the globe, and conclude with an aggregate view of the enterprise scale data infrastructure and knowledge management/decision support systems in industrial R&D. We close the chapter with an overview of the (sorry) state of data infrastructure on university campuses and its impact on current capabilities for training and education.

PARTICLE PHYSICS RESEARCH: PETABYTES PER SECOND

The particle physics research community and their search for the elusive Higgs boson to complete the cast of characters for the standard model of the structure of subatomic matter, seem far removed from the pressing economic concerns of most of humanity and the underlying theme of this SBE&S report – the importance of SBE&S in accelerating and sustaining U.S. economic development and technological advancement. And yet this is the same community whose previous efforts at community-scale integration and organization of scientific workflow were the origins of the World Wide Web and the Internet-based economy that we enjoy today. The current challenge faced by the particle physics community is the distillation, by seven orders of magnitude, of experimental results generated at the rate of petabytes per second, into a globally-accessible, multitier, distributed database of archival results with the associated informatics and visualization tools. Undoubtedly this unprecedented scale will break new ground for data infrastructure and knowledge management, with profound impact(s) that may perhaps rival that of the World Wide Web.

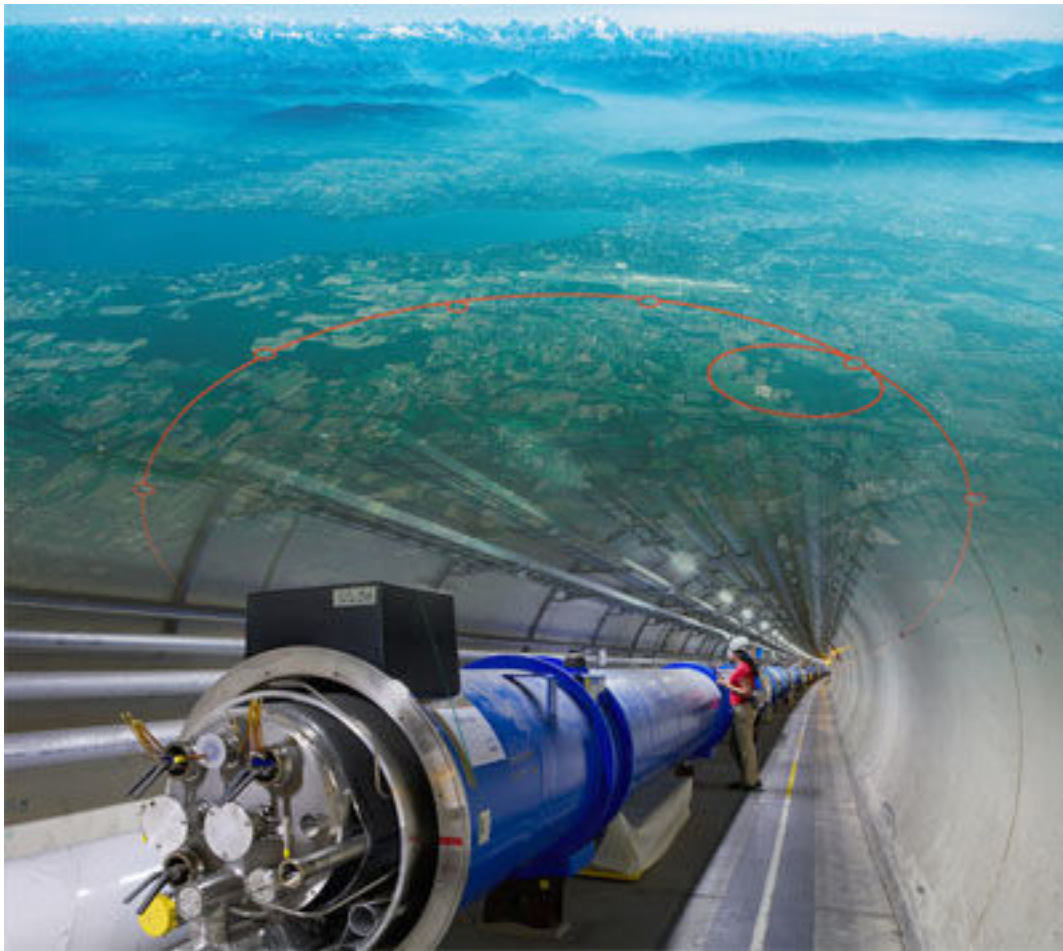


Figure 10.1. Image of the Large Hadron Collider. (Source: <http://lhc-milestones.web.cern.ch/LHC-Milestones/images/photos/ph07-1.jpg>)

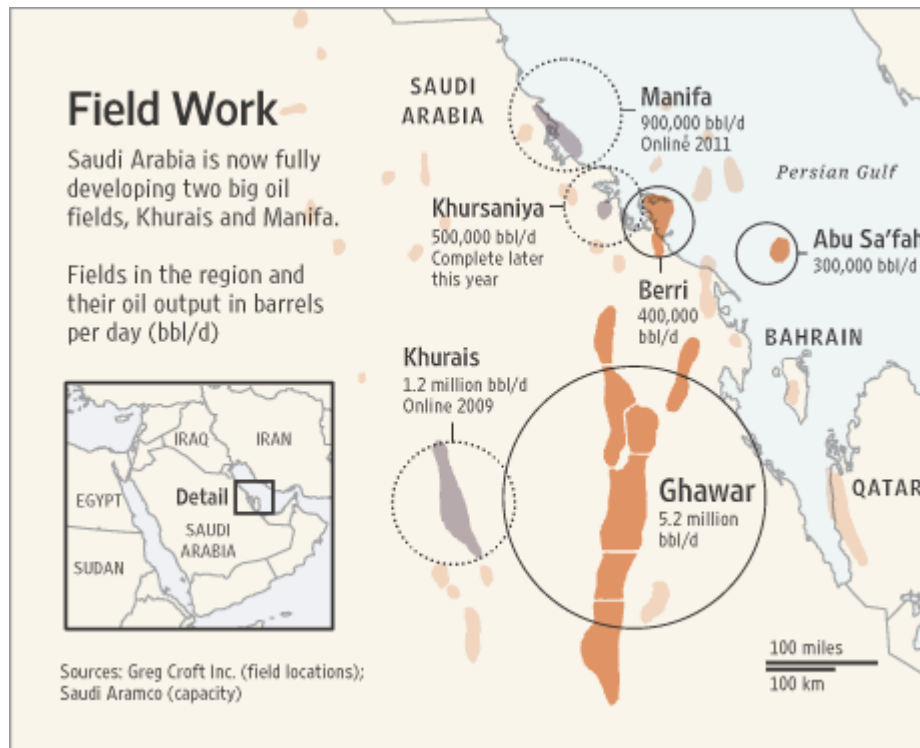


Figure 10.2. Current and future oilfields in Saudi Arabia. (Source: *Wall Street Journal*, April 22, 2008.) The largest undeveloped oil fields including the Khurais (colored light gray) require significant SBE&S resources.

The Large Hadron Collider (LHC) (see Figure 10.1 again) and the global experimental collaboration centered at CERN (European Center for Nuclear Research) are actually a series of interrelated experiments and detector design projects (ALICE, A Large Ion Collider Experiment; ATLAS, A Toroidal LHC Apparatus; and CMS, Compact Muon Solenoid are just three of the six detector experiments). For much of the year, the opposing beams contain protons streaming at 99.999999% of the speed of light, but about one month per year, heavy ions (e.g. lead ions) are used (see Maury). The enormity of the data challenges originates from the fact that the results of scientific interest (either proving or disproving the standard model) are extremely rare events – data architecture and the associated informatics to sift through a data torrent of ion-ion collision tracks and cross-validation across multiple experiments are central to this grand experiment. A summary of the notable informatics tools and integration with data-intensive computing in the open science grid (OSG) is described by Livny. Finally, the enormity of the informatics challenges serves to underscore the end-to-end role of SBE&S: data intensive SBE&S is the closing book-end that pairs with the civil engineers' simulations for tunnel and instrument alignment that preceded the construction phase!

BIG DATA IN LIFE SCIENCES RESEARCH

Ecole Polytechnique Federale de Lausanne, Blue Brain Project - Switzerland

New heights in big data are attained in this collaborative effort between EPFL and IBM to elucidate the cognitive function of the human brain with a 3D digital model (“modeling in every detail”) of the cellular infrastructure and electrophysiological interactions within the cerebral neocortex. Dubbed the “blue brain project,” the data challenges arise from a multiscale model featuring 10,000 neurons, 340 types of neurons, 200 types of ions and 30 million connections. The current capabilities already feature the *in silico* replication at the cellular level, the neocortical column of a rat brain. The ability to perform *in silico* replication of experimental electrical signals as

a function of morphological and synaptic changes would represent a landmark advance in brain research and progress towards the cure for neurodegenerative diseases.

Daresbury Laboratory/Science and Technology Facilities Council, e-HPTX – United Kingdom

As described in our visit to Daresbury Laboratory of the STFC, the United Kingdom's fourth Collaborative Computational Project (CCP4) focused on software development for structure solution and refinement in computational and structural biology (X-ray crystallography). This e-Science and BBSRC (Biotechnology and Biological Sciences Research Council)-sponsored project recognized at a very early stage the importance of data management and workflow as foundational elements of life sciences research: the automation of structure determination, e-HPTX (management of workflow), DNA data collection, and PIMS (laboratory information management system) were all developed as part of CCP4.

Systems Biology Institute - Tokyo, Japan

The research plans at the Systems Biology Institute founded by systems biology pioneer Hiroaki Kitano include data and software infrastructure to support the SBE&S modeling efforts in systems biology. The current plans that impact on "big data" focus on the Systems Biology Markup Language (SBML), Systems Biology Graphical Notation (SBGNL) and Web 2.0 Biology. These elements are all crucial to an organized accumulation of systems-biological knowledge towards progress on major diseases such as cancer, diabetes, and infectious diseases. Although primarily labeled under the bigger umbrella of software infrastructure investments, the data-intensive nature of the activities in this research institute leads us to believe that this is one of the most notable and sustained investments in data cyberinfrastructure in the SBE&S landscape.

Earth Simulator Center and Plans for the Life Simulator – Japan

Jumping in scale from the molecular to the other end in the life sciences (societal scale), new big-data challenges are encountered in Japan's next-generation project, the "life simulator," featuring agent-based models with six billion agents, i.e., a one-to-one mapping to the human population on the planet. How is the mountain of (personal!) data to be gathered and organized to simulate individual profiles and activities on such an awesome scale? What are the implications from confidentiality and privacy issues? These efforts hint at the emerging research opportunities in the social sciences, their impact on policy makers, and the inter-dependence with next generation capabilities in SBE&S as Japan transitions from the Earth Simulator to the Life Simulator. And for the world at large, this may be the first steps in applying SBE&S towards the modeling and (better!) understanding of the complex nonlinear dynamic that we are witnessing in the global financial markets, and thereby providing new tools for governmental financial authorities.

BIG DATA IN INDUSTRY: ENTERPRISE-SCALE KNOWLEDGE INTEGRATION

As noted in the beginning of this chapter, industrial R&D laboratories with their enterprise-scale knowledge integration efforts in the cause of intellectual property (IP) generation constitute the other important group of sustained investments and activities in "big data." In light of the varying but in some cases stringent constraints on attribution of details of in-house research activities, we have aggregated without attribution the activities in big data as it pertains to workflow and knowledge management. Fortunately, such an approach has the fortuitous effect of revealing common patterns that transcend industry types, rather than constricting insights for our study.

In industrial practice, we rarely (if ever) encounter SBE&S as a "point solution" in a manufacturing process or R&D operation. Instead, SBE&S activities appear at various steps in the process or flow of work (i.e. workflow). As the work flows from step to step and functional area (department) to functional area, industrial firms face the challenge of interoperability, or how to translate the (SBE&S) output of one area to the requisite input of the next. In the chemical process industry, the CAPE-Open standard is an example of such an effort including the formation of the COLAN consortium to maintain the open standards as described by Pons (2005). The companies with leading capabilities in SBE&S have taken this problem far beyond the simplistic task of pairwise data conversion (e.g. from CAD tools to CFD simulations) and have proceeded further to implement enterprise-scale

data infrastructure and workflow engines. Notable informatics capabilities include markup languages, the semantic Web (Web 2.0), and data visualization tools, e.g., in the biotechnology and pharmaceutical companies.

Some early successes of investments in data architecture, workflow, and visualization include taking weeks or months out of product development timelines and making SBE&S tools accessible to the nonexperts (companies note that no one person is an expert in all the functional areas of the workflow). We saw this trend across multiple industries including automotive, chemical, consumer products, and pharmaceutical companies. Further progress is impeded by the conflicting goals of the software vendor community and the challenges of interfacing workflow outputs with the regulatory agencies (e.g., for the biotechnology and pharmaceutical industry, see the resource constraints at the U.S. Food and Drug Administration as described in Cassell 2007). Given the high stakes, resolution of these challenges is likely to remain a focus area for elevating the use of SBE&S to new heights in the industrial R&D setting.

BIG DATA – THE ROAD AHEAD FOR TRAINING AND EDUCATION

In this report, we have devoted significant attention to the current state of SBE&S training and education around the world (see Chapter 11). For big data, these same challenges in training and education of the next generation of SBE&S researchers are compounded by the additional complexity arising from inadequate data infrastructure. We contrast the typical state of affairs (e.g., in any major U.S. research university) and then provide a glimpse of the future by describing the best academic example (University of Stuttgart, Simulation Technology Excellence Cluster program) from our global travels.

Consider a graduate student on the university campus of a major U.S. research university today. The standard disk storage allocation for graduate students is typically 500 megabytes, or in some cases, perhaps a generous perk of 1 gigabyte (10^9 bytes) of storage. But the dissertation research may generate a terabyte (10^{12} bytes) of storage each semester! To put this in a computing perspective, this is tantamount to giving the student a portable computer instead of access to a terascale supercomputer. Fortunately, research progresses because the student runs a mini data center and infrastructure at home, using references from the Internet to build a RAID storage farm in the basement. This example underscores the gap in data infrastructure and the steps required to close that gap in terms of education and training for the typical student (i.e., someone who is not a technology enthusiast running a datacenter at home).

The Simulation Technology Excellence Cluster at the University of Stuttgart, benefits from the German Research Foundation (DFG) programs to support collaborative research centers (SFB), trans-region projects (TR), technology transfer units (TFR), research units (FOR), priority program and excellence initiatives. The result is a superbly integrated infrastructure for computation and data management as described in our site report for the university. The integrated data management and interactive visualization environment includes significant new technologies in human-computer interfaces for SBE&S-model setup and sensor networks for real-time control simulations. At Stuttgart, what we saw was not just the integration of data management and computation, but the entire spectrum of SBE&S activities – the success of this approach points to the road ahead for the global SBE&S community.

SUMMARY OF KEY FINDINGS

The role of big data and visualization in driving new capabilities in SBE&S is pronounced and critical to progress in the three thematic areas of this report. We can also interpret the “big data” challenges faced by the SBE&S community as a recursive resurfacing of the CPU-memory “data bottleneck” paradigm from the chip architecture scale to societal scale.

- The locus of large scale efforts in data management correlates by discipline and not by geographical region. The biological sciences (as described in the Chapter 2 on Life Sciences and Medicine) and the particle physics communities are pushing the envelope in large-scale data management and visualization methods. In contrast, the chemical and material science communities lag in prioritization of investments in the data infrastructure. In the biological sciences, in both academic laboratories and industrial R&D centers, there is

an appreciation of the importance of integrated, community-wide infrastructure for dealing with massive amounts of data, and addressing issues of data provenance, heterogeneous data, analysis of data, and network inference from data.

- Similarly, and even within a given disciplinary area, there is a notable difference between industrial R&D centers and academic units, with the former placing significantly more attention to data management infrastructure, data supply chain and workflow, in no small part due to the role of data as the foundation for intellectual property (IP) assets. In contrast, most universities lack a campus-wide strategy for the ongoing transition to data-intensive research and there is a widening gap between the data infrastructure needs of the current generation of graduate students and the capabilities of the campus IT infrastructure.
- Industrial firms are particularly active and participate in consortia to promote open standards for data exchange – a recognition that SBE&S is not a series of point solutions but integrated set of tools that form a workflow engine. Companies in highly regulated industries, e.g., biotechnology and pharmaceutical companies, are also exploring open standards and data exchange to expedite the regulatory review processes for new products.
- Big data and visualization capabilities go hand in hand with community-wide software infrastructure; a prime example is in the particle physics community and the Large Hadron Collider infrastructure networking CERN to the entire community in a multitier fashion. Here, visualization techniques are essential given the massive amounts of data, and the rarity of events (low signal to noise ratio) in uncovering the new science.
- Big data, visualization and *dynamic* data-driven simulations are crucial technology elements in the production of transportation fuels from the last remaining giant oil fields. Global economic projections for the next two decades and recent examples of price elasticity in transportation fuels suggest that *the scale of fluctuation in reservoir valuations would be several orders of magnitude beyond the global spending in SBE&S research.*
- As was the case for the three SBE&S themes and crosscutting topics, “big data” has training and education challenges. Appropriately trained students who are adept with data infrastructure *issues will become increasingly important for research at the SBE&S frontiers.*

REFERENCES

- Anderson, C. 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, June 23.
- Cassell, G. 2007. *FDA science and mission at risk, report of the subcommittee on science and technology*, FDA Science Board.
- “Data-infrastructure-at-home” (including RAID data storage) references: <http://en.wikipedia.org/wiki/RAID>, www.tomshardware.com/us/, www.smallnetbuilder.com/
- Earth Simulator Center (ESC). 2007. *Annual report of the earth simulator center*.
- Hucka, M. et al., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524-531.
- King, N. 2008. Saudis face hurdles in new oil drilling. *Wall Street Journal*, April 22.
- Kitano, H. 2002. Systems biology: a brief overview. *Science* 295:1662-1664.
- Livny, M. LHC SBE&S computational infrastructure. <http://pages.cs.wisc.edu/~miron/>
- Maury, S. Ions for LHC project (I-LHC). <http://project-i-lhc.web.cern.ch/project-i-lhc/Welcome.htm>
- Pons, M. 2005. Introduction to CAPE-Open. <http://www.colan.org>

CHAPTER 11

EDUCATION AND TRAINING

Celeste Sagui

The country that out-computes will be the one that out-competes.

– Council on Competitiveness (<http://compete.org>)

INTRODUCTION

The development and improvement of computer simulations in diverse fields represent one of most important successes in science and engineering in the last thirty years. For the United States to remain competitive in the Sciences and Engineering, the proper teaching of computer modeling and simulation methods and tools in colleges and universities becomes of paramount importance. In the words of Tinsley Oden at the WTEC U.S. Baseline SBE&S Workshop (2007), *“Our educational institutions must be prepared to equip tomorrow’s scientists and engineers with the foundations of modeling and simulation and the intellectual tools and background to compete in a world where simulation and the powerful predictive power and insight it can provide will be a cornerstone to many of the breakthroughs in the future.”*

The United States used to lead the way in education for simulation-based engineering and science (SBE&S). For instance, U.S. certificate programs in Computational Science and Engineering (CSE) were among the first in the world for SBE&S. However many of these programs have not embraced new developments in the last decade. There have been many new efforts throughout the country to fill this vacuum, especially in the form of summer programs, that include schools associated with the main national codes such as AMBER, CHARMM, and NAMD. There are scattered efforts under individual PI and team NSF and DOE grants. The State of Ohio launched the Ralph Regula School of Computational Science in Ohio to focus on “blue-collar” computing, reaching out to K-12 programs and undergraduate students across the state. The Great Lakes Consortium for Petascale Computation’s Virtual School of Computational Science and Engineering, launched last year to help prepare the next generation of researchers to exploit petascale computation and beyond, focuses on graduate and postgraduate education. Nanohub, which supports a virtual community primarily in nanoelectronics, offers educational components along with software access and distribution. In addition, some U.S. institutions are being restructured. For instance, the Institute for Computational Engineering and Sciences at the University of Texas, Austin, offers both education and infrastructure for interdisciplinary programs in CSE and Information Technology (IT). The faculty comes from seventeen departments and four schools and colleges. The institute offers an independent PhD program, with independent faculty evaluations for promotion and tenure.

Despite these new, bold programs, the United States has seen less vigorous growth than other countries in education and training in SBE&S over the past decade. The globalization of higher education is accelerating. According to the *Science and Engineering Indicators* (NSB 2008), the United States continues to attract the largest number and fraction of foreign students worldwide, but these numbers have generally decreased in recent

years.⁷ Some of the reduction in numbers of foreign students studying in the United States is believed to be the result of expanded access to higher education taking place in many countries in Europe and several countries in Asia since 1990. In addition, universities in many countries—including Australia, the United Kingdom, Canada, Japan, and Germany—have actively expanded their enrollment of foreign Science and Engineering (S&E) students. This chapter reports on the WTEC panel’s findings for SBE&S education in the nation and abroad.

WHERE THE UNITED STATES STANDS

Since SBE&S is a relatively new and highly interdisciplinary area of study and application, it is difficult to come up with education statistics belonging exclusively to the area. However, it is still enlightening to examine some general S&E trends for which good statistics exist.

The United States spends more on R&D than any other country in the world, approximately equal to the sum of the other G7 countries—the United Kingdom, Germany, France, Italy, Japan, and Canada—combined (NSB 2008). Figure 11.1 gives a “map” of the countries with the highest R&D investment. The United States has more researchers (OECD 2006) and more patents (NSB 2006) than any other country. On the other hand, a few other countries invest more than the United States on non-defense R&D as a percentage of their GDP (e.g., Japan and Germany). At present, emerging economies (led by the People’s Republic of China) have the fastest rate of growth in R&D investment.

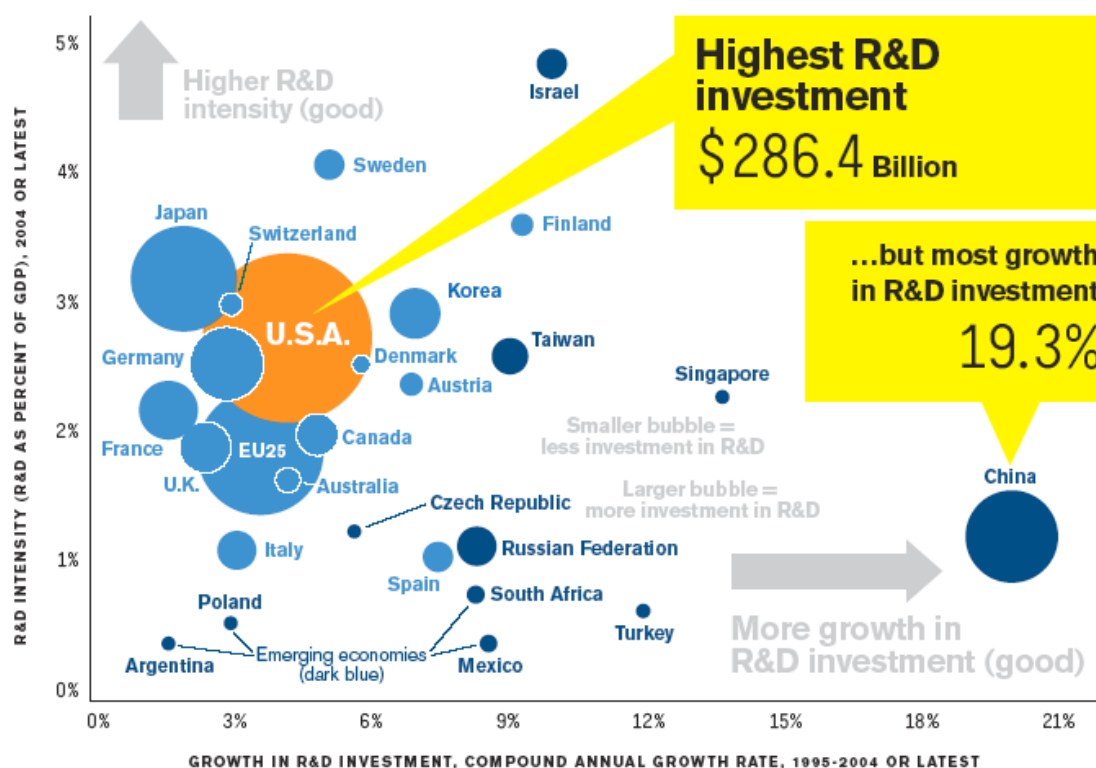


Figure 11.1. U.S. R&D investment remains the world’s largest, but others are increasing their investment faster. This figure has been taken from The Council on Competitiveness, *Competitiveness Index: Where America Stands* (2007). The figure, in turn, was created using data from OECD’s *Main Science and Technology Indicators 2006* (OECD 2006).

⁷ 2007–2008 saw an increase in international student enrollment, attributed to the faltering U.S. dollar and relaxed visa constraints for Chinese students (*World Journal* 2008).

Figure 11.2 shows that the United States has approximately 37% of global R&D spending, 52% of all new patents, 30% of all scientific publications, 29% of all scientific researchers, and 22% of all doctorates in S&E. From these figures, it is clear that the United States continues to be an absolute leader in S&E. However, Figure 11.2 also shows that its global share has fallen as other countries have increased their science and technology initiatives and efforts.

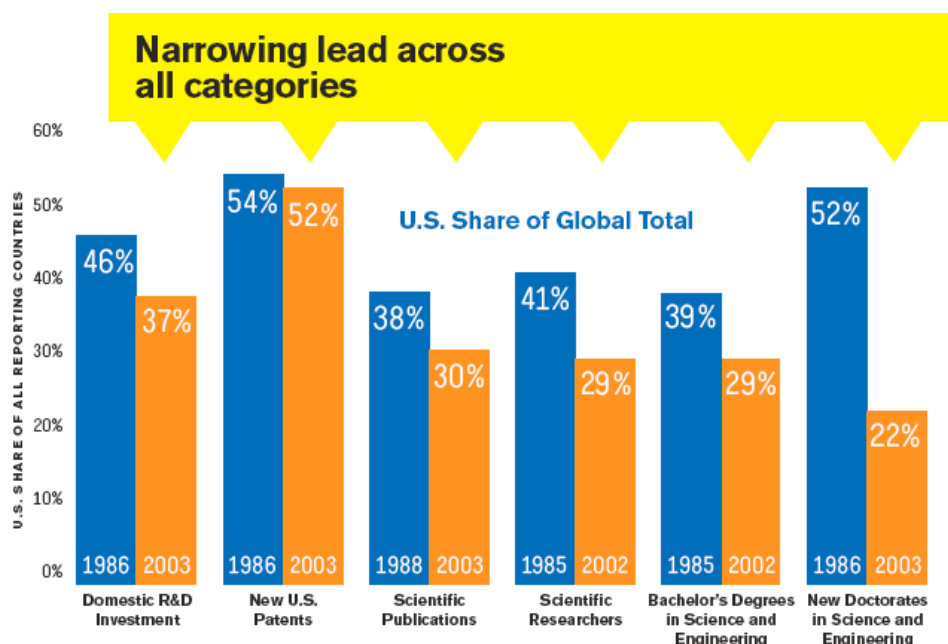


Figure 11.2. U.S. share of global output has fallen across a range of science and technology metrics. This figure has been taken from The Council on Competitiveness, *Competitiveness Index: Where America Stands* (2007). The figure was created using data from OECD's *Main Science and Technology Indicators* (2006), NSF's *Science and Engineering Indicators* (NSB 2006), and the U.S. Patent and Trademark Office.

Although the United States still has the most citations and the top-cited publications, for the first time the European Union has surpassed the United States in total number of scientific publications and the production of PhDs in S&E, according to the data presented in Figures 11.3 and 11.4. In particular, the WTEC bibliometrics study (see Appendix E) shows that the number of SBE&S publications worldwide is double the number of all S&E publications (5% vs. 2.5%), and while in 2007 the United States dominated the world SBE&S output at 27%, China moved up to second place, at 13% (although with low citation indexes). In addition, the U.S. output in SBE&S is less than that of EUR-12 (12 countries of the European Union), with that difference increasing over time.

Problems that especially affect the SBE&S fields include the following:

1. Loss of international students: This is due to both visa restrictions and the diminution of the country's "attractiveness" to foreigners.
2. Loss of PhD students and faculty: The scientific field is no longer attractive to U.S. students; faculty and principal investigators and staff researchers in national labs are being lost to industry and other markets.
3. Lack of adequate training of students in SBE&S: Students are knowledgeable in running existing codes (with visualization), but they are unable to actually write code or formulate a mathematical framework.
4. Rigid "silo" academic structure: The entire academic structure (budgets, courses, curricula, promotion, and tenure, etc.) is aimed at maintaining a vertically structured, relatively insulated disciplinary environment. The system promotes departmental loyalty and highly discourages interdisciplinary work.

5. Serious danger of compromising creativity and innovation: In the present funding environment, many alliances are made for the purpose of seeking funding, independent of the scientific value of the research.

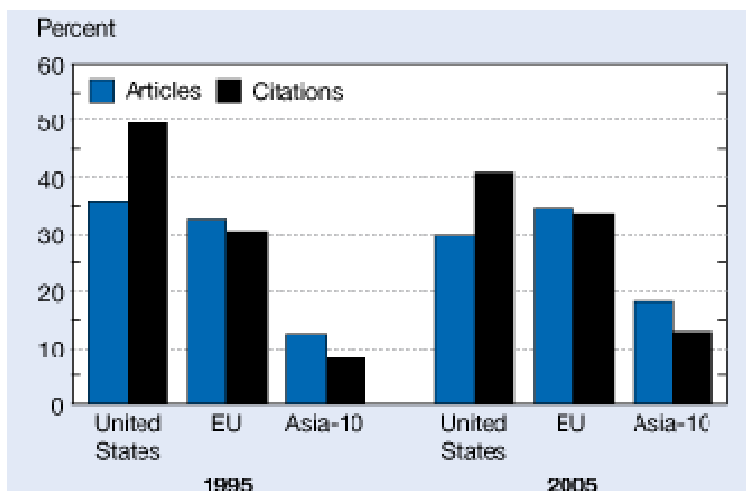


Figure 11.3. S&E articles and citations in all fields, by selected region/country: 1995 and 2005. The figure is extracted from *Science and Engineering Indicators 2008* (NSB 2008). The share of all articles is based on a 3-year period. Article counts are from a set of journals covered by the Science Citation Index (SCI) and Social Sciences Citation Index (SSCI). Articles are classified by the year they entered the database and are assigned to region/country/economy on basis of institutional addresses listed in the article. In the case of articles and citations on a fractional-count basis, i.e., for articles with collaborating institutions from multiple countries/ economies, each country/economy receives fractional credit on the basis of the proportion of its participating institution. Citation data are based on the year the article entered the database. Citation counts are based on a 3-year period with 2-year lag, e.g., citations for 1995 are references made in articles in the 1995 data tape to articles in the 1991–93 data tapes.

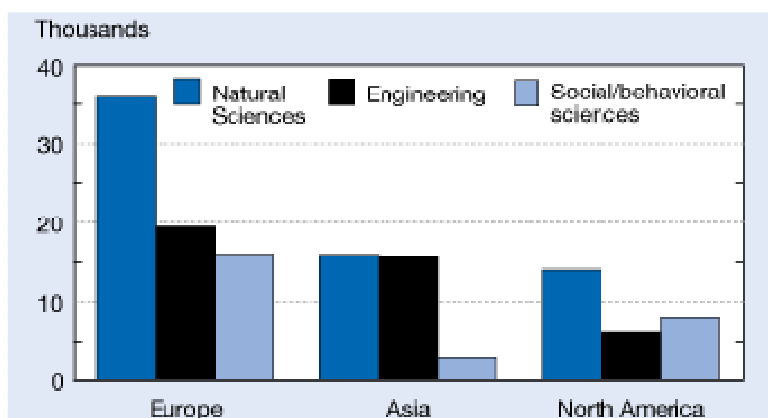


Figure 11.4. S&E doctoral degrees earned in Europe, Asia, and North America, by field: 2004 or most recent year (before 2004). The figure is extracted from *Science and Engineering Indicators 2008* (NSB 2008). Natural Sciences include physical, biological, earth, atmospheric, ocean, agricultural, computer sciences, and mathematics. Asia includes China, India, Japan, South Korea, and Taiwan. Europe includes Western, Central, and Eastern Europe. North America includes the United States and Canada. This figure was compiled based on the following sources: Organization for Economic Co-operation and Development, Education Online Database; United Nations Educational, Scientific, and Cultural Organization (UNESCO), Institute for Statistics database, <http://www.unesco.org/statistics> (April 2007); and national sources.

The U.S. loss of international students and postdocs due to restrictions on visas is far from breaking news. As pointed out by an article in *The Economist* (“Help Not Wanted,” April 10, 2008):

“This [the restriction on visas] is a policy of national self-sabotage. America has always thrived by attracting talent from the world. Some 70 or so of the 300 Americans who have won Nobel prizes since 1901 were immigrants (and many others were direct descendants from immigrants). Great American companies such as Sun Microsystems, Intel and Google had immigrants among their founders. Immigrants continue to make an outsized contribution to the American economy. About a quarter of information technology (IT) firms in Silicon Valley were founded by Chinese and Indians. Some 40% of American PhDs in science and engineering go to immigrants. A similar proportion of all the patents filed in America are filed by foreigners.

“These bright foreigners bring benefits to the whole of society. The foreigner-friendly IT sector has accounted for more than half of America’s overall productivity growth since 1995. Foreigner-friendly universities and hospitals have been responsible for saving countless American cities from collapse. Bill Gates calculates, and respectable economists agree, that every foreigner who is given an H1B visa creates jobs for five regular Americans.”

Other countries are benefiting enormously from the “leak” of talent that is affecting the United States. For instance, instead of the lottery system employed by the United States for assigning H-1B (temporary resident/foreign guest worker) visas, countries like Canada and Australia have a merit system that rewards educational accomplishments. Some companies in New Zealand hand out work visas with their job offers. The European Union is also considering a merit system to allow talented people to become European Union citizens.

Figure 11.5 shows first-time graduate enrollment in Science and Engineering in the United States by citizenship and field from 2002 to 2006. Visa holders constitute approximately 50% of the total S&E enrollment, surpassing in 2004–06 the numbers of U.S. citizens and permanent residents enrolled in Computer Sciences and Engineering. Although other countries (such as the United Kingdom) also have a very large proportion of international students, the absolute number of international students in the United States exceeds by far those from other countries.

Figure 11.6 shows the absolute number of international students per country (left) and the percentage of doctorate degrees earned by foreigners in the United States, the United Kingdom, Germany, and Japan.

According to the most recent *Science and Engineering Indicators* (NSB 2008), students on temporary visas earned more than a third (36%) of all S&E doctorates awarded in the United States in 2005. Temporary residents earned half or more of all U.S. doctorates in Engineering, Mathematics, Computer Sciences, Physics, and Economics in 2005.

The number of temporary resident postdoctoral students in the United States has definitely surpassed the number of U.S. citizen/permanent resident postdoctoral students in S&E (Figure 11.7). The temporary resident numbers among postdoctoral students went from 8,859 in 1985 to 26,975 in 2005, while the number of U.S. citizens/permanent residents among postdoctoral students went from 13,528 in 1985 to 21,678 in 2005, so that temporary visa holders accounted for 55% of all S&E postdocs in academic institutions in the fall of 2005. In particular, the percentage of temporary visa holders in 2005 was 59% in the Biological Sciences, 60% in Computer Sciences, 66% in Engineering, and 64% in the Physical Sciences (NSB 2008).

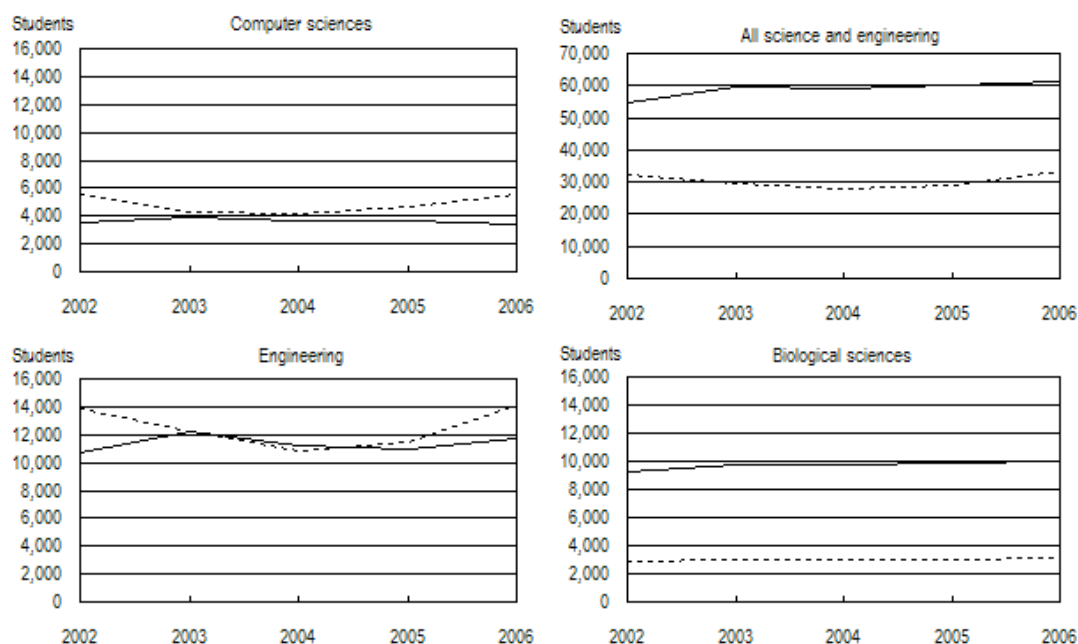


Figure 11.5. First-time, full-time graduate enrollment in S&E, by citizenship and field, 2002–2006. *Solid line:* U.S. citizens and permanent residents. *Dotted line:* visa holders (NSF 2007).

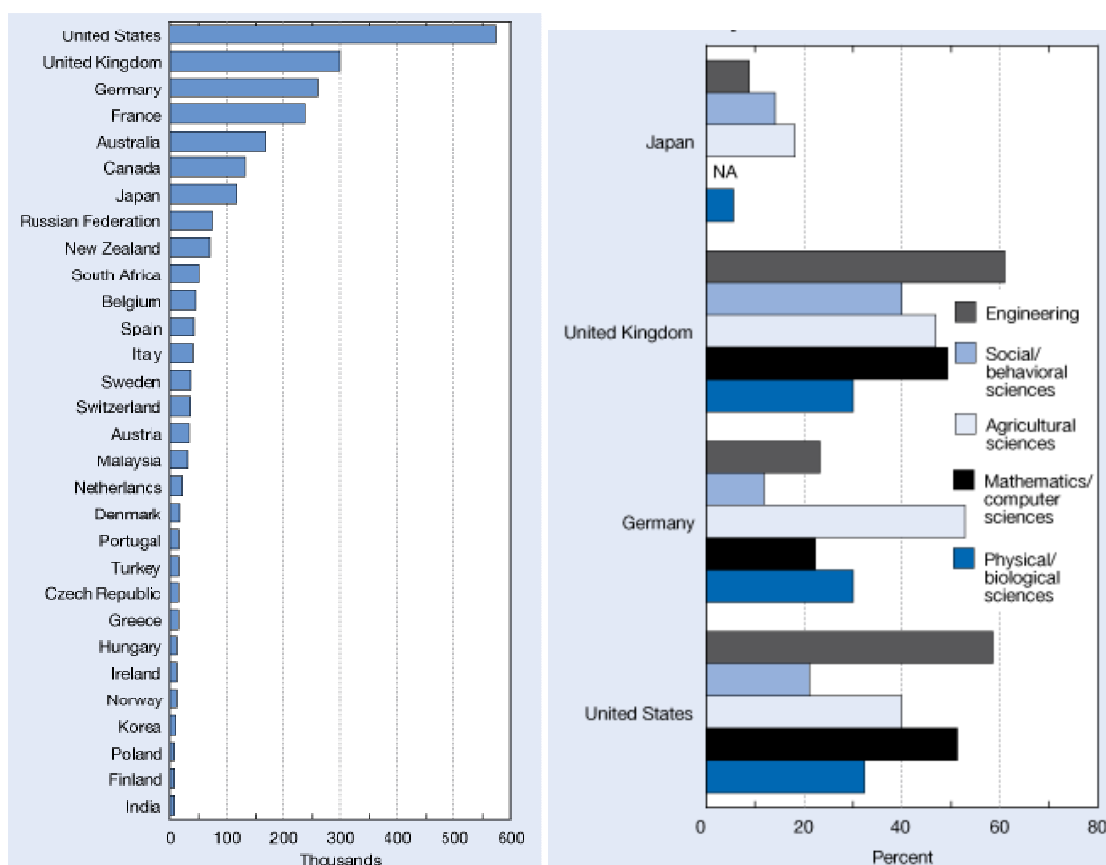


Figure 11.6. *Left:* Foreign students enrolled in tertiary education, by country: 2004. *Right:* S&E doctoral degrees earned by foreign students, by selected industrialized country and field: 2005 or most recent year (before 2005). Both figures are obtained from *Science and Engineering Indicators 2008* (NSB 2008).

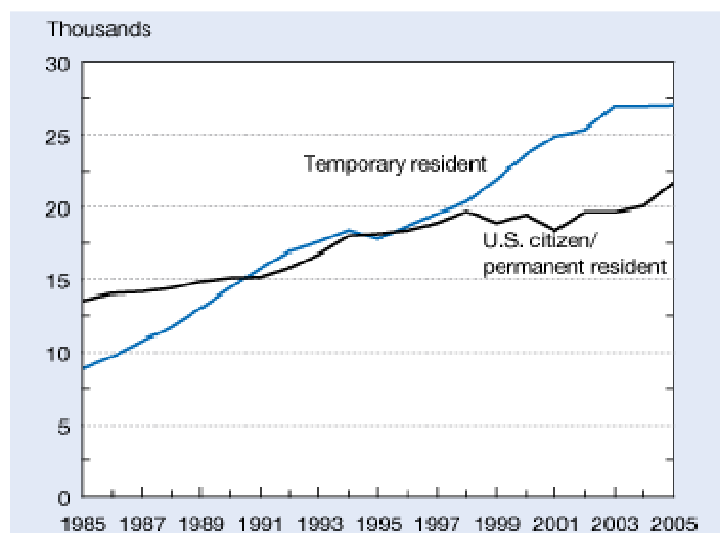


Figure 11.7. S&E postdoctoral students at U.S. universities, by citizenship status: 1985–2005. Figure is taken from *Science and Engineering Indicators 2008* (NSB 2008).

In the United States, the *total* number of S&E doctoral degrees in Engineering, Mathematics, Computer Sciences, and Physical Sciences started falling around 1995–1996 and only began to regain the 1995–1996 values a decade later (NSB 2008). When it comes to *relative* (percentage) performance, the United States performs worse than other countries in some areas of S&E. Figure 11.8 shows the academic R&D share of all R&D for selected countries and economies. The United States lags behind several emerging and developed economies. The right panel shows the percentage of 24-year-olds holding Natural Sciences or Engineering degree by country or economy.

Other alarming statistics obtained from *Rising above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (COSEPUP 2007 and references therein) are the following. The percentage of U.S. undergraduate students with an S&E degree is relatively low: South Korea, 38%; France, 47%; China, 50%; Singapore, 67%; United States, 15%. In the year 2000 an estimated 38% of the total U.S. science and technology workforce was foreign-born. About one-third of U.S. students intending to major in Engineering switch majors before graduation. More chief executive officers of Standard & Poor's 500 companies obtained their undergraduate degree in Engineering than in any other field. There were almost twice as many U.S. Physics BSc degrees in 1956 than in 2004. Federal funding of research in the Physical Sciences, as a percentage of GDP, was 45% less in the fiscal year (FY) 2004 than in FY1976.

HOW OTHER COUNTRIES COMPARE

Faculty and researchers everywhere are more interested in computing and open to the myriad of computing possibilities than ever before. This is due to the increased awareness of simulation successes, important changes in industrial attitude, and the demand for qualified computational scientists in all areas of R&D. Virtual experimentation, with its increased capabilities for explanation and prediction, is rapidly gaining ground across a wide spectrum of industrial and academic activities. Students are attracted to SBE&S because they can easily get good jobs afterwards. Here we report on some general trends that the WTEC panelists witnessed in our site visits. For further details, references, etc., the reader is referred to the site reports in Appendixes B and C.

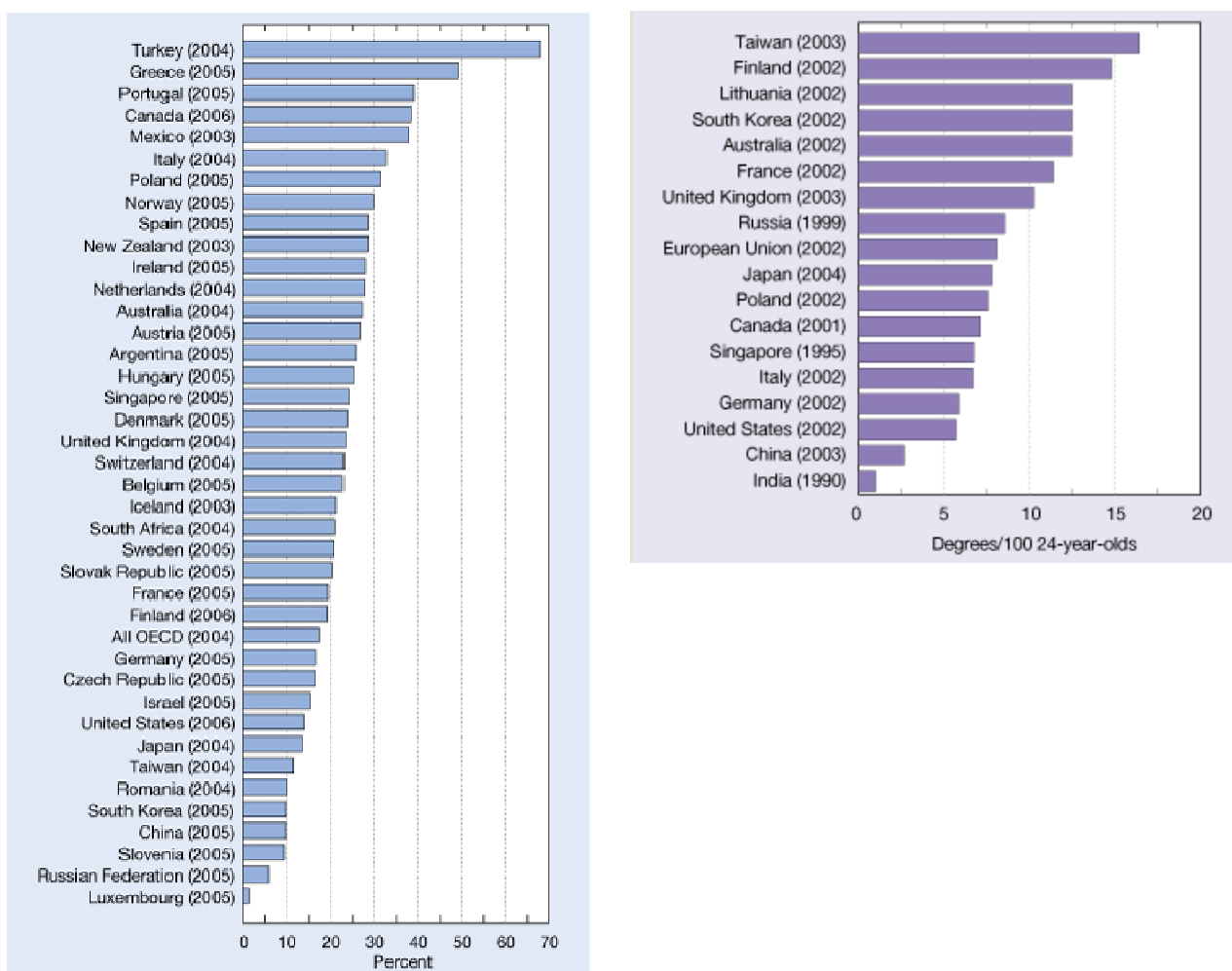


Figure 11.8. *Left:* Academic R&D share of all R&D for selected countries/economies and all OECD. Figure are taken from *Science and Engineering Indicators 2008* (NSB 2008). *Right:* Natural Sciences and Engineering degrees per 100 24-year-olds, by country/economy. Figures are taken from *Science and Engineering Indicators 2006* (NSB 2006).

Finding 1: There is increasing Asian and European leadership in SBE&S education due to dedicated funding allocation and industrial participation.

The United States still provides strong public investment in R&D, but other countries are also taking important initiatives to provide funding for education and to foster industrial collaboration in order to earn an edge in the global SBE&S innovation environment. This trend is clearly seen in Figure 11.2 and also in the somewhat alarming results presented in Figure 11.8 (left) that show the academic share of all R&D for the United States lagging behind that of several important competitors. Although specific statistics for SBE&S do not exist, it is the impression of this panel, as analyzed in various chapters, that the U.S. leadership in application software development in SBE&S is decreasing, due in part to relatively less funding.

Most software development these days takes place at universities and national laboratories. This is the case both in the United States and abroad. Worldwide, industry and private companies find it too onerous to invest in software development within their own working habitat and tend to base their simulation activities on software developed somewhere else. Of course, companies do invest in software when their own specific needs are not met by existing codes, and Chapter 10 provides several examples of this situation, along with a more detailed discussion. The WTEC panel found that the way companies are supporting SBE&S research is by funding intra-

and extramural research at university groups. Examples of this are found in the Institute of Process Engineering in the Chinese Academy of Science, the Department of Engineering Mechanics at Tsinghua University in China, the Center of Biological Sequence in Denmark, CIMNE in Barcelona, and many other institutions. Germany is a particularly striking example of corporate investment, where the share of academic R&D financed by industry considerably exceeds that of any other country (NSB 2008).

Interesting examples of Asian institutions that benefit from healthy government/industry funding in SBE&S are the following:

- *Japan Earth Simulation Center (ESC)*. The primary objective of the ESC is to develop new algorithms, with a focus on multiscale and multiphysics algorithms. It offers industrial participation with a fee structure controlled by the Japanese government. The ultimate goal of the ESC is to simulate physical systems as realistically as possible. One major ESC success is global climate simulation with resolutions of 10 km, and resulting improvement in weather prediction capabilities. The ESC is scheduled to be closed, and the Japanese government is investing heavily in software development to be run on the next supercomputer that will replace the ESC. This is being done in spite of the fact that it not known what form the new supercomputer will take. The goal is that innovation in algorithms will drive at least part of the hardware.
- *The Systems Biology Institute (SBI, Japan)*. SBI was founded by Dr. Hiroaki Kitano, one of the early pioneers of Systems Biology. The institute, which also has experimental labs in the Cancer Institute of the Japan Foundation for Cancer Research, has been funded by the Japanese government for 10 years, including through a grant for international software standards formation. The current research plan focuses on the development of experimental data and software infrastructure. The latter includes Systems Biology Markup Language (SBML), Systems Biology Graphical Notation (SBGN), CellDesigner, and Web 2.0 Biology, designed for the systematic accumulation of biological knowledge. Dr. Kitano believes that it is crucial to invest in software infrastructure because software is critical to the development of systems biology as a field. In recognition that it is difficult to publish software, the merit system in this lab values software contributions as well as publications.
- *University of Tokyo*. Ranked number one in Japan, the University of Tokyo is one of the recipients of the “21st Century Center of Excellence (COE) Program” awards. This program, created by the Ministry of Education, Culture, Sports, Science and Technology, aims to form world-class research and education centers at the universities in Japan; 28 of these centers were established at the University of Tokyo. Many of the original five-year COE projects have ended, but some of these have become part of the Global COE program, which aims to form international research and education excellence programs. The program in Mechanical Systems Innovation directed by Professor Kasagi is highly interdisciplinary, has three main focus areas—energy, biomedicine, and hypermodeling/ simulation—and includes a big education component.
- *Institute of Process Engineering (IPE), Chinese Academy of Sciences*. About 50% of the research funding for IPE comes from industrial sponsors, both domestic and international, with significant funding from the petrochemical industry. IPE also receives significant government funding through the National Natural Science Foundation of China and the Ministry of Science and Technology. Its main SBE&S focus is multiscale simulations for multiphase reactors. In addition to commercial packages, IPE researchers have developed their own codes for particle fluidization and reaction kinetics. Students are especially trained on the use of these codes and packages.
- *Tsinghua University Department of Engineering Mechanics (China)*. Tsinghua’s Engineering Mechanics department is believed to be the largest in the world, and it is ranked number one in China. The department has several high-profile SBE&S collaborations with multinational companies. Of the total annual research budget, 50% comes from the National Natural Science Foundation of China (NSFC), and 30% is from industry-sponsored projects. Even though training in SBE&S-related activities is highly valued by the students because of resulting career opportunities, the researchers regret that the NSFC does not fund code development. However industry often steps into this gap (for instance, simulations of cardiac muscles were sponsored by Medtronic, Inc.).
- *Fudan University, Shanghai (China)*. The work of the Department of Macromolecular Science and the Institute of Macromolecular Science at Fudan University on computational polymer physics is on par with

that of leading groups in the United States, UK, Germany, etc. There is strong emphasis on education. Training on analytical work is required prior to training in computational work, to avoid producing students that view simulation software packages merely as “black boxes.” Prof. Yang is both the director of the Computational Polymer Physics Group and the Vice Minister of Education. In this position, he oversees the “211” program in Beijing, whose aim is to solve problems important for China, to increase the number and impact factors of China’s publications, and to elevate the quality of PhD theses. Strong funding has been allocated for SBE&S and to support 2000 graduate students per year to study abroad.

- *King Abdullah University of Science and Technology (KAUST)*. This university is being built in Saudi Arabia as an international, graduate-level research university with the highest endowment-per-student worldwide. KAUST, scheduled to open in September 2009, is supported by a multibillion dollar endowment, and governed by an independent Board of Trustees. It intends to attract students from around the world, with main research thrusts in (i) Resources, Energy and Environment; (ii) Biosciences and Bioengineering; (iii) Materials Science and Engineering; and (iv) Applied Mathematics and Computational Sciences. These thrusts will be pursued in multidisciplinary Research Centers where SBE&S plays a major role. Recently, KAUST and IBM announced the creation of a research partnership to build and conduct research on the most complex, high-performance computing system in the region and among academic institutions in the world. In a joint statement, they said the new system, named Shaheen, will serve the university’s scientific researchers and put the university on a path to exascale computing in the near future (KAUST 2008). KAUST has also announced a partnership with UC San Diego to build the world’s most advanced visualization center, named Cornea. The Geometric Modeling and Scientific Visualization Research Center (GM&SVRC) will allow researchers to transform raw data into a fully three-dimensional visual experience and will support KAUST’s ambitious plan to deploy state-of-the-art technologies for elite scientific research (*Primeur Monthly* 2008).

Signs of increasing European leadership due to increased funding in SBE&S were observed across a large number of institutions. Some examples follow:

- *Center for Biological Sequence Analysis (CBS, Denmark)*. CBS, part of the Bio-Centrum at the Technical University of Denmark (DTU), is one of the largest bioinformatics centers in the European Union. Its research is focused on the functional aspects of complex biological mechanisms. CBS is funded—in addition to a contribution by DTU—by the Danish Research Foundation, the Danish Center for Scientific Computing, the Villum Kann Rasmussen Foundation and the Novo Nordisk Foundation (U.S.\$100 million), as well as from other institutions in the European Union, industry, and the U.S. National Institutes of Health (bioinformatics, systems biology). CBS success is measured by its large number of publications in high-profile journals (*Science*, *Nature*, *Molecular Cell*, etc.) and its very high citation records.
- *CIMNE, the International Center for Numerical Methods in Engineering (Barcelona, Spain)*. CIMNE is an autonomous research center created in 1987 by the Polytechnic University of Catalonia, the government of Catalonia, and the government of Spain, under the auspices of the United Nations Education Science and Cultural Organization (UNESCO). CIMNE’s mission is three-fold: research, training, and technology transfer, all in the SBE&S area. More than 90% of its funding comes from external sources, with an annual funding of €10 million. CIMNE has strong links to industry, with about 50% of its research funded by industry: aerospace, marine, civil, mechanical, biomedical, and metal companies. CIMNE strongly competes for European Commission projects, with a high success rate, and is a shareholder in three companies specializing in marine engineering, civil and structural engineering, and aeronautics and space applications.
- *Germany*. German SBE&S has undergone spectacular growth due to both government and industry funding. The German Research Foundation (DFG) has provided support for collaborative research centers (SFB), transregional collaborative research centers (TRR), transfer units (TBF), research units (FOR), priority programs (SPP), and “Excellence Initiatives.” Many of these are based on or have major components in SBE&S. The reader is referred to the site reports in Appendix C on the Technical University of Munich, the University of Karlsruhe, and the University of Stuttgart (which will be further discussed below), for a better picture of Germany’s scientific “renaissance.”
- *Fraunhofer Institute for the Mechanics of Materials (IWM) (Germany)*. The Fraunhofer Institutes comprise the largest applied research organization in Europe, with an annual research budget of €1.3 billion and

12,000 employees in 56 institutes. The IWM has a €15.5 million annual budget, with 44% of the budget coming from industry and another 25–30% from the government. It is crucial for the funding model of the IWM that its base funding is a fixed percentage of the industrial funding. The IWM has seen significant growth in recent years (10% per year). At the IWM, fully 50% of the funding supports modeling and simulation (a number that has grown from 30% five years ago). The IWM has a world-class effort in modeling of materials (micromechanical models for deformation and failure/fracture, damage analysis, etc.); simulation of manufacturing processes (pressing, sintering, forging, rolling, reshaping, welding, cutting, etc.); and simulation of components (behavior, upper limits, lifetime, visual testing, etc.). It has a novel method for developing students into potential hires based on awarding competitive €50,000 projects to fresh PhDs to work at IWM in the topics of their choice.

Finding 2: There are a number of new EU centers and programs for education and training in SBE&S—all of them of an interdisciplinary nature.

Here are just a few examples of the many new centers and programs that have been created in Europe:

- *CBS (Denmark)*. The CBS offers an MSc in Systems Biology and another in Bioinformatics that are loosely structured, not linked to any particular department or school. CBS offers realtime Internet training (all lectures, exercises, and exams in cyberspace). Typically, half the students participate onsite and half participate over the Internet. International exchange is highly encouraged; students can take their salary and move anywhere on the globe for half a year.
- *CIMNE (Spain)*. CIMNE organizes courses and seminars on the theory and application of numerical methods in engineering. The attendees are recent university graduates and also professionals. In the last 20 years, CIMNE has organized 100 courses, 300 seminars, and 80 national and international conferences, and has published 101 books, 15 educational software packages, and hundreds of research and technical reports and journal papers.
- *ETH Zürich (Switzerland)*. ETHZ offers campus-wide programs in Computational Science and Engineering, which include the Computational Collaborative, the Institute of Computational Science, and the CSE degree program. The Computational Collaborative provides an umbrella for research across different SBE&S areas and has been very popular with postdocs. The Institute of Computational Science entails interdisciplinary research: Visual Computing, Bioinformatics, Computational Biology, Marine Learning and Bio-inspired Computation, Parallel Computing, Large Scale Linear Algebra, Agent Based Modeling and Simulation, Advanced Symbolic Computation, Multiscale Modeling and Simulation, Systems Biology, Virtual and Augmented Reality, Human-Computer Interface, Pattern Recognition, Nanotechnology, Art, Entertainment, and Finance. The CSE degree program is a pioneering program in Europe offering both BSc and MSc degrees in interdisciplinary areas and combining several departments. The program is also popular with graduate and postdoctoral students who take the senior-level course.
- *Technische Universität München (TUM) and Leibniz Supercomputing Center (LRZ) (Germany)*. TUM offers a wealth of computational programs: (1) the Bavarian Graduate School of Computational Engineering (BGCE) offers a Bavaria-wide MSc honors program (an “elite” program); (2) the International Graduate School of Science and Engineering (IGSSE) offers a PhD-level program that is part of the German Excellence Initiative; (3) the Center for Simulation Technology in Engineering provides research training for PhD students; (4) the Centre for Computational and Visual Data Exploration provides research training for PhD students; (5) there are regular MSc programs in traditional areas of Informatics and also in Mathematics for Engineering, Computational Physics, Computational Mechanics, and CSE; (6) there are two international (in English) Master’s programs in Computational Mechanics and Computational Science and Engineering, with multidisciplinary cooperation involving 7 departments (they also allow for industrial internships); (7) a working group of CSE programs in Germany/Austria/Switzerland (München, Stuttgart, Darmstadt, Frankfurt, Aachen, Braunschweig, Rostock, Erlangen, Bochum, Bremen, Hannover, Dresden, Zürich, Basel, and Graz) that was established in 2005 obtained the legal status of an “eingetragener Verein” (registered voluntary association) in 2008; in addition, TUM has many other programs with other universities and industry. An innovative effort within the BGCE is the Software Project, which promotes the development of software for high-performance computing and computational science and engineering as an educational goal.

Finding 3: EU and Asian education/research centers are attracting an increasing number of international students from all over the world, including the United States.

Examples of this trend in Asia are as follows:

- *Japan.* The main target of the University of Tokyo's Global COE Program is support for international education and research. The National Institute for Materials Science (NIMS) recently created the International Center for Young Scientists, based on four premises: (1) completely international, with English as the language of instruction; (2) independent and autonomous research; and work that is (3) interdisciplinary and (4) innovative. Benefits to participants include a high salary, research grant support, and an ideal research environment. Japan is investing heavily in recruiting foreign students. Slightly below 120,000 overseas students were enrolled in Japanese universities in 2007 and the Prime Minister wants to raise that number to 300,000. The situation in Japan is further described in Finding 7 below.
- *China.* An important aim of China's "211" and "985" programs is to build world-class universities that will also attract foreign researchers. The number of foreign students has increased quite dramatically in China. In 1950, the People's Republic of China received its first group of 33 students from East European countries. By the end of 2000, the total number of international students (1950–2000) in China was 407,000 (Ministry of Education of China 2000). A total of 195,503 international students from 188 countries and regions came to study in China in 2007. They were distributed over China's 31 provinces, autonomous regions, and municipalities (excluding Taiwan Province, Hong Kong and Macao Special Administrative Region), and enrolled in 544 universities and colleges, scientific research institutes, and other teaching institutions. China's top five student source countries were the Republic of Korea (64,481 students), Japan (18,640), the United States (14,758, mainly Chinese-American), Vietnam (9,702), and Thailand (7,306). [*People's Daily Online* 2008]. According to the Observatory on Borderless Higher Education, Malaysia and Singapore are also seeking to attract more foreign students by building up their university systems and offering more programs in English (Jaschik 2007).
- *King Abdullah University of Science and Technology (KAUST).* Saudi Arabia is aflush with money and this is reflected in the creation of KAUST, as mentioned in Finding 1. KAUST, the KAUST-IBM Center for Deep Computing, and the KAUST-UCSD Center for Geometric Modeling and Scientific Visualization are recruiting computational scientists and engineers at all levels, including, of course full-fellowship Master's and Doctoral students. KAUST is perfectly positioned to attract the best and brightest students and researchers from the Middle East, China and India.
- *Australia.* Australia has heavily targeted students in Malaysia and Taiwan for recruitment to its universities, leading to increases in enrollments in recent years. At the same time, the United States and the UK have seen their number of Taiwanese (U.S.) and Malaysian (UK) enrollments decrease (Jaschik 2007).

Examples of this trend in Europe are as follows:

- *United Kingdom.* As shown in Figure 11.6 (left), the United Kingdom ranks second in the world in attracting international students. In all the different university departments that the WTEC panel visited in England (University College London, University of Cambridge, University of Oxford), hosts commented on the number and excellent quality of international students. The reputation of these universities gives them their pick of the best and brightest students and postdocs.
- *CBS (Denmark).* The CBS Internet courses are used to attract international students. They require 20% more effort than traditional courses but bring in lots of revenue and are always oversubscribed.
- *CIMNE (Spain).* (1) CIMNE introduced an international Master's course (Erasmus Mundus Master Course) in computational mechanics for non-European students; it started in 2008 with 30 students. There are four universities involved in this course (Barcelona, Stuttgart, Swansea, and Nantes). (2) CIMNE also has a Web environment for distance learning, where it hosts a Master's course in Numerical Methods in Engineering and other postgraduate courses. (3) It manages a collaborative network of "CIMNE Classrooms," physical spaces for cooperation in education, research, and technology located in Barcelona, Spain, Mexico, Argentina, Colombia, Cuba, Chile, Brazil, Venezuela, and Iran.

- *ETH Zürich (Switzerland)*. The number of international students at ETHZ has increased dramatically. Now there are many excellent Asian students and Russian students who are first-rate in mathematical skills.
- *Vrije University Amsterdam (The Netherlands)*. Half of the Vrije graduate students come from outside the Netherlands, mainly from Eastern Europe. In Theoretical Chemistry, the strongest and most disciplined students come from Germany.
- *TUM (Germany)*. In addition to TUM's international programs mentioned in Finding 2 (international MSc, IGSSE, regional collaborations), it has many other activities on an international scale. It established Computational Engineering in Belgrade, Computational Science in Tashkent/Uzbekistan, Applied and Computational Physics in St. Petersburg/Russia, and it has planned a joint CSE program with the National University of Singapore. Simulation is a very internationally oriented field, and many of TUM's partnerships have their origin in simulation collaborations. About 80% of the SBE&S students in MSc programs come from abroad: the Middle East, Asia, Eastern Europe, and Central and South America.
- *DTU Wind Engineering (Denmark)*. DTU is one of the very few places in the world to offer both MSc and PhD degrees in Wind Engineering. The DTU MSc international program attracts students from around the world.

Finding 4: There are pitfalls associated with interdisciplinary education: breadth versus depth.

An issue that educators and researchers seem to be grappling with is the inescapable fact that educational breadth comes at the expense of educational depth. There is a general feeling that Computer Science students can spend too much time on the “format” of a computer program without really understanding the underlying science. Thus, many faculty and researchers prefer to hire students with a relevant science background (generally Physics or Chemistry) when dealing with research issues that need to find expression in a software program. This view was expressed in many different places: CRIEPI in Japan, ETHZ in Switzerland, and others.

Some groups have come up with an “ideal background” for a student. For instance, Dr. Kitano in the Systems Biology Institute in Japan believes that the ideal educational background for students in Systems Biology may be an undergraduate degree in Physics, Master's degree in Computer Science, and PhD degree in Molecular Biology. ETH Zürich has had success with postdocs having backgrounds in both Science/Engineering and Computing Sciences. Many researchers have noted that in order to solve “grand challenge” problems in a given field, solid knowledge of a core discipline, in addition to computational skills, is absolutely crucial.

Of course, another ubiquitous problem of interdisciplinary research is appropriate evaluation of scientific performance. At present, this is generally carried out within the confines of disciplinary indicators, and it is challenging to come up with a system of credit attribution in interdisciplinary endeavors. Moreover, as pointed out by a participant in the panel of TUM/LRZ in Munich, there is the question of the “hidden innovation phenomenon:” *the prize/profit/recognition goes to the scientist or engineer who runs a given code and not to the person(s) who write(s) the code that enables the scientific breakthrough.*

Finding 5: Demand exceeds supply: academia versus industry.

There is huge demand for qualified SBE&S students, who get hired immediately after earning their MSc degrees and often do not even go into PhD programs. Students go to different industrial settings: to pharmaceutical, chemical, oil, (micro)electronics, IT, communications, and software companies; to automotive and aerospace engineering; or to finance, insurance, environmental institutions, and so forth. This is good insofar as it maintains a dynamic market workforce, but academia would like to see more students continue a tradition of unattached, basic research.

Finding 6: There is widespread difficulty finding students and postdocs qualified in algorithmic and software development.

An important distinction between software and middleware was emphasized in Chapter 10. Because of the U.S. strength in hardware development, the United States leads the world in system-level tools and middleware.

However, the development of open source systems and tools has made system-level software and middleware universally available. Therefore, most countries are not competing for leadership in these areas but for leadership in applications software, where the highest impact on science and engineering is expected to take place.

Unfortunately, postdocs and research staff with appropriate training in the development of algorithms and programs are hard to recruit. Most students in United States and abroad are trained primarily to run existing (generally commercial) codes to solve applied problems rather than learning the skills necessary to develop new codes. Many of the WTEC panel's hosts felt that a big hurdle in programming is insufficient math and physics preparation; another common complaint is that the students are not critical about the results they obtain from a simulation and often do not even know the physical approximations inherent to the codes or underlying algorithms, or the range of validity of these approximations. These issues, in some form or other, were almost ubiquitously expressed during the WTEC panel's site visits: at RICS in Japan, Vrije University in Amsterdam, the Theory of Condensed Matter Group at Cambridge University, IBM Zürich, ETH Zürich, Imperial College London, University College London, the NIMS Computational Material Science Center (CMSC) in Japan, the Institute of Chemistry of the Chinese Academy of Sciences, the Fraunhofer Institute for the Mechanics of Materials in Germany, and others.

Some institutions or groups (for instance, Dalian University of Technology and the Peking University Center for CSE in China) insist that their students develop their own software, for educational purposes, even if it means reinventing the wheel. German institutions seem to have taken a more aggressive approach. For instance, the Technical University of Munich, University of Karlsruhe, and University of Stuttgart all have emphasized software development through such means as their elite programs, degrees in SBE&S, and productive participation in the supercomputer centers.

Indeed, the different SBE&S curricula should always put emphasis on proper software engineering, and—as suggested by the panel's hosts at ETH Zürich and other institutions—development of open source code should be encouraged, as it generally leads to more carefully designed software to which other researchers can contribute. Unfortunately, maintenance of big codes is always a problem in academia, especially when it comes to obtaining the funding for such maintenance.

Although many of the problems described above affect almost every country, it is still eye-opening to the panel to see that U.S. leadership in applications software development for SBE&S has decreased considerably in the last decade, during which time the centers of algorithm development shifted to Europe. This situation, described more quantitatively in Chapters 9 and 10, is partly attributed to the expectation of U.S. funding agencies and university administrators for high-visibility science results that are published in high-profile journals. Software engineering is a long-term endeavor, and such expectations undermine support for long-term progress. Moreover, the pressure of tenure and promotion in the United States actively discourages research that takes a long time to see through to publication. When a faculty member finally reaches a senior level, and in principle is free from these pressures, he or she has to consider the future careers of students/postdocs, again undermining software development in favor of faster accomplishments.

Finding 7: Population matters.

One other factor that affects not only SBE&S but also the entire spectrum of Sciences and Engineering R&D (and indeed, the entire economy of a country) is population. Often disregarded in science (because many of the most populated countries tend to be the least developed scientifically), the reality brought about by dwindling populations in many countries is starting to hit hard.

In this context, there is a stark contrast between the People's Republic of China, currently the country with the largest population, and its Asiatic neighbor Japan. While China's universities are turning out about 400,000 engineers every year (and in addition are dramatically increasing their numbers of international students, with a total of nearly 200,000 in 2007, as noted in Finding 3), Japan is facing a dwindling number of young people entering engineering and technology-related fields (Fackler 2008). The decline is so drastic that industry is going to extremes in advertising and recruiting campaigns to capture engineers or is sending jobs abroad, mainly to

Vietnam and India. Part of the decline is due to a dramatically declining population, since Japan has one of the lowest fertility rates in the world, 1.22 in 2008 (CIA 2008). In addition, "...[A]ccording to educators, executives and young Japanese themselves, the young ... are behaving more like Americans: choosing better-paying fields like finance and medicine, or more purely creative careers, like the arts, rather than following their 'salaryman fathers' into the unglamorous world of manufacturing... The digital technology industry ... is already [estimated to be] short half a million engineers" (Fackler 2008).

Japan is trying to make up for this shortage by recruiting foreigners, but—unlike Americans—the country remains culturally closed, and many foreigners refuse to come. In 1983, Prime Minister Yasuhiro Nakasone presented a plan to raise the annual number of foreign students from 10,428 to 100,000, an accomplishment that took 20 years. For the past few years, however, the number has remained at about 120,000, with 118,498 in 2007 (Ishida 2008). The Japanese University Consortium for Transnational Education was established in March 2006 with participation by fifteen universities. The idea is to send Japanese professors abroad to train future foreign students (up to three years) before they move to Japanese universities as freshmen. This project has worked very well in Malaysia, under a low-interest loan from the Japanese government to finance the program. Now, Prime Minister Yasuo Fukuda wants to raise the number of foreign students to 300,000 (Ishida 2008). In order to achieve this, the government will designate 30 universities to support foreign students, and there will be more overseas offices to recruit and screen prospective students. Half of the foreign students should be able to find jobs in Japan after graduation.

Many countries in Europe are facing similar population crises. The 2008 collective fertility rate for the European Union is 1.5 (CIA 2008). Immigrant fertility rates are much higher, but it is still a matter of much debate whether these rates will stay the same or decrease and whether big sectors of the immigrant population can be culturally assimilated. Like Japan, many European universities and institutions have seen the writing on the wall and try to compensate for the diminishing number of national students by actively recruiting international students, creating international degrees, and/or by fostering excellence in their student body through diverse centers of excellence and elite programs, such as those in Japan and Germany.

The United States is sitting on the magical number of 2.1 for its fertility rate (CIA 2008). Considering that a much smaller percentage of its population joins the Science and Engineering fields than in other countries (see, for instance, Figure 11.8), keeping up the number of students should be a major concern, as has been pointed out in the previous findings. This is particularly true for SBE&S, since leadership in applications software is no longer determined by access to hardware or middleware, but by highly qualified, long-term supported labor. The efforts of the Technical University of Munich, Japan, and CIMNE in establishing offices and a workforce overseas to recruit exceptional students reflect a new trend and drive up competition. The United States is still the country providing most scholarships and fellowships for international students, and when things get tough, it generally rises to the occasion. Thus, prompted by their declining number of international students, dozens of U.S. educational institutions sent representatives to the recent International Education Exhibition in Shanghai. These efforts, plus easier access to visas for Chinese students combined with devaluation of the U.S. dollar have resulted in an increase in Chinese student enrollment in the United States (*World Journal* 2008).

The concept of setting up overseas educational recruiting offices can represent an intriguing new possibility for the United States to consider. As an example, consider the case for recruiting Hispanic students. As of mid-2007, the U.S. Hispanic population represented the largest U.S. minority group, accounting for 15.1% of the total population (U.S. Census Bureau 2007); it is this group that is mainly responsible for keeping the United States at the replacement fertility rate of 2.1. All in all, the United States has been quite positive in fostering the presence of Hispanics in higher education, where their representation is still well below their share of the total population. However, the attitude has been mainly passive: "Since the immigrants [overwhelmingly Mexican] are here, let's work with what we have." Setting up educational recruiting offices or agreements in Latin American countries, on the other hand, would allow the United States to screen for the very best students at a relatively low cost. The "very best" on the other hand, are increasingly being aggressively targeted by European countries. For instance, the "Tour Europosgrados 2008" took place recently in Universidad Nacional de Cuyo, Mendoza, Argentina. Thousands of scientists took part in this "tour," which showcased fellowships, postdoctoral positions, employment opportunities, cooperation programs, etc., sponsored by the European Commission, the German

Academic Exchange Service, the Spanish and French Embassies, the British Council, and the agency CampusFrance [Di Bari, 2008]. U.S.-Hispanic academic collaboration already exists in some obvious areas, like astronomy and paleontology- and archeology-related areas, but other areas (especially SBE&S!) could definitely benefit from enhanced collaboration. An interesting example in the area of experimental nanotechnology is the 2005 agreement signed by Lucent Technology and the Bariloche Atomic Center in Argentina, where Argentinean scientists are trained by Lucent and the company works on Argentinean projects (Sametband 2005).

Ultimately, population matters. In this sense both China, India and now Saudi Arabia, have big winning cards, and even if they are not top competitors right now, chances are that they will be in the not-so-distant future. They are major sources of international students and now are also becoming major emerging destinations for international students in S&E. Countries in Europe and Japan are in a race to get a share of the human capital in the developing world.

CASE STUDY: THE UNIVERSITY OF STUTTGART—A SUCCESS STORY

The University of Stuttgart is a research university with a focus on Engineering and the Natural Sciences. Relevant departments in these areas are Civil and Environmental Engineering; Chemistry; Energy Technology, Process Engineering, and Biological Engineering; Computer Science, Electrical Engineering, and Information Technology; Aerospace Engineering and Geodesy; Engineering Design, Production Engineering and Automotive Engineering; and Mathematics and Physics. Key research areas are modeling and simulation, complex systems, communications, materials, technology concepts and assessment, energy and the environment, mobility construction and living, and integrated products and product design.

The university has a number of collaborative programs and an interdepartmental research structure, with the different departments linked through transfer and research centers interacting with national institutions (e.g., Max Planck Institutes, Fraunhofer Institutes, and the German Aerospace Center), international institutions, and industry. The university sits in one of Europe's strongest economic and industrial regions; Bosch, Fischer, Daimler, HP, IBM, Festo, Pilz Deutschland, Porsche, Trumpf, Stihl, Züblin, and BASF all have facilities in the area.

Stuttgart has a number of superlative technical and computing facilities and projects. These include the facilities of the High Performance Computing Center Stuttgart (HLRS) with a computing power of about 220 Tflops; the Gauss Center for Supercomputing (Europe's most powerful high-performance computing alliance among the universities of Jülich, Munich, and Stuttgart); the Visualization Research Center (VISUS), one of the leading facilities in Europe; the Automotive Simulation Centre Stuttgart (ASCS), in close cooperation with industry since 2007; Archi-Neering (Bangkok's new airport is the result of close cooperation between Civil Engineering and Architecture); Baden-Württemberg Astronautics Center (opening in 2010); the International Center for Cultural and Technological Studies (IZKT); SOFIA, a joint U.S.-German project to build a Boeing 747SP equipped with a high-performance mirror telescope; a wind tunnel that performs tests of aerodynamic and aeroacoustic properties of vehicles up to 265 km/hour; and VEGAS, the research facility for subsurface remediation and for the simulation of contamination processes.

The University of Stuttgart also excels in rankings and funding. It ranks number three in Germany in total amount of external funding (Aachen is first), and number one in Germany in external funding per professor (an average of €400,000 per professor). In 2006 it was one of the top grant university recipients in grant ranking by DFG (Deutsche Forschungs-gemeinschaft, the German Research Foundation; see the sidebar on the next page on DFG-Provided Support for Collaborative Research Centers and Programs in Germany); in the top three for a range of Engineering study programs as ranked by the Centre for Higher Education Development (CHE) (*Der Spiegel*, Focus 2007); number five in the CHE 2006 research ranking; most successful German University in the 6th EU Framework Program, particularly in the fields of simulation technology, energy, and e-health; and number two in Germany for research visits by international scholarship holders and award-winners of the Humboldt Foundation, 2005.

In order to support its research and educational efforts, the University of Stuttgart is creating new professorships and research positions. These include three new professorial positions on Mathematical Systems Theory, Modeling of Uncertain Systems, and Human-System Interaction and Cognitive Systems. In addition, there are 13 new junior professorships and 7 new post-doctoral positions with up to 2 research associates each, tenure-track options for 4 of the junior professors, and a total of 72 scientific projects.

The SimTech Excellence Cluster

The vision of the SimTech Excellence Cluster at the University of Stuttgart, coordinated by Dr.-Ing. Wolfgang Ehlers, is to progress from isolated numerical approaches to integrative systems science. Interestingly, in order to obtain funding, the group gave arguments partly based on the findings of the U.S. NSF Blue Ribbon Panel Report of February 2006: "... [C]hallenges in SBES ... involve ... multiscale and multiphysics modeling, real-time integration of simulation methods with measurement systems, model validation and verification, handling large data, and visualization. ... [O]ne of those challenges is education of the next generation of engineers and scientists in the theory and practices of SBES" (Oden et al. 2006). DFG leadership has agreed wholeheartedly with this report and provided funding accordingly. The SimTech Excellence Cluster brings €7 million/year for 5 years. This plus other sources of funding allow for a long-term sustained agenda.

DFG-Provided Support for Collaborative Research Centers and Programs in Germany

- DFG Collaborative Research Centers (SFB) and Transregional Collaborative Research Centers (TRR): The budget of these centers is about €2–3 million per year for twelve years, in three 4-year terms. Projects include Spatial World Models for Mobile Context-aware Applications, Selective Catalytic Oxidation of C-H Bonds with Molecular Oxygen, Dynamic Simulation of Systems with Large Particle Numbers; Incremental Specification in Context, and Control of Quantum Correlations in Tailored Matter (a transregional project involving the universities of Stuttgart, Tübingen, and Ulm).
- DFG Transfer Units (TFB): These units execute transfer from university to industry. Programs include Simulation and Active Control of Hydroacoustics in Flexible Piping Systems, Development of a Regenerative Reactor System for Autothermal Operation of Endothermic High-Temperature Syntheses, Transformability in Multivariant Serial Production, Rapid Prototyping, Computer Aided Modeling and Simulation for Analysis, and Synthesis and Operation in Process Engineering.
- DFG Research Units (FOR): These units are smaller than the centers, generally consisting of 4–5 researchers. They include Nondestructive Evaluation of Concrete Structures Using Acoustic and Electromagnetic Echo Methods, Development of Concepts and Methods for the Determination of Reliability of Mechatronics Systems in Early Stages of Development, Noise Generation in Turbulent Flow, Multiscale Methods in Computational Mechanics, Specific Predictive Maintenance of Machine Tools by Automated Condition Monitoring, and Positioning of Single Nanostructures—Single Quantum Devices.
- DFG Priority Programs (SPP): These programs bring €1.2–2 million/year, spread over 10–20 projects. Ten new priority programs are opened per year. They include Molecular Modeling and Simulation in Process Engineering, and Nanowires and Nanotubes: From Controlled Synthesis to Function.
- Excellence Initiative: The goal of the Excellence Initiative is to foster excellence in science and research and to raise the profile of top performers in the academic and research community by means of three lines of funding: strategies for the future; excellence clusters, and graduate schools. The University of Stuttgart has been successful in two of the Excellence Initiative's three lines of funding, one of which is the Simulation Technology Excellence Cluster (SimTech Excellence Cluster).

Three new structural research elements of types described in the DFG sidebar above have been founded that provide long-term sustainability for the cluster. They are described below. Compared to traditional university departments with their teaching-oriented "vertical" structure, these research centers are "horizontally" oriented,

thus comprising researchers and their institutions from various departments under the common roof of a research goal.

1. Stuttgart Research Centre of Simulation Technology (SRC SimTech): Opened on April 1, 2007, this research center is the first one at the university and represents both a scientific research unit and a new structural element acting as a research department with its own organizational and administrative structure, including financial resources (€240,000/year) and personnel.
2. SimTech Transfer Unit: This unit bundles all activities of the cluster that require uni- or bidirectional communication with external institutions and industry. It will be embedded in the Stuttgart Transfer Centre, whose role is to transfer research results into application, bundle exchange activities with industrial partners, and provide a basis for all future fundraising activities of individual research centers.
3. Graduate School of Simulation Technology: Part of the Stuttgart School of Science and Technology, the Graduate School integrates the activities of the doctoral students supervised by the members of SimTech.

The SimTech Excellence Cluster has created new lines of elite education, which include both BSc and MSc theses in different SimTech research areas. For the MSc, there is flexible study regulation, with more than one supervisor and at least one from abroad. The MSc student is also required to spend one term abroad. The Graduate School of Simulation Technology also offers many other novelties: (1) there is no more separation in departments, it is completely interdisciplinary; (2) it fosters software skills; (3) it has an international exchange program; and (4) there is joint internal/external and international supervision.

In conclusion, the German Government, the University of Stuttgart (and certainly other universities in Germany) are highly aware of the computational challenges at the root of all meaningful innovation in the S&E fields. Accordingly, they have taken decisive action to implement meaningful changes. Thus, as indicated in the sidebar above, new academic structures have been created, such as the DFG-supported Collaborative Research Centers, Transfer Units, Priority Programs (e.g., Molecular Modeling and Simulation in Process Engineering), Excellence Initiatives (e.g., Simulation Technology Excellence Cluster), and new graduate programs (a new educational model from BSc to PhD at the Graduate School of Simulation Technology). All these programs are strongly supported economically, with several million Euros per year for several years (5–12 years). Both the academics and the government policymakers have understood that a long-term sustained research agenda is the only way this integrative, simulation-based vision can be realized. These initiatives have clearly strengthened the position of the University of Stuttgart as a major global player in simulation technology.

CONCLUSIONS

The conclusions of this chapter are summarized by the findings:

- Increased Asian and especially EU leadership in SBE&S education is due to increased funding and industrial participation. The United States continues to be a leader in most scientific and engineering enterprises, but its lead is decreasing across all S&E indicators: research and development investment, number of scientific publications, number of scientific researchers, etc. For the first time, the number of U.S. doctorates both in the Natural Sciences and in Engineering has been surpassed by those in the European Union and by those in Asia.
- The European Union is investing in new centers and programs for education and training in SBE&S—all of an interdisciplinary nature. New BSc and MSc degrees are being offered in SBE&S through programs that comprise a large number of departments. In many cases, a complete restructuring of the university has taken place in order to create, for instance, MSc degrees, graduate schools (e.g., Stuttgart), or international degrees in simulation technology.
- European and Asian educational/research centers are attracting more international students from all over the world (even from the United States). Special SBE&S programs (in English) are being created for international students. The United States is seeing smaller increases in international enrollment compared to other countries. This is due not only to post-9/11 security measures but also—and increasingly—to active competition from other countries.

- There are pitfalls in interdisciplinary education, including a tradeoff between breadth and depth. In order to solve “grand challenge” problems in a given field, solid knowledge of a core discipline, in addition to computational skills, are absolutely crucial. Another important problem of interdisciplinary research is the evaluation of scientific performance, which currently is generally carried out within the confines of disciplinary indicators.
- Demand exceeds supply. There is a huge demand in the European Union and Asia for qualified SBE&S students who get hired immediately after their MSc degrees by industry or finance: there is both collaboration and competition between industry and academia.
- There is widespread difficulty in finding students and postdocs qualified in algorithmic and program development. Most countries are competing for leadership in applications software development, where the highest impacts on Science and Engineering are expected to take place. Unfortunately, students and postdocs with the appropriate training in the development of algorithms and programs are hard to recruit, since they are trained primarily to run existing codes rather than to develop the skills necessary for programming.
- Still, the center of applications software development has shifted to the European Union. This is due to well-funded, long-term collaboration efforts; students thrive in these environments. The U.S. system for tenure and promotion, which emphasizes high number and visibility of publications, discourages software development. The merit system should value software contributions as well as publications.
- Population matters. Countries in the European Union and Japan are scampering to recruit foreign students to make up for the dwindling numbers of their youths. This is driving fierce competition for international recruiting. The United States, sitting at exactly the “fertility replacement rate” should pay close attention, since only a small percentage of its population joins the Sciences and Engineering professions. China and India, on the other hand, are “sitting pretty” and are likely to become increasingly strong SBE&S competitors.

REFERENCES

- Central Intelligence Agency (CIA). 2008. Rank order: Total fertility rate. In *The World Factbook*. Available online: <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2127rank.html>.
- Committee on Science, Engineering, and Public Policy (COSEPUP) [a joint unit of the National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine]. 2007. *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology. Washington, DC: National Academies Press. Available online: http://www.nap.edu/catalog.php?record_id=11463#orgs.
- Council on Competitiveness. 2007. *Competitiveness index: Where America stands*. Washington, DC: The Council on Competitiveness. Available online: <http://www.compete.org/publications/>.
- Di Bari, V. 2008. Oferta de becas y posgrados europeos en la UNCuyo. *Los Andes Online* August 15. Available online: <http://www.losandes.com.ar/notas/2008/8/15/sociedad-375172.asp>.
- Fackler, M. 2008. High-tech Japan running out of engineers. *The New York Times* May 17. Available online: <http://www.nytimes.com/2008/05/17/business/worldbusiness/17engineers.html>.
- Ishida, I. 2008. Educational renaissance / preparation key for foreign students. *Daily Yomiuri Online* July 31. Available at <http://www.yomiuri.co.jp/dy/features/language/20080731TDY14001.htm>.
- Jaschik, S. 2007. The mobile international student. *Inside Higher Ed* Oct. 10. Available online: <http://www.insidehighered.com/news/2007/10/10/mobile>.
- King Abdullah University of Science and Technology (KAUST). 2008. King Abdullah University of Science and Technology and IBM to build one of the world’s fastest and most powerful supercomputers. Available online: <http://www.kaust.edu.sa/news-releases/king-abdullah-university-and-ibm-build-supercomputer.aspx>.
- Lexington. 2008. Help not wanted. *The Economist* April 10. Available online: http://www.economist.com/world/unitedstates/displaystory.cfm?story_id=11016270.
- Ministry of Education of the People’s Republic of China. 2000. *International students in China*. Available online: http://www.moe.edu.cn/english/international_3.htm.

- National Science Board (NSB). 2006. *Science and engineering indicators 2006*. Two volumes. Arlington: National Science Foundation (volume 1, NSB 06-01; volume 2, NSB 06-01A).
- . 2008. *Science and engineering indicators 2008*. Two volumes. Arlington: National Science Foundation (volume 1, NSB 08-01; volume 2, NSB 08-01A).
- NSF. 2007. First-time, full-time graduate student enrollment in science and engineering increases in 2006, especially among foreign students. Washington, DC: National Science Foundation, Division of Science Resources Statistics, NSF 08-302. Available online: <http://www.nsf.gov/statistics/infbrief/nsf08302/>.
- Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science Foundation blue ribbon panel on simulation-based engineering science*. Washington, DC: National Science Foundation. Available online: http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.
- Organisation for Economic Co-operation and Development (OECD). 2008. *Education Online Database*. Available online at SourceOECD: http://www.sourceoecd.org/vl=1605451/cl=31/nw=1/rpsv/statistic/s4_about.htm?jnlissn=16081250; see also *Education at a glance 2007: OECD indicators*. Paris: OECD Publishing. Available online at <http://www.oecdbookshop.org/oecd/index.asp>.
- . 2006. *Main science and technology indicators 2006*. Paris: OECD (paper and electronic formats). Available online: http://www.oecd.org/document/26/0,3343,en_2649_34451_1901082_1_1_1_1,00.html.
- People's Daily Online*. 2008. Number of international students in China exceeds 190,000 in 2007. March 14. Available online: <http://english.peopledaily.com.cn/90001/6373700.html>.
- Primeur Monthly*. 2008. KAUST announces partnership with UC San Diego to build world's most advanced visualization centre. Available online: <http://enterthegrid.com/primeur/08/articles/monthly/AE-PR-11-08-102.html>.
- Sametband, R. 2005. Argentina invests US\$10 million in nanotechnology. *Science and Development Network* May 12. Available online: <http://www.scidev.net/en/news/argentina-invests-us10-million-in-nanotechnology.html>.
- World Journal*. 2008. Devaluation of U.S. dollar draws international students from China. March 11.
- UNESCO. 2007. United Nations Educational, Scientific, and Cultural Organization.
- U.S. Census Bureau. 2007. News release, U.S. Hispanic population surpasses 45 million; Now 15 percent of total. Available online: <http://www.census.gov/Press-Release/www/releases/archives/population/011910.html>.

APPENDIX A. BIOGRAPHIES OF PANELISTS AND ADVISORS

PANELISTS



Sharon C. Glotzer (Chair)

Sharon C. Glotzer is Professor of Chemical Engineering and Professor of Materials Science and Engineering at the University of Michigan, Ann Arbor. She also holds appointments in Physics, Applied Physics, and Macromolecular Science and Engineering. She received a BS in physics from UCLA in 1987 and a PhD in physics from Boston University in 1993. Prior to moving to Michigan in 2001, she worked in the Polymers Division, Materials Science and Engineering Laboratory at the National Institute of Standards and Technology from 1995-2000 as co-founder and Director of the Center for Theoretical and Computational Materials Science, and from 1993-1995 as an NRC Postdoctoral Fellow.

Dr. Glotzer oversees an active research group working in the areas of computational nanoscience and soft matter simulation, self-assembly and computational materials design. She has published 130 archival publications and presented nearly 200 invited talks and keynote lectures. Dr. Glotzer is a Fellow of the American Physical Society, and the recipient of many awards, including the American Physical Society (APS) Maria Goeppert-Mayer Award, a Presidential Early Career Award for Scientists and Engineers, a Department of Commerce Bronze Medal, the Charles M.A. Stine Award from the Materials Engineering and Science Division of the American Institute of Chemical Engineers (AIChE); and a National Security Science and Engineering Faculty Fellowship. She was a Sigma Xi Lecturer, Allan P. Colburn Memorial Lecturer, Bernard T. Bertman Memorial Lecturer, and Jerome B. Cohen Memorial Lecturer.

Dr. Glotzer has an active record of service on science and technology boards, panels and committees, as well as in professional societies, and she serves on the editorial boards of several journals. Recent service activities includes the National Academies' Solid State Sciences Committee, Technology Warning and Surprise study committee, Biomolecular Materials and Processes study committee, and Modeling, Simulation, and Games study committee; she currently serves on the TIGER Standing Committee on Defense Intelligence.



Sangtae Kim (Vice Chair)

Sangtae Kim is the inaugural Executive Director of the Morgridge Institute for Research, a new nonprofit medical research institute based in Madison, Wisconsin. Previously, he was the Donald W. Feddersen Distinguished Professor of Mechanical Engineering and Distinguished Professor of Chemical Engineering at Purdue University. From February 2004 to August 2005, he was at the National Science Foundation (on loan from Purdue) to serve as the inaugural director of NSF's Division of Shared Cyberinfrastructure and to help guide NSF's formational investments in cyberinfrastructure.

From 1997 to 2003, Dr. Kim served as vice president overseeing the R&D IT departments at two pharmaceutical companies, Eli Lilly and Warner Lambert, during the transition to the data-intensive, post-genomic IT environment in the research-based pharmaceutical industry.

From 1983 to 1997, Dr. Kim was a faculty member in the Department of Chemical Engineering at the University of Wisconsin-Madison, progressing from assistant professor to distinguished chair professor for his work in mathematical and computational methods for microhydrodynamics (now more commonly known as microfluidics). His computational insights into “hydrodynamic steering” played an influential role in 1994-95 in the development of fluidic self assembly (FSA), the dominant process used today for manufacture of low-cost RFID (radio frequency identification) tags. In recognition of his teaching and research accomplishments in high performance computing, he was extended a courtesy faculty appointment in the UW-Madison Computer Sciences Department.

Dr. Kim is a member of the National Academy of Engineering. His research citations include the 1993 Allan P. Colburn Award of the American Institute of Chemical Engineers, the 1992 Award for Initiatives in Research from the National Academy of Sciences and a Presidential Young Investigator award from NSF in 1985. He has an active record of service on science and technology advisory boards of government agencies, the U.S. National Research Council and companies in IT-intensive industries.

A native of Seoul and a product of the “K-11” public schools of Montreal, Dr. Kim received concurrent BSc and MSc degrees (1979) from Caltech and his PhD (1983) from Princeton.



Peter T. Cummings

Peter T. Cummings is the John R. Hall professor of chemical engineering at Vanderbilt University. He is also Principal Scientist in the Center for Nanophase Materials Sciences (CNMS) at Oak Ridge National Laboratory, as well as director of the Nanomaterials Theory Institute within the CNMS. The CNMS is the first of five Department of Energy nanoscience centers, entering full-scale operations on October 1, 2005. Currently, the CNMS has over 200 user projects, with almost 50 of those user projects being in the Nanomaterials Theory Institute.

Dr. Cummings’ research interests include statistical mechanics, molecular simulation, computational materials science, computational and theoretical nanoscience, and computational biology. He is the author of over 300 refereed journal publications and the recipient of many awards, including the 1998 Alpha Chi Sigma award (given annually to the member of the American Institute of Chemical Engineers (AIChE) with the most outstanding research contributions over the previous decade) and the 2007 AIChE Nanoscale Science and Engineering Forum award (given to recognize outstanding contributions to the advancement of nanoscale science and engineering in the field of chemical engineering through scholarship, education or service). He is a fellow of the American Physical Society and the American Association for the Advancement of Science (AAAS).

Dr. Cummings received a BMath degree (First Class Honors) at the University of Newcastle (Australia) in 1976 and a PhD in Mathematics at the University of Melbourne (Australia) in 1980.

**Abhijit Deshmukh**

Abhijit Deshmukh is Professor of Industrial and Systems Engineering, and Director of the Institute for Manufacturing Systems at Texas A&M University. Previously, he was Professor of Mechanical and Industrial Engineering, and Director of the Consortium for Distributed Decision-Making (CDDM) and the Security, Emergency Preparedness and Response Institute (SEPRI) at the University of Massachusetts Amherst. From September 2004 – August 2007, he was a Program Director at the National Science Foundation in the Engineering Directorate and the Office of Cyberinfrastructure. Dr. Deshmukh received his PhD in Industrial Engineering from Purdue University in 1993, and BE degree in Production Engineering from the University of Bombay in 1987.

Dr. Deshmukh was awarded the National Science Foundation Director's Award for Collaborative Integration in 2005, Ralph R. Teetor Educational Award from the Society of Automotive Engineers in 2003, and Milton C. Shaw Outstanding Young Manufacturing Engineer Award from the Society of Manufacturing Engineers in 1999. He was a Lilly Teaching Fellow at the University of Massachusetts from 1999-2000.

Dr. Deshmukh's research group is focused on distributed decision-making, with specific interests in complex systems and complexity in decision-making; coordination and inferencing in distributed sensor networks; multi-scale decision models; negotiation protocols, contract portfolio selection and dynamic pricing in supply chains; multi-agent models and Cyberinfrastructure for extended enterprises; and life-cycle cost estimation and uncertainty propagation in distributed design.

**Martin Head-Gordon**

Martin Head-Gordon is a professor of chemistry at the University of California, Berkeley and is a Faculty Chemist in the Chemical Sciences Division of Lawrence Berkeley National Laboratory working in the area of computational quantum chemistry. He is a member of The International Academy of Quantum Molecular Science.

A native of Australia, Head-Gordon obtained his PhD from Carnegie Mellon under the supervision of John Pople developing a number of useful techniques including the Head-Gordon-Pople scheme for the evaluation of integrals and the orbital rotation picture of orbital optimization. He received the BSc and MSc degrees at Monash University (Australia).

His awards include a 1993 National Science Foundation Young Investigator Award, an Alfred P. Sloan Foundation Research Fellowship (1995-7), a David and Lucile Packard Fellowship (1995-2000), and the 1998 Medal of the International Academy of Quantum Molecular Sciences. Prior to joining the Berkeley faculty, Dr Head-Gordon was a postdoctoral researcher at AT&T Bell Laboratories, where he developed unique electronic structure methods for calculating nonadiabatic energy exchange between molecules and surfaces.

At Berkeley, Martin Head-Gordon supervises a group interested in pairing methods, local correlation methods, dual-basis methods, scaled MP2 methods, new efficient algorithms, and very recently corrections to the Kohn-

Sham density functional framework. Broadly speaking, wavefunction-based methods are the focus of the research. Dr. Head-Gordon has authored over 200 publications and presented over 150 invited lectures.

Martin is also one of the founders of Q-Chem, Inc. which brings commercial, academic and government scientists worldwide in pharmaceuticals, materials science, biochemistry and other fields a comprehensive *ab initio* quantum chemistry program.



George Em Karniadakis

George Em Karniadakis received his SM and PhD degrees from Massachusetts Institute of Technology. He was a Lecturer in the Department of Mechanical Engineering at MIT in 1987, and postdoctoral fellow at the Center for Turbulence Research at Stanford/NASA Ames. He joined Princeton University as Assistant Professor in the Department of Mechanical and Aerospace Engineering with a joint appointment in the Program of Applied and Computational Mathematics. He was a Visiting Professor at Caltech (1993) in the Aeronautics Department. He joined Brown University as Associate Professor of Applied Mathematics and became a full professor in 1996. He has been a Visiting Professor and Senior Lecturer of Ocean/Mechanical Engineering at MIT since September 1, 2000. He was Visiting Professor at Peking University (Fall 2007).

Dr. Karniadakis is a Fellow of the American Physical Society (APS, 2004-), Fellow of the American Society of Mechanical Engineers (ASME, 2003-) and Associate Fellow of the American Institute of Aeronautics and Astronautics (AIAA, 2006-). He is the recipient of the CFD award (2007) by the U.S. Association in Computational Mechanics. His research interests are focused on computational fluid dynamics, stochastic modeling, biophysics, uncertainty quantification and parallel computing. He is the author of three textbooks and more than 250 research papers.



Linda Petzold

Linda Petzold is currently Professor in the Department of Computer Science, Professor in the Department of Mechanical Engineering, and Director of the Computational Science and Engineering Program at the University of California Santa Barbara. She received her PhD in Computer Science from the University of Illinois. She was a member of the Applied Mathematics Group at Sandia National Laboratories in Livermore, California, Group Leader of the Numerical Mathematics Group at Lawrence Livermore National Laboratory, and Professor in the Department of Computer Science at the University of Minnesota.

Dr. Petzold is a member of the U.S. National Academy of Engineering, and a Fellow of the ASME and of the AAAS. She was awarded the Wilkinson Prize for Numerical Software in 1991, the Dahlquist Prize in 1999, and the AWM/SIAM Sonia Kovalevski Prize in 2003. She served as SIAM (Society for Industrial and Applied Mathematics) Vice President at Large from 2000-2001, as SIAM Vice President for Publications from 1993-1998, and as Editor in Chief of the SIAM Journal on Scientific Computing from 1989-1993.

Professor Petzold's research group concentrates on computational science and engineering and systems biology. The computational issues addressed range from development and analysis of numerical methods for problems at

scales from stochastic to deterministic, to sensitivity analysis and model reduction, to the design of software environments to make scientific computation more easily accessible.



Celeste Sagui

Celeste Sagui, associate professor of physics at North Carolina State University, received her doctorate in physics from the University of Toronto and performed postdoctoral work at McGill University and at the National Institutes of Environmental and Health Sciences (NIEHS). She continues to hold an Intergovernmental Personnel Act (IPA) position in the Laboratory of Structural Biology at NIEHS.

Dr. Sagui is exploring properties of condensed matter systems and biomolecules, focusing on methodological developments for the accurate and efficient treatment of electrostatics in large-scale biomolecular simulations, studies of DNA structure and other selected biomolecules, molecular and ion solvation, ordering in modulated condensed matter systems for nanotechnological applications, and organic molecules on surfaces. To explore the interesting properties of these systems, she uses a range of computational methods such as quantum chemistry, density functional theory, classical molecular dynamics and phase field models and hydrodynamical equations.

Dr. Sagui has received many awards, including the NSF CAREER Development Award; NSF POWRE Award; SLOAN Postdoctoral Fellowship in Computational Biology; Ontario Graduate Scholarship (OGS), Toronto; Provost Seeley Fellowship, Trinity College, University of Toronto; International Student Differential Fee Waiver Scholarship, Toronto; Connaught Scholarship, Toronto; and the “Diploma de Honor” awarded by the University of San Luis, Argentina for the highest marks in the graduating class for Physics and Mathematics.



Masanobu Shinozuka

Masanobu Shinozuka, a Member of the National Academy of Engineering since 1978, is currently Distinguished Professor and Chair of the Department of Civil and Environmental Engineering at the University of California Irvine. He is also Norman Sollenberger Professor of Civil Engineering Emeritus at Princeton University. He received his BS and MS degrees from Kyoto University (Japan) and his PhD degree in 1960 from Columbia University.

Professor Shinozuka's research interests include continuum mechanics, stochastic processes, structural dynamics and control, earthquake engineering, reliability of infrastructure and lifeline systems.

Prior to joining the University of California Irvine, he taught at the University of Southern California, Princeton University, the State University of New York at Buffalo where he also served as Director of the NSF National Center for Earthquake Engineering Research (on leave from Princeton University), and Columbia University.

Professor Shinozuka is an Honorary member of ASCE, Fellow of ASME, Senior Member of AIAA, Editor-Emeritus of *The Journal of Probabilistic Engineering Mechanics*, and present and past member of the editorial boards of many reputable technical journals. He is the recipient of various national and international awards and recognition such as ASCE's major research awards, including the Freudenthal, Newmark, von Karman and Scanlan medals, and has been elected to the Russian Academy of Architecture and Construction Science.

Professor Shinozuka continues to be active, currently serving as PI for a \$5.5M joint venture under NIST's Technology Innovation Program entitled, "Next Generation SCADA for Protection and Mitigation of Water System Infrastructure Disaster."

ADVISORS



Tomás Díaz de la Rubia

Tomás Díaz de la Rubia joined LLNL as a postdoc in 1989 after completing his PhD in physics at the State University of New York at Albany. He carried out his thesis research in the Materials Science Division at Argonne National Laboratory and in the Materials Science Department at University of Illinois at Urbana-Champaign. The focus of his scientific work has been the investigation, via large-scale computer simulation, of defects, diffusion, and microstructure evolution in extreme environments.

At LLNL, he first worked on materials issues for the fusion program and then joined the Chemistry and Materials Sciences (CMS) Directorate in 1994. Between 1994 and 1996, he focused his research activities around the development of physics-based predictive models of ion implantation and thin film growth for semiconductor processing in collaboration with Bell Labs, Intel, Applied Materials, IBM and other semiconductor corporations. Between 1994 and 2002, he was also involved in the development of multiscale models of materials strength and aging in irradiation environments and worked in the Advanced Simulation and Computing Program developing models of materials strength.

In 1999, he became group leader for Computational Materials Science and helped build and lead an international recognized effort in computational materials science at LLNL. Between 2000 and 2002, he served as the CMS Materials Program Leader for the National Ignition Facility (NIF), where he focused on optical materials and target development for NIF applications.

Dr. Tomás Díaz de la Rubia was selected as the Associate Director for Chemistry and Materials Science in 2002. Currently, he leads the Chemistry, Materials, Earth, and Life Sciences Directorate, formed in 2007 after merging CMS and the previous Biosciences and Energy and Environment directorates.

Dr. Tomás Díaz de la Rubia has published more than 140 peer reviewed articles in the scientific literature, has chaired numerous international conferences and workshops, and has edited several conference proceedings and special journal issues. He belongs to the editorial board of five major scientific journals, and continues to serve in numerous national and international panels. He was elected a fellow of the American Physical Society in 2002, and is currently the vice chair (chair elect) of the Division of Computational Physics. He is a fellow of the American Association for the Advancement of Science (AAAS), and he served as an elected member of the board of directors of the Materials Research Society between 2002 and 2005.



Jack Dongarra

Jack Dongarra received a Bachelor of Science in Mathematics from Chicago State University in 1972 and a Master of Science in Computer Science from the Illinois Institute of Technology in 1973. He received his PhD in

Applied Mathematics from the University of New Mexico in 1980. He worked at the Argonne National Laboratory until 1989, becoming a senior scientist. He now holds an appointment as University Distinguished Professor of Computer Science in the Computer Science Department at the University of Tennessee and holds the title of Distinguished Research Staff in the Computer Science and Mathematics Division at Oak Ridge National Laboratory (ORNL), Turing Fellow at Manchester University, and an Adjunct Professor in the Computer Science Department at Rice University. He is the director of the Innovative Computing Laboratory at the University of Tennessee. He is also the director of the Center for Information Technology Research at the University of Tennessee which coordinates and facilitates IT research efforts at the University.

He specializes in numerical algorithms in linear algebra, parallel computing, the 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 he is coauthor of several books. He was awarded the IEEE Sid Fernbach Award in 2004 for his contributions in the application of high performance computers using innovative approaches and in 2008 he was the recipient of the first IEEE Medal of Excellence in Scalable Computing. He is a Fellow of the AAAS, ACM, and the IEEE and a member of the National Academy of Engineering.



James Johnson Duderstadt

James J. Duderstadt is President Emeritus and University Professor of Science and Engineering at the University of Michigan.

Dr. Duderstadt received his baccalaureate degree in electrical engineering with highest honors from Yale University in 1964 and his doctorate in engineering science and physics from the California Institute of Technology in 1967. After a year as an Atomic Energy Commission Postdoctoral Fellow at Caltech, he joined the faculty of the University of Michigan in 1968 in the Department of Nuclear Engineering. Dr. Duderstadt became Dean of the College of Engineering in 1981 and Provost and Vice President for Academic Affairs in 1986. He was appointed as President of the University of Michigan in 1988, and served in this role until July 1996. He currently holds a university-wide faculty appointment as University Professor of Science and Engineering, directing the University's program in Science, Technology, and Public Policy, and chairing the Michigan Energy Research Council coordinating energy research on the campus.

Dr. Duderstadt's teaching and research interests have spanned a wide range of subjects in science, mathematics, and engineering, including work in areas such as nuclear fission reactors, thermonuclear fusion, high powered lasers, computer simulation, information technology, and policy development in areas such as energy, education, and science.

During his career, Dr. Duderstadt has received numerous national awards for his research, teaching, and service activities, including the E. O. Lawrence Award for excellence in nuclear research, the Arthur Holly Compton Prize for outstanding teaching, the Reginald Wilson Award for national leadership in achieving diversity, and the National Medal of Technology for exemplary service to the nation. He has been elected to numerous honorific societies including the National Academy of Engineering, the American Academy of Arts and Science, Phi Beta Kappa, and Tau Beta Pi.

Dr. Duderstadt has served on and/or chaired numerous public and private boards. These include the National Science Board; the Executive Council of the National Academy of Engineering; the Committee on Science, Engineering, and Public Policy of the National Academy of Sciences; the Nuclear Energy Research Advisory Committee of the Department of Energy; the Big Ten Athletic Conference; the University of Michigan Hospitals; Unisys; and CMS Energy.

He currently serves on or chairs several major national study commissions in areas including federal science policy, higher education, information technology, and energy sciences, including NSF's Advisory Committee on Cyberinfrastructure, the National Commission on the Future of Higher Education, the AGB Task Force on the State of the University Presidency, the Intelligence Science Board, and the Executive Board of the AAAS.



J. Tinsley Oden

J. Tinsley Oden was the founding Director of the Institute for Computational Engineering and Sciences (ICES), which was created in January of 2003 as an expansion of the Texas Institute for Computational and Applied Mathematics, also directed by Oden for over a decade. The Institute supports broad interdisciplinary research and academic programs in computational engineering and sciences, involving four colleges and 17 academic departments within UT Austin.

An author of over 500 scientific publications: books, book chapters, conference papers, and monographs, Dr. Oden is an editor of the series *Finite Elements in Flow Problems* and of *Computational Methods in Nonlinear Mechanics*. Among the 50 books he has authored or edited are *Contact Problems in Elasticity*, a six-volume series: *Finite Elements*, *An Introduction to the Mathematical Theory of Finite Elements*, and several textbooks, including *Applied Functional Analysis and Mechanics of Elastic Structures*, and, more recently, *A Posteriori Error Estimation in Finite Element Analysis*, with M. Ainsworth. His treatise, *Finite Elements of Nonlinear Continua*, published in 1972 and subsequently translated into Russian, Chinese, and Japanese, is cited as having not only demonstrated the great potential of computational methods for producing quantitative realizations of the most complex theories of physical behavior of materials and mechanical systems, but also established computational mechanics as a new intellectually rich discipline that was built upon deep concepts in mathematics, computer sciences, physics, and mechanics. Computational Mechanics has since become a fundamentally important discipline throughout the world, taught in every major university, and the subject of continued research and intellectual activity. Oden has published extensively in this field and in related areas over the last three decades.

Dr. Oden is an Honorary Member of the American Society of Mechanical Engineers and is a Fellow of six international scientific/technical societies: IACM, AAM, ASME, ASCE, SES, and BMIA. He is a Fellow, founding member, and first President of the U.S. Association for Computational Mechanics and the International Association for Computational Mechanics. He is a Fellow and past President of both the American Academy of Mechanics and the Society of Engineering Science. Among the numerous awards he has received for his work, Dr. Oden was awarded the A. C. Eringen Medal, the Worcester Reed Warner Medal, the Lohmann Medal, the Theodore von Karman Medal, the John von Neumann medal, the Newton/Gauss Congress Medal, and the Stephan P. Timoshenko Medal. He was also knighted as "Chevalier des Palmes Academiques" by the French government and he holds five honorary doctorates, *Honoris Causa*, from universities in Portugal (Technical University of Lisbon), Belgium (Faculte Polytechnique), Poland (Cracow University of Technology), the United States (Presidential Citation, The University of Texas at Austin), and France (Ecole Normale Supérieure Cachan (ENSC)).

In 2004, Dr. Oden was among the seven UT-Austin engineering faculty listed as the most highly cited researchers in the world from 1981-1999 in refereed, peer-reviewed journals, according to the International Scientific Index.

Dr. Oden is a member of the U.S. National Academy of Engineering and the National Academies of Engineering of Mexico and of Brazil. He serves as Co-Chairman of the Accelerated Strategic Computing Initiative (ASCI) Panel for Sandia National Laboratories. He is a Member of the IUTAM Working Party 5 on Computational Mechanics and serves on numerous organizational, scientific and advisory committees for international conferences and symposiums. He is an Editor of *Computer Methods in Applied Mechanics and Engineering* and serves on the editorial board of 27 scientific journals.

Dr. Oden has worked extensively on the mathematical theory and implementation of numerical methods applied to problems in solid and fluid mechanics and, particularly, nonlinear continuum mechanics. His current research focuses on the subject of multiscale modeling and on new theories and methods his group has developed for what they refer to as “adaptive modeling.” The core of any computer simulation is the mathematical model used to study the physical system of interest. They have developed methods that estimate modeling error and adapt the choices of models to control error. This has proven to be a powerful approach for multiscale problems. Applications include semiconductors manufacturing at the nanoscale level. Dr. Oden, along with ICES researchers, is also working on adaptive control methods in laser treatment of cancer, particularly prostate cancer. This work involves the use of dynamic-data-driven systems to predict and control the outcome of laser treatments using our adaptive modeling strategies.



Gilbert S. Omenn

Gilbert S. Omenn is currently Director, Center for Computational Medicine and Biology at the University of Michigan. He served as Executive Vice President for Medical Affairs and as Chief Executive Officer of the University of Michigan Health System from 1997 to 2002. He was formerly Dean of the School of Public Health and Professor of Medicine and Environmental Health, University of Washington, Seattle. His research interests include cancer proteomics, chemoprevention of cancers, public health genetics, science-based risk analysis, and health policy. He was principal investigator of the beta-Carotene and Retinol Efficacy Trial (CARET) of preventive agents against lung cancer and heart disease; director of the Center for Health Promotion in Older Adults; and creator of a university-wide initiative on Public Health Genetics in Ethical, Legal, and Policy Context while at the University of Washington and Fred Hutchinson Cancer Research Center. He served as Associate Director, Office of Science and Technology Policy, and Associate Director, Office of Management and Budget, in the Executive Office of the President in the Carter Administration. He is a longtime director of Amgen Inc. and of Rohm & Haas Company. He is a member of the Council and leader of the Plasma Proteome Project for the international Human Proteome Organization (HUPO). In 2004 he became president-elect of the American Association for the Advancement of Science (AAAS).

Dr. Omenn is the author of 449 research papers and scientific reviews and author/editor of 18 books. He is a member of the Institute of Medicine of the National Academy of Sciences, the American Academy of Arts and Sciences, the Association of American Physicians, and the American College of Physicians. He chaired the presidential/congressional Commission on Risk Assessment and Risk Management ("Omenn Commission"), served on the National Commission on the Environment, and chaired the NAS/NRC/IOM Committee on Science, Engineering and Public Policy.

He is active in cultural and educational organizations, and is a musician and tennis player. Omenn received his BA from Princeton, the MD, magna cum laude, from Harvard Medical School, and a PhD in genetics from the University of Washington.

**David E. Shaw**

David E. Shaw serves as Chief Scientist of D. E. Shaw Research and as a Senior Research Fellow at the Center for Computational Biology and Bioinformatics at Columbia University. He received his PhD from Stanford University in 1980, served on the faculty of the Computer Science Department at Columbia until 1986, and founded the D. E. Shaw group in 1988. Since 2001, Dr. Shaw has devoted his time to hands-on research in the field of computational biochemistry. He is now personally involved in the development of new algorithms and machine architectures for high-speed molecular dynamics simulations of biological macromolecules, and in the application of such simulations to basic scientific research in structural biology and biochemistry and to the process of computer-aided drug design. Although he leads the lab's research efforts in his role as Chief Scientist, his focus is largely technical, with limited involvement in operational and administrative management.

In 1994, President Clinton appointed Dr. Shaw to the President's Council of Advisors on Science and Technology. He is a fellow of the American Association for the Advancement of Science, and was elected to its board of directors in 1998. Dr. Shaw is also a fellow of the American Academy of Arts and Sciences, and serves on the Computer Science and Telecommunications Board of the National Academies.

**Martin Wortman**

Martin Wortman's current research activities focus on technology assessment. His work is concerned with the development of computationally based predictive models that characterize uncertainty associated with the value of risk encumbered technologies; this work appeals to tenets of optimal gambling. In this context, assessment treats the development, design, deployment, and operation of technology as high-stakes wagers having benefits and liabilities that cannot be predicted with certainty. Often, these wagers must be executed when abundant empirical analyses are not available and uncertainty cannot be completely characterized. Key research issues arise when a unique characterization of the probability law on value (i.e., risk) is unavailable (as is often the case with one-and-off-bets). Wortman develops and explores new computational methods, supporting evaluation, assessment, and wagering that are at the interface of probability, stochastic processes, and high-performance computing.

Dr. Wortman teaches advanced course work in Probability, Stochastic Processes, Operations, Reliability & Risk, and Technology Assessment. He teaches introductory graduate courses in Operations Research Methods and Computational Methods. He teaches undergraduate courses in Operations Research Modeling, System Operations, Production Systems, and System Simulation.

Dr. Wortman maintains an active service role in several professional societies. In addition to organizing conference activities and workshops, he has served as Associate Editor for *Operations Research*, Department Editor of Applied Probability and Engineering Statistics for the *IIE Transactions*, Department Editor of Manufacturing for the *IIE Transactions*, and he is past Editor-in-Chief of the *IEEE Transactions on Reliability*.

APPENDIX B. SITE REPORTS—ASIA

Site: Central Research Institute of Electric Power Industry (CRIEPI)
Abiko Site, 1646 Abiko, Abiko-shi
Chiba-ken 270-1194 Japan
<http://criepi.denken.or.jp/en/>

Date Visited: December 5, 2007

WTEC Attendees: L. Petzold (report author), A. Arsenlis, C. Cooper, D. Nelson

Hosts: Dr. Hisashi Kato, Staff Director, International Cooperation Planning Group
Email: kato@criepi.denken.or.jp
Dr. Tetsuo Matsumura, Deputy Director, Materials Science Research Laboratory
Email: matsu-t@criepi.denken.or.jp
Mr. J. Tsutsui, Senior Research Scientist, Environmental Science Research Laboratory
Dr. T. Iwatsubo, Senior Research Scientist, Energy Engineering Research Laboratory
Dr. H. Kaieda, Senior Research Geophysicist, Civil Engineering Research Laboratory
Dr. N. Soneda, Senior Research Scientist, Materials Science Research Laboratory
Dr. K. Nakashima, Research Scientist, Materials Science Research Laboratory
Dr. N. Hashimoto, Research Scientist, Energy Engineering Research Laboratory

BACKGROUND

The Central Research Institute of Electric Power Industry (CRIEPI) is a Japanese nonprofit corporation founded in 1951 with a broad mission of “solving global environmental problems while ensuring energy security.” CRIEPI has an annual budget of ¥36.8 billion,⁸ which comes mostly from the Japanese power industry; the remainder comes from the Japanese government. CRIEPI has 786 employees of whom 656 are classified as researchers. The institute is organized into one center, the Socio-economic Research Center, and 7 research laboratories: Systems Engineering, Nuclear Technology, Civil Engineering, Environmental Science, Electric Power, Energy Engineering, and Materials Science Laboratories. CRIEPI’s permanent research staff members have formal education in electrical, civil, mechanical, chemical, and nuclear engineering as well as in biological, environmental, information, and social sciences. They publish on the order of 1500 papers per year and file on the order of 300 patent applications per year.

SBE&S ACTIVITIES

The largest simulation efforts at CRIEPI are focused around materials science and climate/weather simulations. The materials science simulations focus on the multiscale modeling of irradiation embrittlement processes, light water reactor pressure vessel steels, and the stability and growth of oxide films on silicon carbide. The multiscale aspects of the effort include molecular dynamics simulations of irradiation cascades, copper precipitation in iron, and dislocation irradiation defect reactions; kinetic Monte Carlo simulations of microstructural coarsening during irradiation; and dislocation dynamics simulations of irradiated strength change. The multiscale modeling framework relies on passing relevant information from small length scale simulations to larger length scale simulations, with the final result being a continuum rate model of irradiation damage evolution and corresponding mechanical property change. The oxide stability simulations focus on performing first principles molecular dynamics simulation of interfaces SiO₂ and SiC with defects and observing the chemical reactions of oxygen atoms across the interface. Such interface calculations are at the forefront of computational chemistry and

⁸ Approximately US\$332 million (exchange rate ~¥111/US\$1 December 2007)

required the use of the Earth Simulator to retain stable interfaces and be able to introduce stable point defects at those interfaces.

The CRIEPI researchers, in collaboration with NCAR (National Center for Atmospheric Research) and Earth Simulator staff, led an international team in performing unprecedented global climate simulations on the Earth Simulator that has influenced international policy on global climate change. The effort was sponsored by MEXT, and the porting of the CCSM 3.0 codes to the vector architecture of the Earth Simulator required approximately 10 person-years worth of effort. The result of the effort was the ability to perform multiple 400-year predictions of future temperature and sea level changes with different CO₂ emissions profiles. The simulations required about six months of wall clock time on the Earth Simulator for four different scenarios to be completed. A challenging high-resolution (a grid spacing of 10 km) ocean simulation was also conducted on the Earth Simulator, and eddy-resolved ocean currents were successfully simulated. Since then, the team at CRIEPI has begun performing regional climate simulations of East Asia to assess the regional impacts of global warming, high-resolution weather forecasting simulations, and global models of ocean currents. Several highlights included simulations of a typhoon making landfall in Japan, and predictions of local flooding in areas surrounding Japan's energy infrastructure that required higher resolution than the Japanese Meteorological Agency typically performs.

CRIEPI researchers provide simulation analysis and tools to the Japanese electric power companies in other areas as well. Examples include CRIEPI's Power system Analysis Tool (CPAT) for electric grid stability, CFD analyses of combustion in coal-fired power plants, CFD analysis of transmission line vibration, and seismic analysis of oil tanks. For the most part, CRIEPI employs commercial software in its routine simulations and outsources code development to tailor the commercial tools for problems of interest. However, the simulation tools for the high-end materials science and climate simulations are developed in-house or with a small set of international collaborators. CRIEPI management considers the codes developed in-house to be proprietary; after establishing the proprietary rights the institution shares its codes with the scientific community.

The researchers at CRIEPI working in the materials science, combustion, and climate areas are proposing ambitious simulation projects. In the nuclear materials arena, researchers are interested in simulating the process of stress-corrosion cracking of pressure vessel steels and cladding materials using a multiscale modeling framework. The challenging aspect of this endeavor is that it adds the complexity of surface chemistry and interfacial defects to simulations of mechanical behavior in irradiation environments. In the combustion arena, researchers are interested in predicting the combustion of new fuels and helping to design next-generation power plants. In the climate arena, researchers are interested in increasing the spatial resolution of the ocean current simulations to 0.1 degree such that oceanic eddies could be resolved, in adding complexity and explicit degrees of freedom in their climate models, and in developing a full "Earth system model" that simultaneously models the evolution of the atmosphere, ocean currents, and land masses.

An equally impressive experimental program was presented in the areas of nuclear materials and combustion science. CRIEPI has invested in state-of-the-art experimental hardware that is able to probe material microstructure at the same level as its kinetic Monte Carlo simulations of irradiation damage evolution. Its experimental and simulations efforts are well coordinated, and there appears to be strong feedback between the two. Validation of the kinetic Monte Carlo simulations allows CRIEPI researchers to make credible predictions of material behavior at engineering length scales with knowledge of the proper mechanisms of microstructural evolution. In their combustion experiments, CRIEPI researchers have built small test furnaces to validate the results and CFD simulations, and it appears that the simulators are now driving the experimentalists to make finer measurements of processes than they have had in the past. SBES at CRIEPI appears to have changed the manner in which experimental data is assessed.

In its hiring practices, CRIEPI tends to hire domain specialists when seeking new people for the institution's modeling and simulation activities and does not exclusively seek out candidates with experience in computer science disciplines. The research staff preferred to hire domain specialists and train them to perform computer simulations rather than hire computer scientists and train them in the technical disciplines that require simulation support. There was an opinion shared by the staff that the universities were properly preparing students to perform the sophisticated cross-cutting research needed at industrial institutions.

CONCLUSIONS

CREIPI has an impressive array of simulation and experimental capabilities to address the most challenging problems facing the Japanese power industry. Its high-end simulations capabilities in materials science and global climate change are on par with the premier simulation institutions around the world. The institution's mission focus enables its researchers to build strongly coupled simulation and experimental campaigns enabling them to validate their simulation results and build confidence in predictions. The close coordination of simulation and experiment should serve as an example to the rest of the world.

Site: **Computational Materials Science Center (CMSC)**
National Institute for Materials Science (NIMS)
Tsukuba, Japan

Date Visited: December 7, 2007

WTEC Attendees: S. Glotzer (report author), M. Head-Gordon, S. Kim, J. Warren, P. Westmoreland

Hosts: Dr. Masaki Kitagawa, Vice President, NIMS
 Dr. Taizo Sasaki, First Principles Simulation Group II Leader, CMSC, NIMS
 Dr. Tsuyoshi Miyazaki, Researcher, CMSC, NIMS
 Dr. Masato Shimono, Researcher, CMSC, NIMS
 Dr. Toshiyuki Koyama, Researcher, CMSC, NIMS
 Dr. Shigeru Okamura, International Affairs Office, NIMS

BACKGROUND

The National Institute for Materials Science was formed in 2001 by combining the National Research Institute for Metals (founded in 1956) and the National Institute for Research in Inorganic Materials (founded in 1966). The NIMS mission is (1) fundamental research and generic/infrastructural technology R&D; (2) popularization of research results and promotion of related activity; (3) common use of NIMS facilities and equipment; and (4) training of researchers and engineers and improvement of the quality of these human resources. NIMS is somewhat similar in scope to the Materials Science and Engineering Laboratory at the National Institute of Standards and Technology in the US. Their research is of the highest quality and impact. They have 124 MOUs with countries around the world, and 12 sister institutes in 7 countries. They participate in 10 graduate schools internationally, and collaborate extensively with industry. NIMS ranked 6th in the world in # of citations in materials. Six major research fields for NIMS from 2006-2010 are:

- Nanotechnology-driven advanced materials research
 - Nanotech common key technologies
 - Synthesis and control of novel nanomaterials
 - Nanotech driven materials research for information technology
 - Nanotech driven materials research for biotechnology
- Advanced materials research for social needs
 - Materials research for the environment and energy
 - Materials research for reliability and safety

The Computational Materials Science Center (CMSC) within NIMS aims to develop advanced simulation technologies for nanoscale materials with innovative properties and explore design rules for novel properties and functions. The Center analyzes and predicts properties of nanomaterials and aims to clarify structure-property relationships. Tools include various advanced simulation techniques such as large-scale first-principles simulations, function analysis simulations, strong coupling models, phase-field modeling, and multiscale modeling and simulations.

The director of the CMSC is Dr. Takahisa Ohno. The Center is comprised of four groups, and totals roughly 20 permanent research staff and 20 postdocs and graduate students:

- First Principles Simulation Group I – One group leader, six researchers and engineers, and 12 research fellows, and two administrative staff. Dr. Ohno is also the group leader of the First Principles Simulation Group I.

- First Principles Simulation Group II - – One group leader, five researchers and engineers, four research fellows, and one office assistant.
- Strong Coupling Modeling Group – One group leader, three researchers and engineers, three research fellows, and one office assistant.
- Particle Simulation and Thermodynamics Group – One group leader, six researchers and engineers, four research fellows, and one office assistant.

RESEARCH

The visiting panel heard an overview of NIMS and CMSC research from the Dr. Kitagawa, VP of NIMS, and Dr. Sasaki, Group Leader in CMSC.

SBE&S methods within the Center span multiple scales. First principles simulations include order-N DFT simulations, hybrid methods, and TDDFT. Nanoscale functional analysis includes electron transfer, spin transport, and diffusion quantum Monte Carlo. Strong coupling modeling treats quantum transport, many-body effects, and thermal fluctuations. Multiscale modeling and simulation combines molecular dynamics, Monte Carlo, and phase-field methods. The phase field work is particularly highly regarded within NIMS. Number 2 of the top 3 materials infrastructure accomplishments at NIMS is the structural materials database (see Software below).

The panel was impressed with the quantity and quality of the SBE&S research we heard about. Overall, the computational research is highly sophisticated and on par with the best work in the US and Europe. Current projects within each of the four groups of the CMSC include:

4. First-principles simulation group I:
 - (a) Development of Calculation Method for Large-scale DFT Simulation
 - (b) Investigation of Electron Transport Properties of Nanostructures
 - (c) First Principles Simulation of Redox Reactions
 - (d) Control of the States of Materials
5. First-principles simulation group II:
 - (a) Electron Correlation Effects on Materials Properties
 - (b) Prediction of Properties of Materials
 - (c) Response of Materials (including magnetic, superconducting and optical materials)
6. Strong-coupling modeling group
 - (d) Crystal-Orbit-Spin Coupling in Novel Cuprates and Materials Design
 1. Investigation on Phase Transition and Dynamics of Vortices in Superconductors
 2. Research on Quantum Transport Phenomena
7. Particle simulation and thermodynamics group
 - (e) Prediction of Microstructure Evolution by Phase-Field Method
 - (f) Reaction Diffusion Behavior between Ni-Al-Ir Alloy
 - (g) Research on Nanostructure of Materials by MD Simulations
 - (h) Investigation of Dynamics in Phase Transitions of Superconductors

HPC RESOURCES AND SOFTWARE

The main supercomputer is the Hitachi SR11000/62 with 992 processors and 2 Tb storage. A new major acquisition from a Japanese computer vendor is planned for 2008. They are not currently addressing multicore computing future architectures.

Many of the codes used by the CMSC are developed in house. The CMSC has developed several impressive SBE&S software platforms now used by others. These include:

- A thermodynamics database and simulation software developed in collaboration with AIST, Tohoku Univ., Kyushu Inst. Tech., and InterScience Ltd. The software company name is Materials Design Tech. CO. LTD, a venture company from NIMS.
- MatEX: A Materials Design Platform (<http://matex.nims.go.jp/>). Demonstration programs for the phase-field modeling of microstructure evolution in engineering materials on the web.
- Nanoscale Device Simulation Software (<http://www.rss21.iis.u-tokyo.ac.jp>). They offer services to predict and design the materials properties and functions for the next-generation semiconductor nano-devices on the web.

The materials database involves an exchange of personnel with de la Rubia's group at LLNL.

DISCUSSION

During the talks there was good opportunity for discussion of both scientific issues directly related to the content of the talks, and also more general issues surrounding simulation-based engineering and science. Some of the additional issues that arose included:

- Education and HR in SBE&S: Potential postdocs and research staff with appropriate training in the development of algorithms and programs are hard to recruit. The main cause of this issue is that too many graduate students are being trained primarily to run existing codes to solve applied problems rather than learning the skills necessary to create a new application. NIMS recently created the International Center for Young Scientists (ICYS) to become a more attractive workplace for foreign young scientists. The concept is based upon (1) international: English as working language; (2) independent: autonomous research; (3) interdisciplinary: a fusion of different cultures and fields; and (4) innovative: strategic research. Benefits to participants in this program include a high salary, research grant support (5 M yen/year), and ideal research environment with cutting edge facilities and single occupancy cubicles. In five years, CMSC hired 10 simulators. Most of their backgrounds are in materials science, some physics. Most know how to run codes, but not how to innovate and develop new codes.
- Appreciation of SBE&S: SBE&S activities at NIMS started 15 years ago, and CMSC started in 2001 with support from the Director-General. With new emphasis on nanoscience, experimentalists are becoming more interested in collaborating with simulators. Experimentalists now seek out input from simulators, unlike a several years ago. Simulation now having more and more impact on real problems.
- Interaction with industry: No current interactions with industry.

Site: **Dalian University of Technology**
Department of Vehicle Engineering and Mechanics
Key Lab for Mechanical and Structural Testing
Dalian, 116024, P.R. China
<http://www.dlut.edu.cn/dut-e/main.htm>

Date: December 5, 2007.

WTEC Attendees: S. Kim (report author), S. Glotzer, M. Head-Gordon, J. Warren, P. Westmoreland, and G. Hane

Hosts: Prof. Guiling Ning, Vice President, DUT
 Prof. Hongwu Zhang, Chair Professor and Dean, Faculty of Vehicle Engineering and Mechanics; Email: zhanghw@dlut.edu.cn
 Prof. Wanxie Zhong, CAS Academician
 Prof. Jiahao Lin, CAS Academician
 Email: Jhlin@dlut.edu.cn
 Prof. Hualiang Jiang, Shanghai Institute of *Materia Medica* of CAS
 Prof. Guo Xu; Email: guoxu@dlut.edu.cn
 Prof. Chujie Wu
 Prof. Xikui Li; Email: Xikuili@dlut.edu.cn
 Prof. Ping Hu, DUT School of Automotive Engineering
 Email: pinghu@dlut.edu.cn
 Prof. Yuefang Wang; Email: yfwang@dlut.edu.cn
 Assoc. Prof. Kang Zhan; Email: ahankang@dlut.edu.cn

BACKGROUND

Dalian University of Technology was founded in 1949 as the School of Engineering of Dalian University and became independent as Dalian Institute of Technology the following year. The graduate school was established in 1986, and the present name was adopted in 1988. DUT has received two successive infusions of “211 Project” (higher education construction) infrastructure funds in the 1996 and 2003 five-year cycles, and two “985 Project” (building world-class university structures for the 21st Century) awards in the 2001 and 2004 cycles. Today DUT is one of China’s top engineering and technology-oriented universities, with 1,800 faculty members (of whom 6 are members of the China Academy of Science, CAS, and 4 are members of the China Academy of Engineering, CAE), 15,000+ graduate students (3,100+ of whom are PhD students), and almost 19,000 undergraduate students. Strong ties with industry and a focus on applied innovation are reflected in the makeup of DUT’s 2006 research budget: RMB 518 million,⁹ of which RMB 48.4 million (about 9%) came from the National Natural Science Foundation of China (NSFC).

SBES RESEARCH

SBES research activities at DUT were featured in a symposium held in the morning for the WTEC visiting team, hosted by the Department of Vehicle Engineering and Mechanics, and in afternoon tours of two key laboratories. The symposium included presentations on foundational contributions in mechanics from two CAS Academicians (Wanxie Zhong and Jiahao Lin), and featured recent examples of high-impact work with industry, e.g., SBES for the automotive industry (Prof. Ping Hu) and for the energy/off-shore oil production industries. Prof. Hualiang

⁹ Approximately US\$70 million (Dec. 2007 exchange rate of RMB ~7.4/US\$1).

Jiang of the Shanghai Institute of *Materia Medica* gave a talk on SBES applied to drug discovery and his collaborations with DUT in software engineering.

COMPUTING FACILITIES

Computing resources at DUT are desktop workstations focused on visualization and engineering (CAD/CAE) tools. The WTEC visiting team did not see a high-end cluster in our tour.

WORKSHOP PRESENTATIONS

With the exception of the bioinformatics/drug discovery work, high-end computing did not feature prominently in the SBES work presented at the symposium. Indeed, some of the participants stressed the importance of mathematical analysis to reduce the scope of the computational requirements to fit within computing resources available on the DUT campus. But despite this modest scale of computing, the symposium presented a uniformly strong picture of the impact of SBES in the DUT faculty's prominent contributions to the buildup of national infrastructure (bridges, dams, and buildings) and key industries (automobile design and off-shore oil production).

- *Prof. Wanxie Zhong, CAS Academician* (some issues of computational mechanics in China) presented a talk on symplectic theory, featuring a matrix-mechanics and energy integral version of the classical mathematical (exterior algebra and differential forms) theory of Kang Feng, with applications to the long-time stability and accuracy of FE simulations. His symplectic theory appears to be an application of a unitary, energy-preserving algorithm applied to FEM. Prof. Zhong is interested in having his treatise translated into English.
- *Prof. Hualing Jiang, Shanghai Institute of Materia Medica of the Chinese Academy of Sciences* (some issues of computational mechanics in China) presented an overview of the activities in his institute and its collaborations with DUT in software engineering. The R&D at his institute spans the drug discovery activities from early phase discovery biology to ADME/Tox studies, i.e., the pharmacology R&D pipeline upstream of human clinical trials. Accordingly, the SBES activities of Prof. Jiang and coworkers include curation of bioinformatics databases and cheminformatics tools for target identification (docking and reverse docking), with the high-end computing done at the Shanghai Supercomputer Center. His team members publish in the leading international journals in bioinformatics and protein science and technology. They currently use open-source tools (AMBER, Gromacs) and commercial codes (suite of products from Accelerlys), but they are interested in developing their own next-generation capabilities, e.g., GasDock, InduFitDock for flexibility of proteins. They have applied their suite of toolsets to create a screen for ligands based on increasing the potency of traditional Chinese medicines; this approach has recently generated licensed compounds. The institute leads an 863 Project (China high-tech program) for Drug Discovery/Computational Biology. The institute's advance into clinical trials is limited by budgetary constraints.
- *Prof. Guo Xu (Advances of Computational NanoMechanics in DUT)* presented recent work on multiscale SBES, including heat transfer modeling of MEMS devices, 2D-nonlocal extensions of the Cauchy-Born rule in the fracture mechanics analyses of carbon nanotubes, and stochastic modeling of grain growth. He presented a recent publication and media attention from his group's analysis of the mosquito leg using nanoscience principles. In that vein, future research directions are headed towards biomolecular interactions with nanostructures.
- *Prof. Chujie Wu (Advances in CFD in China)* gave, in lieu of presenting his own work, a most helpful overview of the CFD landscape in China. The talk featured the organization of CFD and fluid mechanics research in China, including the larger-scale collaborative efforts (e.g., Alliance for Fluid Mechanics), the key journals, and funding sources and statistics. From the SBES perspective, he drew special attention to the All Flow/OpenCFD Project. His talk included an interesting demonstration with computer visualization of 3D CFD applied to optimizing the drag on a swimming fish and fish pairs.
- *Prof. Jiahao Lin (High-Efficiency Computing of Random Vibrations and Its Applications in China)* presented the extensive applications of his Pseudo-Excitation Method (PEM) to the buildup of infrastructure in China. This k-space method reduces the complexity of vibrational analyses to desktop computations and

has gained wide usage in China, including within industrial infrastructure (offshore oil platforms of China National Offshore Oil Corporation, CNOOC).

- *Prof. Xikui Li (Numerical Simulation of Porous Media)* presented SBES research applied to composite and porous materials. The first part of his talk featured the chemo-thermo-hygro-mechanical analysis of different failure modes of high-performance concrete vs. normal-strength concrete in building fires. The second part of his talk featured SBES of pollutant transport in clay barriers and its impact on the design of landfills. He concluded with an SBES presentation on non-Newtonian fluid mechanics (the classical 4:1 contraction flow benchmark problem) and the automatics/adaptive identification of the mesh-free region.
- *Prof. Ping Hu (Simulation-Based Design of the Modern Automobile Industry)* presented his research and development of the KingMesh Analysis System (KMAS) and its adoption by the auto industry worldwide, including in the United States and Europe. He had to leave immediately after his presentation for another meeting, but reference materials on KMAS are readily available on the Internet and the Kingmesh website (http://www.kingmesh.com/kmas_en/UG.asp? or [DUT/SAE](http://www.kingmesh.com/kmas_en/DUT/SAE)). Note in particular the digital auto body coding schemes and their role in the future of the auto industry. Prof. Hu's PowerPoint presentation was particularly striking as an example of the DUT's contribution towards globalization of SBES activities; its polish and finish quality would be that expected of the Asia office of a marketing agency with global reach. The DUT School of Automotive Engineering website is <http://sae.dlut.edu.cn/index.aspx>.
- *Prof. Yuefang Wang (Simulation Based Compressor Design)* presented DUT SBES research on the stress analysis of compressor components (bolts). This project is an example of the close collaboration of DUT with industry (Shenyang Blower Company).
- *Assoc. Prof. Kang Zhan (Simulation and Structural Optimization for CAE with Industrial Applications)* presented DUT's twenty-year experience with software engineering in the context of computer-assisted engineering (CAE). DUT researchers are developing JIFEX version 3.0, a finite element package for structural analysis. In the context of DUT's prowess in creating intellectual property (IP), Dr. Zhan expressed concern about the "lack of respect" for IP in China.

LAB TOURS

The WTEC site visit team toured two DUT laboratories in the afternoon. First on our tour was an extremely unique full "fab" facility with relatively state-of-the-art characterization and fabrication tools: micro manipulator, AFM and SEM, plus facilities for semiconductor dopant diffusion, electroforming, photolithography, and spin coating. This facility made an excellent site for professors to pursue research topics and for students to gain both research and vocational training.

The second lab on our tour, the Key Lab for Mechanical and Structural Testing featured an impressive cold-room for the study of ice impact on vibrations of off-shore oil rig platforms, i.e., essentially a *post-facto* research program to manage ice on underdesigned oil platforms.

DISCUSSION HIGHLIGHTS

Research Funding

The WTEC team received from the former university president an insightful overview of the history and current landscape of research funding mechanisms in China, grouped into 5 categories:

1. National projects for higher education and development:
2. The research budget of NSFC, which is RMB 4.5 billion¹⁰
3. "863" funds support projects deemed to be high-tech national priorities

¹⁰ Approximately US\$608 million (Dec. 2007 exchange rate of RMB ~7.4/US\$1)

4. “973” funds (RMB 2 billion total¹¹) support economic development in a number of areas: energy, agriculture, IT, population and health, resources and the environment, materials, coupling areas, and important academic fronts (this last category includes 3 projects in computational mathematics)
5. Funds from industry.

Education and Training of Students

- Undergrads are well prepared for using commercial software but not so well for developing new software. The strong feeling was expressed that students should develop their own software for educational reasons—just learning commercial code is not optimal.
- The feeling was expressed that Dalian students are much better prepared than other universities on this front (using software relevant to SBES).
- Each department at DUT provides its own solution to the software/HPC problem.
- Students graduating from DUT get jobs in both industry and academia. An analysis was presented of who hires for simulation (CAE) expertise and why, with resulting implications for the university and the government:
 - There are changes afoot, as new self-designed products by industry have increased the demand for CAE and simulations.
 - Universities like DUT are well equipped to meet this demand.
 - More national investment is required and expected to arrive.
 - All this has led to more jobs for students in corporate R&D, although job prospects vary by discipline and specialty. For example, CFD is primarily a path to academic jobs, including post-doctoral studies in the United States.
- Issues in and management of state-of-the-art hardware:
 - The preference is for C over Matlab; the lower-level approach gets to the state of the art.
 - Too many architectures.
 - Worry about algorithm dependence of parallelism.
 - More emphasis is needed on SBES.
- IP issues: copyright violations are now viewed as an internal problem, spurring academics to develop new business models for software distribution and development.

¹¹ Approximately US\$270 million.

Site: Fudan University
Department of Macromolecular Science
220 Handan Rd. Yangpu District
Shanghai (200433), P.R.China
<http://www.fudan.edu.cn/englishnew/index.html>
<http://www.polymer.fudan.edu.cn/>

Date Visited: December 6, 2007

WTEC Attendees: S. Glotzer (report author), S. Kim, J. Warren, P. Westmoreland, G. Hane

Host: Prof. Yuliang Yang, Academician CAS,
Director, Dept. of Macromolecular Science,
and Vice Minister of Education, Beijing
Email: ylyang@fudan.edu.cn

BACKGROUND

Fudan University was founded in 1905 with the initial name “Fudan Public School”. Fudan translates into English as “heavenly light shines day after day”. Today the university consists of 17 schools that are further subdivided into 66 departments, and four independent departments (chemistry, physics, environmental science and engineering, and the department that hosted the WTEC visit, macromolecular science). The university has 25,000 full time students and 11,000 continuing education and online students. Foreign student enrollment of 1,650 ranks Fudan with the second highest such enrollment in China. The teaching faculty and research-active faculty (746 engaged in supervising doctoral candidates) numbering 2,300 include 1350 full professors and associate professors. Faculty of special distinction include 25 academicians of the CAS and CAE, 32 chaired professors and 8 lecturer professors of the Cheung Kong Scholars Program and six chief scientists of Project 973. The Department of Macromolecular Science and the Institute of Macromolecular Science of Fudan University were established in May 1993. They originated from the teaching and research group of Macromolecular Science of the Department of Chemistry (1958-1982), the Institute of Macromolecules (1958-1962), and the teaching and research group of the Department of Material Science (1983-1993). They are one of the first academic and research units devoted to Macromolecular Science in China, and also one of the first State-established units that are entitled to granting degrees of Master and Ph.D. in Macromolecular Science as well as providing postdoctoral posts. □□The open laboratory of Polymer Engineering was established under the approval of the former State Education Commission in Feb. 1994, and later in 1999 renamed as the Key laboratory of the State Education Department in Polymer Engineering.

Our host Prof. Yuliang Yang was born in Hai Yan, Zhejiang Province in 1952. He received his doctorate in Science at Fudan University in 1984. His research centers on macromolecular materials. He now has more than ten research programs funded by the National Science Commission and Ministry of Education. He has published hundreds of archival papers in top international journals and has supervised numerous doctoral and masters students. He has held many positions of distinction, including vice president of Fudan University, Special Professor listed in the Cheung Kong Scholars Achievement Award, and Chief Professor of Fudan University. He is currently serving in Beijing as Vice Minister of Education for all of China; in this role he oversees graduate education throughout China.

SBE&S

Prof. Yang’s group is the leading group in computational polymer physics in China. The quality of research in the group is on par with that of the leading computational polymer physics groups in the US, UK, Germany, and elsewhere. They employ state-of-the-art methods in polymer simulation and investigate a range of problems relevant to polymer science, such as phase separation and patterning in polymer blends, block copolymers, etc. Of the 40 current group members, roughly half are experimentalists and half are simulators; a few do both. Graduates of this group go on to postdoctoral positions in the top groups in the US and Europe.

Examples of recent work in the group include the use of time-dependent Ginzburg-Landau methods and related 3D self-consistent field theoretic methods such as the real space Drolet-Fredrickson method developed at UCSB. The Yang group parallelized the Fredrickson code.

DISCUSSION

In lieu of a formal presentation, Professor Yang shared his insights on SBE&S research activities in China, both from his perspective as a professor, as former VP of the Fudan University, and as current vice Minister of Education in Beijing.

Education of Students in SBES

Prof. Yang puts a great deal of emphasis on solid education and training of analytical capabilities. His students all do analytical work prior to doing computational work, to avoid producing students that view simulation software packages as merely “black boxes”. His students take computer courses offered by the computer center, and learn how to program from more senior students and postdocs within the group. In the Yang group, students all learn to program, mostly in Fortran, some in C++, to write their own codes or modify existing codes. Only a few use MPI. Professor Yang involves undergraduate students in research, but this is not common across China. There is no plan yet for teaching students to program new multicore chip architectures; this is not yet something they’re thinking about. Nearly all codes used by group are home-grown.

At Fudan, the undergraduate students are very good, better than at most Chinese universities, with excellent backgrounds in math and programming. Fewer than 10% of the graduate students at Fudan were undergraduates there. In Beijing, at the Ministry of Education, Prof. Yang now oversees the China Scholars program, which supports 2000 graduate students per year to study abroad, with no required commitment to return to China. \$2.5 billion is allocated to this program for the 07-08 academic year. Many students receive their PhD in China, then do a postdoc abroad, and return. Those that return prefer academic positions in eastern China; central and western China universities recognize this and are raising salaries to increase competitiveness.

Funding

The 211 program (<http://www.edu.cn/20010101/21852.shtml>) has been very important. According to a Wikipedia article, “Project 211 is a project of 106 (as of 2007) Key universities and colleges in the 21st century initiated in 1995 by the Ministry of Education of the People's Republic of China. The project aims at cultivating high-level elite for national economic and social development strategies. The project began from the idea that, in the mid 1990s, the 30 elite universities at the time were too low by international research standards. Inclusion in the project means that universities have to meet scientific, technical and HR standards and to offer set advanced degree programs. The figure of 21 and 1 within 211 are from the abbreviation of the 21st century and approximate 100 universities respectively. China now has more than 1,700 standard institutions of higher education, and about 6 percent of them are 211 Project institutions. 211 Project schools take on the responsibility of training 4/5 of doctoral students, 2/3 of graduate students, 1/2 of students abroad and 1/3 of undergraduates. They offer 85 percent of the State's key subjects; hold 96 percent of the State's key laboratories; and utilize 70 percent of scientific research funding (<http://english.people.com.cn/90001/6381319.html>). During the first phase of the project from 1996 to 2000, approximately US\$2.2 billion was distributed.” (Li 2004; http://en.wikipedia.org/wiki/Project_211)

The 211 program is now run by Prof. Yang in his new role in Beijing. The 211 program has raised the level of quality of research substantially, as assessed by (1) solving problems important to China, (2) number and impact factor of papers published, and (3) quality of PhD theses (they awarded 100 PhD thesis awards). Three students from the Yang group have received the best thesis award for their SBE&S research. As vice minister of education, Prof. Yang recognizes that large research projects must include simulation to be successful, and this is a requirement for all successful grants funded by his office.

Program 985 will target lower ranked universities. According to Wikipedia (http://en.wikipedia.org/wiki/Project_985), “Project 985 is a constructive project for founding world-class universities in the 21st century conducted by the government of the People's Republic of China. On May 4, 1998, President Jiang Zemin declared that “China must have a number of first-rate universities of international advanced level”, so Project 985 was begun. In the initial phase, nine universities were given grants in excess of CNY1 billion [China Yuan, also known as Ren Min Bi] each [roughly US\$140 million] over a period of 3 years [<http://english.people.com.cn/90001/6381319.html>]. The second phase, launched in 2004, expanded the programme until it has now reached almost 40 universities. Many participating universities receive tens of millions of yuan each year. [Li 2004] A large part of the funding goes to academic exchange, allowing Chinese academics to participate in conferences abroad, and bringing foreign lecturers to China.” (http://en.wikipedia.org/wiki/Project_211; Li 2004, 17)

Recognition of SBE&S and Interaction in SBES with Industry

Being a State Key Lab, their grant requires them to interact with industry Prof. Yang’s group works on many industrially relevant problems, and he has considerable interaction with Sinopec. Two examples given that helped gain Sinopec interest and support were those of polyethylene film drawing, in which simulations were used to reduce the breakdown time of the machine, and extrusion for plastic pipe manufacturing, where simulations helped show how to balance rigidity and toughness in the material. For the former project, simulations by Yang’s group improved drawing rates from 170 cm/m to 500 m/min. For the latter project, simulations showed how to change polymer composition to avoid hot water pipes splitting. Industry is improving in its appreciation of SBE&S, but slowly. They are not yet hiring many simulators; instead, they work with academic groups to help them do simulations.

The State is improving its appreciation of SBE&S much faster than industry. All of China appears to be recognizing the need to be innovators, and not just the world’s manufacturers, and this requires increased investment in and commitment to research, collaboration, and education, and to SBE&S. There is much pressure placed by the government on intellectual property, which is viewed as critical for innovation. Yang’s group applied for three patents on work based on kinetic Monte Carlo simulations for polymerization. There they invented a new algorithm in 1993 that is now used by many researchers for branching problems in polymers.

HPC Equipment

The Yang group maintains their own 50 cpu cluster. As VP for research of Fudan University, Prof. Yang authorized the purchase of a 259-processor cluster for use across campus. The computer was acquired with funds from the Ministry of Education and is supported by university funds, and thus far researchers on campus are not charged for usage in order to encourage new users. Once the machine becomes heavily used (it was nearly 100% utilized as of December 2007), the university will charge researchers a subsidized fee that will help to pay for a new, larger computer. Prof. Yang noted that this free resource is raising the level of computation and SBE&S on campus. There is no data management plan in place at this time.

REFERENCES

- Aiyar, P. 2006. China hunts abroad for academic talent. *Asia Times Online* February (18).
- Li Lixu. 2004. China’s higher education reform 1998-2003: A summary. *Asia Pacific Education Review V*.
- World Education News and Reviews (WENR). 2006. International rankings and Chinese higher education reform. WENR XIX. Available online: <http://www.wes.org/ewenr/06oct/practical.htm>.

Site: Institute of Chemistry, Chinese Academy of Sciences (ICCAS)
Joint Laboratory of Polymer Science and Materials
No.2, 1st North Street, Zhongguancun
Beijing, 100190, P.R. China
<http://www.iccas.ac.cn/english/>

Date Visited: December 4, 2007

WTEC Attendees: S. Glotzer (report author), S. Kim, J. Warren, P. Westmoreland, G. Hane

Hosts: Prof. Dadong Yan
Email: yandd@iccas.ac.cn
Prof. Hongxia Guo
Email: hxgui@iccas.ac.cn
Prof. Qi Liao
Email: qiliao@iccas.ac.cn
Prof. Xiaozhen Yang
Email: yangx@ppllas.icas.ac.cn
Prof. Charles C. Han (in absentia)
Email: c.c.han@iccas.ac.cn
Prof. Wan Li-jun
Email: w.l.jun@iccas.ac.cn

BACKGROUND

The Joint Laboratory of Polymer Science and Materials was founded in 2003 by Dr. Charles C. Han, who serves to the present as its director, by unifying two laboratories for polymer research housed in the Institute of Chemistry, Chinese Academy of Sciences (ICCAS). ICCAS is a multidisciplinary institute dedicated to basic research in the chemical sciences, and to the development of innovative high-technology application and technology transfer to meet strategic national needs. Its current director is Dr. Wan Li-jun, who gave an overview of the Institute. Major research directions at ICCAS include macromolecular science, physical chemistry, organic chemistry, and analytical chemistry. Major research areas include polymer science and materials; chemical reaction dynamics and structural chemistry; organic solids; photochemistry and photo-functional materials; nanoscience and nanotechnology; colloid, interface sciences and chemical thermodynamics; molecular recognition and selective synthesis; analytical chemistry for life sciences; theoretical chemistry; and high-tech materials. ICCAS has 442 staff, 82 professors, 8 academicians, 117 associate professors, and 171 assistant professors. There is no theory division per se; instead, theorists are embedded within various labs, which increases interaction. Currently 6% of the faculty and staff are theorists/simulators performing SBE&S related work.

The State Key Laboratory of Polymer Physics and Chemistry (SKLPPC), founded in 1956, and the CAS Key Laboratory of Engineering Plastics Laboratory (EPL), founded in 1991 and achieving Key Lab status in 2004, are two Key Laboratories that conduct research on polymers within ICCAS. The Joint Laboratory of Polymer Science and Materials (JLPSM) was created in 2003 to oversee both labs. Dr. Han serves as director and chief scientist of both SKLPPC and JLPSM. The vision of the Joint Laboratory is to unify the experimental, theoretical and computational spectrum of polymer science and technology activities ranging from atomistic scale, to mesoscale materials characterization to macroscale applied polymer rheology and processing. SBES plays an especially foundational role in the Joint Laboratory in that the Lab's experimental program is prioritized to provide insights at the scale-to-scale transitions along the multiscale spectrum, a vision provided by Han. Between 1998-2004, the combined labs published 1038 papers, which includes 643 papers in SCI journals, and 66 papers in journals with impact factors over 3.0. The lab applied for 231 patents during this period, of which 121 were granted. The labs also received four China science prizes, and presented 73 invited lectures at international conferences, during this period.

Professor Charles Han is one of the world's leading polymer scientists. He graduated in Chemical Engineering from the National Taiwan University in 1966. He received his MS in Physical Chemistry from the University of Houston (1969) and PhD in Physical Chemistry from the University of Wisconsin, Madison (1973). The following year he joined the National Institute of Standards & Technology and remained there for the next 29 years as Research Scientist, including Group Leader for the Polymer Blends Group (1985–1995), Group Leader for the Multiphase Materials Group (1999–2002) and NIST Fellow (1995–2002). During his NIST period, Charles Han was Visiting Professor at Kyoto University (Japan) and Changchun Institute of Applied Chemistry at the Chinese Academy of Sciences and Adjunct Professor at Pecking University (China) and the University of Connecticut (a post he still holds). In 2002 Charles Han became a Director and Chief Scientist at the Joint Lab of Polymer Science and Materials of the Institute of Chemistry at the Chinese Academy of Sciences in Beijing. His research interests are in the areas of structure and properties of polymers and polymer blends, scattering in polymers, statics and kinetics in polymer phase separation, and multi-phase phenomena. He is the recipient of several awards, including the Bronze (1980), Silver (1982) and Gold (1986) Medal of the US Department of Commerce, the Dillon Medal (1984) and the Polymer Physics Prize (1999) of the American Physical Society, and Humboldt Senior Research Award from the Alexander von Humboldt Foundation in Germany (1995), and he is a Fellow of the American Physical Society. He is co-author of more than 300 publications in archival journals and books.

DISCUSSION

Dr. Han was on travel, so the spectrum of activities in the laboratory was covered by Drs. Yan and Guo.

- SBE&S research: Examples of SBE&S related research highlighted in the most recent brochure includes molecular dynamics simulations of polyelectrolytes in a poor solvent (Q. Liao, et al., *Macromolecules*, 2003, 36, 3386-3398 and free energy calculations on effects of confinement on the order-disorder transition in diblock copolymer melts. The panel heard short overviews on research examples in SBE&S:
 - Zhigang Shua – organic electronic and photonic materials via ab initio computations. OLED/PLEDs, transport, organic photonics. Uses VASP.
 - Charge mobility for organic semiconductors: intramolecular and intermolecular electron-phonon couplings. Work on Holstein-Peirels model published in *JCP* 127 044506 (2007).
 - O/PLEDs: Focus on aggregation-induced emission and prediction of efficiency. Related work published in *JCP* 126 114302 (2007) and *Adv. Chem. Phys.* 121 (2002).
 - DMRG approaches: EOM-CC methods for non-linear optics. Also “correction-vector” approach (cf. *JPC A* 111, 9291 (2007)). This code is used by a number of other groups, such as Bredas’ group.
 - Hongxian Guo – soft matter at multiple length scales, including molecular dynamics and coarse-grained models.
 - Qi Lao: Molecular dynamics-Lattice Boltzmann hybrid algorithm. Has three students.
 - Xiaozhen Yang: Microscopic and mesoscale simulations of polymers. Develops own codes. Has six students, one asst. professor, and uses part of a 64-node cluster.
 - Dadong Yan: Theory of polymer crystallization and self-assembly.
 - Charles Han: Overseeing major project on integrated theory/simulation/experiment for multiscale studies of condensed phase polymer materials. Funded with 8M RMB for four years, 20 investigators, with roughly 100K RMB per investigator. This is a highly unique and challenging program, inspired by the Doi project in Japan, but goes substantially beyond that project to closely integrate experiment both as a validation tool and to supply needed input and data for models.
- Education of students in SBES: At CAS, there are fewer chances to take courses than at the university in general, and this is also true in simulation. One undergraduate course on introduction to simulation is typical. The graduate school offers one course on molecular simulation where students can learn molecular dynamics. Although many of the researchers’ backgrounds are in physics, this is not true of the students. The student training issues in SBE&S are very similar to elsewhere, with insufficient physics and math preparation.

- Funding: Today, 60-70% of student funding comes from their “NSF”, with almost none coming from the academy. This is a different model than in the past. Increased funding is leading to 20-30% growth per year in students and faculty. 30% of funds are used for equipment.
- Interactions with industry: Chinese industry historically has not done much R&D, and thus innovation has been very difficult. This is starting to change. The State encourages interactions between universities and industry, and the 863 Project supports research with industry specifically. However, simulation has not been popular in 863 projects. One reason is that the industrial timeline is very short term, with results desired within six months. Thus nearly all research in China is funded by the government. The State Key Lab of Polymer Science has funding from the Beijing branch of P&G for a collaboration on detergents for three years.
- HPC resources: Shuai has a 300-cpu cluster and a four-year old 64-cpu cluster. ICCAS has a 128-cpu cluster, and the CAS has a 1024-cpu cluster.

REFERENCES

ICCAS. N.d. Annual report of the Joint Laboratory. Beijing: ICCAS.

- Site:** **Institute for Molecular Science (IMS)**
38 Nishigo-Naka
Myodaiji, Okazaki 444-8585 Japan
<http://www.ims.ac.jp>
- Date Visited:** December 5, 2007
- WTEC Attendees:** P. Cummings (report author), G. Karniadakis, L. Petzold, T. Arsenlis, C. Cooper, D. Nelson
- Hosts:** Prof. Fumio Hirata, Director, Dept. Theoretical and Computational Molecular Science
Group Leader, Developing the Statistical Mechanics Theory in Chemistry and Biophysics
Email: hirata@ims.ac.jp
- Prof. Susumu Okazaki, Director, Research Center for Computational Science
Group Leader, Molecular Dynamics Study of Classical Complex Systems and Quantum Systems in Condensed Phase
Email: okazaki@ims.ac.jp
- Dr. Kenji Yonemitsu, Associate Professor, Department of Theoretical and Computational Molecular Science
Group Leader, Theory of Photoinduced Phase Transitions
Email: kxy@ims.ac.jp
- Dr. Katsuyuki Nobusada, Associate Professor, Department of Theoretical and Computational Molecular Science
Group Leader, Theoretical Studies of Electron Dynamics
Email: nobusada@ims.ac.jp
- Prof. Shigeru Nagase, Dept. of Theoretical and Computational Molecular Science *Group Leader, Theoretical Study and Design of Functional Molecules: New Bonding, Structures, and Reactions*

BACKGROUND

Japan's Institute for Molecular Science (IMS) was founded in 1975. In 2004, it began a process of privatization as part of the newly established National Institutes of Natural Sciences (NINS), an interuniversity research institute corporation comprised of the National Astronomical Observatory of Japan, the National Institute for Fusion Science, the National Institute for Basic Biology (NIBB), the National Institute for Physiological Sciences (NIPS) and the IMS. The NIBB, NIPS, and IMS are all located in Okazaki, in Aichi Prefecture, and from 1981 to 2004 constituted the Okazaki National Research Institutes. The aim of IMS is to investigate fundamental properties of molecules and molecular assemblies through both experimental and theoretical methods. It has a staff of ~75 research personnel and ~35 technical personnel. IMS is also a part of the Graduate University for Advanced Studies (SOKENDAI); therefore, research groups consist of faculty, graduate students, and post-doctoral researchers. The IMS is divided into three sections: Research Departments (of which there are four), Research Facilities (of which there are seven) and a Technical Division. Research groups at the IMS belong to Research Departments or to Research Facilities. All of the IMS research groups, including research groups at the Research Center for Computational Science, support visiting researchers from national, public, and private universities in Japan.

One of the four IMS Research Departments is the Department of Theoretical and Computational Molecular Science (DTCMS), which is closely affiliated with the Research Center for Computational Science (RCCS). The RCCS appears in the IMS organizational chart as being part of a Research Department (along with the DTCMS) as well as being a Research Facility. The stated goal of the DTCMS and RCCS is "to develop theoretical and computational methodologies that include quantum mechanics, statistical mechanics, and molecular simulations in order to understand the structures and functions of molecules in gases and condensed phases, as well as in bio

and nano systems” (IMS 2007, 3). Seven research groups, all conducting research in theoretical and computational molecular science are contained within the DTCMS and RCCS. The WTEC delegation met with five of the seven group leaders, who also included the directors of the DTCMS (Hirata) and RCCS (Okazaki).

RESEARCH ACTIVITIES

The research groups conducting theoretical and computational molecular science research within the DTCMS and RCCS are all well regarded in their respective areas. Information about their research is available at <http://www.ims.ac.jp/eng/researchers/index.html#keisan>. Some of the capabilities within the IMS are unique. For example, the three-dimensional reference interaction site model (3D-RISM) theory for molecular structure is unique to the IMS and represents expertise built up by Fumio Hirata over almost three decades. The 3D-RISM theory can be used to study proteins in solution and is particularly useful for predicting the distribution of water around a protein in solution.

Most of the research codes used with the IMS are in-house codes developed within the institution and are not open-source, although they are shared with collaborators. The RCCS is a substantial computational resource for IMS and external researchers. The website (<http://ccinfo.ims.ac.jp>) is in Japanese only; the most recent data for which usage is publicly available is February 2007, at which time the RCCS was used by 589 scientists in 144 project groups. The RCCS computer systems (March 2008), consisting of Fujitsu PrimeQuest, SGI Altix4700, and Hitachi SR16000, are used to support computational research in quantum chemistry, molecular simulation, chemical reaction dynamics, and solid state physics. These systems are linked to international networks through Super Science Information Network (super SINET).

Most of the WTEC delegation’s discussion at the IMS did not focus on scientific research but on the IMS role in the Next Generation Supercomputing Project (NGSP), discussed in detail in the RIKEN site report. The NGSP will produce a combined vector-scalar supercomputer that is designed to achieve 10 petaflops on the LINPACK benchmark by 2011/2012 at a hardware cost of ¥115 billion. Two major software projects are being funded to develop software to run efficiently (i.e., scale effectively) on the NGSP machine. One project, being pursued by RIKEN with \$15 million/year in funding, will generate software in the life sciences. The other project, funded at \$5 million/year at the IMS, will generate software for the nanoscale sciences. The IMS played a similar role in the National Research Grid Initiative (NAREGI), but it is now focusing on the NGSP project. The issue of software development to scale effectively on petaflop architectures, which will have unprecedented numbers of processor cores, is of great interest in the United States as well as in Japan.

Both the RIKEN and IMS teams developing software for the NGSP face the challenge that their 5-year software development funding began in 2006, while the actual NGSP machine will not be available until 2011, and many of the architectural details of the machine are still being debated. At this time, it appears that there is no simulator for the NGSP machine that would permit software strategies for parallelization to be evaluated before the machine is built. The IMS researchers also stated that they believed that computational scientists would play a limited role in the efficient parallelization on the NGSP hardware. They will collaborate with computational scientists from the University of Tokyo and Tsukuba University in optimizing the codes. This view of the limited role of computational specialists in modifying codes (and, indeed, rewriting them from scratch or completely changing the algorithmic approach) in order to be efficient on large parallel architectures is in contrast to the prevailing view in the U.S. high-performance computing (HPC) community and the U.S. funding model for many large computational projects. For petaflop-and-beyond architectures, this problem is being made even more acute with the proliferation of multicore processors.

The software development activities at IMS and RIKEN are being driven by several computational grand challenges, such as modeling/designing post-silicon electronics, understanding enzymatic catalysis, full-physics simulation of fuel cells, and *ab initio* protein structure prediction, as well as pragmatism about which codes are best ready to migrate to the NGSP-level architecture.

CONCLUSIONS

The IMS is a scientifically first-class institution where researchers have the opportunity to focus on research activities (e.g., no undergraduate teaching duties, as in U.S. national laboratories) while still enjoying an academic environment (e.g., having a graduate program with graduate students, and having tenure with university-style titles and hierarchy). The IMS has an unusual policy of not promoting within the institution—for example, one cannot be promoted from associate professor to full professor within IMS. Thus, an IMS associate professor with aspirations to be a full professor at IMS will have to spend some time at another institution before he/she could be hired as an IMS full professor. This is a policy that clearly is designed to create high standards and to avoid in-breeding, resulting in vitality and breadth in the research programs. However, the IMS is not without challenges. Having a graduate program with no undergraduate program creates the problem that the recruitment of graduate students is more difficult than for regular universities in Japan, who primarily recruit graduate students from the cream of the undergraduate pool of students at their own institutions.

Reforms in the government funding and structure of Japanese science have resulted in the IMS, and the other institutions of the NINS, undergoing a process of privatization since 2004. This was not discussed among the IMS researchers and the WTEC delegation; however, it is clear that these reforms are having significant impact. The IMS Director-General, Hiroki Nakamura, in his introductory remarks for the 2007 IMS brochure accessible from the IMS home page (<http://www.ims.ac.jp>), is unusually frank in saying that, for the future health of Japanese science, "...it would be necessary to reconsider the structure by analyzing the problems we have faced. The present situation of Japanese science policy is unfortunately quite serious. We, scientists, should strive to improve this situation, and at the same time we have to take this adversity as a spring to carry on basic researches of high quality."

The IMS clearly has the opportunity to play a major role in the NGSP, with the funding and the mandate to develop the scalable software base to address nanoscience grand challenges. The schedule of the NGSP deployment and the timetable for funding software development will create a challenge for the IMS researchers.

REFERENCES

IMS. 2007. *IMS (Institute for Molecular Science) 2007 (Annual Report)*. Available online at http://www.ims.ac.jp/eng/publications/ims_2007/index.html.

Site: **Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and Systems Science, Chinese Academy of Sciences
No. 55, Zhong Guan Cun Dong Lu (P.O. Box 2719)
Beijing, 100080, P.R. China
<http://www.cc.ac.cn/>**

Date: December 3, 2007.

WTEC Attendees: M. Head-Gordon (report author), S. Glotzer, S. Kim, J. Warren, P. Westmoreland, G. Hane

Hosts: Prof. Zhiming Chen, Director
Email: zmchen@lsec.cc.ac.cn
Prof. Aihui Zhou, Vice-Director
Email: azhou@lsec.cc.ac.cn
Prof. Zhong-ci Shi, Academician
Prof. Qun Lin, Academician
Prof. Jun-zhi Cui, Academician
Prof. Linbo Zhang
Prof. Li Yuan

BACKGROUND

The Institute of Computational Mathematics and Scientific/Engineering Computing (ICMSEC) is one of the four Institutes that comprise the Academy of Mathematics and System Science (AMSS), of the Chinese Academy of Sciences (CAS). Established in 1995, the institute focuses on fundamental research on efficient and robust numerical methods to address problems that arise from scientific and engineering applications. The results of this research are incorporated in high-performance computer software and programs. The institute presently has 17 professors, 4 associate professors, and 7 assistant professors, about 80 graduate students, and 3 postdoctoral fellows. The professors include 2 academicians of the Chinese Academy of Sciences,¹² and 1 of the Chinese Academy of Engineering.

SBES RESEARCH

An overview of research areas related to SBES at the institute was presented by the Director, Prof. Zhiming Chen, although these areas are only part of the ICMSEC research portfolio.

- The first such area is multiscale simulations, which are conducted with the aim of aiding hydraulic engineering design (dam construction), and also of simulating the mechanical performance of composite materials and structures made from those materials.
- A second area is concerned with developing improved environments for high-performance computing, as exemplified by a project on a 3D adaptive parallel finite element software platform, called parallel hierarchical grid (PHG). This project is targeted at the use of roughly 1000 processors and is built in C using MPI on top of open source packages.
- A third area is the development of a new real-space approach to solving the Kohn-Sham density functional theory equations for molecular and materials problems using adaptive finite elements. This relatively recent development is now applicable to energy calculations on molecules of fullerene size.

¹² For perspective, there are only about 25 mathematicians across China who have been recognized this way, 12 of whom are in the AMSS.

- A fourth area also uses adaptive finite elements to solve problems in computational electromagnetism, as illustrated by transformer design for power stations. There is linkage with industry for this project. Another example from this area relates to the microelectronics industry, where the objective is parasitic extraction.
- The fifth area is computational fluid dynamics, where a present example involves solving the compressible Navier-Stokes equations with chemical reactions occurring.

COMPUTING FACILITIES

The institute houses a teraflop-class cluster of 512 CPUs, with a Myrinet® interconnect, which was installed in 2002. While very much state-of-the-art at the time, this cluster is now approaching the need for replacement. ICMSEC researchers intend to use multicore chips with an InfiniBand interconnect.

DISCUSSION

After the WTEC team's hosts at ICMSEC made their initial presentation, there was an open and extensive discussion of the future opportunities and challenges facing the institute, as well as facing SBES in the United States. Pertinent information includes the following aspects.

- The institute's work is largely fundamental in character, and it does not at present have specific commercial application, although this is possible in the future. Commercial applications face the twin hurdles of meeting concrete performance benchmarks and the need for domain-specific knowledge in order to improve the realism of the model. With the institute's mathematical orientation, it is best equipped to contribute to foundational developments that lay a basis for future applications.
- Students are supported by the Chinese Academy of Sciences. Additional funding comes from a National 973 Project (fundamental research underpinning applications)¹³ on high-performance scientific computing that involves 4 groups of researchers, one of which is from ICMSEC. Present government funding for SBES is weighted towards immediate applications, followed by development of new simulation codes, followed by foundational developments. There is a need for improved support at the foundational level.
- Education is graduate-level only. Students typically enter their graduate program with training in mathematics and not computer science or science or engineering. They are trained in high-performance computing at the institute as part of their graduate studies. Admission is quite selective, with an annual intake of about 15 students per year, based largely on recommendations from leading universities. Most students go on to academic jobs afterwards.
- Industry interactions vary strongly depending on economic conditions, the particular development needs, and external political conditions. An example of the latter is the interaction of institute scientists with oil companies, which has recently resumed as a result of export controls limiting their ability to purchase standard commercial software.
- The institute is facing the challenges associated with the usability of next-generation high-performance computing by the development of appropriate libraries and tools (for instance, the PHG project mentioned above). At the time of the WTEC visit, this development was done primarily with MPI, but it seems likely ICMSEC researchers will be utilizing MPI and OpenMP soon, and in the longer run, they would like computer language developments such as additional structures to permit more transparent use of hierarchical parallel computer systems that blend shared memory processors at the low level with distributed memory at a higher level.

¹³ There were about seventy 973 projects awarded in 2007 in China, across all research areas.

Site: Institute of Process Engineering, Chinese Academy of Sciences
State Key Laboratory for Multiphase Complex Systems
Key Laboratory for Green Process Engineering
Beijing, 100080 P.R. China
http://ipe.ac.cn/ipe2005_english/index.html

Date: December 4, 2007

WTEC Attendees: S. Kim (report author), S. Glotzer, M. Head-Gordon, J. Warren, P. Westmoreland, and G. Hane

Hosts: Prof. Suojiang Zhang, Director, State Key Laboratory for Multiphase Complex Systems
and Key Laboratory for Green Process Engineering
Email: sjzhang@home.ipe.ac.cn
Assoc. Prof. Xiangping Zhang
Asst. Prof. Kun Dong

BACKGROUND

The State Key Laboratory for Multiphase Complex Systems and Key Laboratory for Green Process Engineering are components of the Institute of Process Engineering (IPE) of the Chinese Academy of Sciences (CAS). The IPE was founded in 1958 under the name of the Institute of Chemical Metallurgy, but it was renamed in 2001 in recognition of the broader scope of activities in physical and chemical processing. The IPE has significant interaction with industry, as about 50% of its research funding comes from industrial sponsors, both domestic and international and including significant funding from the petrochemical industry (Sinopec). The balance or government funding includes significant projects of national scope from NSFC (the National Natural Science Foundation of China) and the Ministry of Science and Technology (MOST). The current director of the IPE is Prof. Huishou Liu; his predecessor, Prof. Jinghai Li, is now the Vice President of the CAS. Prof. Li is a CFD expert and under his leadership, the IPE has been a strong proponent of simulations technology. The IPE has 269 faculty members (4 are CAS members and 1 is a member of the Chinese Academy of Engineering), including 44 professors and 55 associate professors. Most of the faculty and students in the IPE are chemical engineers.

The two key laboratories that were the focus of the WTEC team's visit are both under the direction of our host, Prof. Suojiang Zhang, and are active in multiscale experimental and simulations research to support the scale-up of multiphase reactors from bench to plant scales. The two laboratories have 10 faculty members (2 active in simulations) and 20 graduate students.

SBES RESEARCH

SBES activities at these two IPE laboratories initially were in the form of training and use of commercial packages, but the laboratories have developed their own modular components for particle fluidization and reaction kinetics (e.g., for applications in coal gasification and thermal cracking of heavy oils). For these, the research collaboration and sponsorship from Mitsubishi Materials Co. was cited. Researchers in these labs are also developing and maintaining a database on thermophysical properties of new materials, such as ionic liquids, as data inserts for the plant simulators.

COMPUTING FACILITIES

HPC resources within the IPE consist of a 96-CPU cluster, originally acquired in 2005 with 64 processors and upgraded to the current configuration in 2006. Our hosts believe that MOST will support the acquisition of a significant new cluster next year.

DISCUSSION

In lieu of a formal presentation, the entire WTEC visit consisted of an open discussion of future SBES opportunities and challenges in process engineering in China, covering the following key points:

- Students are trained on commercial molecular, CFD, and plant simulations from the vendors (Amber, CHARMM, Fluent, GPRIMS, and Aspen Technology products), they but perform SBES research on the development of modular components to handle particle/fluidized bed technology and reaction kinetics. In chemical engineering undergraduate programs, students have had little experience in code development, so training occurs in the form of workshops and seminars.
- With the emergence of multicore architectures and parallel programming, the IPE is in discussions with several other computationally oriented branches of the CAS concerning collaborations on new courses.
- The IPE leadership views simulations as a strategic area for the future and is working on the development of new multiscale simulations.
- Because of emerging opportunities in SBES research, the laboratories would like to recruit from a broader base than chemical engineering, particularly mathematics, physics, and computer science.
- The IPE has significant international collaborations, including joint programs set up in 2001 with ETH-Zurich in the IPE's Multiphase Reaction Laboratory that includes SBES: "Hydrodynamics, Transport Phenomena, and Numerical Simulation in Heterogeneous Particle-Fluid Systems" and "Multi-Scale Method and Systems Integration for Complex Systems."

REFERENCES

- Dong, K., S. Zhang, D. Wang, and X. Yao. 2006. Hydrogen bonds in imidazolium ionic liquids. *J. Phys. Chem. A*, 110:9775.
- Ge, W., and J. Li. 2003. Macro-scale phenomena reproduced in microscopic systems - pseudo-particle modeling of fluidization. *Chemical Engineering Science* 58(8):1565-1585.
- . 2003. Simulation of particle-fluid system with macro-scale pseudo-particle modeling. *Powder Technology* 137(1-2):99-108.
- He, X., X. Zhang, S. Zhang, J. Liu, and C. Li. 2005. Prediction of phase equilibrium properties for complicated macromolecular systems by HGALM neural networks. *Fluid Phase Equilib.* 238(1):52.
- Li, C., X. Zhang, and S. Zhang. 2006. Environmental benign design of DMC production process. *Trans IchemE, Part A, Chem. Eng. Res. Des.* 84(A1):1.
- Li, C., X. Zhang, S. Zhang, X. Tan, and X. Zhang. 2006. Simulation of multi-component multi-stage separation process - An improved algorithm and application. *Chinese J. Process Eng.* 4(3):247.
- Li, C., X. Zhang, X. He and S. Zhang. 2007. Design of separation process of azeotropic mixtures based on the green chemical principles. *J. Clean. Prod.* 15(7):690.
- Li, J., J. Zhang, W. Ge, and X. Liu. 2004. Multi-scale methodology from complex systems. *Chemical Engineering Science* 59(8-9):1687-1700.
- Liu, X., G. Zhou, S. Zhang, G. Wu, and G. Yu. 2007. Molecular simulation of guanidinium-based ionic liquids. *J. Phys. Chem. B.* 111(20):5658.
- Liu, X., S. Zhang, G. Zhou, G. Wu, X. Yuan, and X. Yao. 2006. New force field for molecular simulation of guanidinium-based ionic liquids. *J. Phys. Chem. B*, 110:12062.
- Lu, J., L. Yu, X. Zhang, and S. Zhang. 2008. Hydrogen product from fluidized bed coal gasifier with *in-situ* fixation of CO₂ part I: Numerical modeling of coal gasification. *Chem. Eng. Technol.*, 31(2):197.
- Ma, J., W. Ge, X. Wang, J. Wang, and J. Li. 2006. High-resolution simulation of gas-solid suspension using macro-scale particle methods. *Chemical Engineering Science* 61:7096-7106.
- Yan, L., X. Zhang, and S. Zhang. 2007. The study of molecular modeling for heavy oil thermal cracking. *Chem. Eng. Technol.* 30(9):1.
- Yu, G., and S. Zhang. 2007. Insight into the cation-anion interaction in 1,1,3,3-tetramethylguanidinium lactate ionic liquid. *Fluid Phase Equilib.* 255:86.

- Yu, G., S. Zhang, G. Zhou, and X. Liu. 2007. Structure, interaction and property of amino-functionalized imidazolium ionic liquids by ab initio calculation and molecular dynamics simulation. *AIChE. J.* 53(12):3210.
- Yu, G., S. Zhang, X. Yao, J. Zhang, K. Dong, W. Dai, and R. Mori. 2006. Design of task-specific ionic liquids for capturing CO₂: A molecular orbital study. *Ind. Eng. Chem. Res.* 45:2875.
- Yu, L., J. Lu, X. Zhang, and S. Zhang. 2007. Numerical simulation of the bubbling fluidized bed coal gasification by the kinetic theory of granular flow (KTGF). *Fuel* (86):722.
- Zhang, S., N. Sun, X. Zhang, and X. Lu. 2006. Periodicity and map for discovery of new ionic liquids. *Sci. China Ser. B*, 49(2):103.
- Zhang, X., C. Li, C. Fu, and S. Zhang. 2008. Environmental impact assessment of chemical process using the green degree method. *Ind. Eng. Chem. Res.* 47:1085.
- Zhang, X., S. Zhang, and X. He. 2004. Prediction of solubility of lysozyme in lysozyme-NaCl-H₂O system with artificial neural network. *J. Cryst. Growth* 264:409.
- Zhang, X.P., S. Zhang, P. Yao, and Y. Yuan. 2005. Modeling and simulation of high-pressure urea synthesis loop. *Comput. Chem. Eng.* 29:983.
- Zhou, G., X. Liu, S. Zhang, G. Yu, and H. He. 2007. A force field for molecular simulation of tetrabutylphosphonium amino acid ionic liquids. *J. Phys. Chem. B.* 111:7078.

Site: Japan Agency for Marine-Earth Science and Technology
Earth Simulator Center (ESC)
Yokohama Institute for Earth Sciences
3173-25 Showa-machi, Kanazawa-ku
Yokohama Kanagawa 236-0001, Japan
<http://www.es.jamstec.go.jp/index.en.html>

Date Visited: December 6, 2007

WTEC Attendees: L. Petzold (report author), P. Cummings, G. Karniadakis, T. Arsenlis, C. Cooper, D. Nelson

Hosts: Dr. Tetsuya Sato, Director-General, ESC
Email: tetsuya@jamstec.go.jp
Dr. Kanya Kusano, Program Director, ESC
Email: kusano@jamstec.go.jp
Dr. Akira Kageyama, Group Leader, ESC
Email: kage@jamstec.go.jp

BACKGROUND

The ground-breaking supercomputer at the Earth Simulator Center (ESC) of the Japan Agency for Marine-Earth Science and Technology was once the fastest computer in the world. It has been operational for six years. The current plan is to shut the machine down in one year and replace it with a commercial machine. The final decision had not been made at the time of this writing.

The Earth Simulator Center has 25 scientists in-house and MOUs with many international groups. Its primary objective is to develop new algorithms, including multiscale and multiphysics algorithms. Resources are allocated by a committee of 24 distinguished researchers. There are programs for industry to use the machine. At first, companies were not interested. Then they got good results for collision analysis. Now some companies are pursuing HPC activities on their own. Companies have to pay for computing services; however, the government has a program to which they can apply for such funds. The initial hesitation of companies in using the Earth Simulator supercomputer was apparently because they were using commercial codes and did not have access to source code. The Earth Simulator Project worked with software companies to optimize their codes for the Earth Simulator. The Japanese government controls the fee structure for using the computer. Five percent of machine use is reserved for industry, but industry presently uses only 1%. More than 50% of the node-hours on the Earth Simulator are used for big jobs. Beginners can use a few nodes. Prospective users must show that their code is optimized before access is granted for more nodes.

R&D ACTIVITIES

The increased speed and performance of the Earth Simulator and the supercomputers that came after it have enabled the simulation of realistic models of whole systems. According to Dr. Sato, Director-General of the ESC, one of the most significant impacts of the Earth Simulator Project has been to stimulate the U.S. and Japanese governments to invest in supercomputer development.

An important lesson learned is that the simulation of physical systems for which models are well-established is usually well-suited to vector machines. On the other hand, problems such as cell dynamics, which require the interaction of experiments and simulation in the development of models, tend to be better suited to scalar machines. Thus, the Riken next-generation supercomputer will feature both vector and scalar capabilities.

The ultimate goal of this project is to simulate physical systems as realistically as possible. This requires multiscale algorithms; this is the main focus of the center.

One of the ESC's big successes has been global climate simulation. The Earth Simulator Project has achieved resolution of 10 km. Validation of such a model is very difficult. Mathematically, the ESC researchers don't have a validation system, but they do compare with historical data. They can obtain 1 km resolution via adaptive mesh refinement.

The group has also achieved some impressive results for weather prediction. Here it is important how quickly they can get the result from the computer. The group can obtain predictions for windstream between buildings in downtown Tokyo for use in urban planning. The software has been used to predict typhoon trajectories, which compare well with past data. For the important problem of cloud dynamics, ESC researchers developed a super water droplet code that uses lumped particles, solving simultaneously for global circulation, condensation, and other variables. The resolution that they can obtain greatly improves results for condensation in particular. An important consideration is sensitivity of the results of the macromodel to the micromodel results. They use this to determine where the micromodel needs adjustment. Load balancing is very important.

The Earth Simulator Project has placed equal emphasis on simulation and visualization. Its researchers make use of a CAVE; the region of interest can be zoomed-in. When something important is identified in the CAVE, further visualization is done on the desktop.

The WTEC team asked Dr. Sato what he envisions as the critical applications for supercomputers in the next 10 years. He answered that social prediction may be more important than physical prediction in the next generation of computing. Perhaps this will rely on agent-based models. The Earth Simulator can deal with 6 billion persons' purchasing habits. Of course, there are many problems that would need to be resolved, including how to obtain the data and the personal profiles. The rules are not yet known. How many people are aggressive or conservative? How do the patterns change when people get more information? Interaction between simulation and "experiment" for updating the rules of individual agents would play an important role.

CONCLUSIONS

The Earth Simulator ushered in a new age of supercomputing in which accurate simulation of whole systems became possible. The impact has been worldwide. Science and industry are beginning to realize and capitalize on the implications of this technology. The world-famous supercomputer has been operational for 6 years; the current plan is for it to be shut down by the end of 2008; however the work of the center will continue with its purchase of a commercial supercomputer.

REFERENCES

- Earth Simulator Center (ESC). 2007. *Annual report of the Earth Simulator Center, 2006–2007*. The Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology.
- . 2006 *The Earth Simulator Center* (brochure). The Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology.
- J. the Earth Simulator* Vol. 6. October 2006.
- J. the Earth Simulator* Vol. 7. June 2007.
- J. the Earth Simulator* Vol. 8. November 2007.
- Mezzacappa, A., ed. 2005. SciDAC 2005, Scientific discovery through advanced computing, San Francisco, USA, 26–30 June 2005. *Journal of Physics: Conference Series* vol. 16.

Site: **Kyoto University**
Yoshida-Honmachi, Sakyo-ku
Kyoto 606-8501, Japan

Date Visited: December 4, 2007

WTEC Attendees: P. Cummings (report author), G. Karniadakis

Hosts: Professor Keiji Morokuma, William Henry Emerson Professor Emeritus, Emory University, Atlanta, Georgia, USA, and Research Leader, Fukui Institute for Fundamental Chemistry, Kyoto University
Email: morokuma@fukui.kyoto-u.ac.jp

Professor Shigeyoshi Sakaki, Department of Molecular Engineering, Graduate School of Engineering, and Director, Fukui Institute for Fundamental Chemistry, Kyoto University; Email: sakaki@moleng.kyoto-u.ac.jp

Dr. Shigehiko Hayashi, Associate Professor, Theoretical Chemistry Group, Department of Chemistry, Graduate School of Science, Kyoto University

Dr. Masahiro Ehara, Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University
Email: ehara@sbchem.kyoto-u.ac.jp

BACKGROUND

Kyoto University (<http://www.kyoto-u.ac.jp>), a major national university located in the imperial city of Kyoto, is the second oldest university in Japan and the second-ranked university in Japan (the University of Tokyo is ranked first). The university has a total of approximately 22,000 students enrolled in its undergraduate and graduate programs. The university is a premier research university, with six Nobel Laureates and two Fields Medalists among its faculties and alumni.

One of the interesting features of Kyoto University is the historically strong engineering program. In fact, the two Nobel laureates in Chemistry who were alumni of Kyoto University (Kenichi Fukui, awarded in 1981, and Ryoji Noyori, awarded in 2001) both graduated from chemistry programs within the Graduate School of Engineering. Kenichi Fukui continued as a faculty member at Kyoto University, eventually serving as Dean of Engineering. Fukui's research focused broadly on the theory of chemical reactions. After Fukui's death in 1998, in 2002 the Fukui Institute for Fundamental Chemistry (FIFC, <http://www.fukui.kyoto-u.ac.jp>) was established to commemorate Fukui's achievements and to become a focal point at Kyoto University for fundamental research in chemistry, with a particular focus on theoretical and computational chemistry. Two of the faculty with whom WTEC met at Kyoto University (Keiji Morokuma and Shigeyoshi Sakaki) are also affiliated with the FIFC; Professor Sakaki is the director of the FIFC.

R&D ACTIVITIES

The WTEC delegation heard research presentations from Keiji Morokuma, Shigeyoshi Sakaki, Shigehiko Hayashi, and Masahiro Ehara.

Morokuma: JST-CREST Program

Keiji Morokuma is one of Japan's most distinguished computational chemists, and was one of Kenichi Fukui's first graduate students, receiving his PhD from Kyoto University in 1957. Among his many accomplishments, he served as the founding director and professor (1977–1992) of the Department of Theoretical Studies and the Computer Center of the Institute of Molecular Science in Okazaki, Japan, and from 1993 to 2006 as the William Henry Emerson Professor of Chemistry and Director of the Cherry L. Emerson Center for Scientific Computation at Emory University in Atlanta, Georgia. Upon his retirement from Emory, he was invited to establish a research

group at the Fukui Institute, which he has accomplished with funding from the CREST (Core Research for Evolutional Science and Technology) program of the Japan Science and Technology Agency, JST. The CREST program supports large-scale team efforts. Morokuma's JST-CREST funding (\$3M in direct costs over 5 years) supports his position as a "daily worker" (in U.S. terms, a consultant paid for days worked), nine post-doctoral researchers, one assistant, and four undergraduate researchers, as well as a 200-core dedicated cluster (with ~150 cores to be added). Morokuma also has access to facilities at the IMS and at Oak Ridge National Laboratory, where he is co-PI on a user project (whose PI is Stephan Irle of Nagoya University), in the Center for Nanophase Materials Sciences. Morokuma was a panelist on the WTEC molecular modeling study (Westmoreland et al. 2002) and so had unique insight into the information this WTEC study was seeking. He particularly praised the JST-CREST program for making available long-term support for activities leading to the development of software. His research presentation focused on his current JST-CREST-supported activities, which are centered around long-time simulation of chemical reactions in complex molecular systems.

Morokuma is well known for the development of the ONIOM method (Svensson et al. 1006; Dapprich et al. 1999; Froese and Morokuma 1999; Morokuma 2002); ONIOM is an acronym for "our own N-layered integrated molecular orbitals and molecular mechanics." It is a multiscale methodology for coupling *ab initio* quantum mechanical (QM) methods with higher levels of description, such as molecular mechanics (MM) and molecular dynamics (MD). It is implemented in a number of quantum chemistry packages, including Gaussian and NWChem. The combination of QM and MM or MD allows the study of relatively large systems that undergo fundamentally quantum transformations, such as reactions. Morokuma's current research includes further development of the ONIOM methodology, combining MD and ONIOM for free-energy calculations, and developing a MD simulation methodology with propagation of electron density at finite electronic temperature (as opposed to Car-Parrinello MD, where the electronic temperature is effectively zero). Morokuma's research also involves application of these methods to the simulation of nanoscale systems—specifically, QM/MD simulation of fullerene formation and reactions (Irle et al. 2003; Zheng, Irle, and Morokuma 2005; Zheng et al. 2007), QM/MD simulation of carbon nanotube growth mechanisms (Irle et al. 2006; Wang et al. 2007), and ONIOM simulation of nanocatalysis. A new research focus is the simulation of enzymatic processes, specifically, active site models for metalloenzyme reactions, ONIOM QM:MM optimization of protein models of metalloenzyme reactions, free energy calculations of enzymatic reactions, and ONIOM QM:MM studies of excited dynamics of biomolecular systems.

Sakaki: Molecular Theory for Science and Technology Group

Professor Sakaki leads the Molecular Theory for Science and Technology (MTST) group (http://www.moleng.kyoto-u.ac.jp/~moleng_02/) within the Department of Molecular Engineering in the University of Kyoto Graduate School of Engineering, as well as serving as Director of the FIFC. In addition to Professor Sakaki, the MTST group consists of one associate professor (Hirofumi Sato), one assistant professor (Yoshihide Nakao), nine PhD students, eight MS students, and four undergraduate students. The focus of the group is on theories of chemical reactions and solvent systems, quantum chemical design of novel reactions, chemical bonding and molecular properties, and statistical mechanics of chemical processes. Sakaki performs theoretical studies of reaction mechanisms mediated by organometallic catalysts (Ray et al. 2007; Ochi et al. 2007). His research is primarily supported by MEXT (Ministry of Education, Culture, Sports, Science, and Technology). The techniques used include CASPT2 (Complete Active Space with Second-order Perturbation Theory); the codes used are Gaussian and GAMESS-US. The primary computing resources are a PC cluster in Sakaki's group and the Altics cluster at the Institute for Molecular Science in Okazaki. The shared memory model of the IMS Altics cluster means that the codes used by Sakaki and co-workers scale well on this machine.

Hayashi: Theoretical Chemistry Group

Shigehiko Hayashi is an associate professor in the Theoretical Chemistry Group (TCG) (<http://kuchem.kyoto-u.ac.jp/riron/indexe.html>), headed by Shigeki Kato, within the Department of Chemistry in the Faculty of Sciences at Kyoto University. In addition to Professor Kato and Dr. Hayashi, this group consists of one assistant professor (Takeshi Yamamoto), one research fellow (Atsushi Yamashiro), twelve PhD students, six MS students, and five undergraduate researchers. Hayashi provided an overview of the TCG research, which could be

summarized as chemical reactions in various environments (Hayashi and Kato 1998; Higashi, Hayashi, and Kato 2007; Yamamoto and Kato 2007)—in the gas phase, where the main tool is quantum dynamics; in solution, where the main tools are integral equation methods and MD; and for proteins, where the main tools are hybrid QM/MM and MD.

Ehara: Nakatsuji Group

Masahiro Ehara is an associate professor in the research group of Hiroshi Nakatsuji (<http://www.sbchem.kyoto-u.ac.jp/nakatsuji-lab/english/index.htm>) in the Department of Synthetic Chemistry and Biological Chemistry within the Graduate School of Engineering. Nakatsuji is also affiliated with the FIFC. In the 1970s, Nakatsuji and Hirao developed the symmetry adapted cluster (SAC) method (Nakatsuji and Hirao 1977; Hirao and Nakatsuji 1978a and b) for the ground states of closed and open-shell electronic structures that was subsequently generalized to the SAC-CI (SAC-configuration interaction) method for excited states by Nakatsuji (1978, 1979a and b). Much of the work in the Nakatsuji group involves further development of the SAC-CI method, including using gradients of the SAC-CI energies (yielding forces) to study dynamics involving ground and excited states, and the application of these methods to increasingly complex systems. The original SAC-CI codes were in-house codes, but they are now available in Gaussian. A guide to the use of SAC-CI is available at the Nakatsuji group website, <http://www.sbchem.kyoto-u.ac.jp/nakatsuji-lab/sacci.html>. Ehara focused on the application of SAC-CI to photofunctional materials—specifically, an artificial fluorescent probe used as a biological chemosensor with emphasis on the photoinduced electron transfer (PET) mechanism, and organic light-emitting diodes (OLEDs) with emphasis on excited-state dynamics and conformational effects. Ehara demonstrated the superiority of SAC-CI as a methodology for the study of these complex systems.

CONCLUSIONS

Kyoto University has a long history in theoretical chemistry and computational quantum chemistry, including the Nobel-prize-winning work of Kenichi Fukui. This tradition is alive and well at Kyoto University, as was demonstrated by the presentations to the WTEC delegation. The work presented was all first-class. Researchers rely on a combination of in-house and external (e.g., IMS) computational resources to accomplish their work. There is clearly a trend towards nanoscience and biological applications in all of the work presented. The WTEC team's hosts were generally upbeat about the current state of funding in Japan for computational research such as theirs.

REFERENCES

- Dapprich, S., I. Komaromi, K.S. Byun, K. Morokuma, and M.J. Frisch. 1999. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *Journal of Molecular Structure-Theochem* 462:1-21.
- Froese, R.D.J., and K. Morokuma. 1999. Accurate calculations of bond-breaking energies in C-60 using the three-layered ONIOM method. *Chemical Physics Letters* 305:419-424.
- Hayashi, S., and S. Kato. 1998. Solvent effect on intramolecular long-range electron-transfer reactions between porphyrin and benzoquinone in an acetonitrile solution: Molecular dynamics calculations of reaction rate constants. *Journal of Physical Chemistry A* 102:3333-3342.
- Higashi, M., S. Hayashi, and S. Kato. 2007. Transition state determination of enzyme reaction on free energy surface: Application to chorismate mutase. *Chemical Physics Letters* 437:293-297.
- Hirao, K., and H. Nakatsuji. 1978a. Cluster expansion of wavefunction — Open-shell orbital theory including electron correlation. *Journal of Chemical Physics* 69:4548-4563.
- . 1978b. Cluster expansion of wavefunction — Structure of closed-shell orbital theory. *Journal of Chemical Physics* 69:4535-4547.
- Irle, S., G.S. Zheng, M. Elstner, and K. Morokuma. 2003. Formation of fullerene molecules from carbon nanotubes: A quantum chemical molecular dynamics study. *Nano Letters* 3:465-470.

- Irle, S., Z. Wang, G.S. Zheng, K. Morokuma, and M. Kusunoki. 2006. Theory and experiment agree: Single-walled carbon nanotube caps grow catalyst-free with chirality preference on a SiC surface. *Journal of Chemical Physics* 125.
- Morokuma, K. 2002. New challenges in quantum chemistry: quests for accurate calculations for large molecular systems. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* 360:1149-1164.
- Nakatsuji, H. 1978. Cluster expansion of wavefunction — Excited-states. *Chemical Physics Letters* 59:362-364.
- Nakatsuji, H. 1979a. Cluster expansion of the wavefunction — Calculation of electron correlations in ground and excited-states by SAC and SAC Ci theories. *Chemical Physics Letters* 67:334-342.
- . 1979b. Cluster expansion of the wavefunction — Electron correlations in ground and excited-states by SAC (Symmetry-Adapted-Cluster) and SAC Ci theories. *Chemical Physics Letters* 67:329-333.
- Nakatsuji, H. and K. Hirao. 1997. Cluster expansion of wavefunction — Pseudo-orbital theory applied to spin correlation. *Chemical Physics Letters* 47:569-571.
- Ochi, N., Y. Nakao, H. Sato, and S. Sakaki. 2007. Theoretical study of C-H and N-H sigma-bond activation reactions by titanium(IV)-imido complex. Good understanding based on orbital interaction and theoretical proposal for N-H sigma-bond activation of ammonia. *Journal of the American Chemical Society* 129: 8615-8624.
- Ray, M., Y. Nakao, H. Sato, and S. Sakaki. 2007. Theoretical study of tungsten eta(3)-Silaallyl/eta(3)-Vinylsilyl and vinyl silylene complexes: Interesting bonding nature and relative stability. *Organometallics* 26:4413-4423.
- Svensson, M., S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, and K. Morokuma. 1996. ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *Journal of Physical Chemistry* 100:19357-19363.
- Wang, Z., S. Irle, G. Zheng, M. Kusunoki, and K. Morokuma. 2007. Carbon nanotubes grow on the C face of SiC (000(1)over-bar) during sublimation decomposition: Quantum chemical molecular dynamics simulations. *Journal of Physical Chemistry C* 111:12960-12972.
- Westmoreland, P.R., P.A. Kollman, A.M. Chaka, P.T. Cummings, K. Morokuma, M. Neurock, E.B. Stechel, and P. Vashishta 2002. *Applying molecular and materials modeling*. Dordrecht, Holland: Kluwer Academic Publishers.
- Yamamoto, T., and S. Kato. 2007. Ab initio calculation of proton-coupled electron transfer rates using the external-potential representation: A ubiquinol complex in solution. *Journal of Chemical Physics* 126.
- Zheng, G.S., S. Irle, and K. Morokuma. 2005. Towards formation of buckminsterfullerene C-60 in quantum chemical molecular dynamics. *Journal of Chemical Physics* 122.
- Zheng, G.S., Z. Wang, S. Irle, and K. Morokuma. 2007. Quantum chemical molecular dynamics study of “Shrinking” of Hot Giant fullerenes. *Journal of Nanoscience and Nanotechnology* 7:1662-1669.

Site: Mitsubishi Chemical Group Science and Technology Research Center (MCRC)
 Yokohama Research Center
 1000 Kamoshida-cho, Aoba-ku
 Yokohama 227-8502 Japan
http://www.m-kagaku.co.jp/english/r_td/index.html

Date Visited: December 6, 2007

WTEC Attendees: P. Cummings (report author), G. Karniadakis, L. Petzold, T. Arsenlis, C. Cooper, D. Nelson

Hosts: Dr. Chihiro Miyazawa, Vice President and Board Member, MCRC
 (location: Yokohama); Email: 2604668@cc.m-kagaku.co.jp

Dr. Takao Usami, General Manager of Polymer Lab and Board Member, MCRC (location: Yokkaichi); Email: 3700460@cc.m-kagaku.co.jp

Dr. Shinichiro Nakamura, Mitsubishi Chemical Research Fellow and Technology Platform Leader, Fundamental Technology Division, MCRC
 (location: Yokohama); Email: shindon@rsi.co.jp

Akio Horiguchi, Production Technology Laboratory and General Manager of Yokkaichi Laboratory, MCRC
 (location: Yokkaichi); Email: 2502537@cc.m-kagaku.co.jp

Dr. Takeshi Ishikawa, Senior Researcher, Polymer Laboratory, MCRC
 (location: Yokkaichi); Email: 3805281@cc.m-kagaku.co.jp

Jun Endo, Senior Research Associate, Polymer Laboratory, MCRC
 (location: Yokkaichi); Email: 3709188@cc.m-kagaku.co.jp

Dr. Tomohisa Nakamura, Group Manager, Planning and Coordination Office, MCRC
 (location: Yokohama); Email: 2203789@cc.m-kagaku.co.jp

Dr. Yuan Chen, Senior Research Associate, Polymer Laboratory, MCRC
 (location: Yokkaichi); Email: 8909684@cc.m-kagaku.co.jp

BACKGROUND

Mitsubishi Chemical Corporation (MCC, http://www.m-kagaku.co.jp/index_en.htm) is one of three subsidiaries of Mitsubishi Chemical Holdings Company (MCHC, <http://www.mitsubishichem-hd.co.jp/english/index.html>), created in 2005. MCC has ten domestic manufacturing plants (at Kurosaki, Yokkaichi, Naoetsu, Mizushima, Sakaide, Kashima/Tobu Zone, Kashima/Hasaki Zone, Tsukuba, Matsuyama, and Odawara); two central research centers (at Yokohama and Tsukuba); plus three additional plant-based research activities (at the Kurosaki, Yokkaichi, and Mizushima locations); and it has major overseas subsidiaries and offices in the United States, Europe, Hong Kong, Singapore, Thailand, and China. MCHC had annual sales of ¥26 trillion in 2006, of which 47% were in petrochemicals, 20% in performance products, 15% in functional products, and 12% in health care.

The Mitsubishi Chemical Group Science and Technology Research Center (MCRC) is the corporate research center for MCC. It has approximately 3000 R&D staff, with its main location (comprising approximately 1100 R&D staff) at the Yokohama Research Center, the location of the WTEC visit. The majority of the research conducted by MCRC (~90%) is focused on healthcare and performance materials businesses. Researchers from the polymer research group, located in Yokkaichi, also participated in the WTEC visit (see location information in hosts section above). MCRC has established a number of strategic partnerships with academic institutions, both within Japan and in other countries (specifically, at MIT and UC-Santa Barbara in the United States, Dalian University of Technology in China, and Imperial College in the U.K.)

R&D ACTIVITIES

The MCRC research activities presented to the WTEC visitors were focused in areas of SBES activities.

SBES Process Tools

The first presentation, by Akio Horiguchi, focused on the use of SBES tools to design, construct, and operate the optimum process. In particular, MCRC uses largely commercial computational fluid dynamics (CFD), finite-element method (FEM), and process modeling and simulation (PMS) codes individually and in combination to design and optimize chemical processes and plants. Process design relied extensively in the past on experimental methods (particularly bench-scale and pilot-plant-scale methods) for verification, but as simulation methods have become more sophisticated, there is less reliance on experiment. MCRC primarily uses commercial software (90%) such as Aspen Tech and gPROMS for process-level modeling and STAR-CD and FLUENT for CFD. Examples were given of MCRC researchers coupling software packages to obtain more detailed and reliable models/simulations—for example, modeling a multitube reactor by combining CFD with gPROMS (via in-house Fortran code) to couple flow and reaction. Most of the calculations are performed on a 100-core blade server. MCRC researchers perform sensitivity analysis by performing as many as 100 runs with parametric variation in design variables and other disturbances (such as feedstock purity and temperature). Since five years ago, MCC management has become comfortable with relying on the predictions of simulation studies for the design and optimization of chemical processes. Challenges for this group are the cost of software licenses, availability of sufficient computational resources, and the difficulty of linking software packages that were not designed for interoperability.

Polymer Modeling

The second presentation, by Takeshi Ishikawa and Jun Endo, focused on MCRC's efforts in polymer modeling. The goal of this effort is to be able to relate molecular structure to polymer properties, including rheology, in a methodology that is truly multiscale (i.e., both upscaling from the molecular level to macroscopic properties, and downscaling from desirable/customer-prescribed macroscopic properties to the corresponding chemical architecture). For equilibrium properties (thermodynamics and structure), Jun Endo described MCRC's effort over a three-year period to develop a capability to relate molecular structure and properties, based on a combination of the molecular self-consistent polymer reference interaction site model (SC-PRISM) and self-consistent field (SCF) theories, utilizing the expertise of David Wu at the Colorado School of Mines and Dilip Gersappe at State University of New York at Stony Brook, respectively. This in-house software is unique to MCRC. The effort to develop this capability was the result of MCRC not obtaining the capability it hoped for from the Octa project (<http://octa.jp>). Unlike, for example, molecular dynamics simulations, SC-PRISM and SCF methods are not heavily computational. Much of the rheology modeling, reported by Takeshi Ishikawa, is based on the primitive chain network model developed by Yuichi Masubuchi (<http://masubuchi.jp/>), now at Kyoto University. In collaboration with Kyushu University, MCRC has developed its own CFD/FEM software for polymer process processing, to apply to problems such as twin-screw extrusion and polymer film blowing.

Computational Science Laboratory

The third and final presentation was given by Shinichiro Nakamura describing activities of the Computational Science Laboratory (CSL) of MCRC. The CSL consists of ~17 members (almost all holding doctorates), and it has in-house computing capabilities consisting of CSL-built clusters of various sizes, the largest containing 800 CPUs. CSL also uses the TSUBAME machine (<http://www.gsic.titech.ac.jp>) operated by the Global Scientific Information and Computing Center (GSIC) at the Tokyo Institute of Technology. In the latest Top 500 ranking (<http://www.top500.org>) of computing speeds (based on actual speeds executing the LINPACK benchmark), TSUBAME is ranked 14th overall and first in Asia. Nakamura gave four examples of research accomplishments by the CSL:

1. Design of a robust (nonphotodegrading) dye for use in printing (Kobayashi et al. 2007)
2. Design of a high-quantum-yield yttrium oxysulfide phosphor for use in television sets (Mikami and Oshiyama 1998; 1999; and 2000; Mikami et al. 2002)

3. Finding an effective additive to improve the performance of Li-ion batteries (Wang, Nakamura, Tasaki, and Balbuena 2002; Wang, Nakamura, Ue, and Balbuena 2001)
4. Development of a new methodology for protein NMR (Gao et al. 2007) based on calculating the NMR shift using the fragment molecular orbital (FMO) methodology developed by Kazuo Kitaura for large-scale *ab initio* calculations.

All of the problems described made use of *ab initio* methods, plus additional methods (e.g., neural networks) where needed. In fact, CSL researchers have contributed to community-based *ab initio* codes (Gonze et al. 2002). The CSL evidently performs fundamental, publishable research, as well as proprietary research that directly benefits MCC and MCHC.

CONCLUSIONS

MCRC has a broad portfolio of SBES research, ranging from fundamental efforts publishable in the general scientific and engineering literature, to highly focused proprietary research that directly impacts current and near-future manufacturing activities of MCC and MCHC. Much computing is done in-house with commercial and community-based codes on in-house clusters, although some of the most demanding calculations are performed on external supercomputers. Adapting commercial codes to provide more detailed models through integration of their inputs and outputs is one of the features of the research being performed at MCRC.

REFERENCES

- Gao, Q., S. Yokojima, T. Kohno, T. Ishida, D.G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura. 2007. Ab initio NMR chemical shift calculations on proteins using fragment molecular orbitals with electrostatic environment. *Chemical Physics Letters* 445:331-339.
- Gonze, X., J.M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.Y. Raty, and D.C. Allan. 2002. First-principles computation of material properties: The ABINIT software project. *Computational Materials Science* 25:478-492.
- Ishikawa, T. S. Kihara, and K. Funats. 2000. 3-D Numerical simulations of nonisothermal flow in co-rotating twin screw extruders. *Polymer Engineering and Science* 40:357.
- Ishikawa, T., F. Nagano, T. Kajiwara, and K. Funatsu. 2006. Tip-clearance effect on mixing performance of twin screw extruders. *International Polymer Processing* 11:354.
- Ishikawa, T., T. Amano, S. Kihara, and F. Kazumori. 2002. Flow patterns and mixing mechanisms in the screw mixing element of a co-rotating twin-screw extruder. *Polymer Engineering and Science* 42:925
- Kobayashi, T., M. Shiga, A. Murakami, and S. Nakamura. 2007. Ab initio study of ultrafast photochemical reaction dynamics of phenol blue. *Journal of the American Chemical Society* 129:6405-6424.
- Mikami, M., and A. Oshiyama. 1998. First-principles band-structure calculation of yttrium oxysulfide. *Physical Review B* 57:8939-8944.
- . 1999. First-principles study of intrinsic defects in yttrium oxysulfide. *Physical Review B* 60:1707-1715.
- . 2000. First-principles study of yttrium oxysulfide: Bulk and its defects. *Journal of Luminescence* 87-9:1206-09.
- Mikami, M., S. Nakamura, M. Itoh, K. Nakajima, and T. Shishido. 2002. Lattice dynamics and dielectric properties of yttrium oxysulfide. *Physical Review B* 65:094302-1-4.
- Wang, Y.X., S. Nakamura, K. Tasaki, and P.B. Balbuena. 2002. Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: How does vinylene carbonate play its role as an electrolyte additive? *Journal of the American Chemical Society* 124:4408-4421.
- Wang, Y.X., S. Nakamura, M. Ue, and P.B. Balbuena. 2001. Theoretical studies to understand surface chemistry on carbon anodes for lithium-ion batteries: Reduction mechanisms of ethylene carbonate. *Journal of the American Chemical Society* 123:11708-11718

Site: Nissan Research Center, Fuel Cell Laboratory
1, Natsushima-cho, Yokosuka-shi
Kanagawa 237-8523, Japan

Date Visited: December 7, 2007

WTEC Attendees: D. Nelson (report author), L. Petzold, T. Arsenlis, C. Cooper

Hosts: Dr. Kazuhiko Shinohara, Senior Manager, Fuel Cell Laboratory
Email: k-shino@mail.nissan.co.jp
Dr. Kev Adjemian, Manager Fuel Cell Laboratory
Email: k-adjemian@mail.nissan.co.jp
Dr. Shyam Kocha, Manager Fuel Cell Laboratory
Email: s-kocha@mail.nissan.co.jp
Kazuo Nagashima, Asst. Manager, Fuel Cell Laboratory
Email: kaz-nagashima@mail.nissan.co.jp
Dr. Noboru Yamauchi, Fuel Cell Laboratory
Email: n-yamauchi@mail.nissan.co.jp
Mitsutaka Abe, Fuel Cell Laboratory
Email: mitsu-abe@mail.nissan.co.jp
Yuichiro Tabuchi, Fuel Cell Laboratory

BACKGROUND

Nissan Research Center conducts wide-ranging R&D on vehicle technology. Headquarters of the Research Center are currently in Tokyo but will be moved to Yokohama. This visit focused on the company's research on fuel cells for vehicular applications. Development is done at the Kanagawa technical center for advanced engineering; Nissan's testing grounds are mostly in Kanagawa Prefecture

Nissan is pursuing fuel cell vehicles because the company promises to reduce CO₂ emissions and because renewable fuels can be used to produce the necessary hydrogen. The Nissan power train roadmap includes, in the short-term, high-efficiency internal combustion engines (ICE), and in the mid- and long-term, introduction of hybrid electric vehicles (HEV) and early introduction of electric vehicles (EV) and fuel cell vehicles (FCV.)

Nissan FCV Development Status

FCV Research began in 1996, with the first actual FCV in 2001. Initial fuel-cell stacks were procured from suppliers. In 2004 Nissan introduced its own stack as well as a high-pressure in-vehicle H₂ storage tank using carbon fiber and aluminum to achieve 10K psi. The Nissan FC stack reduces size, weight, and cost compared with the supplier's version. In 2005, the company placed its first FCV on lease. Its cruising range is the same as for a regular car. Subzero temperature startup is not possible. The car uses a Li-ion battery for acceleration augmentation. Nissan and NEC have a joint venture in Li-ion cells.

FCV development issues include performance, cost reduction, durability, and H₂ storage systems. Basic research issues include reaction mechanism, catalyst selection and processing, proton and H₂O transport. Research on new FC materials includes a non-Platinum catalyst and new polymer membranes. Infrastructure issues include H₂ production and distribution. The primary fuel cell target is a polymer electrode membrane, or proton exchange membrane (PEM) type.

Most Nissan research in this area focuses on the membrane-electrode assembly (MEA). Durability improvements include reduced degradation caused by operation modes (start-stop, load cycle, idle) and reduced degradation by environment (temperature, air pollution). A serious problem is that Pt migrates from the electrode into the

membrane, degrading cell performance. Cost reduction focuses on reducing the volume of Pt (by a factor of one-tenth), cost reduction by innovative material substitution, and system simplification.

SBES R&D ACTIVITIES

Modeling Research for Cell Design (Tabuchi)

Modeling research for cell design is aimed at improvements in (1) cost, (2) power density, (3) longer durability, (4) cold start. Nissan researchers are looking at heat and mass transfer under dry and wet conditions; five types of transport are O_2 , H_2 , heat, electrons, and water. Numerical tools are used as thinking tools, not design tools. Numerical techniques include first-principle, MD, Monte-Carlo, and CFD. The key modeling target is MEA, but the present MEA model is not good enough to meet cost-reduction targets. The bottleneck is meso-nanoscale-level modeling. The models, mostly 1D or 2D continuum models, range from component to MEA to cell in size. They believe they don't need more computer power, rather more "thinking power."

The Nissan model of the catalyst layer is derived from that of M. Eikerling, J. Electrochem. Soc. 153. Oxygen diffusivity vs. water saturation percent is based on a model by C.Y. Wang.

An important simulation problem is stack modeling under cold-start conditions. Thermal and water management with transient state is very important for cold-start. Nissan uses a non-isothermal cold-start model. For example, power density is affected by O_2 and H_2O concentration. A successful model must include frost formation

Membrane Modeling (Abe)

Nissan's membrane material is Nafion. Goals are low cost and compactness, high performance, and high durability; proton conductivity is a limiting factor. Understanding of the mass transport mechanism in the PEM fuel cell is the key for a breakthrough. Issues in mass transport include: H^+ conductivity in membrane, H_2O permeability in membrane, H^+ conductivity in channels, volume fraction of channels, H_2O diffusivity in channels, H^+ concentration, and H^+ diffusivity. Nissan's methods for mass transport are pore model MD for short-range diffusion and dissipative particle dynamics (DPD) for long-range diffusion. The water channel in the Nafion membrane includes some tightly bound water, some loosely bound, and some mobile. Problem size: 10 Å cubed box, 100K's particles. Solved on workstation. Bigger box might help, but the problem is knowledge-limited. Charge transport model: proton diffusivity estimated by MD, including proton hopping model. The model consists of dimensional parameters, angle and distance between H_2O and H_3O^+ , fast diffusion and slow diffusion based on different mobility states of water. Challenges of modeling include incorporating longer-range movement using coarse-grained MD and improved DPD.

Electrode Catalyst Issues (Yamauchi)

The primary issues in this area are cost and durability. The principal goal is reduction of Pt usage or substitution of Pt by a cheaper catalyst. Pt durability problems include Pt dissolution from electrode into membrane and corrosion of Pt by carbon. Nissan researchers are investigating these processes experimentally, but information on the chemical reactions is limited. They need to know the reaction path of the Pt electrode reaction, dissolution process of Pt, etc. Computational chemistry may be helpful here but may not be helpful for developing new catalysts. Challenges include the electric potential effect on electrode reactions and the complex structure of the catalyst surface. There are multiple elementary reactions and reaction pathways—at least 33 elementary reactions at one catalyst. Researchers tried to use the Earth Simulator for this, but it did not have enough computer power. Their plan for a computational chemistry approach includes: (1) calculate performance of catalysts on simple electrode reaction and compare results with experimental results, (2) apply computational chemistry to a more complex reaction (cathode reaction.), and (3) try to predict performance of unknown catalysts, including non Pt.

ADDITIONAL DISCUSSION

The WTEC panel's hosts noted that their ideal candidate for a research position would be trained in computer science because fuel cell science can be learned on the job. Their ability to use simulation is limited by lack of fundamental knowledge. They need to work very closely with experimenters to cross-check results. Verification and validation are primarily by comparison with experiment. Future computation will be in computational chemistry. They don't see any other areas for major investment at this time.

CONCLUSIONS

The Nissan fuel-cell research group is mostly doing small simulations on workstations. Modeling is mainly limited by lack of knowledge, not by lack of computer power. Their main target is the membrane-electrode assembly. They are planning a substantial future project: cost reduction and lifetime extension of the Pt catalyst, and substitution of Pt by a cheaper catalyst. They are considering an application of computational chemistry. Models would include rate equations, molecular dynamics, Monte Carlo transport, and dissipative particle dynamics. They plan to collaborate with an outside party to tackle a series of problems: (1) calculate performance of catalysts in a simple electrode reaction and compare results with experimental results, (2) apply computational chemistry to the more complex cathode reaction, and (3) try to predict performance of alternative catalysts, including non-Pt.

Site: Peking University Center for Computational Science and Engineering
Beijing 100871, P.R. China

Date: December 4, 2007

WTEC attendees: M. Head-Gordon (report author), S. Glotzer, S. Kim, J. Warren, P. Westmoreland

Hosts: Prof. Pingwen Zhang, School of Mathematical Sciences
Email: pzhang@pku.edu.cn
Prof. Wenjian Liu, College of Chemistry and Molecular Engineering
Email: liuwj@pku.edu.cn
Prof. Luhua Lai, College of Chemistry and Molecular Engineering
Email: lhlai@pku.edu.cn
Prof. Jingchu Luo, Center for Bioinformatics
Email: luojc@mail.cbi.pku.edu.cn
Prof. Shaoqiang Tang, College of Engineering
Email: maotang@pku.edu.cn
Prof. Jingping Wang, College of Engineering

BACKGROUND

The Center for Computational Science and Engineering was established in 2001 by Peking University to provide a platform for interdisciplinary research related to large-scale scientific computing, the training of students in computational science and engineering, and the provision of high-performance computing resources to university research groups. At present, 29 faculty members have appointments in the center, in addition to their regular departmental appointments. The university gives the center an annual admission quota of 6–8 graduate students; they join the research groups of the affiliated faculty, who also supervise students admitted by their home departments.

SBES RESEARCH

The WTEC team heard a series of short presentations on the research of several faculty members who are participants in the center. A short summary of each is given below to provide some perspective on the range of research currently performed at the center.

- *Computational biology (Prof. Luhua Lai):* In addition to her center affiliation, Prof. Lai is also a member of the Center for Theoretical Biology, which has the study of biological networks as its central theme. Researchers in her group are active in structure-based drug design, in which they develop their own publicly distributed codes. Prof. Lai has also started computational modeling of the aggregation dynamics of amyloid fibrils, using both atomistic models and statistical potentials. Her research group has collaborations with a software company, but it has no direct links with the pharmaceutical industry at present.
- *Computational fluid dynamics in aerospace research (Prof. Jianping Wang):* Prof. Wang had recently returned to China after more than 2 decades in Japan. He focuses on using CFD simulations (with experimental validation) to increase performance and usage of aerospace components, and additionally to shorten the development cycle and reduce risk. Within CFD, he has developed his own CFD approach, the “finite spectral method,” which he is developing into a multiscale approach.
- *Computational quantum chemistry (Prof. Wenjian Liu):* Prof. Liu’s group has developed the Beijing Density Functional program, which treats the electronic structure of heavy elements where relativistic effects are important. Present research efforts are aimed at unifying the treatment of relativistic and non-relativistic electronic structure, with associated opportunities for developing new methods and algorithms.
- *Bioinformatics (Prof. Jingchu Luo):* Prof. Luo summarized recent research on the rice genome.

COMPUTING FACILITIES

The WTEC team was given a tour of the computer center by the system administrator, Mr. Fan Chun, who also discussed its operation with the team. The center houses a teraflop class machine, based on 128 nodes, each with 2 3.2 GHz Intel Xeon processors, and a total of 8 TB of storage. The interconnect is InfiniBand, and the machine was installed in 2004. It provides computing resources for about 20 groups at the university. The machine was purchased with university funds, and costs for administering the machine are also borne by the university. Users are charged 0.6 RMB per CPU hour to cover just the electricity costs of running the machine. There is a need for more energy-efficient computing, using multicore chips, to reduce this cost. The center provides basic commercial software tools, such as Intel compilers and IMSL software libraries, but otherwise users wishing to employ commercial application codes must purchase licenses themselves. At present a significant fraction of the jobs run on the cluster are serial calculations (perhaps 20–30% of capacity), while typical parallel jobs use moderate numbers of processors (8–16). The center is planning to offer short classes on parallel programming via MPI soon, as an alternative to full semester classes in academic departments.

DISCUSSION

Very little question and discussion time was available, due to the WTEC team's tight schedule. However, in the course of the short presentations, a number of interesting issues were briefly raised:

- In the center, there is an emphasis on programming by students, so that they learn to be capable of building their own computational tools and modifying existing ones. Existing programs may be used in cases where existing methods are adequate, while new ideas are the best basis for the construction of new programs. However for training, some groups do make sure that students go through the process of “reinventing the wheel” by making their own programs.
- There is a cross-cutting issue of increasing computer power leading to a research focus on complex realistic systems, treated by very accurate simulation methods with the by-product of possible loss of physical insight. Several groups discussed this issue.
- Prof. Liu discussed the overall progress of computational chemistry in China, stating that while most activity has traditionally been in applications of established tools, there is now a strong young community that appreciates the importance of developing new theory, algorithms, and software. Beyond Peking University itself, he mentioned specifically several strong groups in quantum dynamics, and the activity of the Nanjing and Xiamen groups in quantum chemistry.
- Peking University's Center for Computational Science and Engineering has a strong focus on international workshops and also on summer schools. Its staff members are additionally establishing a strong network of international collaborations, exemplified by a joint center of excellence in quantitative biomedical research with University of California, San Francisco, and computational quantum chemistry tie-ins that are planned with IACS (Kolkata, India), and the Royal Institute (Stockholm).

Site: **Research Institute for Computational Sciences (RICS)**
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba Central 2, Umezono 1-1-1
Tsukuba 305-8568 Japan
<http://unit.aist.go.jp/rics/index-e.html>

Date: December 7, 2007

WTEC Attendees: M. Head-Gordon (report author), S. Glotzer, S. Kim, J. Warren, P. Westmoreland

Hosts: Atsushi Kirita, Manager, International Relations Office, AIST Email: a-kirita@aist.go.jp
Dr. Tamio Ikeshoji, Director, RICS
Email: t.ikeshoji@aist.go.jp
Prof. Kazuo Kitaura, Kyoto Univ. and Prime Senior Researcher, RICS
Email: kazuo.kitaura@aist.go.jp
Dr. Dmitri Fedorov, Senior Researcher, AIST
Email: d.g.fedorov@aist.go.jp
Dr. Seiji Tsuzuki, Senior Researcher, AIST
Email: s.tsuzuki@aist.go.jp
Dr. Eiji Tsuchida, Researcher, AIST
Email: eiji.tsuchida@aist.go.jp

BACKGROUND

The Research Institute for Computational Sciences (RICS) is one of approximately 20 institutes that form part of AIST, which was founded in 2001 as part of a major reorganization of Japanese government research laboratories. The institutes are long-term entities that coexist with roughly 30 research centers that have fixed 7-year lifetimes. RICS has 29 permanent staff and 13 postdoctoral researchers, who are divided into 5 major groups that focus on quantum modeling, particle modeling, first principles simulations, fundamental analysis, and finally, other topics. Their focus is on simulation at the microscopic rather than the macroscopic level.

RESEARCH

During a 3-hour morning meeting, the WTEC visiting panel heard an overview of RICS research from the Director, Dr. Ikeshoji, followed by technical presentations from Drs. Fedorov, Tsuzuki, and Tsuchida. Summaries of these presentations are given below.

General Aspects

RICS researchers have 3 primary missions: (1) development of computational methodologies and programs, (2) their applications to real systems (biological and nanoscale primarily), and (3) collaborations both inside and outside AIST. As a result, a significant number of programs are developed within RICS, including for the FMO method described by Fedorov below, FEMTEK described below by Tsuchida, and other codes for classical molecular dynamics, first principles molecular dynamics, and quantum simulation of materials. RICS researchers are participants in significant national research projects such as the Next Generation Supercomputing Project, including the fragment molecular orbital (FMO) project described below by Fedorov, a hydrogen storage initiative, and two nanoelectronics projects. Materials applications include calculations on proton transfer in fuel cells, aerosol deposition, and fracture and fatigue using MD.

Fragment Molecular Orbital Calculations

Dr. Fedorov described the basics of the FMO approach, as well as the latest FMO developments (Fedorov and Kitaura 2007). FMO calculations on a large biomolecule are a simplified divide-and-conquer approach in which

calculations are performed on individual residues in the Coulomb field of others, followed by corrections based on calculations on pairs of residues (and, in principle, triples of residues, if needed). The approach is accurate and very efficient, as long as the division of the system is wisely chosen. FMO calculations are manifestly parallel, although there are challenges associated with load balancing. The FMO method was recognized with the Best Technical Paper Award at Supercomputing 05 (the 2005 meeting of the International Conference for High-Performance Computing, Networking, Storage, and Analysis). FMO methods are currently interfaced to the GAMESS program (which is a public domain U.S. code).

Intermolecular Interactions by High-Level Calculations

Dr. Tsuzuki described high-level electronic structure calculations for the interactions of aromatic groups with themselves and also with CH groups (Tsuzuki and Uchimaru 2006). Additionally, he discussed similar calculations of the interactions between the component ions of ionic liquids. These calculations are then used as inputs for the development of force fields that may be used for dynamics.

Linear Scaling DFT Calculations Using FEM

Dr. Tsuchida described the development of a new finite element method (FEM) for density functional theory (DFT), based on tensor product cubic polynomial basis functions (Tsuchida 2007). There are 4 key advances that make this approach promising: (1) the development of adaptive curvilinear coordinates that cause the mesh to deform as the atoms in the molecule move, (2) a fast iterative Poisson solver that uses a multigrid preconditioner, (3) a quasi-Newton solver for the self-consistent field problem, combined with a non-diagonal preconditioner, and (4) a new linear scaling approach based on orbital minimization with additional constraints to ensure linear dependence. The cross-overs for this method appear very competitive with rival methods. The resulting code (FEMTEK) is currently being applied to a first-principles simulation of the dynamics of liquid ethanol.

COMPUTING HARDWARE

The main supercomputer is the “Supercluster,” which at the moment contains a 2144 CPU AMD Opteron cluster that performs at about 6 TFlops, and 2 smaller components (528 processor 1.3 GHz Itanium2, and 536 processor 3.06 GHz Intel Xeon). In 2008 RICS will purchase about 3 more small clusters consisting of about 200 CPUs, each performing in the teraflop range.

DISCUSSION

During the talks there was good opportunity for discussion of both scientific issues directly related to the content of the talks, and also more general issues surrounding simulation-based engineering and science. Some of the additional issues that came up included the following:

- Potential postdocs with appropriate training in the development of algorithms and programs are hard to recruit. The main cause of this issue is that too many graduate students are being trained primarily to run existing codes to solve applied problems rather than learning the skills necessary to create new applications.
- The development of codes within RICS is not always directed at achieving performance that exceeds rival public domain or commercial codes, but is also done to permit quick prototyping of new ideas and new functionality. In other cases, such as the tie-ins to the next-generation supercomputer project, very high (parallel) performance is the target.
- National and international collaborations of RICS researchers are increasing significantly. This is key partly as a mechanism to raise additional research funds, and it is increasingly also seen as part of the mission of the institute. There is also an active program of annual workshops on computational science, some international or regional in scope and others directly primarily at Japanese participants.

REFERENCES

- Fedorov, D.G., and K. Kitaura. 2007. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* 111:6904-6914.
- Tsuchida, E. 2007. Augmented orbital minimization method for linear scaling electronic structure calculations. *J. Phys. Soc. Jap.* 76(3) :034708-1–034708-7.
- Tsuzuki, S., and T. Uchimaru. 2006. Magnitude and physical origin of intermolecular interactions of aromatic molecules: Recent progress of computational studies. *Curr. Org. Chem.* 10(7):745-762.

- Site:** **RIKEN – The Institute of Physical and Chemical Research**
Advanced Center for Computing and Communication (ACCC)
2-1, Hirosawa, Wako-shi
Saitama, 351-0198, Japan
<http://www.riken.jp/engn/index.html>
http://accc.riken.jp/E/index_e.html
- Date Visited:** December 3, 2007
- WTEC Attendees:** G. Karniadakis (report author), P. Cummings, L. Petzold, T. Arsenlis, C. Cooper, D. Nelson
- Hosts:** Dr. Ryutaro Himeno, Director and Senior Scientist, ACCC; Director, Development Group, NGSC R&D Center
 Email: himeno@riken.jp
 Dr. Takayuki Shigetani, Senior Technical Scientist, ACCC
 Dr. Makoto Taiji, Deputy Project Director, Computational and Experimental Systems Biology Group
 Dr. Toshiaki Iitaka, Senior Scientist, Computational Astrophysics Laboratory

BACKGROUND

RIKEN is an independent administrative institution under the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) since 2003. It was originally founded as the first private research foundation in 1917; it was reorganized as a public corporation in 1958 under the *Rikagaku Kenkyusho* Law. Today, RIKEN carries out comprehensive research in science and technology in all fields except social sciences. It has seven campuses in Japan and five campuses abroad (2 in the United States, 1 in the UK, and 1 in Singapore). Its activities focus on brain sciences, accelerator-based research, developmental biology, biomimetic research, terahertz-wave research, genomics, research on allergy and immunology, bioresources, etc. The total funding in FY2007 was ¥89.4 billion (~US\$805 million¹⁴) with about 3,000 researchers, about 1800 visiting Japanese scientists, and 685 visitors from abroad.

The Advanced Center for Computing & Communication (ACCC) is part of RIKEN and provides RIKEN researchers with computer resources and network services. Currently, it operates the RIKEN Super Combined Cluster (RSCC), which includes a mix of computer architectures, scalar, vector and accelerators. In particular, RSCC is composed of three subsystems, five Linux clusters with Xeon 3.06 GHz processors (512 dual nodes and 128 dual nodes x 4), a vector parallel computer (NEC SX-7/32) and an MD-GRAPE3 board (64Tflops). The RIKEN Principal Investigators proposed this hybrid concept to MEXT to develop the Next Generation of Super Computer (NGSC).

COMPUTING FACILITIES

NGSC: Japan's 10 Petaflop Supercomputer

The RIKEN researchers based their proposed design for Japan's next-generation (10 petaflop) supercomputer on the need for multiscale and multiphysics simulation in life science and nanoscience, which, in turn, involves multiple computation components; hence, it requires the interaction of multiple architectures. To this end, they initially proposed a tightly coupled heterogeneous computer consisting of vector, scalar, and MD nodes with a fast interconnect between the heterogeneous nodes. The final design, however, does not have any accelerators.

¹⁴ Dec. 31, 2007 exchange rate of ¥110.88/US\$1.

NGSC Goals: The objective of the NGSC project is the development, installation, and application of an advanced high-performance supercomputer system as one of Japan's "Key Technologies of National Importance." The cost of the project is ¥115 billion (~US\$1.04 billion), and the funding period is FY2006-2012. In particular, the specific goals of the NGSC project are (1) Development and installation of the most advanced high-performance supercomputer system as a national infrastructure component. (2) Development and wide use of application software to utilize the supercomputer to the maximum extent. (3) Provision of a flexible computing environment by sharing the NGSC through connection with other supercomputers located at universities and other research institutes. The WTEC panel members discussed with our hosts the possibility of middleware development as well, but the directors of the project stated that Japan's needs are best accommodated by focusing on application software, while any other required middleware (including compilers) will be obtained from abroad (e.g., the United States), or development of such software can be addressed at a later stage. Specifically, the development of middleware for Grid Computing will be continued at the National Institute for Information (NII). The development of compilers is also conducted by another project. The NGSC will not be involved in the development of the middleware or compilers but will focus on the development of applications software for nanosciences and life sciences. Although research on parallel language is conducted in Japan, it is not part of the NGSC project.

Collaboration with Earth Simulator: It is interesting to compare the NGSC project with Japan's Earth Simulator (ES) project, currently in the sixth year of its operation. ES belongs to the Japan Agency for Marine-Earth Science & Technology (JAMSTEC), which has its own research area; ES is not fully opened to everyone. In contrast, RIKEN will develop and operate NGSC as a national infrastructure facility, with users selected by an independent scientific committee. JAMSTEC had submitted a similar proposal to MEXT for hybrid scalar-vector architecture, but this proposal was rejected in favor of the RIKEN proposal. RIKEN has entered into a partnership with the JAMSTEC in order to make use of the expertise JAMSTEC has gained from running the Earth Simulator. Under their agreement, RIKEN and JAMSTEC will jointly develop application software for the Next-Generation Supercomputer and collaborate in numerous other ways.

Policies: The Office of Supercomputer Development planning of MEXT is responsible for the policies and funding of the NGSC project. The project is coordinated by a project committee with external input from an advisory board. The project leader is Dr. Tadashi Watanabe; the group director of research and development is Dr. Ryutaro Himeno. The three main participants of the project are (1) the RIKEN Wako Institute (focus on life sciences); (2) the National Institute of Informatics (focus on grid middleware and infrastructure); and (3) the Institute of Molecular Science (focus on nanoscience simulation). There is also participation by visiting professors from universities and other national laboratories. In addition, the project has formed a partnership with computer companies and the Industrial Forum for Supercomputing Promotion. The evaluation scheme consists of evaluation committees within MEXT and the Council for Science and Technology Policy (CSTP).

MEXT leadership is committed to sustaining continuous development of supercomputers in Japan. Specifically, MEXT's policy stems from the belief that by computational science, Japan's competitiveness in science and technology can be maintained in a world-leading position and that successive development of supercomputers will be realized by maintaining the necessary technology inside Japan. To this end, the first priority is sustained performance in key applications at the level of at least ten percent of peak performance. The end result of such successive development of supercomputers will be the enhancement of information technology with novel new hardware (e.g., low-power CPUs) but at a reduced cost.

Target Applications and Benchmark Suite: The RIKEN team started selecting applications for candidates in the benchmark suite in January 2006, and at this point, 21 applications have been selected, in the areas of life sciences (6), nanosciences (6), astronomy (2), geophysics (3) and engineering (4). A subset of seven applications will be optimized in the immediate future, with the objective of sustaining at least 1 Petaflop on NGSC (10% of peak performance). These applications include (1) prediction of protein structure, (2) molecular orbital calculations using GAMESS/FMO, (3) multipurpose MD simulations, (4) *ab initio* MD calculation in real space, (5) lattice QCD simulations in studying elementary physics and nuclear physics, (6) atmospheric modeling for global-cloud simulation, and (7) compressible flow simulations around an aircraft and a spacecraft. The targeted

performance numbers are 10 Petaflops for LINPACK and 1 Petaflop for each of the seven selected applications of the benchmark suite.

Roadmap of the Project: The final design was selected by MEXT based on benchmarks on the existing system comprising vector, scalar, and accelerator platforms. The approved design consists only of scalar and vector processors units. It will require 30 MW (including cooling), and its footprint will be 3200 square meters. The computer will be built in the Kobe campus by Hitachi and NEC (vector units) and Fujitsu (scalar units). The panel inquired on the working relationship between the three companies and how closely they can collaborate and share information, but it is not clear at this juncture how close this collaboration will be. The hardware will be in production at the start of 2009 and will be completed by the end of 2010. The system software will be completed by the end of 2008 and will be evaluated by the end of 2010. On the application side, the nanoscience applications will be evaluated by the year 2010, and the life science applications will be evaluated between the years 2011 and 2012. While the latter seems reasonable, the former seems rather unrealistic, since NGSC will not be available during the years 2009-2010 so it is not yet clear on what machine such applications will be evaluated. The panel visited the Institute of Molecular Science (IMS) responsible for the nanoscience applications (see IMS site report) but no further information was obtained on this issue.

Life Science Focus of RIKEN: IMS will focus on nanoscience with three main themes: (1) next-generation energy, (2) next-generation nanoscale biomolecules, and (3) next-generation nanoscale informational materials. Details of these topics are covered elsewhere in this report. Here, we elaborate on the life science core projects of RIKEN. They center on the multiscale modeling of the virtual human, from genes to proteins, to cells, to tissues and organs. This is consistent with the international effort on the Physiome project (<http://www.physiome.org>) and the Europhysiome project (<http://www.europhysiome.org>). Apparently, at this point the RIKEN team will pursue the different scales separately with the long-term objective of coupling all scales. Currently being pursued is a detailed scanning of a live human at a resolution of 1 mm with the objective of achieving 0.3 mm in collaboration with the University of Tokyo. Similar data exist in the United States at the National Library of Medicine, but the data are not for a living human. The problem of data analysis and of data-driven simulation was also addressed in the RIKEN presentation; data assimilation of medical data was identified as an important new topic. New proposals will be solicited by RIKEN next year to address such issues.

CONCLUSIONS

At the time of the WTEC panel visit, the RIKEN team had finalized the detailed hardware design of the 10 Petaflops NGSC in Japan, but no details were given us regarding the systems software, which is apparently the responsibility of the participating companies (Fujitsu, Hitachi, and NEC). The National Institute of Informatics, a partner in the NGSC project, has proposed a grid infrastructure that will connect the new supercomputer to other existing supercomputers. Unlike the Earth Simulator, the new computer will be open to all, with 10% of the total time allocated to use by universities. The new supercomputer center at RIKEN will also serve as an HPC education center and will offer a series of lectures on computer-science-related issues as well as on applications. The staff in center will work closely with universities in trying to affect the curricula in computational science for petaflop scale algorithms, tools, and applications.

The budget of the NGSC project includes funding for applications on life sciences at RIKEN and nanosciences at IMS. There are no plans for validation at this point, although input from experimentalists will be sought for future validations. The PIs of the project appreciate the importance and the difficulties of validating simulations of such complex systems, like the living cell. The emphasis is on applications software, which will be open source, available to all—"There are no borders in science," Dr. Himeno told us. However, no systematic efforts will be undertaken at this point for middleware development, because the Japanese government does not have strong motivation to develop software. The RIKEN PIs see the development of compilers, for example, as a next step after the demonstration of Petaflop performance in the key seven applications selected.

Finally, MEXT is committed to continuous development of supercomputers to advance computational science for competitiveness in R&D and for developing novel low-power CPU components that will eventually find their way to accommodating the increased needs of information technology for consumers.

Site: Shanghai Supercomputer Center
585 Guoshoujing Road
Shanghai Zhangjiang Hi-Tech Park
Shanghai 201203 P.R. China
<http://www.ssc.net.cn/en/index.asp>

Date Visited: December 6, 2007.

WTEC Attendees: M. Head-Gordon (report author) and S. Glotzer

Hosts: Jun Yuan, Vice Director
Tel.: +86 21 50270953; Fax: +86 21 50801265
Email: jyuan@ssc.net.cn
Bo Liu, Scientific Computing Specialist
Dr. Tao Wang, Vice-Manager, Scientific Computing Department
Jiancheng Wu, Vice-Manager, Engineering Computing Department
Ether Zhang, Marketing and International Collaboration Manager

BACKGROUND

The Shanghai Supercomputer Center (SSC) is China's first supercomputer center that accepts applications from the general public. It opened in December 2000 as the result of an initiative of the Shanghai Municipal Government. Faced with the request to acquire a supercomputer for weather-forecasting purposes, local leaders reasoned that the purchase of such a major piece of infrastructure should be used to benefit as broad a range of society as possible. As a result, the first general purpose supercomputer center in China was established. Today it serves some 270 groups and roughly 2000 users from many areas of scientific and engineering research, both in universities and industries. Seventy staff members at present manage the functioning of the center, which is housed in a dedicated building in the Zhangjiang Hi-Tech Park.

The visiting panel met with 5 members of the SSC for approximately 2 hours, during which Vice Director Yuan gave a presentation and there was much discussion in which the three SSC technical representatives also actively participated. The main points covered in the presentation and discussion are summarized below.

R&D/SERVICES

Computing Hardware

The main supercomputer, installed in 2004, is a 532-node Dawning 4000A system, with peak observed performance of 10 TFlops, containing 4 processors (2.4 GHz AMD Opteron 850) per node. Aggregate memory is 4 TBytes, with aggregate disk storage of 95 TBytes, controlled by 16 storage nodes, each with 4 processors. The system interconnect is Myrinet. This machine was designed and assembled in China and represented an investment of approximately 100 million RMB.¹⁵ Upon installation, it was ranked 10th in the "Top 500" list (June 2004). Today it is ranked 3rd within China itself, of the machines for which information is available (the top machine is for meteorology, and the 2nd machine is for oil industry modeling). CPU usage shortly after installation was in the 40% range, but it was over 80% for all of 2006 and 2007.

Future hardware plans call for the installation of a new Chinese-sourced machine by the end of 2008, which will yield peak performance of over 200 TFlops. This machine will be 50% financed by the national government through a large grant from the "863 program." In the longer run, perhaps in 2011 or 2012, a larger machine is anticipated under the 863 program.

¹⁵ Ren Min Bi, this is ~\$135,208,220 at 7.3960 RMB/US\$1, the conversion rate on December 31, 2007.

User Community

The roughly 2000 users are diverse in background and interests. Approximately 53% are from natural science research institutes and universities and consume about 80% of the computer time; approximately 47% are from industry and consume about 20% of the computer time. Most users employ the center for capacity computing (many relatively small jobs), while only a few employ it for very large massively parallel jobs. This is believed to be more due to limitations on the available resources than to intrinsic user demand. Thus, approximately 75% of recent jobs employ 1–4 CPUs, consuming about 20% of the computer time, with roughly similar proportionate CPU use for 5–8, 9–16, and 16–32 CPUs. Most users are not expert simulators and are driven by the need to solve particular applications problems.

Highlights of the results produced at the SSC include

1. Modeling the aerodynamics of the first commercial regional jet produced in China (about 10 million cells)
2. 3D nonlinear analysis of the behavior of a tunnel in an earthquake
3. Flood prediction
4. Specific drug design accomplishments (related to the work discussed separately at Dalian University)
5. First-principles atomistic modeling for a wide range of problems in materials, science and condensed matter physics.

Funding Model

The long-term vision is to be a profitable high-performance computing (HPC) applications service provider (HPC-ASP). This situation is far from a reality at present, because the Shanghai Supercomputer Center is a trail-blazer in China and must first help to create the market for such services, and also because SSC now provides most of its computing resources to the nonprofit natural science research communities. At the moment, the SSC provides users with computer resources and services in exchange for money that covers approximately 30% of operational costs.

Organization and Staff Activities

The technical staff members are organized into 5 technical departments: (1) scientific computing, (2) engineering computing, (3) research and development, (4) technical support, and (5) IT (responsible for networking, PC and server management, and internal information management). Their activities are wide-ranging, including computing infrastructure maintenance, direct support of users, porting and tuning software, providing training to small groups of users, and some software development. The latter can either be independent or in partnership with universities, government, or industry. Center management finds it challenging to recruit good people with appropriate experience, partly because of competitive pressure from corporations. Other possible reasons include the fact that the history of HPC development and application is relatively short in China, and not every university has been prepared to give HPC education to graduates. Therefore, the people the SSC hires are often experts in application areas who must learn about high-performance computing on the job.

U.S. Export Restrictions

In the view of the WTEC team's SSC hosts, HPC is an international activity. However, employees of the SSC are subject to entry restrictions to the United States, which prevents them from visiting U.S. unclassified supercomputing centers or even participating in supercomputing conferences or visiting IT companies and universities in the United States. There are no such restrictions with regard to visiting Japan or Europe. In light of this situation, the WTEC visiting team was grateful for the access and information that our hosts at the Shanghai Supercomputer Center provided us.

Site: Shanghai University
Shanghai, 200072 P.R. China
<http://www.shu.edu.cn/en/indexEn.htm>

Date of Visit: September 6, 2007.

WTEC Attendees: S. Kim (report author), J. Warren, P. Westmoreland, and G. Hane

Hosts: Prof. Qijie Zhai, Assistant President, Shanghai University
Director, Center for Advanced Solidification Technology
Email: qjzhai@mail.shu.edu.cn

Prof. Wu Zhang, High Performance Computing Center (HPCC),
Executive Dean, School of Computer Science and Engineering (CSE),
Shanghai University
Email: wzhang@mail.shu.edu.cn

Changjiang Chair Prof. YueHong Qian,
Institute of Applied Mathematics & Mechanics
Email: qian@shu.edu.cn

Prof. Li Lin, Section Head of Metal Materials, Institute of Material Science & Engineering
Prof. Jinwu Qian, Dean, Sino-European School of Technology

BACKGROUND

Shanghai University (SHU) was founded in 1994 as a merger of four universities in Shanghai, namely, Shanghai University of Technology, Shanghai University of Science & Technology, Shanghai Institute of Science & Technology, and the former Shanghai University. Shanghai University of Science & Technology was established in 1958 based on Chinese Academy of Sciences (CAS), East China Branch. Therefore, it enjoys close ties with the CAS, the President and deans of that university used to be the directors CAS, East China Branch. Today, Shanghai University has a faculty of 3,400, among whom 11 are Academicians of CAS and Chinese Academy of Engineering (CAE), 444 full-time professors and 881 Associate professors. The number of students is 34,000, including 8300 graduates. Research funding of the university is more 500 Million RMB in 2008, ranked top 20 in China.

SBES RESEARCH

An overview of SBES research activities was delivered as presentations (our itinerary allowed only two hours at this site) by selected faculty leaders in the engineering and computer science departments. The most notable and helpful report was the tour de force presentation by Prof. Wu Zhang of the HPCC and the School of CSE with 68-slide PowerPoint presentation delivered within 30 minutes! The size and scale of HPC at this site suggest that our brief visit just barely touch the highlights of the research activities here at the SHU campus. The talks also brought out the high quality of international collaborations in applied research, most notably with industrial firms in both United States and Europe.

Overview of HPC Research and Education at SHU

Prof. Wu Zhang (Computer Science and Engineering) gave an overview of HPC research and education at SHU; after an introduction to the history of HPC at Shanghai University, he covered three topics: infrastructure, algorithms, and applications.

- *History and Background:* SHU has a long history in HPC under the leadship of SHU President and CAS Academician Weichang Qian. The role of simulation and modeling is appreciated not only in science and technology but social sciences, fine arts, economics and management. There are five university-level centers involved in HPC: (1) HPCC; (2) Center for Advanced Computing; (3) Center of CIMS; (4) Center of Multimedia (Visualization) and (5) E-Institution of Grid Technology. Significant HPC activities are found in

the School of Computer Science and Engineering, the College of Science and the Colleges of Engineering, which in this talk was nicely grouped into three slides: numerical methods (Lattice Boltzmann Method, Wavelet and Spectral Methods, Mesh-free Method); computational physics and chemistry (Transport-Diffusion-Reaction Modeling, Geophysics, Meteorology, Earthquake Modeling); and life science and material sciences (Bioinformatics, Control Theory, Material Simulations with grid and parallel computing). SHU is also active in hosting international conferences in HPC.

- *HPC Infrastructure, Past, Present and Future*: The 450GFlops ZQ-2000, built in 1999, was the first cluster at SHU. It has 218 processors (109 nodes), 26GB memory totally, and is connected with PIII 800/Myrinet switches. This system was eclipsed by the 2.15TFlops ZQ-3000 with 392 GB of total memory and 352 (3.06GHz Xeon) processors connected by Infiniband. The ZQ-3000's LINPACK benchmark is 1.51TFlops. For grid computing, SHU/CSE has embraced Open Grid Services Architecture (OGSA), an evolving standard managed by the Global Grid Forum and the Web Service Resource Framework (WSRF) protocol and middleware of Globus Toolkit (GT4). In terms of future infrastructure, there is planned activity and research on optical computing.
- *Algorithms*: There are significant algorithm activities at SHU. Given the time constraints, the talk focused on some key projects in numerical linear algebra and its impact on parallelization of CFD codes. Their researchers' hybrid combination of highly scalable PPD (Parallel Partitioning Diagonal) with K.V. Fernando's Burn at Both Ends (BABE, NAG 1996) factorization forms the basis of their tridiagonal solvers. Achievements are illustrated with several benchmark problems in CFD.
- *Applications*: Dr. Zhang matched the list of HPC activities with departments and schools. A notable example highlighted in his talk is CFD/aircraft design. Many other examples ranged from chip cooling design to wind flow around power plant environments. Examples were summarized quickly due to time constraints.

Other HPC Activities

- Prof. Li Lin presented his CALPHAD application to metallic systems. This work is in close collaboration with top practitioners and automotive industrial firms in Europe and the United States.
- Dr. Z. M. Lu (research interests are in turbulent models and particle transport in air passages of the lunge) shared a document that gave an overview of the SBES research in Computational Mechanics at the Shanghai Institute of Applied Mathematics and Mechanics, featuring the research projects of Professors. Yuehong Qian, Peifeng Weng, Peng Zhang, Wei Feng and Yuming Cheng, whose foci are summarized below:
 - Prof. Yuehong Qian: Lattice-Boltzmann method (LBM) and applications; hypersonic flow simulation; particle flow simulation; nano-fluid simulation; Large Eddy Simulation (LES) with recent variants of the Smagorisky models; LES-LBM implemented with the 19-velocity (D3Q19) lattice model.
 - Prof. Weifeng Chen: Numerical simulation of low-Reynolds number flow with applications to micro-craft flow control; B-B turbulence model, mixed method of LES and RANS for separated flows; aerodynamics of low aspect ratio vehicles; numerical simulations of unsteady viscous flow around rotating helicopter wings.
 - Prof. Peng Zhang: CFD (theory of characteristics, the Riemann problem, simulation of hyperbolic conservation laws) and simulation of traffic flow by high-order model equations
 - Prof. Yuehong Qian: Lattice-Boltzmann method (LBM) and applications; hypersonic flow simulation; particle flow simulation; nano-fluid flow simulation; LES (large eddy simulation) with recent variants of the Smagorisky models; LES-LBM implemented with the 19-velocity (D3Q19) lattice model
 - Prof. Wei Feng and Prof. Yuming Chen: Finite element method (hybrid elements); boundary element method, mathematical theory of meshless methods; nanomechanics (molecular dynamics and the effect of vacancy defects and impurities on elastic moduli); applications to elasticity, elastodynamics, and fracture

COMPUTING FACILITIES

The facility details are as described in Prof. Wu Zhang's discussion of HPC infrastructure.

DISCUSSION

A good part of our discussion focused on SBES education:

- A course has been developed to instruct parallel programming to graduate students in computer science, but the course is considered too advanced for undergraduate students, e.g. non-CS undergraduate science and engineering majors. The course text is from the University of Minnesota (written by Dr. V. Kumar et al. and translated by Dr. Wu Zhang et al).
- A short course has been made available, as needed, for non-CS students and researchers interested in HPC resources and HPC usage.
- There is a biweekly seminar series on high-tech algorithms.
- There is frustration over the disparity of funding for software vs. hardware, and most software packages were viewed as too expensive.

CONCLUSIONS

SHU has comprehensive resources that span the entire spectrum of HPC resources from machines to algorithmic activity.

REFERENCES

- W. Cai and W. Zhang. 1998. An adaptive SW-ADI method for 2-D reaction diffusion equations. *J. Comput. Phys.* 139(1):92-126.
- X.-H. Sun and W. Zhang. 2004. A parallel two-level hybrid method for tridiagonal systems and its application to fast Poisson solvers. *IEEE Trans. Parallel and Distributed Systems.* 15(2):97-106.
- W. Zhang, Z. Chen, R. Glowinski and W. Tong eds., 2005. *Current Trends in High Performance Computing and its Applications*, Berlin: Springer-Verlag.
- D. Zhou, W. Cai and W. Zhang. 1999. An adaptive wavelet method for nonlinear circuit simulation. *IEEE Trans. Circuits 23 and Systems I. Fundamental Theory and Appls.* 46(8): 931-939.

COMMENTS FROM DR. QI-JIE ZHAI, PROFESSOR OF MATERIAL SCIENCE, ASSISTANT PRESIDENT, SHU

The report, *Simulation Based on Engineering Science*, released by NSF in 2006, was very impressive. However, it focuses more on computer simulation than on experimental simulation, which might present current American, even global, academic trends in simulation application. Indeed, the deep understanding on internal mechanisms of simulated problems is fundamental to numerical simulation, while computing only contributes technically to it. We have undoubtedly a long way to go, especially on fields of nonequilibrium, dissipation and multiscale complexity, before our knowledge and realization about many problems could greatly satisfy simulation conditions, though we have already achieved progress in some special cases.

With regard to grand challenge problems, simulation would be the most efficient way. I like to say that numerical or so-called computer simulation should work together with experimental simulation. Unfortunately, there exists obvious divarication between academic and industry. The latter, in fact, pays little attention to numerical simulations in China. For instance, so far we have experimentally simulated the solidification process of tons of steel by only hundreds grams of steel with an experimental simulation device. When enterprises have benefited from our equipment for cost savings and time reduction, they have seemed to ignore the fact that successful experimental simulation partly relied on numerical simulation.

Of course, the limitations of effective experimental simulation appear to be related to its diversification and customization compared with numerical simulation. For this reason, experimental simulation is comparatively difficult to develop and requires high attention from researchers.

Consequently, in order to develop simulation, experimental simulation should be encouraged while developing numerical simulation. As two wheels of one bike, the two sides of simulation need persuading from each other and must develop coordinately.

R&D ACTIVITIES

The PAYAO Web 2.0 Community Tagging System is being developed for the tagging of SBML models. CellDesigner is the gateway, and simulation technology is the back end. This group is focusing on the gateway. Interaction with the system is via a control panel, SBW menu, or by saving to another other format to run the simulation, for example MATLAB or Mathematica. The System can register models, restrict access to all or part of a model, search the tags, tag the models, and link to PubMed IDs. It provides a unified gateway to the biological information network. Dr. Kitano is talking to publishers about sending personalized collections of tags. The alpha release of this system is currently underway.

Why the investment in software infrastructure? Dr. Kitano believes that software is critical to the development of systems biology as a field. Recognizing that it is difficult to publish software, the merit system in this lab values software contributions as well as publications. Funding for the software comes from a Japan Science and Technology and the New Energy Development Organization (NEDO) grant for international software standards formation, which is also funding efforts at Caltech and EBI. There is also funding from various projects in the Ministry of Education, Sports, Culture, Science, and Technology (MEXT).

Computational cellular dynamics efforts are also focusing on techniques for obtaining the data needed to develop and validate the computational models. In particular, microfluidics is used as a way to better control conditions to grow yeast.

Key issues in cellular architecture that are being investigated include the implications of robustness vs. fragility on cellular architecture. Some systems have to be unstable to be robust (for example, cancer). Points of fragility in a system are important to the determination of drug effectiveness. The robustness profile is used to reveal principles of cellular robustness, to refine computer models, and to find therapeutic targets. Yeast (budding yeast and fission yeast) is used as the model organism, and cell cycle is used as the model system. The models are currently ODE models. The limits of parameters are used as indicators of robustness, for example, how much can you increase or decrease parameters without the disrupting the cell cycle? This is a different type of uncertainty analysis than most of the current work in that area in the United States or elsewhere.

Experimental methods are required that can comprehensively and quantitatively measure parameter limits. SBI researchers have developed Genetic Tug of War (gTOW), an experimental method to measure cellular robustness. The gTOW method introduces specially designed plasmid with genes of interest that amplifies itself during growth to see how much it changes the cell cycle. There are implications of this technology for drug development.

The WTEC team's hosts at SBI observed that the problems that will need the RIKEN supercomputer or equivalent computers are likely to be systems biology for drug discovery and docking simulations; redundancy in biological systems, which makes drug targeting difficult; and combinatorial targeting, which will be very computationally intensive.

The WTEC team's hosts also addressed the barriers in systems biology. They noted that researchers in this field still don't have a good methodology for inferring the network from the data. Experimentalists need to understand the needs and limits of computation. Computationalists need to understand what experimentalists can do. The significant questions include how can we mathematically characterize what kinds of experiments can actually be done? Can experiments be created with the computational needs and capabilities in mind? Important milestones will include 3D models of cells, heterogeneous multiscale models, the effects of confined space, and parametric uncertainty with thermodynamic constraints.

CONCLUSIONS

The work of this pioneering group at SBI is characterized by a close coupling of computation with experiment. It has a strong focus on the development of the computational infrastructure for systems biology and on the experimental techniques that will be needed to produce the data that will interact with the computation.

With regard to where will the systems biologists of the future come from, Dr. Kitano believes the best educational background may be an undergraduate degree in physics, masters in computer science, and PhD in molecular biology. A computer science background would work if the student is willing to learn the experimental side.

This group has now been funded for 10 years. It was initially funded on a noncompetitive, nonrenewable 5-year grant, which has been renewed once. This is different from current funding models in the United States. A source of stable, sizeable funding has undoubtedly played a role in the world-class success of this research.

REFERENCES

- Hamahashi S., S. Onami, and H. Kitano. 2005. Detection of nuclei in 4D Nomarski DIC microscope images of early *Caenorhabditis elegans* embryos using local image entropy and object tracking. *BMC Bioinformatics* (May) 24:6(1):125.
- Hucka, M., A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, et al. 2003. The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524-531.
- Kitano, H. 2007. A robustness-based approach to systems-oriented drug design. *Nature Reviews Drug Discovery* 6:202–210 (March). Doi :10.1038/nrd2195.
- . 2000. Perspectives on systems biology. *New Generation Computing Journal*. 18:199–216. Ohmsha, Ltd., and Springer-Verlag, 99-216.
- . 2002a. Computational systems biology. *Nature* 420:206–210.
- . 2002b. Systems biology: A brief overview. *Science* 295:1662–1664.
- . 2004a. Biological robustness. *Nature Review Genetics* 5:826–837.
- . 2004b. Cancer as a robust system: Implications for anticancer therapy. *Nature Reviews Cancer* 4(3):227–235.
- . 2006. Computational cellular dynamics: A network-physics integral. *Nature Reviews Molecular Cell Biology* 7:163.
- . 2007. Towards a theory of biological robustness. *Molecular Systems Biology* 3:137. Doi:10.1038/msb4100179. Published online 18 September.
- Kitano, H., and K. Oda. 2006. Robustness trade-offs and host–microbial symbiosis in the immune system. *Mol. Syst. Biol.* 2(1): msb4100039-E1 (Jan. 17),.
- Kitano, H., et al. 2005. Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology* 23(8):961–966.
- Kyoda, K., and H. Kitano. 1999. Simulation of genetic interaction for *Drosophila* leg formation. *Pacific Symposium on Biocomputing* f99, Hawaii, 77–89.
- Moriya H., Y. Shimizu-Yoshida, and H. Kitano. 2006. *In vivo* robustness analysis of cell division cycle genes in *S. cerevisiae* PLoS Genet. DOI: 10.1371/journal.pgen.0020111.
- Morohashi, M., A. Winn, M. Borisuk, H. Bolouri, J. Doyle, and H. Kitano. 2002. Robustness as a measure of plausibility in models of biochemical networks. *Journal of Theoretical Biology* 216:19–30.
- Oda, K., and H. Kitano. 2006. A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.* msb4100057 (Apr. 18).
- Oda, K., Y. Matsuoka, A. Funahashi, and H. Kitano. 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology* msb4100014., E1–17.

Sites: **Taiwan-U.S. SBE&S Collaborations** (multiple institutions)
Focus on Proposed Collaboration on Transforming Technology through SBE&S and Experimental Investigations

National Cheng Kung University (NCKU)
 No.1, University Road
 Tainan City 701, Taiwan, ROC

National Chung Cheng University (NCCU)
 168 University Road, Minhsiung Township,
 Chiayi County 62102, Taiwan, ROC

National Applied Research Laboratories (NARL)
 3F, No. 106, Ho-Ping E. Road, Sec. 2
 Taipei 106, Taiwan, ROC

National Taiwan University (NTU)
 No. 1, Sec. 4, Roosevelt Road
 Taipei, 10617 Taiwan, ROC

National Science Council (NSC)
 No. 106, Ho-Ping E. Road, Sec. 2
 Taipei 10622, Taiwan, ROC

Academia Sinica
 128 Academia Road, Section 2, Nankang
 Taipei 115, Taiwan, ROC

Dates: June 9–12, 2008

WTEC Attendees: **No panelists attended these meetings. The WTEC panelists are grateful to the following U.S. participants who shared their meeting notes with the panel:**
 Dr. Wing Kam Liu, Department of Mechanical Engineering, Northwestern University
 Dr. J. S. Chen, Department of Civil & Environmental Engineering, University of California, Los Angeles
 Dr. S. C. Max Yen, Materials Technology Center, Southern Illinois University Carbondale

Hosts:

June 9 (NCKU) Dr. Yonhua Tommy Tzeng, National Cheng Kung University
 Dr. Fong-Chin Su, National Cheng Kung University

June 10 (NCCU) Dr. Yeau-Ren Jeng, National Chung Cheng University

June 11 (NARL) Dr. Kuang-Chong Wu, Deputy Director, NARL
 Dr. Michael J. Tsai, Deputy Director, National Nano Device Laboratories (NNDL)

June 11 (NTU) Dr. David Chuin-Shan Chen, National Taiwan University

June 12 (NSC) Dr. Lou Chuan Lee, Minister, NSC
 Dr. M. C. Tsai, Director General of Engineering & Applied Science, NSC
 Dr. Ching-Ray Chang, Director General, Office of International Cooperation, NSC
 Dr. Jennifer Hu, Program Director, Office of International Cooperation, NSC

June 12 Dr. Maw-Kuen Wu, Director General, Institute of Physics, Academia Sinica
(Academia Sinica) Dr. Ting-Kuo Lee, Acting Director, Academic Affairs Office, and Executive Secretary,
 Central Academic Advisory Committee, Academia Sinica

BACKGROUND

From June 9–12, 2008, U.S. and Taiwan scientists held a series of meetings in Taiwan to make presentations and gather input on a proposed U.S.-Taiwan collaboration in SBE&S, entitled, “U.S.-Taiwan Collaboration on Transforming Technology through SBE&S and Experimental Investigations.” The principal scientists involved in the discussions were, representing the United States, Wing Kam Liu (NWU), J. S. Chen (UCLA), and Max Yen (SIUC); and representing Taiwan, Yeau-Ren Jeng (NCCU), Yonhua Tommy Tzeng (NCKU), Fong-Chin Su (NCKU), Chuin-Shan Chen (NTU), and Kuang-Chong Wu (NARL).

Professors Liu, Chen, and Yen had productive meetings with National Cheng Kung University (NCKU) hosts on Monday June 9; with National Chung Cheng University (NCCU) hosts on Tuesday June 10; and with National Applied Research Laboratories (NARL) hosts, and National Taiwan University (NTU) hosts on Wednesday June 11.

On the morning of Thursday, June 12, Professors Liu, Chen, and Yen presented and discussed the proposed U.S.-Taiwan SBE&S Collaboration to the National Research Council (NSC) Director General of Engineering & Applied Science Dr. M. C. Tsai, the Director General of the Office of International Cooperation Dr. Ching-Ray Chang, and the Program Director of the Office of International Cooperation Dr. Jennifer Hu, along with representatives from NCKU, NCCU, NARL, and NTU, among others. A short summary of the proposed U.S.-Taiwan Collaboration and the action items was then presented to NSC Minister Dr. Lou Chuang Lee. He enthusiastically endorsed the action items.

In the afternoon of Thursday, June 12, Professors Liu, Chen, and Yen visited Academia Sinica and met with Dr. Maw-Kuen Wu, Director General of the Institute of Physics, and Dr. Ting-Kuo Lee, Acting Director of the Academic Affairs Office and Executive Secretary of the Central Academic Advisory Committee. They provided the valuable suggestions outlined in the latter part of this report.

SUMMARY

The following complementary capabilities between the U.S. and Taiwan teams were identified in this trip:

- NCKU: Materials characterization facilities; nano medicine
- NCCU: Nanomechanics and biomaterials; cyber infrastructure
- NARL: Enabling devices, nanofabrication and characterization, computing, and selective experimental infrastructures
- NTU: Nano/biomechanics, energy, enabling materials and devices
- US: Software, algorithms, computing, and selected experimental facilities

Through round-table discussions with participating universities and institutions, the following short- and long-term collaboration mechanisms were suggested:

- Joint research program through NSC (Taiwan) and NSF (U.S.)
 - this requires an assembly of a team of experts with complementary multidisciplinary strength
- Global institute: researchers, curricula, and degrees
 - transformative research
 - financial incentive to attract students and co-design of curricula and training programs
 - internship, entrepreneurship, industrial linkage, access to the facilities at national laboratories
 - mirror software (open source) capabilities in Taiwan
- New initiatives

The U.S. and Taiwan participants reached mutual agreement on the following action items:

- NSC and NSF have a mutual understanding that the SBE&S proposal will be an important agenda item to be discussed at a bilateral meeting set for September 18, 2008. NSC and NSF will review the concept of a global institute
- Taiwan SBE&S teams can submit proposals to the NSC funding program and National Priority Program
- Taiwan's National Center for High-Performance Computing (NCHC) may provide infrastructure for the U.S.-Taiwan SBE&S program
- Possible transformative research topics include enabling materials for nanomedicine and energy generation and consumption

CAMPUS VISITS: PARTICIPANTS AND MEETING AGENDAS

National Cheng-Kung University (NCKU), June 9, 2008

Participants

Shih-Hui Chang, Institute of Electro-Optical Science and Engineering, NCKU
 Tei-Chen Chen, Department of Mechanical Engineering, NCKU
 Fei-Bin Hsiao, Institute of Aeronautics and Astronautics, NCKU
 Jang-Yu Hsu, Department of Physics, Department of Engineering and System Science, NCKU
 Chintien Huang, Department of Mechanical Engineering, NCKU
 Ming-Shaung Ju, Department of Mechanical Engineering, Institute of Nanotechnology and Microsystem Engineering, NCKU
 Gwo-Bin Lee, Department of Engineering Science, NCKU
 Dar-Bin Shieh, Director of University Development, Department of Stomatology, NCKU
 Fong-Chin Su, Associate Dean of Engineering, Institute of Biomedical Engineering, NCKU
 Yonhua Tommy Tzeng, President for R&D, NCKU
 Ming-Long Yeh, Institute of Biomedical Engineering, NCKU
 Di-Bao Wang, Department of Aeronautics and Astronautics, NCKU
 Yun-Che Wang, Department of Civil Engineering, NCKU

NCKU Meeting Agenda

08:40	Dr. Di-Bao Wang (Post-Doctor in Department of Aeronautics and Astronautics) picks up U.S. delegates from the hotel
09:00–10:00	Visit Prof. Fei-Bin Hsiao, Department of Aeronautics and Astronautics
10:10–11:00	Visit Prof. Ming-Long Yeh, Nano-Biomechanics
11:10–12:00	Visit Prof. Ming-Shaung Ju, Bioengineering, Neuromuscular Control Systems, BioMEMS, Biomechanics
12:00–13:30	Lunch with Prof. Yonhua Tzeng (Vice President for R&D, NCKU), Prof. Ming-Shaung Ju, Prof. Chintien Huang and Dr. Di-Bao Wang
13:30–14:20	Visit: Prof. Gwo-Bin Lee, Engineering Science Department
15:00–17:00	Round-Table Discussion, Kuang-Fu Campus, NCKU
18:00	Dinner

National Chung Cheng University (NCCU), June 10, 2008

Participants

Sergey Aleksandrov, Department of Mechanical Engineering, Yung-Ta Institute of Technology & Commerce
 I-Ling Chang, Department of Mechanical Engineering, NCCU
 Chun-Ping Jen, Department of Mechanical Engineering, NCCU

Yeau-Ren Jeng, Dean of Research and Development Affairs, Department of Mechanical Engineering, NCCU

Shyi-Long Lee, Department of Chemistry and Biochemistry, NCCU

De-Shin Liu, Department of Mechanical Engineering, NCCU

Shaw-Ruey Lyu, Buddhist Tzu-Chi Da-Lin General Hospital, NCCU

Shu-Wei Wu, Mechanical Engineering Department, National Central University

NCCU Meeting Agenda

09:00	Pick up U.S. delegates from Tayih Landis Hotel Tainan
10:30–12:00	Presentations from US and Taiwan Teams
12:00–14:00	Lunch
14:00–16:00	Round Table Discussion
17:30–19:30	Dinner with Bioenergy Division of Nice Group and Chia-Yi City Government (Nice Plaza Hotel)



Figure B.1. (Left) meeting at NCCU; (right) dinner with NCCU President, Dr. Jyh-Yang Wu (3rd from left) and Mayor of Chia-Yi City, Ming-Hui Huang (5th from left).

National Applied Research Laboratories (NARL), Morning, June 11, 2008

Participants

Chia Ching Chang, Department of Biological Science and Technology, National Chiao Tung University

David C. S. Chen, Department of Civil Engineering, NTU

Franz Cheng, Acting Director, Business Development Division, Headquarters, NARL

Kung-Chong Wu, Deputy Director of NARL, Institute of Mechanics, NTU

Jiann Shieh, National Nano Device Laboratories

Michael J. Tsai, Research Scientist and Deputy Director General, National Nano Device Laboratories

Chih Min Yao, Project Manager, HPC R&D and Application Program, National Center for High-Performance Computing



Figure B.2. (Left) Meeting at NARL; (right) NARL Meeting participants.

NARL Meeting Agenda

10:00–10:05	Welcome, NARL Deputy Director K. C. Wu
10:05–10:35	Introduction of NARL, Franz Cheng, Chih Min Yao, Chia Ching Chang
10:35–11:00	Introduction of US Delegation, Wing Kam Liu, J. S. Chen, M. Yen
11:00–12:00	Open Discussion
12:00–13:30	Lunch

National Taiwan University (NTU), Afternoon, June 11, 2008

Participants

Kuo-Chun Chang, Chairman, Department of Civil Engineering, NTU
 Jeng-Shian Chang, Institute of Applied Mechanics, NTU
 Sheng D. Chao, Institute of Applied Mechanics, NTU
 David C. S. Chen, Department of Civil Engineering, NTU
 Ping-Hei Chen, Department of Mechanical Engineering, NTU
 U. Lei, Institute of Applied Mechanics, NTU
 Tony Wen-Hann Sheu, Department of Engineering Science and Ocean Engineering, NTU
 Kuang-Chong Wu, Deputy Director of NARL; Institute of Mechanics, NTU
 Fu-Ling Yang, Department of Mechanical Engineering, NTU
 Yeong-Bin Yang, Department of Civil Engineering, NTU

NTU Meeting Agenda

14:00–14:05	Welcome, K.C. Chang, Chair of Civil Engineering Department
14:05–14:10	Introduction of US and NTU Delegation, K. C. Wu, David C. S. Chen
14:10–14:30	Introduction of SBE&S, W. K. Liu, J. S. Chen, M. Yen
14:30–15:00	Experience Sharing: research activities and international collaboration at NTU, S. D. Chao, P. H. Chen, Tony Sheu
15:00–16:30	Round Table Discussion
16:30–18:00	Break
18:00–20:00	Dinner

National Science Council (NSC) and Academia Sinica, June 12, 2008

NSC Meeting Participants

Ching-Ray Chang, Director General, International Corporation Department, NSC

David C. S. Chen, Department of Civil Engineering, NTU
 Jennifer Hu, Program Director, International Cooperation Department, NSC
 Yeau-Ren Jeng, Dean of Research and Development Affairs, Department of Mechanical Engineering, NCCU
 Lou-Chang Lee, Minister of NSC
 Dar-Bin Shieh, Director, University Development, Department of Stomatology, NCKU
 M. C. Tsai, Director General, Engineering & Applied Science Department, NSC
 Kuang-Chong Wu, Deputy Director of NARL; Institute of Mechanics, NTU

Academia Sinica Meeting Participants

David C. S. Chen, Department of Civil Engineering, NTU
 Yeau-Ren Jeng, Dean of Research and Development Affairs, Department of Mechanical Engineering, NCCU
 Ting-Kuo Lee, Acting Director, the Academic Affair Office and Executive Secretary, Central Academic Advisory Committee Institute of Physics, Academia Sinica
 Dar-Bin Shieh, Director, University Development, Department of Stomatology, NCKU
 Kuang-Chong Wu, Deputy Director of NARL; Institute of Mechanics, NTU
 Maw-Kuen Wu, Director, Institute of Physics, Academia Sinica

Meeting Agenda

10:00 – 11:30 NSC Meeting with Directors of Engineering and International Affairs, Presentation of US Team on US-Taiwan Collaboration on Transforming Technology through Simulation Based Engineering and Science (SBE&S) and Experimental Investigations
 11:30 – 12:00 Meeting with NSC Minister Lee
 14:30 – 16:00 Meeting with Maw-Kuen Wu and Ting-Kuo Lee, Academia Sinica



Figure B.3. Meeting at NSC, from left: K. C. Wu, M. Yen, J. S. Chen, NSC Minister Lee, W. K. Liu, Y. R. Jeng, C. S. Chen, D. B. Shieh.

SUMMARY OF MEETING DISCUSSIONS

Why SBE&S in Taiwan?

- Complementary (or validation) of the current experimental capability (“enabling nanotechnology” in Taiwan) with the simulation-based engineering and science capability (from United States)
- There is less focused SBE&S program in Taiwan

- Shorten the time from research & development toward transformative impact (delivery of technology)
- Break the get-to-know barrier in building international partnership with unique compatibility
- Expand the concept of “materials and systems by design”

Implementation

- Review existing/desired one-on-one (project-to-project) collaboration between Taiwan and US involved in SBE&S
- Discuss the concept of “SBE&S” Institute: charter, functionality, benefits, and impact
- Discuss the role of SBE&S in (1) medicine & healthcare, (2) energy and environmental sustainability, (3) enabling materials
- Solicit ideas/suggestions and action plan toward forming “International Network on SBE&S Institute”

Potential Benefits

- Joint research program through NSC and NSF with a focus on integrated experiments and SBE&S for enabling materials, nano medicine, and energy harvesting. This requires an assembly of a team of experts in basic science, engineering, and medicine with complementary multidisciplinary and multiscale strength
- US will provide open source software and algorithm for team research
- Exchange researchers, joint curriculum, dual degree
- Develop financial incentive to attract foreign students with selected universities
- Engage in co-design of curriculum and training program, degree programs with global compatibility
- Each country (institute) serves as an expeditor to create international internship, entrepreneurship, industrial linkage
- Provide access to the facilities at national laboratories, build mirror (software) capability

Synergy and Logistics

- Develop a list of potential members of “SBE&S institute” in Taiwan
- Create a new national track (within the exiting priorities) in NSC
- What is the current capability and infrastructure of SBE&S in Taiwan (as compared to US and Europe)?
- Who in chain of command of the decision making that the US delegation should visit?
- What will be the roadmap to form a funding program that supports the concept of SBE&S in Taiwan and how much?
- Who in the grass root (individuals and academia) should push for this?

ACTION ITEMS CONCLUDED FROM NSC AND ACADEMIA SINICA MEETINGS

NSC Meeting Action Items

- NSC and NSF have mutual understanding that SBE&S program will be an important agenda to be discussed in the September 18, 2008, bilateral meeting; NSC and NSF will review the concept of a global institute
- Taiwan teams in SBE&S can submit proposals to NSC funding program and National Priority Program
- National Center for High-Performance Computing (NCHC) may provide infrastructure for the U.S.-Taiwan SBE&S program
- Possible transformative research topics will include enabling materials for nano-medicine and energy generation and consumption

4.2 Academia Sinica Meeting Action Items

- Need an education plan to promote SBE&S for next generation

- National Center for High-Performance Computing (NCHC) should take a leading role and have a meeting before the September NSF-NSC meeting
- Need stronger endorsement from NSC
- Taiwan needs to promote SBE&S
- Consider integrating other fields (e.g., Social Science) into SBE&S
- Talk to Minister of State, Jin-Fu Chang, and Deputy Minister of NSC, Lih. J. Chen, to champion for SBE&S
- Energy and environmental programs in Taiwan should consider implementation of SBE&S into the program
- Show concrete examples in the proposed plan
- Get Jim Chang (U.S. Air Force) involved to support SBE&S
- Taiwan should perform a review of the current status of SBE&S

CONTACT INFORMATION

Name	Organization	Telephone	E-Mail
Wing Kam Liu	Northwestern University, Department of Mechanical Engineering	847-491-7094	w-liu@northwestern.edu
Jiun-Shyan (J. S.) Chen	University of California, Department of Civil & Environmental Engineering,	310-267-4620	jschen@seas.ucla.edu
S. C. Max Yen	Southern Illinois University at Carbondale, Materials Technology Center	618-536-7525	myen@siu.edu
Yeau-Ren Jeng	National Chung Cheng University, Department of Mechanical Engineering	886-5-2428189 0933-278212	imeyrj@ccu.edu.tw
Chun-Ping Jen	National Chung Cheng University, Department of Mechanical Engineering	886-5-2720411, Ext. 33322	imecpj@ccu.edu.tw
De-Shin Liu, Scott	“ “	886-5-2720411, Ext. 33305	imedsl@ccu.edu.tw
I-Ling Chang	“ “	886-5-2720411, Ext. 33319	imeilc@ccu.edu.tw
Shyi-Long Lee	National Chung Cheng University, Department of Chemistry & Biochemistry	886-5-2428305	chesll@ccu.edu.tw
Michael Chan	National Chung Cheng University, Department of Life Science and Institute of Molecular Biology	886-5-2720411, Ext. 66510	biowyc@ccu.edu.tw
Cheng-Chung Chou	“ “	886-5-2720411, Ext. 66506	bioccc@ccu.edu.tw
Shaw-Ruey Lyu	Buddhist Tzu-Chi Da-Lin General Hospital	886-5-2648000, Ext. 5912	srlyu@seed.net.tw
Sergey Aleksandrov	Yung-Ta Institute of Technology & Commerce, Department of Mechanical Engineering	0972-043-374	Sergei_alexandrov@yahoo.com
C-S David Chen	National Taiwan University, Department of Civil Engineering	886-2-3366-4275 0912-259-495	dchen@ntu.edu.tw
Yeong-Bin (YB) Yang	“ “	886-2-3366-4245	ybyang@ntu.edu.tw
Kuo-Chun Chang	“ “	886-2-3366-4232	ciekuo@ntu.edu.tw
Liang-Jenq Leu	“ “	886-2-3366-4263	ljleu@ntu.edu.tw
Kuang-Chong Wu	National Taiwan University, Institute of Applied Mechanics	886-2-3366-5695	wukc@spring.iam.ntu.edu.tw

Name	Organization	Telephone	E-Mail
U Lei	“ “	886-2-3366-5673	leiu@spring.iam.ntu.edu.tw
Jeng-Shian Chang	“ “	886-2-3366-5678	jschang@spring.iam.ntu.edu.tw
Sheng Der Chao	“ “	886-2-3366-5066	sdchao@spring.iam.ntu.edu.tw
Ping-Hei Chen	National Taiwan University, Department of Mechanical Engineering,	886-2-3366-2689	phchen@ntu.edu.tw
Fu-Ling Yang	“ “	886-2-3366-2683	fulingyang@ntu.edu.tw
W.H. Tony Sheu	National Taiwan University, Department of Engineering Science and Ocean Engineering	886-2-3366-5791	twhsheu@ntu.edu.tw
Kuang Chong Wu	National Applied Research Laboratories	886-2-27378017	wukc@narl.org.tw
Franz Cheng	“ “	886-2-66300620	franz.cheng@narl.org.tw
Chih Min Yao	National Applied Research Laboratories, National Center for High-Performance Computing	886-3-5776085, Ext. 355	yao@nchc.org.tw
J. H. Michael Tsai	National Applied Research Laboratories, National Nano Device Laboratories	886-3-5726100, Ext. 7588	jhtsai@ndl.org.tw
Chia Ching Chang	National Chiao Tung University, Department of Biological Science and Technology	886-3-5712121, Ext. 56958	ccchang01@faculty.nctu.edu.tw
Shih-Hui Chang	National Cheng Kung University, Institute of Electro-Optical Science and Engineering	886-6-2757575, Ext. 65290	gilbert@mail.ncku.edu.tw
Tei-Chen Chen	National Cheng Kung University, Department of Mechanical Engineering	886-6-2757575, Ext. 62168	ctcx831@mail.ncku.edu.tw
Fei-Bin Hsiao	National Cheng Kung University, Institute of Aeronautics and Astronautics	886-6-2757575, Ext. 63667	fbhsiao@mail.ncku.edu.tw
Chintien Huang	National Cheng Kung University, Department of Mechanical Engineering	886-6-2757575, Ext. 62190	chuang@mail.ncku.edu.tw
Jang-Yu Hsu	National Cheng Kung University, Department of Physics	886-6-2757575, Ext. 65208	jyhsu@phys.ncku.edu.tw
Ming-Shaung Ju	National Cheng Kung University, Department of Mechanical Engineering	886-6-2757575, Ext. 62163	msju@mail.ncku.edu.tw
Gwo-Bin Lee	National Cheng Kung University, Department of Engineering Science	886-6-2757575, Ext. 63347	gwobin@mail.ncku.edu.tw
Dar-Bin Shieh	National Cheng Kung University, Department of Stomatology	886-6-2353535, Ext. 5376	dshieh@mail.ncku.edu.tw
Yonhua Tommy Tzeng	National Cheng Kung University, Department of Electrical Engineering	886-6-2757575, Ext. 50901	tzengyo@mail.ncku.edu.tw
Fong-Chin Su	National Cheng Kung University, Institute of Biomedical Engineering	886-6-2757575, Ext. 63422	fcsu@mail.ncku.edu.tw
Ming-Long Yeh	National Cheng Kung University, Institute of Biomedical Engineering	886-6-2757575, Ext. 63429	mlyeh@mail.ncku.edu.tw
Di-Bao Wang	National Cheng Kung University, Institute of Aeronautics and Astronautics	886-6-2757575, Ext. 63642	dibao.wang@gmail.com
Yun-Che Wang	National Cheng Kung University, Department of Civil Engineering	886-6-2757575, Ext. 63140	yunche@mail.ncku.edu.tw

Meeting:	Taiwan–U.S. Workshop on Simulation-Based Engineering and Science (SBE&S) for Emerging and Transforming Technology
Meeting Site:	National Cheng Kung University (NCKU) Institute of Innovations and Advanced Studies (IIAS) No.1, University Road, Tainan City 701 Taiwan, Republic of China
Sponsors:	Taiwan National Science Council (NSC) U.S. National Science Foundation (NSF)
Date:	February 18–21, 2008
Meeting Reporter:	Wing Kam Liu, Northwestern University (also ASME and NSF)
Technical Chairmen:	Dr. Wing Kam Liu, Northwestern University Dr. Yonhua (Tommy) Tzeng, National Cheng Kung University Dr. Max Yen, Southern Illinois University Carbondale
Hosts:	Dr. Ching-Ray Chang, Director General, Department of International Cooperation, Taiwan National Science Council Dr. Jennifer Hu, Program Director, Taiwan National Science Council Dr. Ken Chong, National Science Foundation, Program Director and Engineering Advisor Dr. Anne Emig, Program Director, East Asia and Pacific Program, OD/OISE, NSF
Contributors and Speakers:	Dr. Elisa Budyn, University of Chicago at Illinois Dr. J. S. Chen, University of California, Los Angeles Dr. Chia-Ching Chang, National Chiao Tung University Dr. Ching-Ray Chang, National Taiwan University; Director General, NSC Dr. Tei-Chen Chen, National Cheng Kung University Dr. Hsin Chu, Taiwan and Institute of Physics, Academia Sinica Dr. Ken Chong, NSF Program Director and Engineering Advisor Dr. Woei-Jer Chuang, National Cheng Kung University, College of Medicine Dr. Joel Collier, Department of Surgical Research, University of Chicago Dr. David Eddington, University of Illinois at Chicago Dr. Eliot Fang, Sandia National Lab Dr. Dean Ho, Northwestern University Dr. Jennifer Hu, Taiwan NSC Dr. Jang-Yu Hsu, National Cheng Kung University Dr. Bing Joe Hwang, National Taiwan University of Science and Technology Dr. Yeau-Ren Jeng, National Chung Cheng University Dr. F. J. Kao, National Yang-Ming University Dr. Adrian Kopacz, Northwestern University Dr. Michael M.C. Lai, President National Cheng Kung University Dr. Jun Liu, Pacific Northwest National Lab Dr. Wing Kam Liu, Northwestern University Dr. Tso-Pin Ma, Yale University Dr. Tijana Rajh, Argonne National labs Dr. Fong-Chin Su, National Cheng Kung University Dr. Mike Teitell, University of California, Los Angeles Dr. Michael Tsai, National NanoDevice Lab, Taiwan Dr. Yun-Che Wang, National Cheng Kung University Dr. Tomasz Wiltowski, Southern Illinois University Dr. Max Yen, Southern Illinois University

BACKGROUND

The Taiwan–U.S. Workshop on Simulation-Based Engineering and Science (SBE&S) for Emerging and Transforming Technology took place February 18–21, 2008, at National Cheng Kung University (NCKU) in Tainan, Taiwan, with grants and leadership from the U.S. National Science Foundation and the Taiwan National Science Council. This unique workshop brought together researchers from outstanding Taiwanese and U.S. universities and national laboratories in an intimately collaborative environment. The aim of the workshop was to enhance collaboration between research teams in emerging technology in Taiwan and the United States through an intensive exchange of related state-of-the-art research being carried out in both countries. The discussions and interactions at the workshop identified research activities and roadmaps to remove the challenges associated with focused applications. It is hoped that the interactions among researchers and students will lead to meaningful research partnerships, curriculum building, and technology transfer between these countries.

The workshop focused primarily on nanoengineered biomedicine aspects of simulation-based engineering and science, on enabling materials and energy harvesting issues, on nanodevices, and on related nanoengineering and nanomanufacturing issues.

Given the rapid advances in genomics, informatics, sensing, wireless communication, and microelectromechanical systems (MEMS) and nanoelectromechanical systems (NEMS) technologies, it is not difficult to envision that future medical treatments will be patient-specific and handled by integrated nanoscale devices that can sense, think, communicate, and act. These integrated nanodevices can be equipped with various functions to perform multiple special-purpose tasks. It is envisioned that an integrated medication process could potentially include the use of nanoengineered devices and processes. For example, a self-guided, or remotely steered, sensing device could be injected into the human body to be able to move around for better disease detection and cell sample collection. Another application could be a special device designed to deploy a site-specific precise dosage of medicine; the nanoengineered therapeutics could incorporate special coatings to tailor the release profile of a given drug. A nanosensor could also be added to monitor the effectiveness of the medication in real time. Each of these methods could improve health care and disease management for many patients.

Although clinical researchers and biologists are very interested in the medication processes described above and are currently ambitiously pushing for the technology developments necessary to make them a reality, there are several basic engineering challenges. These challenges can be categorized into 3 areas: materials, energy, and manufacturing processes.

Materials selection is critical for these applications. In order to perform appropriate functions, some components may need to be strong but highly flexible. There also might be other components that require a certain level of stiffness along with shape-memory capabilities. Coating materials to minimize stiction and friction, interaction between organic and inorganic materials, energy absorption, tribology, and Radio Frequency (RF) transmitting capabilities are all important consideration factors in nanodevice designs.

Providing power to a nanodevice to support movement and communication functions, and minimizing power consumption could be among the most difficult challenges that nanoengineers face. While energy in the general sense is not the first thing that many think of when addressing nanomedicine, it is certainly a critical requirement for novel implant technologies. New nanotechnology-enabled energy sources must be discovered, and storage systems must be identified.

In addition, a high level of attention must be placed on development of new delivery systems. It is expected that during the delivery of drugs via nanoscale devices, all the nanomedicine will be consumed and most of the nanodevice disposed of. Therefore, robust and high-throughput nanomanufacturing processes have to be developed for making nanomedicine and fabricating nanodevices. It is well known that there are many nanoscale building blocks (e.g., nanotubes, nanowires, nanorods, and nanoparticles) that have special properties and can be used for special functions. However, how to control the growth of these building blocks to obtain preferred sizes and shapes, and how to manipulate and arrange them to the desired organized patterns still remain major

engineering challenges. Correlation of the nanopatterns with material properties also requires an intensive research effort.

The above focus areas serve as the integrative foundations for the emerging field of nanoengineered medicine with respect to device development. The coalescence of these integrative foundations serve as a “means to an end” approach to achieving critical advancements, whether a novel diagnostic system, organ-addressing implant, or active biomaterial.

To best address the focus areas of the workshop, it was divided into three stages: technical talks, technology tours, and panel discussions. The panel discussions were held at the end to assess the success of this workshop. The Panel Discussions section of this report reviews the goals and achievements of the workshop as well as suggestions for improvement and potential points of collaboration.

TECHNICAL TALKS

Overview

The technical talks given at this workshop targeted the following topics: biomedicine and drug development; nanomaterials and nanodevices; nanomanipulation and nanomedicine; nanoenergy; and nanoengineering; for all topics, speakers addressed the implications for simulation-based engineering and science for emerging and transforming technology, which was the overarching theme of the workshop. The field of nanoengineering encompasses all engineering practices on the nanoscale. Because of the interdisciplinary nature of this field, this topic was addressed in several lectures, including those on multiscale and multiresolution analysis, in particular, the lecture given by Dr. Eliot Fang in the nanoengineering session.. Each lecture discussed the idea of modeling and simulation-enabled nanoengineering, or a smaller branch of nanoengineering, and its emerging applications and technologies.

In the biomedicine and drug development category, presenters discussed simulation-based biomechanics, models of DNA molecules and their biosensor applications, the design for potent and selective integrin drugs, and the use of a stochastic multiresolution mathematical analysis framework for molecular bioregenerative engineering.

In the category of nanomaterials and nanodevices, presenters discussed molecular dynamics studies of nanostructures using nanomechanics to model human cortical bone and computational nanophotonics, molecular dynamics study of phase transformations for silicon under nanoindentation and its applications, Moore’s Law and high-k gate dielectrics for advanced transistor technology, functionalized nanomaterials with applications in biology and technology, and the study of cell dynamics using force propagation and live cell interferometry.

In the category of nanomanipulation and nanomedicine, presenters discussed nanofabrication of carbon nanocones as indenter tips, the use of bioregenerative engineering in the treatment of human disease, using time-resolved microscopy for molecular dynamics imaging and simulation, work towards scaffolds for regenerative medicine, and laser-induced transition of nanodiamonds. The presenters on nanoenergy discussed nanostructured materials and nanoenergy and their technological and scientific barriers and economical opportunities.

Speakers and Topics

Session One: Biomedicine and Drug Development

To kick off the technical talks, Dr. Ching-Ray Chang, Professor of Physics at National Taiwan University and Director General of the Department of International Cooperation of NSC briefly reported on the long-term partnership between NSC and NSF. Flagship projects include the FORMOST-s/Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) project in atmospheric sciences; International Long-Term Ecological Research; the TAIGER project involving integrated investigations of the geodynamics of the Taiwan Orogeny, the Pacific Rim Application and Grid Middleware Assembly (PRAGMA) project; as well as the ALMA North America. With respect to the development of nanoscience and nanotechnology, the NSC takes the leading

role in Taiwan in developing many emerging technologies that correspond to nanoengineering, energy harvesting, nanomanufacturing, and nanomedicine/biology. Simulation-Based Engineering and Science (SBE&S) in Enabling Transformative Technology is a major program theme at NSF. Dr. Chang expressed the expectation that the holding of this workshop can lead to many possibilities and impacts across the engineering and applied sciences.

Dr. Ken P. Chong gave the audience an “NSF Program Overview on Research in SBE&S, Solid Mechanics, and Materials.” He reviewed the NSF programs generally as well as those in simulation-based engineering sciences, including the recommendations by the Blue Ribbon Panel chaired by Tinsley Oden (see Oden et al. 2006). He then explained the main importance of mechanics and materials in all of the transcendent technologies and gave examples of research and education opportunities and challenges in them.

Dr. Wing Kam Liu talked about “Stochastic Multi-resolution Mathematical Analysis Framework-Integrated Material Design to Molecular Bio-Regenerative Engineering.” He explained the sophisticated theory of the assembly process of nanowires and the process of endothelial cell adhesion at the molecular and cellular level using multiresolution mathematical analysis framework.

Dr. Fong-Chin Su explained the concept of “Simulation-Based Biomechanics,” specifically the concept of a graphic visual interactive musculoskeletal system (VIMS) and its applications.

Dr. Jiun-Shyan Chen presented “Predictive Models for Atomic to Continuum Modeling of DNA Molecules and Biosensor Applications.” He proposed a multiscale coarse graining method and multilevel homogenization formulation for numerical simulation of DNA and the application of this model to nanometer-scale biosensors for DNA sequencing.

Dr. Woei-Jer Chuang talked about “Design and Selective Integrin Drugs,” focused mainly on the family of RGD/KTS-containing proteins. The applications and uses of integrins were discussed.

Session Two: Nanomaterials and Nanodevices

Dr. Yeau-Ren Jeng explored “Mechanical Property and Interfacial Phenomena of Structures Using Nanomechanics.” He examined the mechanical properties of carbon nanotubes and the nanomechanical properties of the hard tissue of a tooth, in particular the properties of fluoride-treated enamel surfaces.

Dr. Elisa Budyn investigated “A Multi-Scale Approach to Assess the Fracture Strength of Human Cortical Bone Microstructure.” She proposed a new procedure to investigate such properties and to diagnose any pathological modification of the morphology or the mechanics of the material used.

Dr. Jang-Yu Hsu talked about “Molecular Dynamics Studies of Nanostructures.” He studied the nanoparticle construct, the carbon-related structures, and water dynamics.

Dr. Max Yen presented “An overview on Nanomaterials and Nanodevices Research at the Materials Technology Center, Southern Illinois University Carbondale”. His research was focused on sensors, functional polymers/glass, drug delivery system, fabrication of carbon-carbon nano-tubes and rule-based simulation of nano matters.

Dr. Tijana Rajh discussed “Energy Transduction at NanoBio Interfaces,” focusing mainly on exploring new approaches for electronic linking of semiconductor and metallic nanoparticles to electroactive moieties such as conductive polymers.

Dr. Tei-Chen Chen presented “A molecular dynamics study of phase transformation in mono-crystalline Si under nanoindentation.” In his study, Dr. Chen identified the phases of BC8 and R8, similar to the phase Si-I, and extracted them from the deformation region during unloading. He also investigated the phenomenon of pop-out during unloading.

Dr. David T. Eddington presented “Tiny Transformation Technologies,” explaining the beneficial phenomena of these technologies at many scales, including mesoscale (e.g., platform to rapidly quantify the quality of transplantation tissues), microscale (e.g., the system to facilitate high through hypoxia experimentation) and nanoscale (e.g., the use of cnidocytes as a functional material in devices).

Session Three: Nanodevices

Dr. Tso-Ping Ma talked about “Moore’s Law and Modern IC Technology.” He answered the questions of what makes the silicon chip thick and why is the electronic industry growing so fast. He also gave an overview of silicon IC technology and its applications.

Dr. Michael J. Tsai presented “The Research of the Group of Simulation and Modeling at the National Nano Devices Laboratories” that included an overview of Taiwan’s National Nano Devices Laboratories. In addition, he discussed the findings from the simulation of a gated single carbon nanotube field emitter with magnetic focusing.

Dr. Dean Ho lectured on “Functionalized Nanomaterials at the Interface of Biology and Technology” and the progress his lab has made towards utilizing nanomaterials in energy and medicine, specifically, in using biomimetic membranes as matrices for protein reconstitution and copolymer and diamond-based nanomaterials for drug-delivery applications.

Dr. Michael A. Teitell’s talk on “Cellular Dynamism during Force Propagation Revealed by Live Cell Interferometry” familiarized the audience with a new approach to characterizing diseases by using live-cell interferometry to detect cytoskeletal remodeling behaviors.

Dr. Shih-Hui Chang presented “Computational Nanophotonics: Plasmonic Phenomena in Tailored Metallic Nanostructures.” The speaker discussed his recent work, which utilized a fully parallel 3D Finite-Difference Time-Domain method to study nanoplasmonic systems.

Session Four: Nanomanipulation and Nanomedicine

Dr. Yun-Che Wang presenting “Nanofabrication with Carbon Nanocones as Indenter Tips.” A molecular dynamics simulation method was used to show that using a hollow nanocone is preferable to using a solid nanocone for performing nanoindentation.

Adrian Kopacz presented “Simulation and Prediction in Vascular Systems- Employing Bioregenerative Engineering in Human Pathology Impediment.” He discussed how a newly developed immersed finite element method was applied to simulate shear flow in endothelial cells, as well as to study the effect of red blood cell deformation on pulmonary diffusion capacity.

Dr. Fu-Jen Kao’s talk, “Implementing Time-Resolved Microscopy for Molecular Dynamics Imaging,” stressed the informative applications of time-resolved microscopy on virus infection and apoptosis of cells with a special focus on autofluorescence.

Dr. Joel H. Collier presented “Self-Assembling Polymer-peptide Conjugates: Towards Scaffolds for Regenerative Medicine.” The focus of the lecture was to discuss two approaches: the assembly of peptide-polymer conjugates and co-assembling peptides that form betasheet fibrillar gels.

Dr. Chia-Ching Chang gave a talk titled, “Laser Induced Popcorn-like Conformational Transition of Nanodiamonds,” emphasizing that the popcorn-like transformation of the sp^3 diamond structure of nanodiamonds to sp^2 graphite could be applied to non-conventional therapeutic strategies in the biomedical field.

Dr. Chao-Cheng Kuan gave a talk on “Conductance of Single Molecular Junctions,” demonstrated the conductance of Au-alkanedithiol-Au molecular junctions from first principles and also discussed the implications of the findings.

Session Five: Nanoenergy

Dr. Jun Liu opened the session with a talk titled, “Nanostructured Materials and Nanoenergy.” Dr. Liu discussed how the current challenges regarding energy may be overcome with the implementation of nanocrystalline materials. In addition, he discussed the needs in large-scale atomistic treatments of charge/ion transport dynamics through porous nanocrystalline networks.

Dr. Bing-Joe Hwang spoke on “Atomic Manipulation and Characterization of Bimetallic Nanoparticles.” Dr. Hwang presented an X-ray absorption spectroscopy-based methodology for determining the alloying extent and surface composition of bimetallic nanoparticles. He also discussed the effects these properties have on the electrochemical activity of the nanoparticles.

Dr. Tomasz Wiltowski gave a lecture titled, “Energy Research; Technological and Scientific Barriers and Opportunities.” Dr. Wiltowski’s talk hit on some of the primary concerns regarding energy production and consumption. In addition, Dr. Wiltowski discussed some of his research group’s current research interests, including the utilization of nanomaterials as catalysts and reactive separation agents in hydrogen production.

Session Six: Nanoengineering

Dr. H. Eliot Fang’s talk was titled, “Modeling and Simulation Enabled Nanoengineering—Moving from Nanotechnologies to Emerging Applications.” Dr. Fang’s presentation emphasized the lack of parity between existing theoretical methods for the atomic or macro scales, and those theoretical methods necessary for a fundamental understanding of the behaviors of nanoscale materials. Also, Dr. Fang reviewed the successes and remaining challenges regarding the advance of nanoscience and nanotechnologies, techniques to manipulate nanomaterials, and computational modeling of nanoscale materials.

TECHNOLOGY TOURS

In accordance with the workshop focus on SBE&S for enabling transforming technologies, two days of the workshop were devoted to technical tours of two of Taiwan’s federally assisted industrial parks, the Hsin-Chu Science-Based Industrial Park and the Central Taiwan Science-Based Industrial Park. Each science industrial park is affiliated with several premier universities, national laboratories, and hi-tech industries. The Hsin-Chu Science-Based Industrial Park serves as the model for industry-government-academia entrepreneurship primarily focused on electronics. The Central Taiwan Science-Based Industrial Park is primarily focused on R&D in optoelectronics. The technical tours visited key national research facilities on the National Cheng Kung University campus, as well as two Industrial Park businesses, Taiwan Semiconductor Manufacturing Company, Limited (TSMC), and Taiwan Chi Mei Optoelectronics Corporation (CMOS).

Taiwan Semiconductor Manufacturing Company, Ltd.

TSMC has the broadest range of technologies and services in the industry. It created a dedicated semiconductor foundry industry when it was founded in 1987, and strives to provide superior semiconductor manufacturing services for worldwide customers and to cultivate mutually beneficial, long-term partnerships. The company has readied its most advanced 12-inch GigaFabs for 45 nm production, having the capacity to generate tens of thousands of 12-inch wafers per month. One of the GigaFabs, Fab-14, became the world’s most advanced process technology in the foundry industry and is the only 100%-automated production line foundry in the world. A comprehensive tour of the foundry was provided, explaining the production process of highly integrated, very small, and very low power devices for a wide variety of markets.

Taiwan Chi Mei Optoelectronics Corporation

CMOS is one of the world's largest suppliers of liquid crystal displays for flat screens for television sets and monitors. It was founded in 1998 as one of the first Taiwanese companies using thin-film-transistor liquid crystal display (TFT-LCD) for its own color filter production. CMOS also provides leading-edge technology for organic light-emitting diode (LED) displays, with its largest display being presented in its showroom. CMOS's Eco TV, a highly efficient dynamic contrast display was also showcased in the company showroom.

The technical tours spurred much in the way of informal discussions and networking among the workshop participants. Through this close interaction, the researchers from the United States and Taiwan began to establish exchanges and collaborations based on their research emphases.

PANEL DISCUSSIONS

The last section of the workshop was dedicated to a panel discussion between Mike Teitell (University of California, Los Angeles), Joel Collier (University of Chicago), Tommy Tseng (National Cheng Kung University), Elliot Fang (Sandia National Laboratory), S. H. Chen (National Cheng Kung University), and Wing Kam Liu (Northwestern University), as well as the audience members present.

Goals Achieved and Future Ideas

The goal of this panel was to discuss the success of this workshop and offer suggestions for potential improvements. The main objectives of this international workshop were to bring together leading experts in nanotechnology to discuss research progress and their future perspectives, to serve as a catalyst for international collaboration and partnership in the areas of Simulation-Based Engineering and Science (SBE&S) in Enabling and Transforming Technology, and to tour Taiwan's R&D and entrepreneurship infrastructure in nanotechnology. These goals were more than achieved through the diligent efforts of the organizing committees.

The workshop described ongoing R&D work with results that can serve as the building blocks of the SBE&S concept. Experimental, computational, and theoretical research were presented in areas of nanoengineering approaches, nanoengineered medicine and health care, nano materials and devices, and energy harvesting. Since SBE&S is an overarching concept for "product realization" or "virtual prototyping" in medicine, biology, energy, nanoengineering and many other fields, it is vital that it be coupled with experimental investigations. This coupling has the potential to realize foundational and universally applicable methodologies for fundamentals-driven technology fabrication.

It was suggested that there could be an advantage to introduce both experimental and computational procedures to researchers in one field who are relatively unfamiliar with the specifics of the other. This may lead to a better understanding among researchers of what the other discipline, either experimental or computational, has to offer. This knowledge could lead to a better basis for collaboration between experiments and computations.

Other suggestions included a Web-based mailing list where attendees of this workshop could keep in touch and continue to collaborate. Since NCKU has an existing mailing list similar to the one being suggested, it is possible that the workshop mailing list could be built on the existing list. Another idea was to continually rotate the location of the workshop in upcoming years. This will make the workshop accessible to more researchers as well as initiate diverse ideas and collaborations.

The experiences gained from this workshop can be utilized as a working model for international collaboration and organization for the advancement of nanotechnology and nanoengineering. The next steps are to define a transforming and enabling technology that can arise from the medical, scientific, mathematical, computational, and engineering expertise that was present for this collaboration. Translatable solutions will be targeted for emerging issues in energy, biomedicine, materials, and devices. The end goal is to achieve an important research discovery or methodology in transforming technology with a particular reference to SBE&S.

Potential Collaborations

Many potential areas of interactions were noted during this workshop.

For nanomedicine, three major areas for focused R&D collaboration were outlined:

8. Knowledge of interfacial phenomena at the nanoscale through better understanding of the drug, protein, and polymer adsorption and desorption in complex biological milieus, as well as better understanding of cell attachment and integrin binding
9. Knowledge of biomacromolecular assemblies, including DNA topology and gene activation, integrin binding with disintegrins and ligands, and self-assembling peptides, proteins, and polymer coatings
10. Cell stiffness and behavior in live cell interferometry, matrix stiffness and cell adhesion, and matrix stiffness and macrophage activation, which could expand multiscale, multiphysics aspects of biological systems

In nanoengineering, four areas of future collaboration were defined:

11. The study of nanomaterials, including the shape and dimension control and the interaction between dissimilar materials; design criteria and performance expectations can be used to better understand the correlation between substructure and properties
12. Expanded nanoenergy research that includes a focus on new energy sources and storage systems as well as delivery systems that could greatly advance the development of nanodevices
13. In the field of nanomanufacturing, improve material manipulation as well as robust and high-throughput manufacturing processes; these developments are critical for continued growth
14. Modeling and simulation research needs reliable methods to deal with multiphysics and multiscale issues in space and time; there is room for growth in the realm of appropriate science-based materials models and numerical algorithms as well as in the quantification and margin of uncertainties of models

There were many important achievements of this workshop. It provided an understanding of the R&D roadmap and interests of Taiwanese academic high-tech institutes as starting points for collaborative discussions. Specific investigators were identified with committed interests towards collaborating with members of the U.S. delegation. In addition, current and upcoming NSC and university initiatives were overviewed to establish a roadmap for extended collaboration and larger program initiatives.

From these many achievements, many benefits are to be had. There was a great deal of international exchange of expertise and broader understanding of research program administration across the globe. This workshop also provided a key understanding of the importance of integrated SBE&S and experimental investigation, and it encouraged multidisciplinary interfacing as a staging point for the creation of new fields of research.

Future Collaborations between Taiwan and the United States

The following collaboration approaches were suggested:

15. Immediately, the investigators will identify mutual research topics between the two countries, such as those given during the workshop. The U.S. and Taiwan investigators can request supplemental support to existing NSF and NSC projects.
16. There are initial talks with institutes in both the United States and Taiwan for research. Again, both U.S. and Taiwanese investigators have to agree on mutual research topics, ways of collaboration and education, amount of funding per year, and project duration.
17. In the long term, both U.S. and Taiwanese investigators will work together with NSF and NSC as well as other funding agencies to come up with new initiatives for research.
18. The above concepts can be extended to other countries, once there is evidence that they are working.

NATIONAL CHENG KUNG UNIVERSITY RESPONSES TO WTEC SBE&S SURVEY

The survey was completed on December 1, 2007, by Dr. Yonhua (Tommy) Tzeng (tzengyo@gmail.com), Vice President for R&D, National Cheng Kung University.

(Note: only those questions that were applicable to NCKU are included here.)

General

(i) What are the major needs, opportunities or directions in SBE&S research over the next 10 and 20 year time frames?

Directions

- Cell and molecular engineering
- Property prediction of artificial (functional) materials
- Creation of super computing centers is a must for the next 10 to 20 years.
- To close the gap between simulations with length and time scales of different order of magnitude is the key challenge.

Opportunities

- Scientific computing could be an important software industry in Taiwan due to the advantages of Taiwan's PC industry and PC cluster infrastructure.

(ii) What are the national and/or regional funding opportunities that would support research to meet these needs, and/or take advantage of these opportunities? Are these funding opportunities expanding?

- Government funding institutions include National Science Council, Industrial Technology Research Institute and National Health Research Institute. With the rise of semiconductor business, industrial funding also is becoming available.

Materials/ Energy & Sustainability/ Life Sciences and Medicine

(i) What major breakthroughs in these fields will require SBE&S; which are you and/or your colleagues pursuing?

- Materials: First-principles calculations to determine properties of nanoscale and biological objects
- Life Sciences: Cell/molecular mechanics, bone remodeling, and dental mechanics.

(ii) Within your institution, region, or country, are there identified targets of opportunity for applications of simulation either for scientific research or for engineering applications in these fields?

- Drug design, performance evaluation of MEMS devices, and other related fields

(iii) Which problems could benefit most from a 1-2 order-of-magnitude increase in computational power?

- Cell/molecular engineering and biological interactions.

(iv) What are examples of major SBE&S successes or failures in these fields?

- There are many successful stories in the SBE&S fields, such as the startup of the many computing software companies.

(v) *Do investigators, laboratories, and institutions receive any financial compensation for patented inventions derived from their simulations?*

- Yes, but few so far.

(vi) *Have any start-up companies spun-off based on simulation efforts in your lab? If so, please describe them.*

- No.

Validation, Verification, and Quantifying uncertainty

Describe efforts and advances within your institution to validate and verify codes and to quantify uncertainty in simulation-based predictions?

(a) In engineering studies, we have done dispersion analysis to validate the stability of new algorithms and predict their accuracy limits.

(b) In human movement, we validated codes using simplified experimental setup and indirect clinical evidence.

Simulation Software

(i) *What percentage of code used in your group is developed in-house? What percentage of code is commercial? What percentage is open-source? What percentage has been developed by others (e.g., under contract or by acquisition)?*

- It's about 50–50: 50–60% of code used is developed in-house; 40–50% is commercial. However, 100% of the code for the research investigating computational electromagnetics is developed in-house.

(ii) *What are the biggest issues in using models/simulations developed by others? How easy/difficult is it to link codes to create a larger or multifaceted simulation environment?*

- The simulation codes developed by others may not fit advanced and specific problems; however, adopting new models and algorithms doesn't seem to be a problem. It is generally difficult to link codes from different providers for larger simulations

(iii) *Who owns the intellectual property rights (IP) to the codes developed in your group?*

- In most cases, the programming authors own the copyright; the National Science Council (NSC) has rights to its utilization.

(iv) *How do you deal with liability issues for products developed with codes from other sources?*

- Most researchers avoid this problem by only using in-house codes, commercial codes, or free software.

Big Data and Visualization

What type of data and visualization resources do you need (and have access to) for SBE&S research?

- HDF data structure and parallel I/O for data storage and parallel I/O for visualization. Most SBE&S researchers rely on existing software packages to aid visualization.

Engineering Design

(i) *What type of models/codes do you use, develop, or conduct basic research on pertaining to different phases of engineered system design (conceptual, parametric optimization, operational/control)?*

- Currently, we are developing the following models for high-frequency, nanoelectronics, and photonics device applications:
 - Boltzmann and quantum transport models of carriers (electrons and holes) in semiconductor devices

- Particle simulation by Monte Carlo method
- Electromagnetic field simulation by using FDTD, FEM methods.

Nevertheless, several genetic-type algorithms (Genetic Algorithm, Neural Network, Particle Swing, etc.) were used to develop a computational intelligence in engineering optimization design

(ii) What are the data requirements for a comprehensive life-cycle model of an engineered system? Do you have repositories in place that provide the data?

- Currently we don't have such environment at National Nano Device Laboratories (NDL; <http://www.ndl.org.tw/web/eng/index.html>); however, we are in the process of setting up a project to collaborate with the researchers at NCHC (National Center for High-Performance Computing) that can provide us not only a better storage and backup facility but also the computing environment.

(iii) How do you couple output of the models/simulations to the decision making process (including quantification of uncertainty/error in the predictions)?

- We are still in a small group of numerical modeling at NDL (there is a shortage of well-trained students in this area); however, we have good opportunity to work with researchers in various backgrounds such as the nanofabrication and metrology scientists at NDL and university professors. As a result, there is a unique opportunity here to develop more realistic design, simulation, or prediction in semiconductor devices.

(iv) What is the curriculum for training doctoral students in all aspects of designing engineered systems using simulation/modeling tools?

- None.

(v) Are there efforts within your institution to couple physics-based models with macroscopic logic or econometric models?

- No.

Next-Generation Algorithms and High-Performance Computing

(i) Would you characterize the SBE&S research in your lab as needing and/or using primarily desktop, teraflop, or petaflop computing resources?

- Most SBE&S research extends from using desktop for code/algorithm development, to small simulation based on lab-size clusters within 32 nodes, and in the production phase, extends to teraflop computation performed in the national supercomputer center. However, due to the limited computational resources in Taiwan, teraflop or petaflop computation is quite limited or impossible.

(ii) What are the computational bottlenecks in your simulation problems? What solutions exist now and in the near future, either from a hardware perspective, algorithm perspective, or both?

- The first problem is always the algorithm. Once the code is well developed, then the hardware is of concern in the production mode. But for academic work, we tend to be more in the development phase than in the production phase. For large-scale simulation, data analysis using parallel I/O and the parallel I/O visualizations already lead to certain achievement in recent years. Recent development in hardware accelerated computation provides some alternative solutions for modest size simulation, but for true large scale simulation, the development of petascale supercomputer and parallel I/O data analysis and visualization tool is the obvious way to go.

(iii) What is the state-of-the-art in your community in discrete event simulations (model size, parallelization or thread limits, special techniques - time warp, rollback, etc)?

- We rely on local PC cluster to do our job. We may access large PC cluster or main frame at NCKU or NCHC to do simulation. As far as the community goes, the heavy users tend to be computational chemists and computational physicists utilizing codes such as VASP, CASTEP or Gaussian. PIC and MD also could

run into huge resource issues. In addition, the state-of-the-art in Electromagnetic simulation in discrete event simulation has been extended to terascale, the parallelization is quite linear for larger scale simulation, the checkpoint and time warp can be done easily with modern cluster tools.

Education and Training

(i) Is there a formal scientific computing or computational science and engineering graduate program at your institution?

- Scientific computing courses are constantly offered, but the focus is not physical science or cross-discipline-oriented. For students doing theoretical work, the algorithm and programming languages are part of their major training. In National Cheng Kung University, there is a university-level computer center providing services to researchers in scientific computing. However, if students do not join computational research groups, they seldom have opportunities to benefit from the computer center.

(ii) What level of computing expertise do your incoming graduate students or employees arrive with? Do you feel that they have the necessary background to conduct research in SBE&S, or do they need extensive preparation after arriving before they can be productive researchers?

- Incoming graduate students, on average, do not possess sufficient programming skills. They have taken a one-year undergraduate computer programming course in their first year of college, and that's it. They need extensive preparation before being able to do anything.

(iii) What kind of training is available in your organization for simulation and high-performance computing? Is it adequate? Do you believe it will be adequate to address computing on multicore and petascale architectures? What plans are in place for training programs in next-generation computing?

- Different research groups provide graduate-level courses for high-performance computing; however, there is still room for improvement to make the training more efficient. For multicore and petascale architectures, it's the future of scientific computing, and hence teaching these should be encouraged. The NCHC occasionally provides an informative type workshop for high-performance computation. To take advantage of the tera/peta-scale architectures, a longer workshop or summer school is necessary. For next-generation computing, research-oriented studies are underway for familiarizing ourselves with future computing ideas.

(iv) What fraction of students in your institution study/work in the area of SBE&S, and how has this fraction changed over the past 5–7 years?

- Approximately 5%–30 of students are working on SBE&S related work. Compared with the past 5–7 years, this fraction remains roughly about the same.

(v) What fraction of graduate students/postdocs in SBE&S come from abroad? From which country do these researchers originate? How many students in your country would you estimate go abroad to do their PhDs in SBE&S?

- Maybe less than 1% of students come from abroad, mainly from southeastern Asia. Some professors have postdocs from Russia or India. About 30% of students may go abroad for their PhD in SBE&S.

(vi) After completing their PhDs in SBE&S-related fields, what is the route your students take? Do they take postdoctoral positions in your country or abroad? What fraction eventually achieve permanent jobs in SBE&S? Is this a desired career path for students? Do they perceive many job opportunities related to their training in SBE&S? How does industry view students with training in SBE&S?

- Most students choose academic teaching and research jobs, if possible. Otherwise, they may serve as postdocs in either Taiwan or other countries. They receive reasonable job offerings. Industries welcome students with SBE&S backgrounds because they usually have received solid training.

Funding, Organization, and Collaboration

(i) *What are the roles of academic, government, and industrial laboratories in SBE&S research in your country?*

- In Taiwan, the academic field initiates all kinds of SBE&S research, but major funding is provided by government. In addition, NCHC provides the computing resources beyond small-scale PC clusters. It is a general practice that academia sends proposals to the government agencies, such as the National Science Council, for possible funding opportunities. Although government and industrial laboratories also perform scientific research, most innovative research is done in academic laboratories.

(ii) *Who pays for, and/or is responsible for, the development and sustainability of SBE&S infrastructure (including long-term data storage costs, code maintenance, open-source software, etc.)?*

- In term of infrastructure, lots of resources go to NCHC in Taiwan. With manpower around 300 and the annual budget approximately 25 million USD, NCHC is responsible for the national high-performance computing and networking infrastructure, in particular for local academic use.

(iii) *What is the funding situation for SBE&S in your country? What are the major sources of funding? Over the past 5 years, has funding of SBE&S increased, decreased, or remained roughly constant relative to funding of all of science and engineering research? Is most SBE&S research funded as single-investigator grants, small team grants, or large teams and/or initiatives? What is the typical duration over which proposals in SBE&S are funded?*

Funding for SBE&S in Taiwan is mainly from government (particularly the NSC), with some funding from industry (e.g., Chi-Mei Electronics), amounting to around \$400–700 thousand NTD per year for each PI. Over the past 5 years, funding for SBE&S has been generally increasing compared to other research fields. Most SBE&S research is funded as single-investigator grants. The typical duration for the funds is about one to three years.

REFERENCES

- Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science Foundation blue ribbon panel on simulation-based engineering science*. Arlington, VA: National Science Foundation.

- Site:** Toyota Central R&D Labs, Inc.
41-1, Aza Yokomichi, Oaza Nagakute,
Nagakute-cho, Aichi-gun, Aichi-ken, 480-1192, Japan
<http://www.tytlabs.co.jp/eindex.html>
- Date Visited:** December 6, 2007
- WTEC Attendees:** P. Cummings (report author), A. Arsenlis, C. Cooper, L. Petzold, G. Karniadakis,
D. Nelson
- Hosts:** Dr. Satoshi Yamazaki, Research Advisor and Manager
Technology and Systems Laboratory
E-mail: syamazaki@mosk.tytlabs.co.jp
- Dr. Shiaki Hyodo, Research Manager
Computational Physics Laboratory, Materials Department
Email: e0668@mosk.tytlabs.co.jp
- Dr. Hitoshi Washizu, Researcher
Tribology Laboratory, Mechanical Engineering Department
Email: washizu@mosk.tytlabs.co.jp
- Dr. Tomoyuki Kinjo, Researcher
Computational Physics Laboratory, Materials Department
Email: e1308@mosk.tytlabs.co.jp
- Dr. Seiji Kajita, Researcher
Tribology Laboratory, Mechanical Engineering Department
e1389@mosk.tytlabs.co.jp

BACKGROUND

The Toyota Central R&D Labs (TCRDL), Inc., was established in 1960 for the purpose of carrying out basic research for Toyota Group companies and thus contributing to company growth and the advancement of science and technology. The major product of the Toyota Group companies is automobiles, at which Toyota is singularly successful. In the first quarter of 2007, Toyota became the leading producer of cars worldwide, surpassing General Motors for the first time. Other Toyota products, overshadowed by the automobile segment, include sewing machines (<http://www.sewtoyota.com>) and prefabricated houses (http://www.toyota.co.jp/en/more_than_cars/housing/index.html).

The research activities of the TCRDL focus on issues related to automobile production, such as resource and energy conservation, environmental preservation, enhancement of comfort and safety, and advanced information processing. There are five R&D divisions at TCRDL. In the areas of *energy and the environment*, TCRDL is researching new catalysts; analyzing combustion mechanisms in internal combustions engines; and developing next-generation batteries, solar cells, and fuel cells. In the *safety/human engineering* area, TCRDL is studying human body dynamics to determine the kind of injuries people are likely to suffer in auto collisions in order to help design safer automobiles. In the *mechanical engineering* area, TCRDL contributes to the development of vehicle technology that improves the performance, efficiency, and active safety of automobiles through research on control, kinetic and vibration analyses, and tribology. In the *systems engineering and electronics* area, TCRDL carries out research on high-speed digital mobile telecommunication technologies for man-machine interfaces and display devices; on power control devices for electric vehicles; and on information processing technologies for the improvement of design, manufacturing, and logistics processes. In the *materials* area, TCRDL is developing functional materials, including metallic composites, organic-inorganic molecular composites, and high-performance cell materials, as well as developing plastics and rubber recycling technologies. Funding for these activities is drawn from the Toyota Group corporations, and projects are selected from the technical needs of the Toyota Group and from proposals from the TCRDL staff.

R&D ACTIVITIES

Simulation-based engineering and science activities support all five divisions of the TCRDL. The role of SBES at the laboratory is in the development and application of simulation tools to aid in understanding phenomena that are observed by experimentalists at the facility. Some highly publicized simulation packages at TCRDL include the following:

1. THUMS: Total HUMAN Model for Safety is a suite of detailed finite element models of the human body that can be subjected to impacts to assess injury thresholds. Models included detailed meshing of musculoskelature and internal organs.
2. EVAS: Engine Vibration Analysis System is an engine design tool that is helping to develop lighter and quieter engines by providing highly accurate vibration predictions.
3. Aerodynamic Noise Simulator COSMOS-V is used to calculate temporally fluctuating airflow and predict noise such as wind noise and wind-throb, as well as vibrations caused by pressure fluctuations.
4. Mill-Plan: Software designed to determine the optimal number of machining processes and the tool form and cutting conditions for each process by closely examining the machining procedures and evaluating their overall efficiency.

The highly publicized simulations of the structural response of Toyota automobile frames to impacts, and simulations of the wind resistance of auto bodies, have reached a level of maturity such that they are no longer conducted at the TCRDL but at the Toyota Motor Company. In addition to the enumerated list, there are significant simulation efforts at TCRDL focused on the chemistry of combustion in internal combustion engines, on the multiscale modeling of fuel cells, on the performance of catalysts in oxidizing carbon monoxide, and on predicting the viscosity of synthetic lubricants.

It was evident that there is a growing interest in modeling human behavior and interaction with machines, and social network interactions to understand disruptions to Toyota's production lines and/or to predict the behavior of markets towards disruptive technologies such as the hybrid automobile. The discussion in this area brought up the issue of uncertainty, and the view of uncertainty that was offered by the TCRDL staff was one in which uncertainty was an emergent property of complex interacting systems whose individual dynamics are well understood but whose synergistic behavior may become chaotic. As pointed out by the TCRDL researchers, the introduction of a hybrid automobile more than a decade ago (the *Prius*) was a very nonlinear, highly uncertain event that has paid off handsomely for Toyota.

Most of the simulation work is performed on stand-alone workstations. However, if more computing power is needed, TCRDL has an NEC SX-5 on site and has mechanisms in place to buy computer time at the Earth Simulator and at other locations. In terms of CPU time, TCRDL spends the most time in performing materials science simulations to uncover material structure-property relationships, followed in decreasing order by time spent on structural simulations of vibrations and acoustics, magneto-electric simulations in support of LIDAR (Light Detection and Ranging) applications, and CFD/combustion simulations of internal combustion engines.

Impressive simulation results were presented in TCRDL's effort to predict the viscosity of synthetic lubricants. The results showed that molecular dynamics simulations of the viscosity of known molecules with small simulation volumes could reproduce the trends observed in the laboratory, but the absolute value of the viscosity predicted by the simulation was orders of magnitude off because of the difference in the film thicknesses and shear rates of the experiment versus the simulation. Large-scale molecular dynamics simulations were conducted on Japan's NAREGI national project supercomputer system; in those simulations, the film thickness and shear rate were comparable to the experimental measurements, and the two came into agreement.

Publication of the research conducted at TCRDL is promoted when possible, and the laboratory publishes a quarterly journal entitled *R&D Review of Toyota CRDL* that it makes freely available on the Internet. The TCRDL also sponsors small (~20 invited speakers) closed conferences for its scientists to interact with leading

academics and learn about the latest research; for example, a workshop in the area of SBES entitled "Decision Making and Uncertainty in Nonlinear Complex Systems" was held in Copenhagen, Denmark, in November 2006.

The researchers at TCRDL identified several grand challenge problems for future SBES activities. The first is to perform first-principles molecular dynamics simulations of fuel cells to understand the mechanisms of hydrogen atom transport within those systems. The second, in the area of catalysis, is to understand the nature of carbon-oxygen bonding and resonances in the vicinity of catalyzing surfaces.

CONCLUSIONS

The depth and breadth of SBES activities at TCRDL are impressive, and the research atmosphere is comparable to, if not better than, other corporate R&D centers around the world. Although most of the lab's simulation requirements are satisfied with OTS hardware, its researchers have access (as part of a national policy on industrial utilization of national high-performance computing facilities) to the largest national computing facilities in Japan, if they are required to obtain meaningful results. This interaction between government and corporate facilities in SBES appears to be beneficial to both parties in Japan. The cost of computer time at the national supercomputing facilities was not discussed.

Site: Tsinghua University Department of Engineering Mechanics
 School of Aerospace
 Beijing, 100084, P.R. China
<http://hy.tsinghua.edu.cn/English/about/page9.asp>

Date: December 3, 2007

WTEC Attendees: S. Kim (report author), S. Glotzer, M. Head-Gordon, J. Warren, P. Westmoreland, G. Hane

Hosts: Prof. Quanshui Zheng, Yangtze Chair, Professor, and Chairman, Department of Engineering Mechanics
 Email: zhengqs@tsinghua.edu.cn
 Prof. Zhuo Zhuang, Deputy Dean, School of Aerospace
 Email: zhuangz@tsinghua.edu.cn
 Prof. Gexue Ren
 Email: Rengx@mail.tsinghua.edu.cn
 Prof. Xiong Zhang, Director of Institute of Dynamics and Control
 Email: xzhang@tsinghua.edu.cn
 Dr. Baohua Ji, Associate Professor
 Email: bhji@tsinghua.edu.cn
 Dr. Bin Liu, Associate Professor
 Email: liubin@tsinghua.edu.cn

BACKGROUND

The Department of Engineering Mechanics at Tsinghua University, believed to be the largest such department in the world, is ranked #1 in China in its field. Research projects in the department span the entire spectrum of activities in mechanics, including significant SBES activities involving global collaborations. Thanks to its reputation, the department has several high-profile SBES collaborations with multinational companies. The department has 89 faculty, 240 undergraduate students, 260 graduate students, and 19 postdoctoral fellows. The annual research budget is 20 million RMB¹⁶, of which 50% is from the National Natural Science Foundation of China (NSF-China; <http://www.nsf.gov.cn/Portal0/default106.htm>) and 30% is from industry-sponsored projects. The professors have numerous awards and are active on editorial boards of the top international journals in Mechanics. From 1999 to 2006, approximately 100 PhD theses in all fields (natural and social sciences, engineering, and so on) were selected for China's National Distinguished PhD thesis award; 16 of these were awarded in Mechanics, and Tsinghua's Department of Engineering Mechanics received eight of these coveted awards.

SBES RESEARCH

A departmental overview and background context pertaining to SBES was presented by the Department Chair Prof. Quanshui Zheng. This was followed by highlights of research from the department's SBES portfolio, presented by the lead investigators: multiscale simulations of materials (Z. Zhuang); multibody dynamics with applications (G. Ren); coarse-grained MD simulations with applications to modeling of protease inhibitors (B. Ji); and bead-spring and bead-rod polymeric chain dynamics (B. Liu).

¹⁶ About \$2.7 million at the time of the WTEC team's visit (approximate exchange rate of 7.396 RMB/US\$1).

Multiscale Simulations

Multiscale simulations, in particular bridging continuum and molecular scales via combination of finite element and molecular dynamics simulations was an overriding theme that emerged across all research examples presented in the overview. The department's strong roots in solid mechanics were also evident in the presentations. The broader societal impact of the research applications included development of stronger filaments of carbon nanotubes from multiscale understanding of fracture mechanisms, multibody dynamics simulations of the operation of telescopes, coarse-grained models of protease inhibitors for HIV drug development, and acceleration of large-scale simulations of polymeric chains.

COMPUTING FACILITIES

In November 2007, the department acquired a 2 TFlops 32-node cluster. Each node has 2 quad-core processors, which are interconnected by a 10 G infiniband.

DISCUSSION

After the presentations, there was an open, albeit brief, discussion of future SBES opportunities and challenges facing the department, which touched on the following points:

- The research program features many international collaborations in SBES.
- The department has a computational dynamics course using FEAP from Prof. Taylor of UC Berkeley. With the emergence of multicore architectures, mechanics students will have to take advantage of courses from other departments (e.g., computer science) to receive their training in parallel computing. Currently, there is relatively little emphasis in formal training of students on data management.
- Despite the department's significant research activities and excellent international reputation, relatively little funding currently comes in the form of the large-scale "973" and "863" national funds for high-visibility projects; this impacts the budget available for computational and data infrastructure. Furthermore, NSFC does not fund code development.
- Only about 20% of undergraduates go directly to careers in industry; most go on to graduate studies. Graduate students go on to a balance between industry and academic careers. Training in SBES-related activities is highly valued by the students because of resulting career opportunities.
- Industry interactions with industry R&D centers within China and also on a global scale with leading multinational companies, as noted earlier, are very strong. The research sponsored by Medtronic, Inc., in the simulation of cardiac muscles was cited as an illustrative example.
- The faculty values the accessibility of open source software in furthering their research goals in SBES-related fields. The WTEC team's hosts also noted that this aspect is counter-balanced by researchers' desire to maintain a competitive edge by developing their own codes.

REFERENCES

- Ji, B.H., and H.J. Gao. 2004. Mechanical properties of nanostructure of biological materials. *Journal of the Mechanics and Physics of Solids* 52:1963-1990.
- Gao, H.J., B.H. Ji, I.L. Jäger, E. Arzt, and P. Fratzl. 2003. Materials become insensitive to flaws at nanoscale: Lessons from nature. *Proc. Natl. Aca. Sci. USA* 100:5597-5600.
- Liu, B., Y. Huang, H. Jiang, S. Qu, and K.C. Hwang. 2004. The atomic-scale finite element method. *Computer methods in applied mechanics and engineering* 193(17-20):1849-1864.
- Wang, L.F., Q.S. Zheng, J.Z. Liu, and Q. Jiang. 2005. Size dependence of the thin-shell model for carbon nanotubes. *Phys. Rev. Lett.* 95:105501.
- Zheng, Q.S., B. Jiang, S.P. Liu, Y.X. Weng, L. Lu, Q.K. Xue, J. Zhu, Q. Jiang, S. Wang, and L.M. Peng. 2008. Self-retracting motion of graphite microflakes. *Phys. Rev. Lett.* 100:067205.

- Site:** **University of Tokyo**
Room 31A1, Eng. Bldg. #2, Hongo Campus, Bunkyo-ku
Tokyo 113-8656, Japan
- Meeting Scheme:** **Workshop on R&D in Simulation-Based Engineering and Science**
- Sponsor:** 21st Century Center of Excellence Program on Mechanical Systems Innovation
 School of Engineering, The University of Tokyo
- Date Visited:** December 7, 2007
- WTEC Attendees:** G.E. Karniadakis (report author), P. Cummings, L. Petzold, T. Arsenlis, C. Cooper,
 D. Nelson
- Hosts:** **University of Tokyo Members**
 Prof. Nobuhide Kasagi, Leader of the 21COE Program,
 Dept. of Mechanical Engineering
 Email: kasagi@thtlab.t.u-tokyo.ac.jp
 Prof. Chisachi Kato, Institute of Industrial Science
 Prof. Shigeo Maruyama, Dept. of Mechanical Engineering
 Prof. Yoichiro Matsumoto, Dean, Dept. of Mechanical Engineering
 Prof. Hiroshi Okuda, Research into Artifacts, Center for Engineering
 Prof. Shinsuke Sakai, Dept. of Mechanical Engineering
 Prof. Satoshi Watanabe, Dept. of Material Engineering
 Prof. Shinobu Yoshimura, Dept. of Quantum Engineering
- Visitors**
 Prof. Shunichi Koshimura, Disaster Control Research Center, Tohoku University
 Prof. Michael A. Leschziner, Dept. of Aeronautics, Imperial College London
 Prof. Masaru Zako, Dept. of Management of Industry and Technology, Osaka University

BACKGROUND

Professor Nobuhide Kasagi of the University of Tokyo's Department of Mechanical Engineering generously organized an extremely informative workshop to introduce the visiting WTEC panelists to computational science in Japan and, in particular, to the 21st Century Center of Excellence (COE) Mechanical Systems Innovation program at the University of Tokyo that focuses on hyper-modeling/simulation. The WTEC team heard presentations from a number of the academic researchers working under this program.

The University of Tokyo (UT) is ranked number one in Japan; it was established in 1877 as the first national university in Japan. It offers courses in essentially all academic disciplines at both undergraduate and graduate levels and conducts research across the full spectrum of academic activity, including medicine. The University of Tokyo has a faculty of over 4,000 and a total enrollment of about 29,000, evenly divided between undergraduate and graduate students. There are currently more than 2,000 international students and over 2,500 foreign researchers at UT for both short and extended visits. The University of Tokyo is known for the excellence of its faculty and students; ever since its foundation many of its graduates have gone on to become leaders in government, business, and academia.

In June 2001, the Ministry of Education, Culture, Sports, Science & Technology (MEXT) issued a report on a new policy for structural reform of the national Universities in Japan and established a new program on Centers of Excellence (COE) for the period 2002–2007, the “21st Century COE Program.” The aims of this program are to form world-class research and education centers at universities in Japan, to cultivate creative human resources

capable of improving research standards and leading the world, and to promote the building of universities of global appeal. Twenty-eight centers have been established at the University of Tokyo. Each is a five-year project. Therefore, projects begun in FY2002 have already ended; but some of them have developed into Global COE projects. The Global COE Program is a project that has inherited the fundamental concepts of the 21st Century COE Program and is given targeted support for forming internationally excellent education and research centers by MEXT. Applications for the Global COE Program will be accepted in each field from FY2007 to FY2011. In FY2007, six projects were adopted at the University of Tokyo.

The leader of the 21st Century COE Program at UT focusing on Mechanical Systems Innovation (MSI) is Professor Kasagi, who was the main organizer of the workshop. There are many researchers from different departments and institutes of UT participating in the 21st Century COE, including the Departments of Mechanical Engineering, Engineering Synthesis, Environmental & Ocean Engineering, Aeronautics & Astronautics, Quantum Engineering & Systems Science, Geosystems Engineering, Nuclear Engineering and Management, the Institute of Industrial Science, and the Center for Disease Biology and Integrative Medicine. There are three main research focus areas of the MSI program: (1) Energy, (2) Biomedicine, and (3) Hyper-modeling/simulation. The last one aims at advancing the art in multiscale/multiphysics modeling and the synergistic use of simulation and experiment. In addition, MSI has an educational aim in fostering creative young researchers. Professor Kasagi had invited some additional scientists from outside UT to enrich discussions at the workshop. In the following, we briefly describe the presentations given by the Japanese researchers.

RESEARCH PRESENTATIONS

Satoshi Watanabe, “Quantum Simulation on Nanoscale Measurements of Materials Electrical Properties”

This is a five-year program funded by CREST-JST at the level of ¥190 M and it is a joint effort between UT, Tokyo University of Science, Nihon University, and Kobe University. The main objective is to develop quantum simulators to understand nanoscale electron transport for helping accurate interpretation of nanoscale measurements. These simulators include single- and multiple-probe measurements, capacitance evaluation, and analysis on effects from other external factors. A grand challenge remains the simulation of non-steady dynamics of atoms and electrons in non-equilibrium open systems. The computational resources employed are both PC clusters as well as supercomputers at UT.

Yoichiro Matsumoto, “Multiscale Modeling of Molecular Collision Dynamics”

The focus of this presentation is on multiscale modeling. A new method for molecular collisions and evaluation of physical properties (e.g., viscosity, conductivity) was presented that takes into account rotational energy. The results shown suggested great improvement in conductivity predictions. In addition, topics on multiscale modeling of surface reactions, bubble-bubble interactions, and nano-bubbles were discussed. Finally, the human simulator was described as a joint project with RIKEN.

Shigeo Maruyama, “Molecular Dynamics Simulation of Nucleation Process of Carbon Nanotubes”

Various applications of single- and multi-walled carbon nanotubes (CNT) were discussed, including applications at Toyota. Results with 2 nm standing CNT were presented, apparently the first such achievement. The objective of this project is to understand metal-carbon interaction inside a CNT using MD. Issues associated with MD acceleration via molecule positioning were discussed. There is a close MD-experiment interaction. A typical MD simulation on a small (5-PC) cluster may take up to six months.

Nobuhide Kasagi, “Role of Simulation in Flow and Energy Systems”

Prof. Kasagi is a leader of the 21st COE Program. He first presented a comprehensive view of Japan’s strategic basic plan on science & technology, and the goal of UT programs’ on mechanical systems innovation and its history. The third period of the national basic plan is underway (2006–2010), funded at the level of ¥25 trillion. It supports research with emphasis on eight fields such as life sciences, information sciences, nanosciences and the

environment, and the education of young scientists. It also supports two key technological areas, namely the next-generation supercomputer at RIKEN and the space transportation system at JAXA. Prof. Kasagi then proceeded to describe his own work, which comprises two main areas: energy systems and shear flow control. Interesting concepts related to a microturbine 30 kW energy generation system combined with a solid oxide fuel cell were presented, and the power supply needs for mobile systems were addressed. On the shear flow control, direct numerical simulations at the highest currently possible Reynolds number were presented, and MEMS-based techniques were shown to effectively control near-wall turbulence.

Shinobu Yoshimura, “Multiscale and Multi-physics Simulation for Predicting Quake-Proof Capability of Nuclear Power Plants”

The main objective of this project is full-scale simulation of components but also of buildings subject to earthquakes. Most of the simulations presented involved one-way coupling on flow-structure interactions. A fully coupled simulation will require access to the next-generation supercomputer system at RIKEN.

Masaru Zako, “Disaster Simulation in Chemical Plants”

The purpose of this project is the development of a simulation capability for disaster propagation in chemical plants considering the effects of tank fire and wind speed and direction. Another key question is how to estimate associated damages using the results of simulations in conjunction with information from the Geographical Information System (GIS). Monte Carlo methods of estimating radiation heat transfer were presented, and examples from tank fires and gas explosions in realistic scenarios were discussed. A validation example was presented, but it is clear that long-term prediction of such events remains a very difficult task. This project is well funded by a MEXT special project on earthquake disaster mitigation in urban areas, by the Petroleum Association of Japan, and by industry.

Shunichi Koshimura, “Developing Fragility Functions for Tsunami Damage Estimation Using Numerical Models and Post-Tsunami Survey Data from Banda Aceh, Indonesia”

The purpose of this project is simulation of tsunamis and also the estimation of damages caused, based on simulation results and survey data. Advanced methods in solving the shallow water equations were developed for realistic terrains, and a “fragility function” that estimates the probability of damage was constructed for first time for Banda Aceh.

Hiroshi Okuda, “Middleware for Parallel Finite Element Applications”

The objective of this project is to develop middleware for the main generic operations in finite elements, including global assembly and solvers, for a JST-CREST (2006-2011) project focusing on predictive capability for earthquakes and tsunamis. The usual parallel libraries such as MPI and OpenMP are hidden from the user, and optimization is pursued both on scalar and vector architectures. Very good performance results were achieved on the Earth Simulator for the structural analysis code, but the viscous flow codes does not scale as well. The software analyzer of the structure code (FrontSTR) is selected as a target software package to be optimized at RIKEN’s next-generation supercomputer.

Chisachi Kato, “State of R&D on Leading-Edge Computational Science & Engineering Simulation Software”

The purpose of the “Revolutionary Simulation Software (RSS)” project is to develop Japanese-made software for life sciences, nanosciences, engineering, and urban environments. The project spans the period 2005-2007, it is supported at ¥1.2 billion per year, and it involves 120 researchers. The software developed is free to be used in Japan or abroad, and to the date of this writing, 40,000 downloads have been reported. Continuous improvement and maintenance of software will be accomplished through commercialization; currently, seven companies, including Advantech Corporation, Mizuho Information & Research Institute, Inc., and NEC Soft, Ltd., are pursuing such commercialization.

CONCLUSIONS

Research at the University of Tokyo (UT) is of very high quality. The 21st century COE Program has provided the resources to develop new simulation capabilities in energy systems, disaster modeling and damage evaluation, life sciences, nanotechnology, and urban environments. The Institute of Industrial Science at UT is supported generously by MEXT to work closely with researchers conducting fundamental research in order to develop free software with the potential of commercialization by the private sector. Initial results of this collaboration scheme are very encouraging, and the simulation capability developed in diverse areas is very impressive. Some of the researchers reported isolated efforts in verification and validation, but no separate funding and hence no systematic effort is underway to address this issue.

APPENDIX C. SITE REPORTS—EUROPE

Site: Autonomous University of Barcelona and
Materials Science Institute of Barcelona (ICMAB-CSIC)
Research Center for Nanoscience and Nanotechnology
Campus de la UAB, 08193
Bellaterra, Catalonia, Spain

Date Visited: February 29, 2008

WTEC Attendees: C. Sagui (report writer), G. Karniadakis, A. Deshmukh, G. Lewison, P. Westmoreland

Hosts: Professor Dr Pablo Ordejón, Group Leader, Theory and Simulation
Edifici CM7; Facultat de Ciències
Email: ordejon@icmab.es
Dr. Alberto García
Email: albertog@icmab.es
Dr. Eduardo Hernández
Email: ehe@icmab.es

BACKGROUND

The Research Center for Nanoscience and Nanotechnology (Centre d'investigació en nanociència i nanotecnologia), also known as CIN2, at the Autonomous University of Barcelona (Universitat Autònoma de Barcelona) is a mixed research institute that has as partners the nationally funded Spanish National Research Council (Consejo Superior de Investigaciones Científicas; CSIC; <http://www.csic.es>) and the local (Barcelona) Catalan Institute of Nanotechnology (Institut Català de Nanotecnologia; ICN; <http://www.nanocat.org>).

CSIC is Spain's largest national institute for research, comprising about 100 institutes between the sciences and the arts. The central government provides money for the basic functions of these institutes, but research money must come from external funding. CSIC allocates budgets for a four-year period. People hired under CSIC are “public servants” with tenure for life, although there is talk now to substitute this mode for long-term positions followed by evaluation.

The legal structure of the ICN, on the other hand, is that of a private foundation with public capital. ICN was founded by the Catalan government in 2003 and has two partners, the Ministry of Universities, Research, and Information Society of the Catalan Government, and the Autonomous University of Barcelona. ICN has the flexibility to hire personnel as needed, and institute management decides the salaries and contracts.

From the point of view of scientific research policy, the CIN2, launched in September 2007, is a relevant example of collaboration between Catalonia and the central government of Spain. Its formal and legal structure allows exploiting the synergies between both partners (CSIC and ICN), following the principles of scientific excellence that are key to the CIN2's foundational deed. CIN2 combines the best of both worlds: the stability of the central government with the funding and flexibility of ICN.

RESEARCH

CIN2 activities are focused on the theory and simulation of processes at the nanoscale. They aim at understanding the basic physics behind the behavior of nanoscale systems at the atomistic level. The techniques focus on the electronic and atomic structure and dynamics, and their evolution in response to external conditions. CIN2 activities include the development of methods and simulation tools for the description of these systems, from semiempirical to fully first-principles approaches. Current interests include, among other topics, the understanding of electronic transport at the nanoscale (including inelastic effects and strong electronic

correlations), the simulation of STM images and STS spectra, the interaction of molecules with surfaces, and different aspects of nanostructures such as nanoclusters and nanotubes.

The CIN2 group is also involved in the SIESTA project. This was developed as an approach to compute the electronic properties and perform atomistic simulations of complex materials from first principles. Very large systems, with an unprecedented number of atoms, can be studied while keeping the computational cost at a reasonable level. The SIESTA code is freely available for the academic community (<http://www.uam.es/siesta>), and this has made it a widely used tool for the study of materials. It has been applied to a large variety of systems, including surfaces, adsorbates, nanotubes, nanoclusters, biological molecules, amorphous semiconductors, ferroelectric films, low-dimensional metals, etc. There are strong interactions with the Electronic Structure Lab at the Materials Science Institute of Barcelona (Institut de Ciència de Materials de Barcelona; ICMAB-CSIC; Prof. Canadell et al.).

A more “itemized” research list includes:

- Methods and codes for atomic scale simulations: SIESTA
 - Electronic structure methods & molecular dynamics
 - Quantum electronic transport (TranSIESTA)
- Electronic transport
 - Fundamental problems (NDR, Inelastic transport—weak and strong coupling, Kondo, ferroelectric tunnel junctions, and so on)
 - Applications (STM, molecular electronics, nanowires, surfaces, and so on)
- Molecules on surfaces
 - Molecule-molecule and molecule-substrate interactions
 - Mainly metallic surfaces
- Simulation of electronic dynamics processes in dye-sensitized thin-film solar cells
 - Electron excitations in dye molecule
 - Electron transfer processes
- Ion dynamics in oxides
 - Oxygen/vacancy mobility in nanostructured functional oxides
 - Electrochemical devices, fuel cells, gas sensors

FUNDING

At present, the group has funding from the following sources:

- The coordinated Spanish Ministry of Education and Science (MEC) project (2006–2009) provides a total of €532,000, with collaborators in the Autonomous University of Madrid, the University of Oviedo, and Materials Science Institute of Madrid (ICMM-CSIC). The topic is “Quantum-mechanical simulations and scanning probe microscopy in current problems in surfaces, complex materials, biomolecules and nanostructures.”
- A “CONSOLIDER project” (support for supercomputing and e-science) has a budget of €5 million for 5 years and involves 23 groups around the country. The title of the project is “Strategic research action to advance in the development of supercomputing environments and their application to grand challenges projects in the areas of Life Sciences, Earth Sciences, Astrophysics, Engineering and Materials/ Condensed Matter Physics.” CIN2’s task is to address some of the challenges in Materials Sciences and Condensed Matter Physics and to carry out massive parallelization of SIESTA. Most of the money goes to post-docs’ salaries.

SIESTA

The Siesta program has the following key characteristics:

- Density Functional Theory (DFT) code
- logarithmically designed for efficiency, scales linearly with system size
- The code has become standard in the Quantum community and competes with other codes (VASP, CASTEP, CPMD, Gaussian, etc.)
- A large development team from several institutions (8 researchers just in the restricted “code team”)
- Very long-term project with several sources of funding
- Freely available for the academic community (~3500 users worldwide)
- Nonacademic license available for a fee (Motorola, Sumitomo Chem., Samsung, Air Products, Fujitsu, and others)
- Currently 300+ publications referencing the code per year
- Currently SIESTA has excellent performance in serial machines but poor parallel scaling
- Current developments
 - Optimization and parallelization, in collaboration with the Barcelona Supercomputing Center
 - Implementation of new tools and methods (hybrid QM/MM, electronic transport with inelastic effects, free-energy issues)

ELECTRONIC STRUCTURE OF MATERIALS AT ICMAB-CSIC

The group investigating the electronic structure of materials works in close collaboration with CIN2. The main lines of research are organic conductors; low-dimensional metals; electronic transport; fullerenes, nanotubes, and other “nano”; magnetoelectrics; multiferroics; computational crystallography; materials for hydrogen storage; phase behavior and phase diagrams; TD-DFT; and novel simulation techniques.

The research tools of the group include expertise in both first-principles and semi-empirical electronic structure; empirical force fields, molecular dynamics, Monte Carlo, and structural relaxation.

The main sources of funding are the Spanish Ministry of Science and Education; the Catalan regional government, and several EU funding programs.

COMPUTING INFRASTRUCTURE

- Barcelona Supercomputing Center: 10240 processors, 94 TFl; 3rd largest in Europe, 13th largest in the world; funded by the Spanish central government, the Catalan government, and through special pricing from IBM
- Diverse clusters are shared by the group

EUROPEAN CONNECTIONS

- There have been efforts to coordinate things with other EU groups in order to consolidate a European “scientific power.”
- Psi-K network: EU funded; this was very important in getting groups together for collaboration and dissemination of information; however, the organization was not interested in continuation of projects, only in new projects
- Psi-K network will transition into a nonprofit private entity with funding from research groups that use the tools (<http://psi-k.dl.ac.uk>)

- The European Centre for Atomic and Molecular Computations, CECAM, has played a very important role in disseminating information; there is an agreement with Psi-K to put new effort into code development (<http://www.cecarn.fr>)
- Research funding in Spain is increasing; EU funding has had a significant impact (FAME; Center for Multi-Functional Materials)

EDUCATION

- These groups and institutes are not academic; they carry out research but not teaching
- They get only PhD students and postdocs; there is no formal training—the students learn through the assignments
- Students do research in CSIC, but the degrees are granted through universities
- PhD students generally take 4 years (there is no funding after 5 years); 1 or 2 years are devoted to courses for credit; there is an initiative to put everything together and give the material in two months (very structured and intensive courses)
- There is no graduate degree for computer simulation; there is something on informatics, and there is a MSc on Computational Science at the Politechnic University of Catalonia
- There are no career opportunities in Spain for computational support staff; the WTEC team's hosts hope that this will happen in the future following the establishment of the CSIC
- International students: most of the students and post-docs are from outside Spain—several from Latin America; the main problem for international students is bureaucratic barriers for visas and working permits for foreigners outside the European Union.

INDUSTRY

- Very little industrial investment in research in Spain
- Some pharmaceutical research
- RAPSOL
- Carburos Metálicos: bought by the United States, this company has built a center for research in materials in CSIC and wants to establish a collaboration on materials aspects of gas storage; MATGAS (materials for gases) will have as main activities research projects, a big chunk carried out via simulations (the United States pays for the research).

Site: **BASF – The Chemical Company**
Carl-Bosch-Strasse 38
67056 Ludwigshafen, Germany

Date Visited: February 26, 2008

WTEC Attendees: A. Deshmukh (report author), G. Karniadakis, C. Sagui, G. Lewison, P. Westmoreland

Hosts:

Dr. Christoph Grossmann (contact person), Senior Research Manager
Distillation, Reactive Distillation, Adsorption, Ion Exchange
Tel: +49 621 60-56440; Email: christoph.grossmann@basf.com

Dr. Ekaterina Helwig, Science Relations and Innovation Management
Tel: +49 621 60-92892; Email: ekaterina.helwig@basf.com

Dr. Horst Weiss, Senior Research Manager
Polymer Physics, Molecular Modeling
Tel: +49 621 60-22445; Email: horst.weiss@basf.com

Dr. Ansgar Schaefer, Team Leader, Quantum Chemistry
Physical Chemistry and Informatics, Scientific Computing
Tel: +49 621 60-78376; Email: ansgar.schaefer@basf.com

Dr. Anna Schreieck, Head of Scientific Computing
Tel: +49 621 60-78253; Email: anna.schreieck@basf.com

Dr. Michael Schaefer, Research Manager
Particle Measuring Technology, Process Engineering
Tel: +49 621 60-79264; Email: michael.schaefer@basf.com

Dr. Lars Vicum, Research Manager
Precipitation, Process Engineering
Tel: +49 621 60-58308; Email: lars.vicum@basf.com

Dr. Markus Linsenbuehler, Research Manager
Particle Separation and Aerosol Technology, Process Engineering
Tel: +49 621 60-97668; Email: markus.linsenbuehler@basf.com

Dr. Jan-Martin Loening, Senior Research Manager
Adsorption, Extraction, Physical Properties, Process Engineering
Tel: +49 621 60-56543; Email: jan-martin.loening@basf.com

Dr. Wolfgang Gerlinger, Research Manager
Fluid Dynamics, Process Engineering
Tel: +49 621 60-91721; Email: wolfgang.gerlinger@basf.com

Dr. Jobst Ruediger von Watzdorf, Head of Research Group
Chemicals Research and Engineering, Reaction Engineering
Tel: +49 621 60-54578; Email: jobst-ruediger.watzdorf-von@basf.com

BACKGROUND

BASF is the world's leading chemical company. The company employs over 95,000 total people worldwide, with more than 8,600 engaged in research and development activities. In 2007, BASF posted sales of €58.0 billion and income of approximately €7.6 billion. BASF's product portfolio comprises of five business segments:

1. *Chemicals*. This segment consists of Inorganics, Catalysts, Petrochemicals, and Intermediates divisions.
2. *Plastics*. BASF is one of the world's leading producers of plastics. In 2007, the segment was organized in three divisions: Styrenics, Performance Polymers, and Polyurethanes.
3. *Performance Products*. This segment is made up of the Construction Chemicals, Coatings, Functional Polymers, and Performance Chemicals divisions.

4. *Agricultural Products and Nutrition*. This segment comprises the Agricultural Products and Fine Chemicals divisions. BASF Plant Science conducts research in the field of plant biotechnology.
5. *Crude Oil and Natural Gas*. BASF's oil and gas activities are pooled in the Wintershall Group, which is active in the exploration and production of crude oil and natural gas, as well as in the trading, transport, and storage of natural gas sectors.

The percentage sales by each of the five major business segments are as follows: (1) Chemicals 24.44%, (2) Plastics 23.29%, (3) Performance Products 20.18%, (4) Agricultural Products & Nutrition 8.61%, and (5) Oil & Gas 18.15%.

MODELING AND SIMULATION ACTIVITIES

Simulation and modeling tools are broadly used in most of BASF's business segments. A few examples of research and development projects using simulation and computational models are listed below.

- BASF's computational chemists are studying three-dimensional structures of proteins, protein-ligand complexes, specialty chemicals, and how these structures interact on a molecular level. Currently, many processes, especially in molecular biology, can be visualized in a highly simplified manner using a lock-and-key analogy. Knowing the three-dimensional structure of the molecules involved is a basis for using a substitute key to prevent the real key from fitting into the lock, finding a better key, or changing the lock so that a completely different key will fit.
- BASF scientists are using virtual screening by computer to test thousands of candidate compounds and determine whether their three-dimensional and electromagnetic properties would allow them to interact favorably with target proteins. Hands-on experiments need only be performed on substances that pass this screening process. *De novo* design takes things one step further. The computer screens all synthetically accessible building blocks and selects those that best interact with the binding site.
- Biocatalytic processes involve enzymes that catalyze chemical reactions of the substrate. Despite knowing the ligands and enzymes involved, scientists still want to know the mechanism underlying the reaction and how the enzyme catalyzes the chemical reaction efficiently at room or body temperature. BASF researchers are using quantum mechanical computations to follow biochemical reactions and thus provide insights into reaction kinetics and thermodynamics. The information obtained from these calculations can help scientists to optimize fermentative production processes and develop more efficient enzymes.
- BASF researchers use bioinformatics in all activities performed at a molecular biological level. This is the case in the areas of biocatalysis, fine chemicals, and agrochemicals. The scientists have in-house access to all public molecular biology databases and internal BASF databases to assist them in their work. Collaboration with leading bioinformatics firms is another important element of this work.
- Agrochemicals used to protect crops from weeds, fungi, and insects work by blocking specific metabolic processes, thus maximizing efficacy and minimizing chemical use. The molecular sites of action of crop protection substances in the organisms to be combated—the targets—are proteins that are essential to the survival and development of fungi and weeds. The search for new active substances is more effective if the targets have been identified; hence, researchers concentrate their efforts on tracking down the relevant proteins in order to inactivate them with the help of tailored active substances. Researchers who want to identify the key proteins in an organism must first elucidate their functions, at least in the vital segments of the metabolic process. BASF researchers are using Functional Genomics strategy, where they draw upon the knowledge of the genome of the living organism under study. Analysis of sequencing data and comparison with the genetic information of other organisms enables scientists to predict or establish hypotheses on parts of the metabolism and the role of specific proteins. Functional characterization of these proteins is supported by bioinformatics methods.
- Vitamins and amino acids can often be produced more efficiently using microorganisms than by conventional synthesis methods. For example, BASF researchers have used a bacterium called *Corynebacterium glutamicum* to produce the amino acid L-lysine by fermentation. Their goal is to alter the metabolism of the bacterium so as to optimize the production of lysine. Bioinformatics played a key role in

genome sequencing and analysis and in elucidating the metabolism of the bacterium. Out of the 3300 genes identified, the scientists involved in the project have established the function of 2100 previously unknown genes. They used bioinformatics to analyze the data on already investigated genes from other microorganisms with similar gene sequences. A factor complicating analysis of the *Corynebacterium* genome was the presence of a whole series of genes whose function could not directly be ascertained because there were no known genes with a similar sequence. The researchers analyzed approximately 100 other genomes to compare the position of the genes in relation to each other and to compare the number of genes that corresponded to the enzymes of a metabolism segment. This method unearthed a promising gene that is thought to code for a regulator of lysine biosynthesis.

CRITICAL NEEDS IN MODELING AND SIMULATION

The following points highlight areas where the WTEC visiting team's hosts indicated that BASF researchers see critical needs in the modeling and simulation domain in order to achieve breakthroughs in materials, energy, and life-sciences:

- *Validation of models.* BASF researchers emphasized the difficulty in experimental validation of models. They noted that models are often constructed with insufficient data or physical measurements, leading to large uncertainty in the input parameters. For example, missing data results in parameters estimation errors for industrial-scale mass- and heat-transfer models, resulting in unreliable predictions even when the understanding of the physics is there. They highlighted the need to consider the economics of parameter estimation and model refinement. Figure 1 shows (qualitatively) the tradeoff between the cost of parameter estimation and cost of errors in prediction.

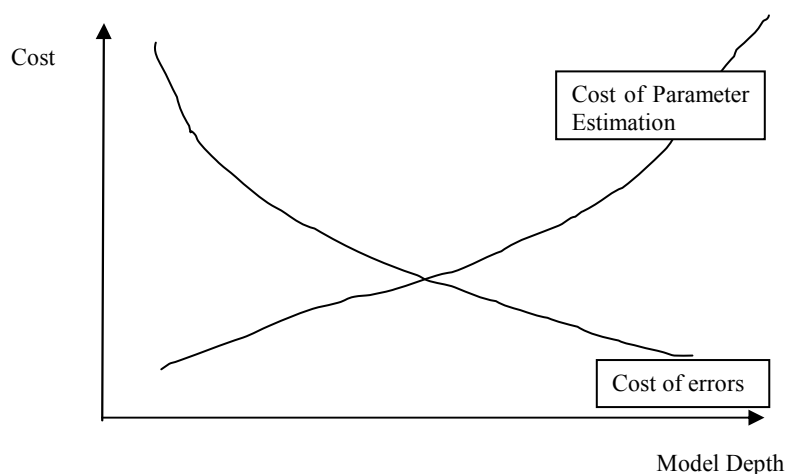


Figure C.1. Tradeoffs noted by BASF researchers between costs of parameter estimation and costs of errors.

- *Predictive modeling.* BASF researchers noted that current modeling and simulation methods work well for existing products. However, they are not ideally suited for developing new products that are not derivatives of current ones. Currently, models are mostly used to understand/explain the experimental observations. They emphasized that predicting product properties, not just process characteristics, is important. Moreover, they noted that most of the design software is deterministic, and understanding of stochastic uncertainty needs to be included in the analysis. As a result of modeling and simulation tools, BASF is conducting fewer physical experiments. Its scientists often use models to design experiments. However, in many cases there is still reluctance to develop new products and equipment purely based on computational models.
- *Large data and visualization.* BASF researchers noted that all aspects of data handling are challenging, ranging from acquisition, archiving, retrieving, to formalizing. They noted that the real bottleneck is not the

availability of computational resources but data availability and using large data in a meaningful way. Availability of experimental data was an issue for BASF at the time of the WTEC visit, especially in the reaction engineering area. Situation awareness was a critical challenge when the information was distributed. For example the team's hosts noted that visualization, archiving, and retrieval of HTP information is a problem.

- *Workflow methods.* BASF researchers noted that simulation models are very complicated and not very easy to use by non-simulation experts. They emphasized that workflow methods are needed that connect models with analysis and visualization, which could save weeks to months out of the product realization process. In order to create flexible workflow models, interfaces across different levels need to be standardized. These interfaces must be accepted by all the major software developers and users. For example, in order to develop multiscale models from the atomistic to device levels, models describing particle dynamics need to be coupled with continuum mechanics and chemical models. BASF is collaborating with Sandia National Labs on hydrodynamics of particles and coupling continuum mechanics with chemical models.
- *CAPE Open.* This is an effort in the chemical engineering domain to create standardized interfaces. BASF has participated in CAPE Open, which was launched with EU Funding, since its inception. The EU funding for this effort ended in 2005, and the project is currently being supported by consortium fees. A key drawback with the CAPE Open project is that major software developers, like Aspen Technology, have not completely adopted the standards developed under this effort; hence, these interfaces have not resulted in actual industry-wide standards in the chemical engineering community. Another issue in the workflow area is the lack of integration of process models with enterprise level models that use tools such as SAP-APO, CPLEX, GAMS, Dash Optimization, or other discrete event simulations.
- *New algorithms.* BASF researchers emphasized the need for new methods and algorithms in order to take advantage of the improvements in computational hardware resources. For example, they noted that the industry still uses the same methods to analyze catalysis that were developed more than 15 years ago, just on faster computers. In order to be able to go from particle based methods like DPD to continuum models, new methods and algorithms need to be developed. Moreover, they identified the need for parallel algorithms to be able to scale up to larger computational resources, especially in areas such as quantum chemistry, where 2 orders of efficiency improvements are needed in order to have an impact in that area. Overall, they characterized the impact factors of the key improvements as follows: 30% computing, 30% algorithms, 40% physics.

MODELING AND SIMULATION INFRASTRUCTURE

BASF has significant human and computational resources in-house devoted to modeling and simulation activities. They have 15 full time staff in scientific computing, who are also trained mathematicians, physicists, or chemists. The scientific computing staff supports other divisions in modeling and simulation activities. Other modeling groups are in polymer reactions (10), atomistic and molecular modeling in polymer research (15), process engineering/reaction modeling (8), CFD (10), and bioinformatics (15). BASF also funds extramural research with universities and companies. For example, it has a project with Heidelberg University that is jointly funded by the German Research Foundation (DFG), and BASF funds projects in the area of process scheduling at Berlin University and Princeton University. The intellectual property policy of the company allows the algorithms to stay with the developers. BASF management sees the value for the company in using these algorithms for internal use. The company has several Linux clusters in the range of 100 to 200 CPUs per group in-house. They feel that more computational power may not help unless new methods are developed.

There are several funding mechanisms for modeling and simulation activities. The internal resources for research activities come in part from business units and also from central funds. Overall, BASF's R&D expenditure was about €1.38 billion in 2007 with increasing tendency (forecast 2008: €1.45 billion). The company also participate in several projects funded by BMBF and under the EU 7th Framework Program. For example, BASF is the coordinator for the NanoModel project. The company is also working on two EU platforms, one in theoretical chemistry and another called FP3 Factory.

BASF management feels that it is important for simulation and modeling experts to have an interdisciplinary background. They need expertise in the domain, processing, analysis, systems, and modeling. The WTEC team's hosts emphasized the need for changes in academia to train scientists and engineers who have all the interdisciplinary skills needed to be successful simulation experts in industry.

CONCLUSIONS

BASF is the world's leading chemical manufacturer. Modeling and simulation activities are pervasive in most of its business segments. The key issues of interest to BASF in this area seem to revolve around coupling experimental data with modeling methods to get better estimates of model parameters, developing predictive capabilities for new products, developing new methods to handle and visualize large amounts of data, and creating workflow tools to easily link multiscale, multiphysics models.

Site: Center for Atomic-Scale Materials Design (CAMD)
Technical University of Denmark Department of Physics
Bldg. 311, DK-2800 Lyngby, Denmark
<http://www.camd.dtu.dk>
<http://www.dtu.dk>

Date Visited: February 27, 2008

WTEC Attendees: G. Karniadakis (report author), A. Deshmukh, C. Sagui, P. Westmoreland, G. Lewison

Hosts: Professor Jens K. Nørskov, CAMD Director
Tel: +45 4525 3175; Fax: +45 4525 2399
Email: norskov@fysik.dtu.dk; website: <http://dcwww.camp.dtu.dk/~norskov/>
Professor Karsten W. Jacobsen, CAMD Assistant Director
Tel: +45 4525 3186; Fax: +45 4525 2399
Email: kwj@fysik.dtu.dk; website: <http://dcwww.camp.dtu.dk/~kwj/>
Assistant Professor Jan Rossmeisl, CAMD Group leader, Theoretical Electrochemistry
Tel: +45 4525 3166; Fax: +45 4593 2399
Email: jross@fysik.dtu.dk; website: <http://dcwww.camd.dtu.dk/~jross/>
Assistant Professor Thomas Bligaard, CAMD Group leader, Theoretical Surface Science & Materials Informatics
Tel: +45 4525 3179; Fax: +45 4593 2399
Email: bligaard@fysik.dtu.dk; website: <http://dcwww.camd.dtu.dk/~bligaard/>
Assistant Professor Kristian S. Thygesen, CAMD Group leader, Molecular Electronics
Tel: +45 4525 3188; Fax: +45 4593 2399
Email: thygesen@fysik.dtu.dk; website: <http://dcwww.camd.dtu.dk/~thygesen/>
Ole Holm Nielsen, Head of Computer Services at CAMD
Tel: +45 4525 3187; Email: ole.h.nielsen@fysik.dtu.dk
website: <http://dcwww.camp.dtu.dk/~ohnielse/>

BACKGROUND

The Center for Atomic-scale Materials Design (CAMD) plays a leading role internationally in the development and use of molecular modeling and simulation for catalysis and material synthesis. In an earlier form, it was the Center for Atomic-Scale Materials Physics (CAMP), which had been established in 1993 through the Danish National Research Foundation at the Technical University of Denmark (Danmarks Tekniske Universitet) and the University of Aarhus. CAMD was established by The Lundbeck Foundation, which provides approximately \$1 million per year of funding for five years. It is also funded by external research grants and company collaborations, including an industrial affiliates program with yearly membership fees. Total annual funding is about \$3 to \$4 million per year.

CAMD also operates one of the four national supercomputing sites. Originally there had been a National Computing Center, which now has been demolished and replaced with four university clusters. CAMD operates Niflheim, a 1600-cpu Linux cluster with 2836 Gb storage and a peak performance of 9 teraflops. A typical problem size is an electronic-structure calculation on 200 to 300 atoms with every data point taking about a week on 16 to 32 processors. They are tied to Nordunet, the Scandinavian high-speed Internet backbone (<http://www.nordu.net/ndnweb/home.html>). At present, CAMD is not doing any grid computing, although the intent is to link decentralized resources over the grid. Part of the reason is that local computing is oversubscribed by 300 to 400%, leaving no resources available to share over the grid.

The current staffing level is five senior scientists, three PhD physicists staffing the computer center, 13-post-docs, 18 PhD students, and 10 project students. CAMD also works with the Danish National Research

Foundation's Center for Individual Nanoparticle Functionality (CINF), which focuses on experimental surface and nanomaterials physics (<http://www.cinf.dtu.dk>).

RESEARCH AND DEVELOPMENT

The stated objective of CAMD is to solve the inverse problem of designing materials based on desired properties; i.e., to develop methodology for systematic computational design of new materials and functional nanostructures while gaining insights into materials science. Theory development and experimental data are integrated with materials informatics and electronic-structure calculations, mainly using electronic density-functional theory (DFT). Two example projects were described:

- Simulation transformed the understanding of ammonia synthesis on ruthenium catalysts as occurring at step sites (crystallographic edges) rather than on open crystallographic faces (Honkala et al. 2005). Better, more affordable catalysts were sought. It has long been recognized that catalytic activity typically has a maximum among different metals when plotted against the heat of adsorption, due to competition between adsorption rate, which rises with increasing temperature, and equilibrium adsorption extent, which decreases with temperature. A first-principles kinetics model was developed using DFT-GGA and Monte Carlo simulation of surface coverage. Calculated energies of dissociative adsorption extended the work to methanation, revealing that the alloy Fe₃Ni had the optimal properties (Anderson et al. 2006). That finding has been verified and implemented industrially by Haldor Topsøe.
- The second achievement was in electrolytic generation of hydrogen (Greeley et al. 2006). This study applied computational high-throughput screening to search for improved materials. Full DFT calculations were performed for 736 symmetrically distinct surface models as metal slabs, involving 16 elements. The analysis indicated that a surface alloy of platinum and bismuth was optimal.

Work on enzyme catalysis was mentioned, including analysis of nitrogenase and hydrogenase functions.

Development of electronic-structure methods is another central activity of CAMD. The CAMD Open Software project (CAMPOS) provides an "Atomic Simulation Environment" that uses Python-based scripts and interfaces:

- Asap, CAMD's classical molecular-dynamics code
- Dacapo, CAMD's planewave ultra-soft-pseudopotential code
- GPAW, a grid-based projector-augmented wave DFT method
- MMTK, the open-source Molecular Modeling Toolkit of Konrad Hinsen (<http://dirac.cnrs-orleans.fr/MMTK/>) that provides a library of molecular-simulation capabilities
- SIESTA, the Spanish Initiative for Electronic Simulations with Thousands of Atoms, a large-scale DFT code due to Pablo Ordejón at the Universidad Autónoma de Barcelona (see site report) and his coworkers

A recent important step has been development of force fields using DFT and approximations with error bars. Validation and verification use internal benchmarking against other electronic-structure codes like VASP and Gaussian. CAMD researchers have also developed a Java-based "Virtual Materials Design Framework" that they consider to be the first systematic theory-based tool to search for materials based on desired properties. It includes a set of databases, filters, and visualization tools. The license to use it is sold cheaply to academic institutions, and databases are open-source.

DISCUSSION

Educational approaches for simulation at DTU were discussed. In this department there are BS, MS, and PhD degrees offered in Physics and Nanotechnology. From the beginning of their studies, physics students are oriented to modeling and to programming. The introductory courses use MATLAB/Maple, and students take a programming course in Java and an introductory numerical methods course. All courses at the MS and PhD levels are in English so long as there is at least one non-Danish student in the class.

However, there is no degree or program in simulation. Development of major new codes is done by staff, not by students, and this is considered an important part of their advances and continuity. Three people with Physics or Chemistry PhDs were hired by the department to develop codes, and the belief is that without this approach, CAMD could not do what it is doing. This approach is prevalent across the university. Of course, students are involved in working with the codes and helping refine them.

REFERENCES

- Andersson, M.P., T. Bligaard, A. Kustov, K.E. Larsen, J. Greeley, T. Johannessen, C.H. Christensen, and J.K. Nørskov. 2006. Toward computational screening in heterogeneous catalysis: Pareto-optimal methanation catalysts. *J. Catal.* 239:501.
- Honkala, K., A. Hellman, I.N. Remediakis, A. Logadottir, A. Carlsson, S. Dahl, C.H. Christensen, and J.K. Nørskov. 2005. Ammonia synthesis from first principles calculations. *Science* 307:555.
- Greeley, J., T.F. Jaramillo, J. Bonde, I. Chorkendorff, and J.K. Nørskov. 2006. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature Materials* 5:909.

Site: CERN (European Organization for Nuclear Research)
 CH-1211, Geneva 23, Switzerland
<http://public.web.cern.ch/Public/Welcome.html>
<http://press.web.cern.ch/press/PressReleases/Releases2008/PR01.08E.html>

Date Visited: February 28, 2008

WTEC Attendees: S. Glotzer (report author), L. Petzold, C. Cooper, J. Warren, V. Benokraitis

Hosts: Dr. James Shank, Research Professor, Center for Computational Science, Boston University, and Executive Program Manager for US-ATLAS Computing
 Tel: (617) 353-6028; Email: shank@bu.edu

Dr. Homer A. Neal, Samuel A. Goudsmit Professor of Physics, University of Michigan, and Director, UM-ATLAS Collaboratory Project
 Tel: (734) 764-4375; Email: haneal@umich.edu

Dr. Steven Goldfarb, Assistant Research Scientist, University of Michigan
 Tel: +41 (22) 767 1226; Email: Steven.Goldfarb@cern.ch

BACKGROUND

CERN, The European Center for Nuclear Research, is the world's largest particle physics research center. Here, scientists use giant machines—particle accelerators and detectors—to study the smallest objects in the universe. CERN's large accelerator rings reside 100 meters (320 feet) underground, beneath vineyards and pastureland along the French/Swiss border. They are elegant and vital tools for researchers pursuing questions about the origins of matter and the universe. CERN, with headquarters in Geneva, is one of the world's leading laboratories for particle physics, if not the leading laboratory. At present, its member states are Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, The United Kingdom, India, Israel, Japan, the Russian Federation, the United States, Turkey, the European Commission, and UNESCO have Observer¹⁷ status.

The ATLAS project is a worldwide collaboration comprising over 2100 scientists and engineers from 167 institutions in 37 countries and regions, including Argentina, Armenia, Australia, Austria, Azerbaijan, Belarus, Brazil, Canada, Chile, China, Colombia, Czech Republic, Denmark, France, Georgia, Germany, Greece, Hungary, Israel, Italy, Japan, Morocco, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Taiwan, Turkey, United Kingdom, and the United States. On February 29, 2008, the ATLAS collaboration at CERN celebrated the lowering of its last large detector element. The ATLAS detector is the world's largest general-purpose particle detector, measuring 46 m long, 25 m high and 25 m wide; it weighs approximately 7000 tons and consists of 100 million sensors that measure particles produced in proton-proton collisions in CERN's Large Hadron Collider (LHC; http://en.wikipedia.org/wiki/Large_Hadron_Collider).

The first piece of ATLAS was installed in 2003; since then, many detector elements have journeyed down the 100 m shaft into the ATLAS underground cavern. Known as the “small wheel,” this is the final element to complete the ATLAS muon spectrometer. There are two ATLAS small wheels; though small in comparison to the rest of the ATLAS detector, each detector is 9.3 m in diameter and weighs approximately 1 MN, including massive shielding elements. Each detector is covered with sensitive detectors to identify and measure the momentum of particles that will be generated in the LHC collisions. The entire muon spectrometer system contains an area of greater than 13,000 square meters, including 1.2 million independent electronic channels. The detector has the ability to track them with a positional accuracy of a few tens of micrometers. In the process, the subatomic particles are passed through a magnetic field produced by superconducting magnets so as to cause the

¹⁷ For more information about CERN and roles of member and observer states, please see <http://public.web.cern.ch/Public/en/About/Global-en.html>.

particles to bend; the trackers allow measurement to be made of the extent of the bending, and this allows researchers to determine the momentum of the observed particles

With this final component of the collider/detector in place, experiments may commence. The ATLAS collaboration will focus now on commissioning work in preparation for the start-up of the LHC this summer. Experiments at the LHC will allow physicists to take a big leap on a journey that started with Newton's description of gravity. Gravity is ubiquitous since it acts on mass; to date, however, scientists have been unable to explain why particles have the masses they have. Experiments such as ATLAS may provide the answer. LHC experiments will also probe the mysterious dark matter and energy of the Universe, they will investigate the reason for nature's preference for matter over antimatter, probe matter as it existed close to the beginning of time and look for extra dimensions of space-time. Dr. Shawn McKee, Associate Research Scientist at University of Michigan (smckee@umich.edu; 734-764-4395), and CERN host, Dr. James Shank, lead Tier2 ATLAS projects at University of Michigan and Boston University, respectively.

SBE&S RESEARCH

At the beginning of construction of the LHC, simulation in the form of finite element modeling (FEM) was used to estimate the deflection expected in the detector, itself. Using this approach, the maximum deflection (sag) in the 22 m diameter detector, caused by its own weight of 7000 tons (~9 kN), was predicted to be 25 mm, which required compensation during the design stage. This was the first application of simulation to the LHC project. Accurate estimation of deflection to the cabin was critical to achieve the required alignment of the proton beam of the LHC, and thus this example of civil engineering, in which the mass and volume of the detector was compensated by the removal of bedrock, is critical to the success of forthcoming experiments. Similarly, in the design of the electromagnetic (EM) accelerator, simulation of beam transport included EM forces on beams, a requirement in the design of the system that transports the proton beam around the 27 km-circumference accelerator.

Since completion of the design of the collider, scientific simulation has been used extensively to predict the outcome of many high-energy physics (HEP) experiments that are planned following completion of the construction project. Indeed, the entire LHC project would not be possible without simulations to predict the outcome of events which will then be compared against actual data as the experiment goes online.

In 1998, the Monarch project, led by H. Newman of Caltech, was established to distribute computing for the LHC, the goal of which was the generation of a general solution rather than one that was specific to HEP. While this began as its own dedicated grid, it has evolved to become a virtual grid that is supported by the Open Science Grid (<http://www.opensciencegrid.org/>) in the United States and its counterpart in the EU, Enabling Grids for E-science project (<http://www.eu-egee.org/>). This led to the establishment of worldwide collaborations in HEP and consumed the majority of HEP resources in the EU and US and resulted in the development—by computer scientist Miron Livny (<http://pages.cs.wisc.edu/~miron/>), colleagues at the University of Wisconsin-Madison, Ian Foster of Argonne National Laboratory and the University of Chicago., and many others—of a virtual toolkit and distributed computing grid, the capacity of which currently is 60,000 cores but which is expected to grow, in the near-term, to 100,000 cores.

Moreover, Monte Carlo simulations are being applied to predict detector output. In general, the researchers who participate in the experimental phase of the LHC project are seeking events that violate or as yet evade elucidation by the HEP Standard Model (http://en.wikipedia.org/wiki/Standard_Model), such as the Higgs boson (http://en.wikipedia.org/wiki/Higgs_boson). The Higgs boson is the only Standard Model particle not yet observed experimentally, but its presence would help explain how otherwise massless elementary particles still manage to construct mass in matter. In particular, it would explain the difference between the massless photon and the relatively massive W and Z bosons. Elementary particle masses, and the differences between electromagnetism, caused by the photon, and the weak force, caused by the W and Z bosons, are critical to many aspects of the structure of microscopic matter. As of April 2008, no experiment has directly detected the existence of the Higgs boson, but this may change as the Large Hadron Collider (LHC) at CERN becomes operational.

GEANT4 (<http://en.wikipedia.org/wiki/Geant4>), shorthand for GEometry ANd Tracking, is a platform for the simulation of the passage of particles through matter. It uses large, open-source Monte Carlo code and is an integral element of the international LHC initiative. GEANT4 now contains all current accumulated HEP knowledge that describes the interaction of high-energy particles with materials, which is based largely on experimental data. In addition, the GEANT4 data repository is applicable to medical physics, in that it provides insight into the interaction of ion beams with cells and tissues and is useful in guiding clinicians in radiation treatment of certain cancers. The GEANT4 initiative at CERN is augmented by that at Fermilab (<http://www.fnal.gov/>), the latter of which includes a radiography program, and the Stanford Linear Accelerator Center (SLAC; <http://www.slac.stanford.edu/>).

Data challenges at CERN are immense, and include management of the results from very large-scale simulations and the comparison of that data to experimental data. The simulation of 100 million events (collisions), for example, produces 1 PByte of data.

In calendar year 2007, ATLAS used 6000 CPU-years for simulation, which is only a fraction of the computing resources that will be required for the collection of data once experiments begin. LHC experiments are expected to generate 100 PB of data per second per experiment. Dedicated electronics is used to significantly reduce this data volume by filtering it to remove data that merely corroborate previously observed phenomena and the Standard Model. It is expected that the resulting, archived data will be reduced to tens of PB per year. Simulation is used to calibrate existing data and to advance the understanding of the sensitivity of the detector to new physical phenomena. The various LHC experiments plan to simulate between 20% and 100% of the equivalent amount of real data that they will gather each year, requiring 10 of thousand's of CPU-years to produce.

The discovery of Chi B b,d,s quarks, first reported in June 2007, although motivated by planning in ATLAS, was derived from data taken at Fermilab by Neal's D0 group¹⁸ at the University of Michigan (Drs. Neal, Qian, and Burelo). It was the first observation of a particle containing quarks from each of the so called quark "generations" (consisting of a b-quark, a d-quark and an s-quark). This particle, called the Xi_b hyperon, was predicted by the Standard Model but had never been seen. Some hidden combinatorial rule might have prevented its existence and thus challenged the Standard Model. But, in fact, Neal's group confirmed the Standard Model prediction. These results were recently published in *Physical Review Letters* (D0 Collaboration 2008). In this discovery (driven by the Michigan D0 Group), millions of events were simulated in the D0 detector, assuming that there were no unusual correlations between the resulting particles. The suggestion that a new particle had been found came from the observation that correlations did exist between certain final state particles in the actual data, even though such correlations were not predicted in the simulated data. Deployment of visualization software helped the researchers to further confirm that what they were seeing was indeed the genuine cascaded decay of a new particle. This is an example of how simulations can help scientists find unexpected behavior and then be led to explore more deeply what might be the underlying physical origin.

For the training of next-generation high-energy physics (HEP) scientists, simulation is very important because of the paucity of experimental data. HEP scientists are in great demand by industry; for example, 18 PhD-level HEP scientists from the D0 experiment have been hired recently by Lucent to work in areas such as database management and simulation.

REFERENCES

D0 Collaboration. 2008. Measurement of the forward-backward charge asymmetry in top-quark pair production. *Physical Review Letters* 100: 142001–142007.

¹⁸ The D0 experiment (<http://www-d0.fnal.gov/>; see also Wikipedia) is a collaboration of scientists consisting of several groups around the world who conduct research on the fundamental nature of matter, with experiments based at the Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois (USA). The D0 collaboration is named for one of the proton/antiproton interaction regions, where beams on the Tevatron synchrotron ring intersect.

Site: CIMNE (International Center for Numerical Methods in Engineering)
Edificio C1, Campus Norte UPC
Gran Capitan, s/n. 08034 Barcelona, Spain
<http://www.cimne.com>

Date Visited: February 29, 2008

WTEC Attendees: G. Karniadakis (report author), A. Deshmukh, C. Sagui, P. Westmoreland, G. Lewison

Host: Professor Eugenio Onate, CIMNE Director
Email: onate@cimne.upc.edu; Tel: 34 932 057 016
Mr. Pere-Andreu Ubach de Fuentes, R&D Project Strategies
Email: ubach@cimne.upc.edu; Tel: 34 934 017 399

BACKGROUND

CIMNE was created in 1987 under the auspices of UNESCO by the autonomous government of Catalonia. It is the second oldest independent research center in Catalonia. CIMNE has its own juridical status as a consortium between the government of Catalonia and the University Polytechnic of Catalonia (UPC), and now Spain's federal government will join the consortium. The headquarters of CIMNE are located at the department of Civil Engineering of UPC. CIMNE has also offices for training in 15 classrooms distributed around the world. CIMNE's mission is three-fold: (1) Research, (2) Training activities, and (3) Technology transfer. Currently, more than 90% of its funds come from external sources, and its annual funding level is close to €10 million. There are currently 190 employees, and with the new national government funding, 40 more employees will be added. Specifically, about 45 employees are permanent research staff, 15 are administrative personnel, and the rest are researchers in training.

RESEARCH AND DEVELOPMENT

CIMNE has been one of the main research centers focusing on the development and application of finite element methods (FEM) in computational mechanics applications—in particular, stabilized FEM. A unique method developed at CIMNE is a hybrid FEM/particle approach that allows modeling of free surfaces and moving boundaries more easily; impressive results were shown for marine engineering applications. While the long-term focus has been structural and fluid mechanics, more recent work has been on electromagnetics, design and optimization, food engineering, and other multidisciplinary problems. The focus and strength of CIMNE researchers are multiphysics rather than multiscale applications. There is no particular emphasis on parallelization of their codes as they can currently fit up to 12 million FEM on a high-end PC; all of the codes used at CIMNE were developed in-house, including pre- and postprocessing, and many of their codes are now commercial software products. For example, GiD (www.gidhome.com) is a commercial PC-based pre/postprocessor with a very wide user base in industry and academia, and it is interfaced with many commercial codes, such as NASTRAN.

CIMNE is also focusing on *verification and validation* research, and currently it funds an experiment that will quantify forces on a concrete block by flow motion. This will be used as a benchmark for validating flow-structure interactions models that it has been developing. However, there is no particular emphasis on uncertainty quantification based on stochastic modeling techniques.

CIMNE has a very strong interactions and links with industry, with about half of its research and development projects funded by industry, e.g., aerospace, marine, civil, mechanical, biomedical, and metal forming companies. CIMNE researchers bid for about 150 proposals per year and in their 20 years' history have had 141 European Commission projects, 217 national projects, and 452 industry projects. CIMNE is a shareholder in three companies, specializing in marine engineering, civil and structural engineering, and aeronautics and space applications.

EDUCATION, TRAINING, AND DISSEMINATION

CIMNE organizes courses and seminars related to the theory and application of numerical methods in engineering. The attendees are recent university graduates and also professionals. Since 1987 CIMNE has organized about 100 courses and 300 seminars. A new program is the *Erasmus Mundus Master course*, which is an international course for masters in computational mechanics for non-European students. In its first year at the time of the WTEC team's visit (2008), the course had 30 students. One of the requirements is that the students spend equal time at least in two of the four different universities involved in the course (Barcelona, Stuttgart, Swansea, and Nantes).

CIMNE has developed a Web environment for distance learning education via the Internet. The Virtual Center of Continuing Education of CIMNE gathers information early on in a course to facilitate the registration process. Teachers can also follow the students and carry out different tutorials and exercises. This center hosts the Master Course in Numerical Methods in Engineering and other postgraduate courses of CIMNE (<http://www.cimne.upc.edu/cdl>).

CIMNE has been a leader in disseminating information on computational mechanics by organizing more than 80 national and international conferences; 13 of these conferences took place in 2007. CIMNE also publishes books, journals, monographs, scientific computing reports, and educational software related to theory and applications of numerical methods in engineering. In its first 20 years, CIMNE published 101 books, 15 educational software packages, 296 research reports, and 492 technical reports. CIMNE researchers publish about 60 papers per year.

CIMNE classrooms are physical spaces for cooperation in education, research, and technology development, created jointly by CIMNE and several other universities. They are located in Barcelona but also in other cities in Spain and around the world, including Mexico, Argentina, Colombia, Cuba, Chile, Brazil, Venezuela, and Iran.

CONCLUSIONS

CIMNE is a unique research and development center that has done a lot in its 20-year history to promote and foster both fundamental and technological advances associated with computational methods for engineering problems. Although its current emphasis is on multiphysics applications, CIMNE researchers still publish papers on fundamental mathematical topics, e.g., new formulations and error estimation and bounds in FEM, but also on optimization and design of complex systems, and they are branching out into new applications with great potential impact, e.g., food engineering and decision support systems. CIMNE's education history and recent development of a four-campus MS program for international students as well as its distance learning and training activities are impressive and are having great impact in Spain, in Europe, and around the world.

Site: Ecole Polytechnique Fédérale de Lausanne (EPFL)
 Institute of Analysis and Scientific Computing (IACS)
 Department of Modeling and Scientific Computing (CMCS)
 Department of Mathematics
 SB, IACS-CMCS, Station 8
 CH-1015 Lausanne, Switzerland
http://iacs.epfl.ch/index_e.html

(Remote site report) Politecnico di Milano (Polimi)
 Laboratory for Modeling and Scientific Computing (MOX)
 p.za Leonardo da Vinci 32
 20133 Milan, Italy
<http://mox.polimi.it/>

Date Visited: February 23, 2008.

WTEC Attendees: S. Glotzer (report author), L. Petzold, C. Cooper, J. Warren, V. Benokraitis

Hosts: Professor Alfio Quarteroni, Professor and Chair of Modeling and Scientific Computing at the Institute of Analysis and Scientific Computing of EPFL;
 Professor of Numerical Analysis, Polimi, and Scientific Director, MOX
 Email: alfio.quarteroni@epfl.ch
 Dr. Simone Deparis (post-doc), EPFL
 Email: simone.deparis@epfl.ch
 Marko Discacciati (post-doc), EPFL
 Email: marco.discacciati@epfl.ch

BACKGROUND

This site report is both a direct report of the system at the Ecole Polytechnique Fédérale de Lausanne (EPFL) Department of Mathematics and a remote site report for the Politecnico di Milano Laboratory for Modeling and Scientific Computing (Modellistica e Calcolo Scientifico, or MOX) in Italy, as Professor Quarteroni holds positions at both these institutions. He has been Chair of Modeling and Scientific Computation (CMCS) at EPFL/CMCS since 1998, and director of MOX in Milan since 2002 (as well as Professor of Numerical Analysis at Polimi since 1989). Prof. Quarteroni was kind enough to prepare overviews of both institutions for the WTEC team that visited EPFL.

Professor Quarteroni studied mathematics at the University of Pavia, Italy, and at University of Paris VI.¹⁹ In 1986 he became full professor at the Catholic University of Brescia, later professor in mathematics at the University of Minnesota at Minneapolis, and professor in numerical analysis at Politecnico di Milano. He was appointed full professor at EPFL in 1998. His team carried out the aerodynamic and hydrodynamic simulations for the optimization of *Alinghi*, the Swiss sailing yacht that won the last two America's Cup races in 2003 and 2007.

In the Swiss system (such as EPFL), the position of chair comes with tenure, while other faculty with up to four-year contracts work with professors. Even these temporary positions are considered to be good jumping-off points for careers in SBE&S by those the WTEC team members talked to. It is typical that scientists obtain their PhDs abroad but return to become professors at EPFL. The international outlook at EPFL is reflected in the makeup of its student body and faculty, as foreign nationals comprise one-third of the undergraduate students, two-thirds of the graduate students, and three-fifths of the professors. Indeed, EPFL has been hiring full

¹⁹ Pierre & Marie Curie University, one of the 13 universities descended from the medieval University of Paris.

professors from the United States in increasing numbers. Professors can count on having research support for several funded graduate students indefinitely. The excellent infrastructure is a strong attraction for researchers.

In the Italian system, a hierarchy of professors who all work in a more coordinated fashion, with direction from the full professor. MOX was founded in 2002 with three faculty members, a budget of \$100,000, and 120 freshmen. It created a new, now very popular, BS/MS curriculum in mathematical engineering, focused one-third on basic engineering, one-third on fundamental math, and one-third on applied math. Students who work on an externally funded project can generally obtain higher salaries. Funding in MOX comes, roughly equally, from industry, public funding agencies, private funding agencies, and public European projects.

Eight years ago, all countries in Europe participated in what is known as the Bologna reform; it is now a universal 3+2 (BS/MS equivalent) program. This uniformity allows students mobility throughout Europe. As the program is so good, the job prospects are excellent for individuals with master's degrees, which has the unintended consequence of making talented PhDs much more difficult to obtain.

SBES RESEARCH

EPFL/CMCS

The CMCS projects focus on fluid flow problems with numerous areas of application. Medicine and water sports are primary interests.

At CMCS there is a project for extracting geometry from raw data, to get a 3D “map” of the arterial system. These reconstructions are then turned into a finite element structure that can provide the basis for detailed simulations of blood flow.

Another project is attempting a classification of aneurisms by their geometrical structure. *A posteriori* error analysis is difficult in this problem because of the complexity of fluid-structure interaction. Differences in the metabolic responses of different patients add a lot of uncertainty to the model. It has been found that the flow field is very sensitive to the geometry.

Particular attention is paid to difficulties in attaching different length and time scales (the so-called multiscale problem) to the cardio-vascular system. They have developed a 3D flow model, while the major arteries and veins are sensibly modeling in 1D, and currently the capillary network is treated as 0D; interfacing these different dimensionalities has proved nontrivial. The models have substantial similarities to electronic circuit models (not surprisingly as they are governed by underlying current conservation laws).

Perhaps the most spectacular modeling presented by Professor Quarteroni concerned the CMCS work in mathematical models for sport, specifically “Multiphysics in America’s Cup,” a tour-de-force in fluid flow modeling. This was a formal collaboration with the *Alinghi* designers, with substantial proprietary information being shared by the designers. Commercial codes were used, initially Fluent, then CFX, because of the perceived superiority of the included turbulence models. The implementation required a supercomputer. (In order to iterate the design with the developers, a turnaround time of 24 hours was required.) The simulations had approximately 20 million elements, with 162 million unknowns in the solution. The work was rewarded with two America’s Cup victories.

MOX

The MOX program has as its mission to develop, analyze, and apply numerical models for engineering problems and real-life applications, enhance the scientific cooperation with other departments of Polimi, foster the collaboration with industrial partners as well as national and international institutions through funded research projects, and to organize short courses open to industrial and academic researchers. The MOX research group, based in the Polimi Mathematics Department, is comprised of 3 full professors, 10 assistant professors, 2

associates, 4 post-docs, 4 post-grad research assistants, and 13 PhD students. There are over a dozen industrial partners, with closely integrated collaborative research on a number of topics.

MOX maintains a number of both fundamental and applied research areas. Fundamental research is being pursued in (i) Adaptive numerical methods and error control for PDEs; (ii) control and shape optimization; (iii) numerical methods for fluid dynamics; (iv) level set methods; and (v) fixed, hybrid, and discontinuous finite elements for PDEs. Collaborators include Massachusetts Institute of Technology, Virginia Tech, EPFL, University of Texas Austin, University of Maryland, University of Santa Fé (Argentina), University of Trento (Italy), INRIA-Roquencourt (France), and Imperial College (UK).

MOX is pursuing a number of applications, many of which are addressing the state-of-the-art in continuum modeling, both in algorithmic development and in the application of HPC to real-world engineering applications. One application that was discussed, Geophysical modeling of the evolution of sedimentary basins, required calculations that involve large space and long time scales, necessitating adaptive meshes, Lagrange tracking of interfaces, topological changes, and methods for non-Newtonian fluids. This application required the development of production codes for use by the oil industry.

Other applications of interest are in the life sciences, in particular, drug delivery from nano/microstructured materials and controlled drug delivery with application to drug-eluting stents (funded by Fondazione Cariplo and Istituto Italiano di Tecnologia, IIT). Additionally, a highly sophisticated analysis is being done of the complex fluid-dynamics of blood flow in the cardiovascular system.

An impressive collaboration with surgeons in London has evolved. Research involves multiscale simulations of the circulatory system. The surgeon can use the simulation to help decide which procedure to use on the implant of cavopulmonary shunts on infants with congenital heart diseases. Professor Quarteroni observed that it, appropriately, takes a long effort to gain the surgeon's trust. To initiate such a collaboration requires especially motivated doctors, and the modeler needs to prove his techniques. However, after a number of successes, he has the enviable position of having too many good opportunities and must be selective—many doctors want to work with him. Often the bridge between the modeler and the doctor is a bioengineer.

Students who are involved in industrial collaborations receive a lab course credit. Indeed, there is a large number of industrial collaborations:

- ENI SpA, Divisione E&P: Modeling geological structure evolution in complex situations in 2D and 3D
- Arena Italia SpA: Numerical models for the performance analysis of a swimmer
- CTG Italcementi Group: Numerical simulation of an air-combustion gas mixer
- Filippi Lido Srl: Mathematical and numerical models for the dynamics of rowing boats
- RSI Tech: Mathematical modeling and numerical simulation of the propagation of a georadar signal in a hydraulic fracture
- INDAM: Mathematical and numerical modeling of the electro-fluid-mechanics of the heart
- MIUR–Cofin: Numerical models in fluid dynamics with application to the simulation of the cardiovascular system and the environment
- IIT (Italian Institute of Technology): Nanobiotechnology models and methods for local drug delivery from nano/micro structured materials
- Fondazione Cariplo: Development of mathematical and numerical models for optical devices based on nanostructured electrochemical produced materials; Mathematical modeling of drug release from drug eluting stents into the arterial wall.
- Siemens, Fondazione Politecnico: Numerical and statistical methods for assessing the risk of rupture in cerebral aneurisms

Software

The Polimi system allows for the development of a substantial code base. Multiple research foci, derived from industrial collaborations, medical research collaborations, as well as in-house motivated basic research interests, provide a critical mass of customers, and a large, motivated faculty team that maintains the code base. Substantial funding for these projects, including overhead, comes from industrial grants.

A number of software tools have been developed at MOX. They include, among several others,

- LifeV (www.lifev.org), a finite element library in C++ developed in collaboration with EPFL and INRIA
- Stratos, a finite element code for 3D free barotropic and baroclinic fields
- Steam2D/Steam++/Steam3D, a suite of computer codes for the simulation of the evolution of sedimentary basins (non-Newtonian Stokes flow with interfaces)
- Stratos-Wave, a simulation of boat dynamics by a variational inequality approach
- Kimè, a simulation of the dynamics of a rowing scull
- Glow1D/Glow2D, a simulation of the functioning of a glow plug for diesel engines
- MM1, a section-averaged model for river flow and sediment transport developed in collaboration with CUDAM—University of Trento
- MOX also has a number of software teaching modules that leverage off widely available software.

MOX has managed the difficult trick of maintaining a substantial code-base by ensuring continuity of the software development. The group is large enough, with a number of senior members. Professors can increase their salary by up to 50% via industrial money.

REPRESENTATIVE PUBLICATIONS OF THE EPFL/CMCS AND MOX GROUPS

- Abbà, A., C. D'Angelo, and F. Saleri. 2006. A 3D shape optimization problem in heat transfer: Analysis and approximation via BEM. *Math. Models Methods Appl. Sci.* 16(8):1243–1270.
- Agoshkov, V.I., P. Gervasio, and A. Quarteroni. 2006. Optimal control in heterogeneous domain decomposition methods for advection-diffusion equations. *Mediterr. J. Math.* 3(2):147–176.
- Babūška, I., F. Nobile, and R. Tempone. 2005. Worst case scenario analysis for elliptic problems with uncertainty. *Numer. Math.* 101(2):185–219.
- Badia, S., A. Quaini, and A. Quarteroni. 2008. Splitting methods on algebraic factorization for fluid-structure interaction. *SIAM J. Scientific Computing* 30:1778–1805.
- Burman, E., A. Quarteroni, and B. Stamm. 2008. Stabilization strategies for high order methods for transport dominated problems. *Boll. Unione Mat. Ital.* 1(1):57–77.
- Burman, E., and B. Stamm. 2008. Symmetric and non-symmetric discontinuous Galerkin methods stabilized using bubble enrichment. *C.R. Math. Acad. Sci. Paris* 346(1-2):103–106.
- Burman, E., and P. Zunino. 2006. A domain decomposition method based on weighted interior penalties for advection-diffusion-reaction problems. *SIAM J. Numer. Analysis* 44(4):1612–1638.
- Canuto, C., M.Y. Hussaini, A. Quarteroni, and T.A. Zang. 2006. *Spectral methods. Fundamentals in single domains*. Berlin: Springer-Verlag.
- Canuto, C., M.Y. Hussaini, A. Quarteroni, and T.A. Zang. 2007. *Spectral methods. Evolution to complex geometries and applications to fluid dynamics*. Berlin: Springer.
- D'Angelo, C., and A. Quarteroni. 2008. On the coupling of 1D and 3D diffusion-reaction equations. Application to tissue perfusion problems. *Math. Models Methods Appl. Sci.* (publication pending).
- Dedè, L., and A. Quarteroni. 2005. Optimal control and numerical adaptivity for advection-diffusion equations. *Math. Model. Numer. Anal.* 39(5):1019–1040.
- Deparis, S., M. Discacciati, G. Fourestey, and A. Quarteroni. 2006. Fluid-structure algorithm based on Steklov-Poincaré operators. *Comput. Methods Appl. Mech. Engrg.* 195:5797–5812.

- Deparis, S., M.A. Fernández, and L. Formaggia. 2003. Acceleration of a fixed point algorithm for fluid-structure interaction using transpiration conditions. *Math. Model. Numer. Anal.* 37(4):601–616.
- Di Pietro, D.A., S. Lo Forte, and N. Parolini. 2005. Mass preserving finite element implementations of the level set method. *Appl. Numer. Math.* 56(9):1179–1195.
- Discacciati, M., A. Quarteroni, and A. Valli. 2007. Robin-Robin domain decomposition methods for the Stokes-Darcy coupling. *SIAM J. Numer. Anal.* 45(3):1246–1268.
- Doğan, G., P. Morin, R.H. Nochetto, and M. Verani. 2007. Discrete gradient flows for shape optimization and applications. *Comput. Methods Appl. Mech. Engrg.* 196(37-40):3898–3914.
- Formaggia, L., and F. Nobile. 2004. Stability analysis of second-order time accurate schemes for ALE-FEM. *Comput. Methods Appl. Mech. Engrg.* 193(39-41):4097–4116.
- Formaggia, L., E. Miglio, A. Mola, and N. Parolini. 2007. Fluid-structure interaction problems in free surface flows: application to boat dynamics. *Internat. J. Numer. Methods Fluids* 56(8):965–978.
- Formaggia, L., J.F. Gerbeau, F. Nobile, and A. Quarteroni. 2001. On the coupling of 3D and 1D Navier-Stokes equations for flow problems in compliant vessels. *Comp. Methods Appl. Mech. Engrg.* 191(6-7):561–582.
- Formaggia, L., J.F. Gerbeau, F. Nobile, and A. Quarteroni. 2002. Numerical treatment of defective boundary conditions for the Navier-Stokes equations. *SIAM J. Numer. Anal.* 40(1):376–401.
- Formaggia, L., S. Micheletti, and S. Perotto. 2004. Anisotropic mesh adaption in computational fluid dynamics: application to the advection-diffusion-reaction and the Stokes problems. *Appl. Numer. Math.* 51(4):511–533.
- Massimi, P., A. Quarteroni, F. Saleri, and G. Scrofani. 2007. Modeling of salt tectonics. *Comput. Methods Appl. Mech. Engrg.* 197(1-4):281–293.
- Micheletti, S., and S. Perotto. 2006. Reliability and efficiency of an anisotropic Zienkiewicz-Zhu error estimator. *Comput. Methods Appl. Mech. Engrg.* 195(9-12):799–835.
- Micheletti, S., S. Perotto, and M. Picasso. 2003. Stabilized finite elements on anisotropic meshes: A priori error estimates for the advection-diffusion and the Stokes problems. *SIAM J. Numer. Anal.* 41(3):1131–1162.
- Migliavacca, F., F. Gervaso, M. Prosi, P. Zunino, S. Minisini, L. Formaggia, and G. Dubini. 2007. Expansion and drug elution model of a coronary stent. *Comput. Methods. Biomech. Biomed. Engin.* 10(1):63–73.
- Miglio, E., A. Quarteroni, and F. Saleri. 1999. Finite element approximation of quasi-3D shallow water equations. *Comput. Methods Appl. Mech. Engrg.* 174(3-4):355–369.
- Parolini, N., and A. Quarteroni. 2005. Mathematical models and numerical simulations for the America's cup. *Comput. Methods Appl. Mech. Engrg.* 194 (9-11):1001–1026.
- Parolini, N., and E. Burman. 2005. A finite element level set method for viscous free-surface flows. Applied and industrial mathematics in Italy, 416-427. *Ser. Adv. Math. Appl. Sci* 69, Hackensack, NJ: World Sci. Publ.
- Prosi, M., P. Zunino, K. Perktold, and A. Quarteroni. 2005. Mathematical and numerical models for transfer of low-density lipoproteins through the arterial walls: A new methodology for the model set up with applications to the study of disturbed luminal flow. *J. Biomechanics* 38:903–917.
- Quaini, A., and A. Quarteroni. 2007. A semi-implicit approach for fluid-structure interaction based on an algebraic fractional step method. *Math. Models Methods Appl. Sci.* 17(6): 957-983.
- Quarteroni, A., A. Veneziani, and P. Zunino. 2002. A domain decomposition method for advection-diffusion processes with applications to blood solutes. *SIAM J. Sci. Computing* 23(6):1959–1980.
- Quarteroni, A., and A. Veneziani. 2003. Analysis of a geometrical multiscale model based on the coupling of ODEs and PDEs for blood flow simulations. *Multiscale Model. Simul.* 1(2):173–195.
- Quarteroni, A., and F. Saleri. 2006. *Scientific computing with MATLAB and OCTAVE*, 2nd ed. Berlin: Springer-Verlag.
- Quarteroni, A., and G. Rozza. 2003. Optimal control and shape optimization in aorto-coronary bypass anastomoses. *Math. Models Methods Appl. Sci.* 12(13): 1801–1823.
- Quarteroni, A., and G. Rozza. 2007. Numerical solution of parametrized Navier-Stokes equations by reduced basis methods. *Num. Meth. Part. Diff. Eq.* 23(4):923–948.
- Quarteroni, A., R. Sacco, and F. Saleri. 2007. *Numerical mathematics*, 2nd ed. Berlin: Springer-Verlag.
- Restelli, M., L. Bonaventura, and R. Sacco. 2006. A semi-Lagrangian discontinuous Galerkin method for scalar advection by incompressible flows. *J. Comput. Physics* 216(1):195–215.

- Rozza, G., and K. Veroy. 2007. On the stability of the reduced basis method for Stokes equations in parametrized domains. *Comput. Methods Appl. Mech. Engrg.* 196(7):1244–1260.
- Saleri, F., and A. Veneziani. 2005. Pressure correction algebraic splitting methods for the incompressible Navier-Stokes equations. *SIAM J. Numer. Analysis* 43(1):174–194.
- Vergara, C., and P. Zunino. 2008. Multiscale modeling and simulation of drug release from cardiovascular stent. *SIAM J. Multiscale Modeling Simul.* 7(2):565–588.

Site: Ecole Polytechnique Fédérale de Lausanne (EPFL), Blue Brain Project
SV-BMC / AA-B118
CH-1015 Lausanne, Switzerland
<http://bluebrain.epfl.ch/>

Date Visited: February 26, 2008

WTEC Attendees: L. Petzold (report author), S. Glotzer, J. Warren, C. Cooper, V. Benokraitis

Hosts: Dr. Robert Bishop, EPFL, Chairman of Advisory Board to Blue Brain Project
Email: bob.bishop@epfl.ch
Dr. Sean Hill, IBM, Blue Brain Project Manager for Computational Neuroscience
Email: sean.hill@epfl.ch
Dr. Felix Schuermann, EPFL, Blue Brain Project Manager
Email: felix.schuermann@epfl.ch

BACKGROUND

In June 2005, IBM and the Brain Mind Institute at the Ecole Polytechnique Fédérale de Lausanne (EPFL) in Switzerland announced a plan to create a digital 3D replica of a neocortical column. The column is the elementary building block of the mammalian neocortex, which represents about 80% of the brain and is believed to house cognitive functions such as language and conscious thought.. Named after the IBM Blue Gene supercomputer it relies on, the Blue Brain Project in its first phase has the goal of modeling, in every detail, the cellular infrastructure and electrophysiological interactions of a single neocortical column. Professor Henry Markram, who founded the Brain Mind Institute, directs the Blue Brain Project. Dr. Felix Schuermann manages the project and its international team of 35 scientists.

R&D ACTIVITIES

The Blue Brain Project was started by Henry Markram, a biologist, motivated by his experimental biology work. The Neocortex takes up 80% of the human brain. The project started out to reverse-engineer the neocortex. This involved collecting experimental data for different types of cells, types of electrical behavior, and types of connectivity. The model includes 10,000 neurons, 340 types of neurons, detailed specific ion channels, and 30 million connections. This is a detailed multiscale model, related to variables as measured in a laboratory. It is a huge system of ordinary differential equations, modeling electrochemical synapses and ion channels. The main objective of the model is to capture the electrical behavior. A major effort has been put into developing a modeling and simulation workflow so that the model can be automatically generated from biological measurements, tested against experimental results, and revised with additional biological data in an iterative and ongoing process. The current capability is a faithful *in silico* replica at the cellular level of a neocortical column of a young rat. A goal is to capture and explain electrical, morphological, synaptic, and plasticity diversity observed in the lab. Building a database of experimental results is a major part of this effort.

Some of the long-term objectives of this effort are to model a normal brain and a diseased brain to identify what is going wrong, and to evaluate emergent network phenomena. A demonstration was given in which the model replicated slice experiments that exhibit network-scale phenomena. The simulation showed low-frequency cortically generated oscillations emerging from changing extracellular potassium. In this way, it may someday be possible to do *in silico* pharmacology experiments.

The Blue Brain Project models the 10,000 neuron cells, involving approximately 400 compartments per neuron, with a cable-equation for the passive properties and Hodgkin-Huxley models for the active ion channel contributions. There are 30 million dynamic synapses. The computing platform used is a 4-rack blue gene/L with a 22.4 TFlop peak and 8,192 processors. Additionally, an SGI Prism Extreme, with 300 GB shared memory and

16 graphics cards, is used for visualization. The project features an impressive integration of experiment, modeling, high-performance computing, and visualization.

SELECTED PUBLICATIONS

Druckmann, S., Y. Banitt, et al. 2007. A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Frontiers in Neuroscience* 1(1):7–18.

Hines, M., H. Markram et al. 2008. Fully implicit parallel simulation of single neurons. *Journal of Computational Neuroscience*. (In press; epub available ahead of print.)

Kozloski, J., K. Sfyakis, et al. 2008. Identifying, tabulating, and analyzing contacts between branched neuron morphologies. *IBM Journal of Research and Development* 52(1/2).

Markram, H. 2006. The blue brain project. *Nat. Rev. Neurosci.* 7(2):153–60.

Site: **Eni SpA**
Via Felice Maritano 26
20097 San Donato, Milan, Italy
http://www.eni.it/en_IT/home.html

Date Visited: February 28, 2008

WTEC Attendees: A. Deshmukh (report author), G. Karniadakis, G. Lewison, P. Westmoreland

Hosts: Dr. Francesco Frigerio, Physical Chemistry Department, Refining & Marketing Division
Tel: +39-02-52056443; Fax: +39-02-52036347
Email: francesco.frigerio@eni.it

Ing. Alfredo Battistelli, Environmental Remediation and Land Conservation
Snamprogetti SpA (now Saipem SpA)
Tel: +39-0721-1682467; Fax: +39-0721-1682984
Email: alfredo.battistelli@saipem.eni.it

Dr. Luciano Montanari, Technical Manager
Physical Chemistry Department, Refining & Marketing Division
Tel: +39-02-52046546; Fax: +39-02-52036347
Email: luciano.montanari@eni.it

Ing. Fabrizio Podenzani, Development of Technology, Process Engineering
Refining & Marketing Division
Tel: +39-02-52056717; Fax: +39-02-52036116
Email: fabrizio.podenzani@eni.it

BACKGROUND

Eni is one of the most important integrated energy companies in the world operating in the oil and gas, electricity generation and sale, petrochemicals, oilfield services construction and engineering industries. Eni is active in around 70 countries with a staff of more than 70,000 employees. Eni is composed of a Corporate and three major business divisions: Exploration & Production, Gas & Power, Refining & Marketing, as well as several affiliated and subsidiary companies, such as Snam Rete Gas, Saipem, and Polimeri Europa.

Eni R&D activities are housed mainly in the Refining & Marketing Division and in the Corporate Strategy Division. Eni Research & Development is one of the major centers of excellence for industrial research in Europe. It works in the entire oil and gas technology chain and in renewable sources, while safeguarding the environment, with a sustainable development perspective. The focus areas for technological innovation are: upstream oil and gas; natural gas conversion and bituminous sources; downstream gas; downstream oil—refinery and petrochemical processes; downstream oil—products development (fuels, lubricants, and specialties); advanced energy systems and renewable sources (photovoltaic); and environment and sustainability.

Eni Research & Development operates through all stages of innovation, from technology monitoring and scenario studies to applied research, as well as from the development and evaluation of technologies to the commercial exploitation of R&D results. Examples of research projects involving modeling and simulation include design of nanomaterials for catalysts and solar cells (nanowires), simulation of recovery process, noise transmission, and environmental impact studies in the Health, Safety, and Environment domain. A recent project involving simulation of product shape selectivity by zeolites produced an efficient nonquantitative screening method for the selection of a suitable alkylation catalyst. Snamprogetti, which is a former subsidiary of Saipem (SpA), focuses on simulation and modeling for structural analysis, CFD, and reservoir engineering issues, such as migration of contaminants in the subsurface and remediation of contaminated sites, underground injection of acid gases and geological sequestration of greenhouse gas mixtures, and diagenetic evolution of sedimentary basins by simulation of mass and energy transport in porous media. They primarily use conventional continuum codes, such as those belonging to the TOUGH2 family of reservoir simulators developed by the Lawrence Berkeley National Lab. Snamprogetti developed new code capabilities within the TOUGH2 architecture both in collaboration with

LBNL as well as within R&D projects sponsored by Eni SpA. New TOUGH2 codes developed in collaboration with LBNL are distributed to the public by the Energy Science and Technology Software Center (ESTSC) and the Nuclear Energy Agency (NEA). Eni also has a large R&D investment in renewable energy resources, particularly solar energy through a recent collaboration with MIT. They also have collaborations with PNL and use the codes developed there.

MODELING AND SIMULATION NEEDS

The WTEC team's research hosts at Eni highlighted some of their key issues in modeling and simulation:

- *Accessibility of simulation tools.* Currently the simulation tools available in industry are used by experts in computational methods, not people who are domain experts. In many cases, these two groups do not overlap. There is a severe need to develop simulation tools that can be used by domain experts, possibly with help or tailoring from the simulation experts.
- *Validation.* Eni researchers use data from internal company sources and/or literature. Most validation efforts are initiated using published data, which are then followed up with internal company data or new experiments (this does not happen often). The researchers do not often use field data in model validation.
- *Multiphysics, multiscale codes.* The Eni researchers identified this as a critical issue for the company. For example, they are interested in simulation and prediction of molecular behavior in industrial processes (molecular parameters that mainly influence product formation). Currently they have a small research project with Politecnico Milano to integrate Fluent and Ansys together. Another project focuses on COMSOL, which is a multiphysics framework developed by a Swedish company.
- *Knowledge management.* This was identified as another critical issue for Eni. It has some efforts underway in the area of analytical chemistry where IT tools have been used to exchange, archive, and retrieve knowledge generated in different projects across the organization.
- *Decision-making and enterprise-level modeling.* The R&D group is pretty far from the decision-makers and does not participate actively in the use of simulation models in decision-making. They provide feedback to the decision-makers when asked for more data or higher fidelity predictions. The materials and process-level modeling efforts are not tightly coupled with system or enterprise-level models.

Research Infrastructure

- The company does not generally get funding from the Italian government for research projects. It has some EU funding, but not specifically in simulation and modeling activities, with few exceptions related to the development of tools for the modeling of migration of organic mixtures in contaminated sites.
- Eni has direct relationships with universities, where the company provides funding for projects. Current collaborations with universities include MIT (solar energy), Politecnico Milan, Politecnico Turin, University of Turin, CNR Pisa, Bologna (modeling), University of Padova, University of Washington and Brescia. There are also collaborations ongoing with the National Research Council (IGG-CNR, Pisa).
- Eni typically owns all intellectual property developed under projects funded by the company.
- The research group had a unit devoted to computing when part of EniTecnologie. In a new organizational structure, a centralized computing service does not exist; each project works on its own computational needs and resources. Typical computations are done on small clusters of ~10 processors. There is some sharing of computational resources across projects. They do not use national super-computing resources, which are owned/managed by a consortium, due to the costs. The group agreed that improvements in raw computing power, algorithms, and implementation would greatly help their projects.
- The WTEC team's hosts noted that there is a gap in the educational background of incoming researchers/staff. They stressed that domain experts are not trained in simulation tools or don't understand the tool capabilities, and the simulation experts are very often not experts in the domain.

CONCLUSIONS

Eni is one of the largest petrochemical enterprises in the world. Its simulation and modeling activities are being conducted primarily in the Research & Development group. The key issues identified by Eni researchers revolve around model validation, accessibility of simulation tools, knowledge management, and development of new multiphysics, multiscale modeling methods.

REFERENCES

- Battistelli, A. 2008. Modeling multiphase organic spills in coastal sites with TMVOC V.2.0. *Vadose Zone Journal* 7:316-324.
- Battistelli, A. 2004. Modeling biodegradation of organic contaminants under multiphase conditions with TMVOCBio. *Vadose Zone Journal* 3(3):875-883.
- Battistelli, A., C. Calore, and K. Pruess. 1997. The simulator TOUGH2/EWASG for modelling geothermal reservoirs with brines and a non-condensable gas. *Geothermics* 26(4):437-464.
- Falta R.W., K. Pruess, S. Finsterle, and A. Battistelli. 1995. T2VOC User's Guide. Lawrence Berkeley Laboratory Report LBL-36400. Berkeley, CA, p.155.
- Frigerio, F. and L. Montanari, L. 2007. Characterisation of the surfactant shell stabilising calcium carbonate dispersions in overbased detergent additives: Molecular modelling and spin-probe-ESR studies. *Lecture Notes in Computer Science*, Vol. 4488/2007, Springer Berlin/Heidelberg, p. 272-279.
- Giorgis T., M. Carpita, and A. Battistelli. 2007. 2D modeling of salt precipitation during the injection of dry CO₂ in a depleted gas reservoir. *Energy Conversion and Management* 48(6):1816-1826.
- Millini, R., F. Frigerio, G. Bellussi, G. Pazzuconi, C. Perego, P. Pollesel, and U. Romano. 2003. A priori selection of shape-selective zeolite catalysts for the synthesis of 2,6-dimethylnaphthalene. *Journal of Catalysis* 217:298-309.
- Millini R. and C. Perego. 2008. The role of molecular mechanics and dynamics methods in the development of zeolite catalytic processes. *Topics Catal.* accepted.
- Pruess K. and A. Battistelli. 2002. TMVOC, a numerical simulator for three-phase non-isothermal flows of multicomponent hydrocarbon mixtures in saturated-unsaturated heterogeneous media. Report LBNL-49375, Lawrence Berkeley National Laboratory, Berkeley, CA.

- Site:** ETH (Swiss Federal Institute of Technology) Zürich
http://www.ethz.ch/index_EN
 Computational Science and Engineering (RW/CSE), <http://www.cse.ethz.ch/>
 Computational Science and Engineering Laboratory, <http://www.cse-lab.ethz.ch/>
 Institute of Fluid Dynamics, <http://www.ifd.mavt.ethz.ch/>
 Department of Materials, <http://www.mat.ethz.ch/>
 Institute for Building Materials, <http://www.ifb.ethz.ch/>
 Computational Physics for Engineering Materials, <http://www.ifb.ethz.ch/comphys/>
 Institute for Theoretical Physics, <http://www.itp.phys.ethz.ch/>
- Date Visited:** February 28, 2008
- WTEC Attendees:** S. Glotzer (author), L. Petzold, C. Cooper, J. Warren, V. Benokraitis
- Hosts:** Prof. Dr. Domenico Giardini, Director of the Swiss Seismological Service, Chair of Seismology and Geodynamics, Swiss Seismological Service
 HPP P 6.1, CH-8093 Zurich, Switzerland
 Tel: +41-44-633-2610; Secretariat: +41-44-633-2605
 Email: d.giardini@sed.ethz.ch
- Prof. Dr. Hans Jürgen Herrmann, Institut f. Baustoffe (IfB) (Institute for Building Materials), HIF E 12 Schafmattstr. CH-8093 Zürich, Switzerland
 Tel: +41 44 633 27 01; E-Mail: hans@ifb.baug.ethz.ch
- Prof. Dr. Ralf Hiptmair, Seminar of Applied Mathematics
 Email: hiptmair@sam.math.ethz.ch
- Prof. Rolf Jeltsch, Seminar of Applied Mathematics
 Email: rolf.jeltsch@sam.math.ethz.ch
- Prof. Dr. Leonhard Kleiser, Institute of Fluid Dynamics, Department of Mechanical and Process Engineering, ETH Zurich ML H 36, Sonneggstrasse 3, CH-8092 Zürich, Switzerland
 Email: kleiser@ifd.mavt.ethz.ch
- Prof. Petros Koumoutsakos, Chair of Computational Science Universitätsstrasse 6, CAB H 69.2, CH-8092 Zürich, Switzerland
 Tel: +41 1 632 52 58; Fax: +41 1 632 17 03
 Email: petros@ethz.ch
- Prof. Dr. Markus Reiher, Associate Professor for Theoretical Chemistry, Laboratory of Physical Chemistry HCI G 229, Wolfgang-Pauli-Str. 10, 8093 Zürich, Switzerland
 Tel: +41 44 633 43 08
 Email: markus.reiher@phys.chem.ethz.ch
- Prof. Dr. Christoph Schwab, Seminar of Applied Mathematics
 Email: christoph.schwab@sam.math.ethz.ch
- Prof. Dr. Matthias Troyer, Computational Physics, Institute for Theoretical Physics, HPZ E7 Schafmattstr. 32, CH-8093 Zürich, Switzerland
 Tel: +41 44 633 25 89; Fax: +41 44 633 11 15
 Email: troyer@itp.phys.ethz.ch
- The WTEC panel is additionally grateful to
 Prof. Dr. Peter Chen, ETH Zurich Vice President Research

BACKGROUND

The Swiss Federal Institute of Technology Zürich, ETHZ, is one of the premier universities in Europe, on par with the very top universities in the United States. The panel met with faculty from a number of departments, including physics, chemistry, mathematics, computer science, and several engineering departments, and had an extensive and fruitful discussion. Of particular relevance to SBE&S are campus-wide programs in Computational

Science and Engineering (CSE), including the Computational colLaboratory (CoLab), the Computational Science and Engineering Laboratory, and the Computational Science and Engineering degree program. There are currently two chaired professorships at ETHZ in Computational Science held by Professors Michele Parrinello and Petros Koumoutsakos.

The StrategischeErfolgsPositionen program of ETHZ (2001–2007) in CSE provided the opportunity to develop this strategic area at ETHZ with the aim towards international leadership and appeal. The Computational/Collaborational Laboratory in CSE (CoLab) was a central part of this program. It was active between the period of 2002 and 2007. Key elements included a visiting faculty program, a post-doctoral program, and summer workshops.

The Interdisciplinary Curriculum in Computational Science and Engineering (RW/CSE), based in the Departments of Mathematics and Physics, is managed by a committee where approximately four different departments are represented. It started in 1997 as a diploma degree program. In the meantime, it provides Bachelor's and Master's degrees to students at ETHZ. According to the program website (www.cse.ethz.ch or www.rw.ethz.ch), "The CSE Bachelor Program consists of three years of studies (6 semesters, first, second and third year). It is also possible to enter after a first year of basic studies at ETH Zürich or elsewhere. The basic exams after the first year are counted for 57 credit points (ECTS). In the following two years the students have to acquire 123 ECTS, half of them in mandatory Basic Courses, the other half mainly in mandatory Core Courses and in eligible Fields of Specialization and Elective Courses and with a Bachelor Thesis, respectively." Furthermore, the "CSE Master Program at ETH Zürich (which started in 2005) consists of one year of studies (2 semesters) followed by a Master Thesis. The Master Program is based on the CSE Bachelor Program and its objective is to prepare students for a successful professional career in research in industry and business and/or on a university level. The Master students have to acquire 90 ECTS mainly in mandatory Core Courses, in eligible Fields of Specialization and Elective Courses as for the Bachelor Program and with a Term Paper and a Master Thesis, respectively."

In addition to the main Bachelor's and Master's CSE program, the Departments of Computer Science and Mechanical Engineering offer Master's specializations in the fields of Computational Science and Engineering (CSE).

SBE&S RESEARCH

SBE&S research at ETHZ spans nearly all disciplines and in many areas represents the state of the art in the field. Examples of the research areas of some of our hosts follow.

Leonhard Kleiser's main research areas are turbulent flows and laminar-turbulent transition. Turbulent and transitional flows are investigated by direct numerical simulations. Here no turbulence models are employed, but the basic equations of fluid dynamics are numerically integrated on large computers whereby all relevant length and time scales of the flow must be resolved. Such investigations contribute to a better understanding of the fundamental phenomena and mechanisms of transition and turbulence, to the conceptual exploration of methods for flow control, and to the development of improved models for practical calculation methods.

Petros Koumoutsakos' research activities (<http://www.cse-lab.ethz.ch>) are in the areas of particle methods, machine learning, biologically inspired computation, and the application of these techniques to problems of interest in the areas of Fluid Mechanics, Nanotechnology, Biology, and their interfaces. The common patterns that he researches are in multiscale modeling, simulation, design and optimization, and high-performance computing as applied to problems in nanotechnology, fluid mechanics, and life sciences. Research problems in his group range from the design of nano syringes, to tumor-induced angiogenesis and aircraft wakes.

Research interests at the Seminar for Applied Mathematics (SAM) (Proff: Hiptmair, Jeltsch & Schwab) include computational finance in particular efficient pricing under multiscale stochastic volatility models; computational electromagnetics; high-dimensional finite elements for elliptic problems with multiple scales; tensor product approximation & anisotropic Besov regularity of elliptic PDEs; FEM for elliptic problems with stochastic data; design of an hp-adaptive FE code for general elliptic problems in 3D; Lévy models in finance, and numerical

analysis and computation; numerical solution of operator equations with stochastic data; simulations for high current arc plasmas; sparse tensor product methods for radiative transfer; and derivative pricing with additive driving processes; numerical methods for hyperbolic conservation laws with and without source terms, simulations in hyperbolic conservation laws include applications to Euler equations for compressible flow, elastic plastic wave equations, Navier-Stokes equations, conservation laws with intrinsic constraints, involutions, such as Magnetohydrodynamics equations

Hans Jürgen Herrmann works on problems in granular materials, including dunes, Apollonian packings, density waves, fragmentation, stratification, segregation, sedimentation, dissipative gases, the shape of a sand pile, the dip under the heap, nonlinear elasticity of packings and shear bands, SOC on small-world lattices and the brain, compactification, fibers, cellular automata, complex networks, percolation, kinetic gelation, cluster-cluster aggregation, traffic, mineral dendrites, superplasticity, Potts models, fracture, growth phenomena and geometrical critical phenomena.

Matthias Troyer is a leader in the development of new algorithms for quantum Monte Carlo for fermionic systems. His main research fields in computational physics are the development of generic parallel simulation algorithms for strongly interacting quantum systems, the investigation of quantum phase transitions, and the simulation of strongly correlated electron systems.

Markus Reiher's current research interests comprise topics from relativistic quantum chemistry, bioinorganic and coordination chemistry, theoretical spectroscopy, and the foundations of chemistry.

For the research of all computational scientists at ETH see the CSE annual reports (www.rw.ethz.ch/dokumente/index_EN)

SBE&S EDUCATION

Computational Science & Engineering (RW/CSE) was launched at ETHZ as a unique educational program in 1997, with an emphasis on both research and training in computational science and engineering. A CSE curriculum was developed, and since its inception there have been two chairs of CSE at ETH Zurich. Existing courses were combined into an educational program, and thus the program was built at virtually no cost. The program was initially launched as a Master's program, and has now been extended downward to the Bachelors degree. In 2008 the first freshman students will begin with a major in CSE. Nearly 30 faculty members participate in the program. The host department is mathematics and physics, which handles the administration. Some of the courses were borrowed from Electrical Engineering (students will take calculus and linear algebra, for example, with the Electrical Engineering students). As an alternative to the CSE degree program, students may do a specialization in CSE (Master's) while enrolled in a traditional degree program.

At the time of the WTEC visit, the ETH RW/CSE program had 40 undergraduates and 28 Master's students. In computer science, approximately 20 students take a Master's with a specialization in CSE, rather than taking a Master's degree in CSE directly. A report on the CSE educational program is available online at <http://www.cse.ethz.ch>. Walter Gander, Martin Gutknecht, Rolf Jeltsch, Kaspar Nipp and Wilfred van Gunstern started the program. Many graduate students and postdocs take the senior level undergraduate CSE courses, particularly parallel programming. There is a new class in Computer Science on high-performance computing. The number of students has increased steadily each year, because there is an increasing computing content in many of the traditional engineering disciplines. The WTEC team's hosts indicated there has been radical change in the culture of both Engineering Departments and industry in the direction of more and more computation. Also the Mathematics Department has become more open to computing and how it can influence the direction of research. Computing is today more explicitly involved in experimental laboratories, for example, in the design of experiments.

Faculty at both campuses of ETHZ are more interested in and open to computing today than previously, due in part to new hires, increased awareness of computational science, and a change in industrial culture. Students see that they can get good jobs if trained in computational science and engineering. There has also been a major shift

towards virtual experimentation in many fields, due to the increased predictive and explanatory capabilities of SBE&S.

The Computational CoLaboratory (<http://www.colab.ethz.ch>) was founded and financed in the context of the ETHZ Strategic Positions (SEP) between 2002 and 2007. Postdocs in the CoLab were supported either fully by the CoLab or they received support shared by individual faculty and by the CoLab. This was a highly successful program leading, among other achievements, to 7 faculty appointments among its 20 postdoctoral fellows. The program was not continued after the end of the ETHZ SEP program.

The WTEC team's ETHZ hosts commented that the ETHZ CSE Master's program was a pioneering program in Europe. The Bachelor's program also is now copied elsewhere in Europe. For example, there is a Bachelor's CSE program in Stuttgart based on the ETH Zurich model. Students from Physics and Chemistry are taking the undergraduate CSE courses even though they are not CS majors. This has also had a big impact. When asked which type of students the computational faculty members prefer to work with, one responded: "For research, the physics students; for engineers or software developers, the CSE majors. A potential problem is that the students' backgrounds are very broad, but not deep. There are courses at the computer centers on how to use the supercomputers, that one can easily send their students to." There is some difference of opinion among researchers about student preparation for this field. There is an excellent overall impression of the preparation and focus of the CSE students.

COMPUTING FACILITIES

Computational scientists at ETH Zurich have access to large scale computing platforms at the Swiss National Center for Scientific Computing (CSCS, www.cscs.ch) and at the IBM Rüschlikon Research Lab (IBM BG-L).

ETHZ provides flexibility in how faculty members support computational clusters. One group (Reiher) has a 370 core computer cluster. Faculty described good experiences with the campus computer center, operating a 6,000-core cluster. However, others opt, in addition, to have computer clusters directly available for their groups. ETHZ has a budget for internal equipment and supports the acquisition of shared computer clusters by faculty. Funds are awarded competitively, with 60% success rate for internal proposals. Individuals can get on the order of 250,000 Swiss Francs every 4–5 years for computers. Some cost-sharing (about 10%) is required. WTEC team members noted that this program seems unique to ETHZ.

Simulation codes used are both black box and home grown, depending on the group.

DISCUSSION/OTHER COMMENTS

Members of the first group from the Department of Mathematics graciously provided written answers to questions provided in advance by the WTEC panel, and the rest of the group also provided verbal answers during our highly fruitful and engaging discussion. Both the written and verbal answers are included below.

GENERAL

One very important need for SBE&S is the development of much better techniques for programming in scientific computing. Better programming languages and mechanisms, debuggers, and analyzers lag far behind hardware. There is an effort to bring advances in Computer Science such as software engineering and databases into SBE&S. The successful bridging of these concepts into state of the art problems and techniques in SBE&S can have a tremendous impact on the efficient usage of available computational power. The need to maintain a balance between hardware, numerical methods and software will be an everlasting problem of SBE&S research.

Energy and Sustainability

The effective use of computers can have a significant impact on energy and sustainability as computers and computation are becoming ubiquitous.

Start-up Companies

ETHZ strongly supports the development of startups. There are few efforts, however, in startups based on SBE&S. Consulting is very popular among advanced ETHZ students.

Simulation Software

The amount of in-house development is probably normal here. A number of packages are modified and supplemented rather than written from scratch, while at the same time, several groups develop and maintain in-house software.

One faculty mentioned as issues to address in the coming years: Black boxes used to solve PDEs, where probability densities are fed in to obtain risk range of solution. Black boxes need to be re-engineered to solve today's problems; you can't use what you use for single inputs. Needed are a multilevel discretization, and adaptivity of the black boxes. High dimensionality will become insufficient. Transparency of black boxes must be increased.

At the science level there is not enough expertise in computer science embedded within science disciplines, and this is hurting our ability to develop new simulation software.

Algorithms and software have provided new tools for many disciplines and problems, such as strongly correlated materials. Examples include density of states of plutonium. Such problems could not be solved by MC before but can be now done on a laptop using continuous time MC solvers by Troyer. These now can be used for many materials that we couldn't do before, such as plutonium, which is the most strongly correlated simple metal around. Since this year (2008), our algorithm is used at the LANL and LLNL U.S. National Labs. Thus, there are huge opportunities and progress in faster MC methods for QM. QMC has no problem dealing with bosons, but there are big problems with fermions. The problem is NP-hard, so, you cannot use classical computers to solve problems with fermions. We have new methods, but these don't scale well on vector and other machines, but they are already millions of times faster than previous methods due to algorithmic advances. We have problems with code longevity when students who developed the code leave.

In the CoLaboratory there is a consistent effort to adopt rigorous software engineering practices in developing scalable and maintainable software. The group develops software in multiscale particle methods and distributes the software as open source (see www.cse-lab.ethz.ch/software.html). The software has enabled state of the art simulations in CFD using billions of particles (see Chatelain, P. et al. 2008. Billion vortex particle direct numerical simulations of aircraft wakes, *Computer Methods in Applied Mechanics and Engineering* 197, 1296-1304).

Big Data/Visualization

So much of science is now data-driven. We have so much data, we're not looking at it all...less than 1% of all data gathered is looked at. The situation is further extenuated by the increasing size of available processors and the consequent problem scale-up. This will require a major effort in data-driven simulations. There are efforts in the CSE lab to use problem specific data compression algorithms to access the simulation data.

Big data is very important to high-energy physics. There is an extremely good visualization group at CSCS, Manno. Jean Favre is first rate and has a highly competent team.

Engineering Design

In Mechanical Engineering there are several groups (among others Hora, Riener, Mueller, D'Andrea) that rely on computational science for developing effective engineering designs.

Next-Generation and HPC

The WTEC team's hosts pointed out that we (the field) don't always exploit the full power of our computers. Algorithms and hardware are equally important. "Of 12 orders of magnitude in speedup, 6 came from

algorithms.” This was a notion emphasized by Norm Schryer at least as far back as 1980. It’s the people that matter. Good algorithms can be equivalent to tens of thousands of processors. The challenge, of course, is to coordinate advances in hardware and software.

When asked what they would do with increased computing power, our hosts commented that error increases as well, and this must be considered. For example, will 64-bit precision be enough for grand challenge problems? For some problems, yes; for many, no.

Problems that could be attacked with petascale computing include cancer, socioeconomic systems (e.g., model entire countries with customers down to individual levels), traffic flow (e.g., in all of Switzerland), problems in neurobiology (e.g., reengineering the structure of the brain), treating molecules embedded in solvents, and problems in seismology (including mantle dynamics).

Giardini mentioned the CGI project in the United States. The EU has a roadmap workshop on e-infrastructure for seismology. A 40k processor machine is used at San Diego State University San Diego Supercomputing Center) for visualization. Seismology is a data driven discipline. Lots of simulations are behind the design of the Yucca Mountain facility. The next challenge is real-time data streams—before, for example, the location and time of an earthquake. Now we want to run scenarios, using data-driven dynamic simulations. The goal is a 5-second warning of an event (e.g., earthquake). This is driving all seismology simulations. In seismology, it is now half modeling, half experiment, which is a big change from the past.

The bottleneck to advances in SBE&S remains programming. Software still lags behind hardware, and as long as the emphasis is on buying expensive machines rather than on people, this problem will persist.

Education and Training

To fully leverage investment in large computers, we need a matching effort in human resource development, to train students and researchers to use machines and to develop algorithms and software.

Would joint postdocs between science/engineering and computing sciences be a solution? ETHZ had a program that did this (the CoLab); it was quite successful and produced a high percentage of future faculty members.

ETHZ has several programs for SBE&S education and training. For example, the CSE Bachelor’s and Master’s degree curricula at ETHZ (RW/CSE, www.cse.ethz.ch) are good programs but need more people resources. ETHZ students who seem best prepared for certain programming tasks come from Physics Faculty in addition to Computer Science (www.inf.ethz.ch). The opinion of Petros Koumoutsakos is that Computer Science students are very well educated in programming, but they lack the background and possibly the interest in Engineering and Natural Sciences. This may be attributed to the scope of Computer Science as a discipline so far. There is tremendous room for improvement by breaking down traditional disciplinary barriers.

Incoming graduate students are poorly prepared in programming, but not worse than in any other comparable institution. One faculty mentioned that “In the Mathematics Department, there remains a bias: analysis is highly regarded, whereas programmers remain only programmers. Alas, we need to do both extremely well. Interdisciplinary work is desirable until hiring/respect are demanded, then we revert to bad old habits. It has been so in other places I’ve been: Bell Labs and ETHZ are similar in this regard. My students from Physics have a much more balanced perspective, it seems to me. My Chemistry students seem to be weak programmers. Both Physics and Chemistry students are of very high quality, but in programming, Physics does better.”

It’s become evident that in recent years the number of foreign students has increased dramatically. This is also true in Swiss Mittelschule. Where do the CSE students go when they graduate? Banks, insurance, software, pharmaceutical, oil companies, universities. They have absolutely no problem finding a job. The problem is of losing students to industry and master’s level.

Students should NOT be comfortable using codes that exist.

What is the publication standard in SBE&S? To what extent can the computational experiments be replicated? It is important to train the students to be skeptical and question the results of the computation and their relevance to the scientific or engineering discipline. Key issues such as validation, verification, and efficiency must become parameters of one's education.

We should put more weight on proper software engineering in the CSE curriculum. Open source is an inducement to producing good software that students and other researchers can add to. Maintenance of code is a big problem in the academic environment. There is a bigger role for Computer Science in CSE in the future.

Evaluation of scientific performance is now completely geared towards disciplinary indicators. We should give more value to integration, just as we have always done with depth. How do you assign credit in an interdisciplinary research project?

REFERENCES

A report on the CSE educational program is available online at <http://www.cse.ethz.ch>.

SIAM Working Group on CSE Education. 2001. Graduate education in computational science and engineering. *SIAM Rev.* 43:163–177.

Site: **Fraunhofer Institute for the Mechanics of Materials (IWM)**
Wöhlerstrasse 11
79108 Freiburg, Germany
http://www.iwm.fraunhofer.de/englisch/e_index.html

Date Visited: February 29, 2008.

WTEC Attendees: S. Glotzer (report author), L. Petzold, C. Cooper, J. Warren, V. Benokraitis

Hosts: Prof.-Dr. Peter Gumbsch, Director of the IWM
 Email: peter.gumbsch@iwf.fraunhofer.de
 Prof.-Dr. Hermann Riedel, Leader, Materials-Based Process and Components Simulation;
 Email: hermann.riedel@iwf.fraunhofer.de

BACKGROUND

Professor Gumbsch is head of the Fraunhofer Institute for Mechanics of Materials (Institut Werkstoffmechanik, IWM), with locations in Freiburg and Halle/Saale, Germany. He is a full professor in Mechanics of Materials and head of the Institute for Reliability of Systems and Devices (IZBS) at the Universität Karlsruhe (TH). Before that he was head of the Research Group for Modelling and Simulation of Thin Film Phenomena at the Max-Planck-Institut für Metallforschung in Stuttgart, Germany. His research interests include materials modeling, mechanics and physics of materials, defects in solids, and failure of materials.

The Fraunhofer Institutes comprise the largest applied research organization in Europe, with a research budget of €1.3 billion and 12,000 employees in 56 institutes. The institutes perform their research through “alliances” in Microelectronics, Production, Information and Communication Technology, Materials and Components, Life Sciences, Surface Technology and Photonics, and Defense and Security. Financing of contract research is by three main mechanisms: institutional funding, public project financing (federal, German Länder, EU, and some others), and contract financing (industry).

The Freiburg-based Fraunhofer Institute for the Mechanics of Materials, hereafter referred to as IWM, has 148 employees (with another 75 based at its other campus in Halle), and a €15.5 million budget (€10.7 million for Freiburg, €4.7 million for Halle). A remarkable 44% of the budget comes from industry. Another 25–30% base funding derives from the government. *It is crucial to the funding model of the IWM that base funding is a fixed percentage of the industrial funding.* The IWM has seen significant growth in recent years (10% per year), a figure that is currently constrained by the buildings and available personnel. At the IWM *fully 50% of the funding supports modeling and simulation* (a number that has grown from 30% five years ago).

The IWM has 7 business units: (1) High Performance Materials and Tribological Systems, (2) Safety and Assessment of Components, (3) Components in Microelectronics, Microsystems and Photovoltaics, (4) Materials-Based Process and Components Simulation, (5) Components with Functional Surfaces, (6) Polymer Applications, and (7) Microstructure-Based Behavior of Components.

SBES RESEARCH

The IWM has a world-class effort in applied materials modeling. The WTEC team’s hosts noted that a canonical example illustrated how automobile crash simulations could be improved. The IWM has developed microscopic, physics-based models of materials performance and then inserted these subroutines into FEM codes such as LSDYNA, ABAQUS, and PAMCRASH. The IWM enjoys strong collaborations with German automobile companies, which support this research.

Research areas that the WTEC team’s hosts discussed with us included modeling of materials (micromechanical models for deformation and failure, damage analysis), simulation of manufacturing processes (pressing, sintering,

forging, rolling, reshaping, welding, cutting), and simulation of components (prediction of behavior, upper limits, lifetime, virtual testing).

The IWM has a strong program in metronomy at small scales, developing models of failure/fracture, coupling data from nano-indentation with engineering- and physics-based models of same. Models of dislocation-structure evolution seem competitive with state-of-the-art efforts at other institutions. The IWM has a substantial simulation effort in powder processing/sintering, and demonstrated the full process history from filling to densification, and subsequent prediction of materials properties. Not surprisingly, given its strong affiliation with the automobile industry, the IWM has sophisticated program in modeling springback in stamped components, rolling, forming, and friction.

A detailed, so-called “concurrent,” multiscale model of diamond-like-carbon thin-film deposition was presented. This model is truly multiscale and presents an alternative to some of the other integrated approaches in the literature. The model integrates models from Schrödinger’s equation up to classical models of stress-strain (FEM) and captures such properties as surface topography, internal structure, and adhesion.

COMPUTING FACILITIES

The IWM employs mostly externally developed codes for its simulations (with some customized in-house codes that couple to the externally developed codes). All the codes run on parallel machines. It was noted that privately and/or industry-funded projects cannot use the national supercomputers. Thus, industry-funded projects are done with small clusters in-house, and the researchers also have access to two shared (among five Freiburg-based institutes) clusters with 256 and 480 nodes. At the largest (non-national) scale, the Fraunhofer institutes have a shared 2000-node cluster, of which the IWM uses approximately 25% of the capacity. Because of its disproportionate demand for computational power, the IWM plans to purchase its own 300-node machine soon.

EDUCATION/STAFFING

In our conversations it was observed that it is challenging to find materials scientists with a theoretical background strong enough to make able simulators. The typical training is in Physics, and materials knowledge must then be acquired through on-site training. Atomistic simulations are uniformly performed by physicists and theoretical chemists. Currently it is difficult to obtain staff skilled in microstructure-level simulation, as they are in high demand by industry.

A particularly novel method for developing students into potential hires was discussed. A one-week recruitment workshop in the Black Forest was held where twenty €50,000 projects were awarded through competition to fresh PhDs to do the work at the institute of their choosing. It is expected that once these students begin developing in their new environments they will then craft new research proposals in-house.

Site: IBM Zurich Laboratory, Deep Computing
Säumerstrasse 4
CH-8803 Rüschlikon, Switzerland
<http://www.zurich.ibm.com/>
<http://www.zurich.ibm.com/deepcomputing/>

Date Visited: February 28, 2008

WTEC Attendees: L. Petzold (report author), S. Glotzer, J. Warren, C. Cooper, V. Benokraitis

Hosts: Prof.-Dr. Wanda Andreoni, Program Manager, Deep Computing Applications
Email: and@zurich.ibm.com
Dr. Alessandro Curioni, Manager Computational Science
Dr. Costas Bekas

BACKGROUND

Deep Computing (DC) aims at solving particularly complex technological problems faced by IBM, its customers, and its partners by making use of (large-scale) advanced computational methods applied to large data sets. In 1999, an organization called the Deep Computing Institute was created in IBM Research. Its task was to promote and coordinate DC activities, which imply advances in hardware, software, and development of innovative algorithms, as well as the synergy of all three components.

The group that the WTEC visiting team visited at IBM Zurich is Deep Computing Applications. Its head, Dr. Wanda Andreoni, is also a member of the IBM Deep Computing Council, which coordinates all Deep Computing activities at IBM.

R&D ACTIVITIES

The work of this DC research group is problem-driven. It is aimed at patents of new materials or processes. Its researchers do a considerable amount of work with other IBM organizations and with outside companies, and they take a pragmatic vision in these collaborations. Where do the topics and collaborations come from? Many are addressing internal needs where they can make a difference. Industry comes to them for their unique expertise on electronic structure calculation and molecular dynamics. Sometimes they seek out industrial contacts. Now there is an interest in computational biology, and there are collaborations with Novartis and Merck. The contacts usually come from scientist to scientist, but sometimes they come from higher-level management. Many industries have realized that computational science can help them, and they approach IBM for help. There is also an Industry Solution Lab (ISL) to put the customers in contact with them. Most collaborations, however, come from direct interaction between scientists. The information generated by the ISL has been helpful in identifying some of the computational trends and needs of the future. Members of the group remarked that to really put a method to the test, big problems from industry are invaluable. The applications to work on are selected on the basis that they need to be of real research interest to the group, and also to the business of IBM. The money comes from the companies, or from IBM, on a case-by-case basis. This group has 60% core funding from IBM.

There is a big success story here. This group went from a situation where the IBM technology group was coming to it with a supercomputer and asking the team to exploit it, to a situation where the computer architecture is now driven by the needs of important applications. The Blue Gene project is an example. How was this transition accomplished? The key was to first build trust and credibility. They now work directly with the computer architects. They have direct communication with the architects regarding parameters across the architecture. There has been an enormous change in the culture at IBM regarding Computational Science and Engineering (CSE). In the beginning, this DC group was a “luxury group.” Now it is well-integrated. The group believes that the key to their success in the long term is their flexibility and careful choice of projects. They are now resting on a foundation of credibility built in previous years. The group currently consists of 4 permanent researchers and 2-

6 students and postdocs. Substantial growth is planned in the near future. The biggest bottleneck to growth is in finding appropriately educated people who will meet their needs.

The group regularly works with students, coming from the areas of physics, chemistry and bioinformatics. Geographically, they come mainly from Switzerland and Germany.

They have found that many of this generation of students do not have a good foundation in programming, programming for performance, or massively parallel computing. To find very good people for permanent positions is not easy at all. Sometimes good people come there, and then they get big money offers from the financial industry and leave. Potential employees need to be able to work well as a team and to love science.

How does this group interact with their counterparts at IBM in the United States? In the United States, the major DP research efforts are in systems biology, protein folding, bioinformatics, and machine learning. The computational biology center (around 10 permanent staff plus postdocs) in Yorktown is a virtual center of which this group is a partner. There is also a group in dislocation dynamics in Yorktown.

With regard to future trends in CSE, the group remarked that massively parallel is here to stay, unless a completely new technology comes along. The software dilemma is: where do you focus your attention – general programming methodologies for parallel computers, or getting all the speed you can for a particular application? Algorithmic research is needed in many areas; most current algorithms are not built to scale well to 1000+ processors. One of the problems mentioned was graph partitioning on irregular grids, for massively parallel processors.

With regard to CSE education, the group believes that the educational focus in the United States is sometimes too broad, and traditionally the European focus has been too narrow. Something in the middle seems to be required. Students need to know more than just how to use a code. They have no basis for judging the quality of the results produced by the code. Students do not know about the method they are using, and are ill-equipped to write their own codes. They do not know the limitations of the codes they are using. The application scientist needs to understand the simulation methods and the theory, and be capable of modifying or extending a code. At the same time, many of the codes are poorly written and not well documented. Students need to know the science, the algorithms and theory, and programming and software development techniques.

During this visit we also found out about an expanded mission of CECAM++ (see related CECAM trip report) from Dr. Andreoni, who represents Switzerland in the CECAM organization. Our IBM hosts argued that coordination is needed to solve the big problems facing science and engineering, and that the focus of the community should not be to simply produce publications. CECAM has for several decades provided important education and training opportunities in modeling and simulation. A new incarnation of CECAM will return to its initial mission as a coordinating body within the European community. Funds are provided by currently 10 countries—roughly a million Euros per country per year. The dominant community of CECAM has been molecular dynamics, but it will broaden to include computational engineering and applied mathematics. An analogy with KITP was mentioned.

CONCLUSIONS

The big story here is the influence of important applications on computer architecture, and the direct collaboration between these two groups. This group has also been very successful in collaborating both internally, and with outside industries. They derive a significant fraction of their funding from projects with outside industries. Their success rests on a foundation of credibility, and a well-developed area of expertise. They are slated for considerable growth in the near future. The bottleneck to that growth is the availability of properly educated people who will meet their needs.

Site: Imperial College London and Thomas Young Centre
(London Centre for Theory & Simulation of Materials)
South Kensington Campus, London SW7 2AZ, UK
<http://www3.imperial.ac.uk/materials>
<http://www.thomasyoungcentre.org/index.html>
<http://www3.imperial.ac.uk/materials/research/centres/tyc>

Date Visited: February 29, 2008

WTEC Attendees: M. Head-Gordon (report author), P. Cummings, S. Kim, K. Chong

Hosts: Prof. Mike Finnis, Dept. of Materials, Dept. of Physics
Email: m.finnis@imperial.ac.uk
Prof. Adrian Sutton, Dept. of Physics
Email: a.sutton@imperial.ac.uk
Dr. Peter Haynes, Dept. of Materials, Dept. of Physics
Email: p.haynes@imperial.ac.uk
Dr. Nicholas Harrison, Dept. of Chemistry
Email: nicholas.harrison@imperial.ac.uk
Dr. Patricia Hunt, Dept. of Chemistry
Email: p.hunt@imperial.ac.uk
Dr. Andrew Horsfield, Dept. of Materials
Email: a.horsfield@imperial.ac.uk
Dr. Arash Mostofi, Dept. of Materials, Dept. of Physics
Email: a.mostafi@imperial.ac.uk
Dr. Paul Tangney, Dept. of Materials, Dept. of Physics
Email: p.tangney@imperial.ac.uk
Prof. Robin Grimes, Dept. of Materials
Email: r.grimes@imperial.ac.uk

BACKGROUND

Imperial College has a very strong tradition in materials science research, through the Department of Materials (in Engineering), the Department of Physics, and to a lesser extent, the Department of Chemistry. The Chair of Materials Theory and Simulation is held by Professor Mike Finnis, who has joint appointments in Materials and Physics. A recent significant development for materials modeling and simulation community at Imperial, and indeed in London generally is the establishment of the Thomas Young Centre (TYC), which is an umbrella organization for London-based activities in this area.

RESEARCH

During a 3-hour mini-symposium, the visiting WTEC team's hosts presented short talks on their research. Brief summaries follow. Due to time constraints and the interests of those present, the presentations were mainly concerned with the atomic scale work in the TYC, although Imperial, like the other London Universities, has materials theory and simulation in progress at all length scales. A poster display covering more material was presented by some of the postdocs and students during the lunch break.

- *Fusion materials and fuel cells as targets for multidisciplinary and multiscale simulation.* Prof. Finnis observed that theory and simulation of materials is most effective as a collaborative activity between theorists and experimentalists and increasingly requires a wide range of theoretical expertise. He illustrated these considerations with reference to the Thomas Young Centre, and two or three exemplar collaborative

projects in progress, including creep resistant design for turbine blades, and nonadiabatic dynamics of high energy ions in materials.

- *Grand challenges in theoretical and computational materials research, education, and training.* Prof. Sutton suggested that the future of modeling interfaces will be the introduction of grand-canonical simulations to minimize free energies through variations of local compositions and planar atomic densities in single component systems, which will require a multiscale approach using interatomic potentials for the grand-canonical simulations, validated by density-functional theory.
- *Linear-scaling algorithms for density-functional theory and the ONETEP code.* Dr. Haynes has led development of the ONETEP code with the twin aims of true linear scaling (time to science) and controlled accuracy (up to the level of the traditional plane-wave pseudopotential approach). This involves new algorithms and methods designed for parallel computers, which was illustrated for the case of the design of synthetic inhibitors for the zinc enzyme carbonic anhydrase II. The code has been released commercially by Accelrys, Inc.
- *Strongly interacting electrons and magnetism in pure carbon materials.* Dr. Harrison described the CRYSTAL code, which is capable of very large simulations of strongly-correlated electronic systems. He illustrated these capabilities with applications to pure carbon materials such as defects in graphene and nanopeapods with potential applications in quantum computing.
- *Are all noses electronic?* Dr. Horsfield described current work on the hypothesis that humans recognize odorants on the basis of their vibrational frequencies, which are detected by inelastic electron tunneling.
- *Model Hamiltonians with first-principles accuracy.* Dr. Mostofi described how large-scale *ab initio* electronic structure calculations and the maximally localized Wannier function (MLWF) approach are combined in order to study the electronic properties of complex nanostructures such as silicon nanowires and DNA strands. MLWFs provide an accurate, localized, and minimal basis set in which to diagonalize the Hamiltonian. In the MLWF basis, Hamiltonians for large, complex systems are constructed directly from the short-ranged Hamiltonians of smaller constituent units, for extremely high efficiency.
- *Structure and diffusion in liquids and glasses: simulations at several length scales.* Dr. Tangney described studies of silicate liquids and glasses with interatomic potentials that have been parameterized without reference to experimental data by using density functional theory, and with a close coupling of continuum modeling of ion diffusion and Secondary Ion Mass spectrometry (SIMS).
- *Simulation of nuclear materials.* Dr. Grimes discussed applications of atomic-scale computer simulations to nuclear fuel performance and waste containment and how such knowledge can be used by the nuclear industry.

DISCUSSION

During the talks there were some short opportunities for discussion of both scientific issues directly related to the content of the talks, and also of more general issues surrounding simulation-based engineering and science. Several additional issues came up:

- Regarding education, no departments of physics or materials in the UK are teaching the theoretical materials physics that is needed for PhDs in computational materials science. Led by Prof. Sutton, a new 2-year course on materials physics within the Physics undergraduate course has been developed at Imperial that will fill this vacuum if sufficient teaching resources can be provided. To date these have not been forthcoming.
- [Note added after meeting: In the Spring of 2008 the research council EPSRC announced a call for 40 Doctoral Training Centres (DTCs). Each of them will have a significant component of education and training, and the PhD will be extended to 4 years to allow for this. Led by Prof. Sutton, Imperial has submitted an outline proposal for a DTC in the area of theory and simulation of materials; many of the courses on theoretical materials physics developed for the undergraduate degree in Physics will be used, together with more advanced options. The outcome of the bid will be known by the Fall.]
- There was general discussion of funding for development of simulation code for materials modeling. In the UK context, successful ventures such as OneTEP (Haynes) have succeeded despite challenges in obtaining

appropriate funding. In the case of CRYSTAL, UK researchers have contributed to an international collaboration with Italian researchers.

Materials modeling and simulation has no shortage of grand challenges for the coming decade and beyond. A few of the areas discussed were the problems of bridging length and time scales, data management and visualization, interoperability of legacy, as well as new codes. Some specific problems in which these issues arise were already summarized in the research discussed above. Additionally, the simulation of the first nanosecond of radiation damage induced by a 1 MeV gamma ray was mentioned, as well as plastic deformation of glassy polymers and polymer nanocomposites.

Site: **Institute Français du Pétrole (French Petroleum Institute)**
1 & 4, avenue de Bois-Préau
92852 Rueil-Malmaison Cedex, France
<http://www.ifp.com>
<http://www.ifp.fr/>

Date Visited: February 28, 2008

WTEC Attendees: S. Kim (report author), P. Cummings, M. Head-Gordon, K. Chong

Hosts: Dr. Hervé Toulhoat, Assistant Director, Scientific Direction
Email: herve.toulhoat@ifp.fr
Dr. Anthony Wachs, Applied Mechanics Division
Dr. Christian Angelberger, Energy Application Techniques Division
Dr. Jean-Marc Gratien, Technology, Computer Science and Applied Mathematics Division
Dr. Carlos Nieto, Applied Chemistry and Physical Chemistry Division
Dr. Pascal Raybaud, Catalysis and Separation Division
Dr. Diego Klahr, Data Processing Services and Telecommunications Division

BACKGROUND

The French Petroleum Institute (Institute Français du Pétrole, IFP) is a state-owned industrial and commercial establishment (EPIC) with the mission of advancing research in energy, transportation, and the environment, and catalyzing the transfer of technology from fundamental research to industrial development. Founded in 1944, the institute today has grown to a staff of 1,735 full-time employees; of this total, 65% are in R&D at IFP Rueil-Malmaison and IFP Lyon. There are 219 doctoral and post-doctoral researchers, representing over 50 disciplines, including geological sciences and automotive engineering. Research output features over 200 scientific publications per year, and the institute has 12,500 active patents. The budget for 2007 was €301.5 million, of which €241.3 million was for R&D.

COMPUTING HARDWARE

The WTEC team's hosts gave a comprehensive presentation on the IFP's HPC strategy. Notes below on the presentation by Dr. Klahr give more details, but we note here an important strategic aim of the institute, historically attained, that its HPC facility must track by one order of magnitude the capability of the top system in the world; current plans maintain this with a system going online in March 2008 with a peak speed of 17 TFLOPS (114 quadri processor servers of quadri core AMD Barcelona 2 GHz nodes with 32 to 64 GB memory per node; new DDR Infiniband interconnect). This HPC upgrade brings 4.5X peak performance using less space and with the same heat dissipation (85 kW).

PRESENTATIONS

The opening presentation by Dr. Toulhoat set the stage by providing the background information on IFP summarized above. Upon turning to the research activities, he presented the institute's five complementary strategic priorities; this portion of the presentation and the main themes from the subsequent speakers are summarized below. In his closing remarks, he also stressed the dynamic nature of IFP's SBES directions. Prioritized efforts involving SBES can span multiple decades and into the future (e.g., reservoir simulation) and can come and go (e.g., atmospheric chemistry 1988–1996). From the present to beyond 2010, SBES priorities include reservoir simulation, basin modeling, structures and fluid mechanics, turbulent combustion in IC engines, IC engines control, CFD for process engineering, and Equation-of-State and molecular modeling.

Dr. Hervé Toulhoat: Innovating for Energy

IFP's scientific R&D activities are driven by the following five complementary strategic priorities: Extended Reserves; Clean Refining; Fuel-Efficient Vehicles; Diversified Fuels; Controlled CO₂. Each subsequent presentation fit into this strategic framework and amplified the introductory talk with a more detailed exposition of the SBES issues. Dr. Toulhoat's talk concluded with a discussion of structures for research-industry partnership.

- Extended Reserves is based on the reasonable assumption that oil and other fossil fuels will remain the dominant source of transportation fuels and chemical feedstock. R&D themes for this strategy targets increased success rate in exploration, improving the recovery ratio in reservoirs, and developing new fields in extreme environments.
- Clean Refining focuses on obtaining the highest possible yields of transport fuels from a unit basis of raw materials in an environmentally responsible fashion. The research themes are the production of high-quality fuels; the conversion of heavy crudes, residues, and distillates; and the production of petrochemical intermediates.
- Fuel Efficient Vehicles recognizes the importance of reducing fuel consumption and the development of new powertrain systems for alternative fuels (e.g., biofuels). The four R&D themes are development of highly efficient engine technologies, including conventional and hybrid powertrains; development of pollutant after-treatment technologies; development of electronic control strategy and onboard software; and validation and specification of alternative fuels (e.g., biofuels and NGV) with low CO₂ emissions.
- Industrial outlets for R&D results are achieved by a combination of complementary research-industry partnerships. These technology transfer routes include strategic subsidiaries; shorter-term arrangement with companies for the industrial application of an R&D result; the sale of R&D studies or conduct of a joint research project; development support of SMEs (small and medium enterprises) via spin-off of startup companies with IFP employees and IFP assistance, using R&D discovered by IFP; and transfer of know-how to startup or developing companies via Demeter and 3E funds.

Dr. Anthony Wachs: Direct Simulation of Particulate Flows with Collisions.

This talk presented an overview of the institute's interests in computational modeling of fluid-solid interactions with a view to the applications in multiphase flow in production fluidized beds in chemical engineering processes. Recent work features a novel collision model that allows for incorporation of shape effects (nonspherical shapes including sharp edges and corners), applicable even to dense, concentrated suspensions. This is a significant advance in the field of suspensions modeling. Extension of this promising approach to full three-dimensional simulations with large number of particles is a timely opportunity for the next growth spurt in SBES/HPC resources. (It is estimated that 2500–5000 particles provide statistically meaningful results, but even in a 2D domain, these runs take on the order of 10 days with the current serial algorithm; the goal is to push towards 100,000 to 1 million particles with a parallelized version with full MPI and HPC scale resources.) The simulation results for sedimentation of dense swarms (20% solid fraction in a 2D rectangular domain) reproduces experimentally observed features such as mean settling times and the transition to chaos at higher Reynolds numbers. The computational method features a fixed grid with particles moving on the grid and distributed Lagrange multipliers to impose particulate rigid body motions. The numerical method is implemented as the IFP proprietary software GRIFF (GRains in Fluid Flow).

Recent Papers

- Yu, Z., A. Wachs, and Y. Peysson. 2006. Numerical simulation of particle sedimentation in shear-thinning fluids with a fictitious domain method. *Journal of Non-Newtonian Fluid Mechanics* 136:126–139.
- Yu, Z., X. Shao, and A. Wachs. 2006. A fictitious domain method for particulate flows with heat transfer. *Journal of Computational Physics* 217:126–139.
- Wachs, A., and Y. Peysson. 2006. A distinct Element Granular Solver/Fictitious Domain Method for the numerical simulation of particulate flows. Presentation at the Fifth International Conference on Computational Fluid Dynamics in the Process Industries, Melbourne, 13–15 Dec.

Yu, Z., and A. Wachs. 2007. A fictitious domain method for dynamic simulation of particle sedimentation in Bingham fluids. *Journal of Non-Newtonian Fluid Mechanics* 145:78–91.

Wachs, A. 2007. A DEM-DLM/FD method for direct numerical simulation of particulate flows: Sedimentation of polygonal isometric particles in a Newtonian fluid with collisions. Submitted to *Computers & Fluids*, November.

Dr. Christian Angelberger: Large Eddy Simulation Techniques Applied to IC-Engines Design

R&D activities towards the development of cleaner and more efficient piston engines is a significant activity of the institute and numbers 200 full-time employees (FTEs). SBES activity within this group is housed in the department of Engine CFD and System Simulation (40 FTEs), with the CFD focus on RANS (Reynolds Averaged Navier-Stokes) and LES (Large Eddy Simulation) approaches for modeling of flows, fluid injection, and combustion. Previous software developed by this department has been commercialized in the AMESim platform. The current focus, a migration from RANS to LES (solve large flow scales, model small ones), permits a more realistic modeling of the variance observed in the real engine cycle. In addition to cyclic variability, the benefits of the LES approach include modeling of misfires, cold starts, fast transients, and related pollutant levels. This research CFD code is jointly developed and co-owned with CERFACS (European Center for Research and Advanced Training in Scientific Computing) as an initiative to support industrial and gas turbine burner applications. The numerics feature an unsteady, fully compressible reactive solver; second- and third-order finite volume and finite element convective schemes; explicit time advancement; unstructured moving meshes (ALE and CTI); and NSCBC boundary conditions. Parallelism is achieved via MPI (MPL library) and ported on all major processors on the market and demonstrated linear speed-up to 4000 processors (BlueGene/L). Code performance is illustrated by a ten-cycle simulation of the PSA XU10 engine (4-valve, SI, PFI) at 120 hours/cycle on a 32-Xeon processor system.

In terms of education/workforce development, the department collaborates actively with several universities and since 2000 has averaged one PhD graduate per year in this CFD area. Since 2002, two post-doctoral researchers have also been trained. The SBES/CFD capabilities of the IFP in engine research have strategic significance for its role in helping automotive manufacturers meet ever stricter guidelines for cleaner and more fuel-efficient engines.

Dr. Jean-Marc Gratien: Software Platforms for Oil and Gas Exploration and Powertrain Engineering: Addressing Supercomputing Challenges

(The title of his Powerpoint presentation was “Supercomputing at IFP Developing New Efficient Parallel Business Application.”) This talk featured trends in hardware and applications, new challenges in HPC, and research axes in HPC to overcome the challenges and software policy at IFP. The trend analysis focused on the evolution of number of nodes and the emergence of multicore architecture and the resulting challenge of distributed memory. Using CO₂ sequestration models as an illustrative application, he illustrated the multilength-scale (well area, reservoir area, basin area) and multidisciplinary (geomechanics, geophysics, fluid dynamics) challenges of these SBES problems. The large problem size, coupling of different physics models, plus the multiple time scales as well as spatial scales are challenges for HPC systems comprised of 2048 CPUs and 16 GB of memory per node. The challenge is to guarantee high scalability to a large number of processors (load balancing, communication vs. CPU cost, data partitioning), manage large data sets (high performance parallel I/O), and perform visualization and analytics on large data sets. To this end, HPC research focuses on efficient parallel partitioners, load balancing algorithms, and parallel data management (parallel data servers). Numerical analysis research features robust parallel linear solver, domain decomposition algorithms, and multiscale time and space steps.

The software platform policy recognizes the presence of at least three conceptual layers from the software engineering perspective: the lowest level—computer science algorithms; the middle level—numerical algorithms; and the complex physical model. The group has developed two platforms (Arcane, OpenFlow) to facilitate a “plug-in” modularity technique to exploit HPC advances in a specific component. Arcane is a platform to design parallel 2D and 3D finite volume/element applications. In particular, it features the use of shared common services for low-level services and mesh services; numerical services (solvers, discretization scheme); and “business” services (thermodynamics, geochemical, etc.). OpenFlow is a java business environment managing the

data model (persistence services and visualization plug-ins efficient even on large meshes) and workflow using business applications. Workflow application management includes launching and linking parallel applications, providing business data to cluster applications, and getting results from cluster applications and storing them in a unique business data model independent of each application. The group collaborates with computer science departments at other institutes (CEA, BRGM), universities, and the private company GIP.

Dr. Carlos Nieto-Draghi: Applications of Molecular Simulation in New Energy Technologies Research

This presentation featured two applications of SBES and molecular simulations: (1) the reduction of emissions from flexi-fuel diesel engines and (2) capture and sequestration of CO₂. Efforts to develop cleaner and more fuel-efficient engines includes significant R&D interest in the automotive industry in optimizing new designs of flexi-fuel HDi (high-pressure, direct injection) diesel engines that can run on biodiesel. But standard correlations for the thermophysical properties, e.g., viscosity, of such complex fuels are out of range at the high pressures (0.1 MPa to 250 MPa at temperatures of 293.15K to 700K) encountered in HDi Engines. As diesel fuels contain more than 2000 chemical constituents, molecular simulations are the logical option to obtain the required properties data. The presentation highlighted the role of lumped models to account for the chemical functional groups for a range of expected biodiesel fuels and thereby manage the scale of the computational challenge. The SBES success is illustrated by a favorable match of computed (MD) vs. experimental values for the kinematic viscosity biodiesel component rapeseed methyl esters: $5.7 \pm 0.9 \text{ mm}^2/\text{s}$ vs. $5.1\text{--}5.6 \text{ mm}^2/\text{s}$, and computational runs ranging up to 250 MPa.

For the second theme, separation and sequestration of CO₂, SBES plays a role in both aspects of CO₂ technology. For the separation of CO₂ by capture from flue gas, new CO₂-absorbant nanoporous materials are integral to a successful design. This is especially important in view of the dangerous compounds (acid gases) present in the capture stream. But the huge parameter space of known MOF structures makes SBES a logical option to explore for the best structures (simulations of CO₂ adsorption with the Grand Canonical Monte Carlo method). (MOFs or metal organic frameworks are crystalline compounds consisting of metal ions coordinated to rigid organic molecules to form 1D, 2D or 3D structures that are porous; a class known as isoreticular MOFs or IRMOFs—as published in *Nature* 1999 by Yaghi and O’Keeffe—is of great interest to the chemical sciences community because they exhibit high storage capacity for gases.) For CO₂ sequestration, the transport modeling for the CO₂ reservoir requires data for the transport properties (e.g., the diffusion coefficients in the Stefan-Maxwell relations) of multicomponent mixtures of CO₂ with other residual reservoir fluids. Simulations (Monte Carlo methods) to predict these properties are integral to reservoir design and operations.

These IFP projects are in collaboration with Paris-Sud University, ENSCP (Prof. A. Fuchs, also on our schedule), University of Tarragona, University of Clermont-Ferrand, Universidad Simón Bolívar in Venezuela, and the oil industry (TOTAL S.A.). In view of the huge parameter space entailed and the number of chemical species to achieve the more realistic simulations, these projects illustrate the potential impact of future increases in HPC and SBES capabilities.

Recent Papers

Ungerer, P., C. Nieto-Draghi, B. Rousseau, G. Ahunbay, and V. Lachet. 2007. Molecular simulation of the thermophysical properties of fluids: From understanding toward quantitative predictions. *Journal of Molecular Liquids*. 134(1–3):71–89; doi:10.1016/j.molliq.2006.12.019.

Nieto-Draghi, C., T. de Bruin, J. Pérez-Pellitero, J.B. Avalos, and A.D. Mackie. 2007. Thermodynamic and transport properties of carbon dioxide from molecular simulation. *Journal of Chemical Physics* 126(6):064509; doi:10.1063/1.2434960.

Bonnaud, P., C. Nieto-Draghi, and P. Ungerer. 2007. Anisotropic united atom model including the electrostatic interactions of benzene. *Journal of Physical Chemistry B, Condensed Matter* 111(14):3730–3741; doi: 10.1021/jp067695w.

Nieto-Draghi, C., B. Creton, A. Bocahut, T. de Bruin, and V. Lachet. 2007. Molecular simulations of oolycyclic aromatic compounds. Presentation at the AICHE 2007 Annual Meeting, 3–9 November 2007. See <http://aiche.confex.com/aiche/2007/techprogram/P84103.htm>.

Galliero, G., C. Nieto-Draghi, C. Boned, J.B. Avalos, A.D. Mackie, A. Baylaucq, and F. Montel. 2007. Molecular dynamics simulation of acid gas mixtures: A comparison between several approximations. *Industrial & Engineering Chemistry Research* 46(15):5238–5244; doi:10.1021/ie061616l.

Nieto-Draghi, C., P. Ungerer, and B. Rousseau. 2006. Optimization of the anisotropic united atoms intermolecular potential for n-alkanes: Improvement of transport properties. *Journal of Chemical Physics* 125(4):044517–044517.

Dr. Pascal Raybaud: Applications of DFT in the Rational Design of Industrial Refining Catalysts

This research program is motivated by the following challenges faced by the refining industry: production of ever cleaner fuels; heavier petroleum feedstocks; and CO₂ greenhouse gas issues, including fuels from biomass. Each of these can be impacted favorably by the development of better nanoscale catalysts, which in turn provide exciting opportunities for SBES in the form of density functional theory (DFT) molecular modeling. After a brief introduction to the structure of metal-supported catalysts, the presentation focused on DFT issues: chemical events (bond breaking, bond formation) at catalytic sites of complex organic/inorganic materials. The time scales are ps-ns, and length scales are 1-10 nm; already available are atomic-scale descriptions of the active sites and the role of the reaction conditions (T, p). The main software packages are VASP, Gaussian, Materials Studio, Medea, and these are employed to simulate the energy landscape and the resulting microkinetic modeling (BEP, volcano curves) of new catalytic materials. The expected increase in HPC and SBES capabilities in the future will allow increasing systems size (beyond 500 atoms) and increased system complexity (multiple phase systems with solvent effect on heterogeneous and homogeneous catalysts by QM-MM methods).

Recent Papers

Raybaud, P., D. Costa, M. Corral Valero, C. Arrouvel, P. Sautet, and H. Toulhoat. 2008. First principles surface thermodynamics of industrial supported catalysts in working conditions. *Journal of Physics: Condensed Matter* 20:064235.

Daudin, A., A.F. Lamic, G. Perot, S. Brunet, P. Raybaud, and C. Bouchy. 2008. Microkinetic interpretation of HDS/HYDO selectivity of the transformation of a model FCC gasoline over transition metal sulfides. *Catalysis Today* 130:221–230.

Raybaud, P. 2007. Understanding and predicting improved sulfide catalysts: Insights from first principles modeling. *Applied Catalysis A: General* 322:76–91.

Dr. Diego Klahr: Supercomputing at IFP: Capabilities and Plans for the Future.

HPC has a history from the 1960s of being recognized in the IFP as essential for oil and gas research. The HPC lineup from the 1960s to 2003 featured the CDC 3600, 6600, and 7600; Cray XMP 1S, Convex C2; and Fujitsu VP2400 and VPP500. More recently, from 1997 to 2003, the lineup has included NEC SX5 and SGI Origin2000. The institute has about 110 active users of HPC facilities; 80% of the internal applications originate from the oil and gas research units of the institute. Profiles of the usage show programming in Fortran77, Fortran90, Fortran96, C/C++, and Java. Parallelism in the applications shows up as OpenMP, pthreads, MPI, and hybrid methods. Applications run the entire gamut of memory bound, I/O bound, and MPI bound. Some of the more popular commercial packages run on the facility include Fluent, Abaqus, Gaussian, and VASP. At a high level, the oil and gas codes exhibit the following profiles:

- Upstream Market
 - Seismic: I/O bounded from 1 to 50 TB of data, big files, parallel I/O, memory bound
 - Geology: memory bound and MPI bound
 - Reservoir modeling: memory bound, MPI bound
- Downstream Market
 - Molecular Dynamics: very long simulation time
 - Car engine simulations: memory bound, MPI bound, or OpenMP limited

In 2007, the HPC facility delivered 3.4 Mhrs and 99.5% availability, up from 0.1 Mhrs in 2003. The main system featured a 3.6 FLOPS peak performance using 200 mixed nodes AMD (2.4, 2.6 single and dual cores) Itanium2 and Power5, Infiniband SDR interconnect, and a 14 TB GPFS parallel file system.

Beyond the March 2008 upgrade, future plans for hardware call for investigation of hybrid capabilities via graphics cards, FPGAs, the cell processor. From a system management perspective, the greater scale will require proactive supervision to anticipate hardware and software failures. HPC R&D interests also continue in programming models and validation of cluster components, all under the X1848 project umbrella.

DISCUSSIONS AND ADDITIONAL COMMENTS

The IFP presentations, while meant to be a sample overview of the IFP's research activities in SBES, demonstrate that the institute is deeply grounded in the fundamental mathematical, computational, science, and engineering concepts that form the foundation of SBES projects in the service of technological advances. The coupling of this sound foundation with the institute's significant and ongoing contributions to the solution of "real world" problems and challenges of energy, environment, and transportation form an excellent showcase for the present capabilities of SBES and its even greater future potential. This site offered many excellent examples for illustrating the main themes of the WTEC panel's report chapters.

This site visit and discussion gave the WTEC visiting team an opportunity to observe the interplay between government policy (e.g., green policy, Kyoto accords, etc.) and a government-owned institute's applied research with a long-time horizon. The IFP research is more applied than that typically found in universities, and its time horizon starting from fundamental science is of longer duration than that typically found in industry development projects. Use of SBES to develop thermophysical properties tables of complex fuel mixtures (biodiesels) at high pressures is a good example of precompetitive research and a platform that benefits the entire automotive industry and all bio-fuel industry participants. The strategy of developing an extensive intellectual property (IP) portfolio for the refining of heavy crude with the view of a refinery feedstock mix shifting in that direction (e.g., increasing share of transportation fuels originating from tar sands) is another illustration of the long time horizon of the IFP.

The IFP provides interesting examples of an emerging model for funding public goods R&D. Thanks to its growing IP portfolio and permitted ownership (typically minority stakes) of companies that are spun-out of its research discoveries, in the future, the institute expects to generate at least 40% (as recalled from oral statements) of its budget from such sources. This mechanism should be of wide, global interest as a win-win incentive structure to accelerate the pace of research discoveries in government laboratories.

The institute is interested in recruiting high-quality graduate students and postdocs from all over the world. Opportunities are also available, among others, for visiting faculty on sabbaticals.

Site: **Institute of Fluid Mechanics of Toulouse (IMFT)**
Unite Mixte de Recherche 5502 CNRS-INPT -UPS
Allée du Professeur Camille Soula, 31400 Toulouse, France
<http://www.imft.fr>

Date Visited: February 25, 2008

WTEC Attendees: G. Karniadakis (report author), A. Deshmukh, G. Lewison

Host: Dr. Marianna Braza
 Tel: 33 5 6128 5839; Fax: 33 5 6128 5899
 Email: braza@imft.fr
 Guillaume Barbut, PhD Student
 Rémi Bourguet, PhD Student

BACKGROUND

The Institute of Fluid Mechanics of Toulouse (Institut de Mécanique des Fluides de Toulouse, IMFT) is a public research Institute belonging administratively to the CNRS, the Institut National Polytechnique of Toulouse (INPT) and the University Paul Sabatier (UPS). It was founded in 1918 with focus on flow experiments in small-scale models but today its activities are equally divided in computational and experimental fluid mechanics. It employs about 200 permanent staff while more than 120 faculty and 100+ PhD students and postdocs are involved in the research activities. There are six research groups focusing on turbulence, aeronautics (including controls, optimization, sensitivity analysis, shape optimization, reduced order modeling), combustion, two-phase flows, environmental flows, and hydrology. IMFT has one of the most advanced fluids experimentation facilities in France, including a historic experimental subsonic wind tunnel designed by G. Eiffel, established in 1938 and still in use, that General de Gaulle visited in 1959. IMFT is the biggest fluid mechanics laboratory in France (there are other ones: in Lyon – LMFA (smaller 80 faculty); Marseille – IRPHE (phenomena beyond equilibrium); LABM, (Laboratoire d’Aérodynamique et de Biomécanique du Mouvement), Ecole Centrale Nantes (biomechanics & aerodynamics) – LEA (Laboratoire d’Etudes Aérodynamiques – Poitiers), IMFL (Institut de Mécanique des Fluides de Lille) and LADYX, Ecole Polytechnique Paris). There are also other public research Institutes doing CFD as the INRIA (Institut National de Recherche en Informatique et Automatique).

RESEARCH

The goal of IMFT is to advance research in the mechanics of fluids with close experimental validation for applications in the areas of energy, processes, the environment and health. The group of Dr. Braza, Director of Research CNRS, has been working on direct (DNS) and large-eddy simulations (LES) of turbulent prototype flows for a long time, and more recently on the detached eddy simulation (DES) of industrial-complexity flow applications. Specifically, DES was the central theme of a European initiative with the name DESider, (Detached Eddy Simulation for Industrial Aerodynamics), (<http://dd.mace.manchester.ac.uk/desider>) – a 10 M Euro program motivated by the increasing demand of the European aerospace industries to improve their CFD-aided tools for turbulent aerodynamic systems with massive flow separation. DES is a hybrid of LES and statistical turbulence modeling and it is designed as a multiscale framework for modeling and capturing the inner and outer scales of turbulence. DESider involved many academic and industrial groups with EADS (European Aeronautics & Defense Systems) acting as the coordinator of the program and IMFT a main participant providing turbulence simulations but also unique experimental data for validation of DES codes. Other companies participating are Dassault, Volvo, Peugeot/Citroen, Airbus, etc. All together, 18 organizations were involved (8 from industry, 5 from research institutions, and 5 from Universities) from 8 different European countries. DESider was supported by the 6th Framework program (2002-2006) on “Aeronautics and Space.”

Furthermore, IMFT participates in the UFAST (Unsteady Effects in Shock-Wave-induced separation) project of the European Sixth Framework Programme, which is coordinated by the Polish Academy of Science (Gdansk) Institute of Fluid-Flow Machinery, and which includes about 18 partners. The objectives are to perform *closely coupled experiments and numerical investigations concerning unsteady shock wave boundary layer interaction (SWBLI)* to allow for feeding back numerical results to the experiments, and vice versa, for the sake of physics and modelling of compressibility effects in turbulent aerodynamic flows. Using RANS/URANS and hybrid RANS-LES methods, UFAST aims at assessing new methods for turbulence modelling, in particular for unsteady, shock dominated flow. UFAST investigates the range of applicability between RANS/URANS and LES for transonic and supersonic flows.

A follow up-to DESider is the project EMORPH, a multiphysics project (called FP7-AAT-2008-RTD-1/CP-FP) that couples smart materials (electro-active morphing) with CFD and with structural mechanics. Its objective is the development of integrated design tools, controller design and flight control (use distributed actuators that use vibration energy). IMFT is the lead of the project in collaboration with Airbus Toulouse and the LAPLACE (Laboratoire Plasma et Conversion d'Énergie) laboratory that has expertise on new electro-active and energy-efficient materials. IMFT has very close collaborations with Airbus Toulouse and very often Airbus engineers co-supervise PhD students associated with IMFT. However, Airbus researchers have their own turbulence code (ELSA, developed by ONERA) and do not use directly the in-house codes of IMFT.

In terms of validation and uncertainty quantification, in addition to grid tests for numerical accuracy, turbulence model accuracy is evaluated by comparisons against unique 3D particle image velocimetry (PIV) data gathered from experiments at IMFT. Such extensive experimental databases are required in validating a new tensorial turbulence, eddy-viscosity model that researchers at IMFT have developed. In addition, researchers conduct tests for upstream noisy inflow conditions and have recently published joint work with Airbus on sensitivity of lift and drag forces due to such random disturbances.

COMPUTING INFRASTRUCTURE

IMFT is the biggest user of supercomputing cycles in France. It uses primarily the three national supercomputing facilities (Paris IDRIS funded by CNRS, 207 Tflops; Montpellier CINES with mainly IBM machines since they have a factory nearby, 50 Tflops; and Toulouse CALMIP in Université Paul Sabatier). Overall, the computing resources are adequate, and allocation of resources is done through grants. Other available supercomputing resources include CEA in a suburb of Paris (Atomic Energy) and Project Grid 5000—connecting clusters all over France.

EDUCATION & EXCHANGE PROGRAMS

Of the 100+ PhD students affiliated with IMFT, most are working on simulation work and are supported by IMFT and national fellowships (Ministry of Education and Research). The Ministry of Defense (DGA-Délégation Générale pour l'Armement, <http://www.recherche.dga.defense.gouv.fr>) provides fellowships for postdocs on selected topics such as aeroelasticity, biomechanics, and fluid mechanics (about 20 nationally in FM). There is a recent emphasis by the French government to encourage doctoral and postdoctoral training abroad through the Lavoisier program (<http://www.egide.asso.fr>). In addition, CNRS funds permanent staff and senior researchers to visit abroad for six months or so.

CONCLUSIONS

IMFT is a leading fluid mechanics laboratory in Europe with a focus on both fundamental research on simulation and also industrial applications. It is closely interacting with Airbus Toulouse and other companies (Peugeot, Renault) and has been involved in interesting European projects involving many universities, companies, and other research institutions. Its unique experimental facilities serve the simulation work very well, as they provide very useful pointwise measurements required for model validation, especially for unsteady turbulent flows with massive separation.

Site: **IRIT (Institut de Recherche en Informatique de Toulouse), and
ENSEEIH (Ecole Nationale Supérieure d'Electrotechnique, d'Electronique,
d'Informatique, d'Hydraulique et des Télécommunications)**
(At ENSEEIH) 2, rue Charles Camichel B.P. 7122,
31071 Toulouse Cedex 7, France
<http://www.irit.fr/sommaire.php3?lang=en>
<http://www.enseeih.fr/en/index.html>
<http://gridtlse.org>

Date Visited: February 25, 2008

WTEC Attendees: A. Deshmukh (report author), G. Karniadakis, G. Lewison

Hosts: Prof. Patrick Amestoy, Head, Parallel Algorithms and Optimization (APO) Team
IRIT – Computer Science and Mathematics Laboratory
ENSEEIH – Institut National Polytechnique de Toulouse
Tel: (33) 05 61 58 83 85; Fax: (33) 05 61 58 83 06
Email: Patrick.Amestoy@enseeih.fr

Dr. Ronan Guivarch, Assistant Professor; Referent, Grid-TLSE project
IRIT – Computer Science and Mathematics Laboratory
ENSEEIH – Institut National Polytechnique de Toulouse
Tel: (33) 05 61 58 84 08; Email: Ronan.Guivarch@enseeih.fr

Daniel Ruiz
IRIT - Computer Science and Mathematics Laboratory
ENSEEIH - Institut National Polytechnique de Toulouse
Tel: (33) 05 61 58 83 35; Email: ruiz@enseeih.fr

Victoria Moya, Visiting Student from Zaragoza, Spain

Prof. Michel Daydé (not present)
Vice-Head, IRIT – Computer Science and Mathematics Laboratory
ENSEEIH – Institut National Polytechnique de Toulouse
Tel: (33) 05 61 58 82 70; Email: Michel.Dayde@enseeih.fr

BACKGROUND

The WTEC visiting team visited IRIT to talk primarily about Grid-TLSE with several members of the collaborative Grid-TLSE team from ENSEEIH, IRIT, and INPT.

ENSEEIH (Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications)

In 1907, the city of Toulouse assisted its seven-century-old university in the creation of an institute dedicated to teaching and research, the Institute of Electrotechnology and Applied Mathematics of the University of Toulouse. Its principle goals were to train the engineers and managers necessary for the hydraulic and electric renovation of the southwest region of France and to contribute to the technological development of these scientific disciplines. The increasing number of students and technical progress led to the creation of the Electrical Engineering and Hydraulics Departments in 1955 followed, in 1956, by the Electronics Department, and in 1959, for the first time in a French engineering school, a Department of Applied Mathematics. In 1967, it became the Department of Computer Science. Today, presided over by Professor Alain Ayache, ENSEEIH is structured into 5 departments: Electrical Engineering and Automation; Electronics and Signal Processing; Computer Science and Applied Mathematics; Hydraulics and Fluid Mechanics; and Telecommunications and Networks. There are 5 closely linked research labs.

IRIT (Institut de Recherche en Informatique de Toulouse)

IRIT was founded in 1990. Since then, it has played a prominent role in Toulouse computer science research. It brings together more than 400, members among which are 300 researchers, faculty members, and PhD students affiliated with CNRS (Centre National de la Recherche Scientifique), INPT (Institut National Polytechnique de Toulouse), UPS (Université Paul Sabatier) and UT1 (Université Toulouse1 Sciences Sociales). Research at IRIT covers most of the fields where computer and information science is in progress, be it in its core, ranging from computer architecture to software engineering and computer networks, or in its most contemporary developments like artificial intelligence and cognitive systems, multimedia man-machine interaction, and image interpretation and synthesis. IRIT promotes interdisciplinary research so as to cross-fertilize information science and technology with other disciplines, such as linguistics, psychology, ergonomics, and neurosciences, which in turn can benefit from new information-driven concepts and tools. The researchers at IRIT are involved in major national and European research projects.

INPT (Institut National Polytechnique de Toulouse)

INPT is part of a consortium of three Institut National Polytechniques in France referred to as "grandes écoles d'ingénieurs" in Grenoble, Nancy, and Toulouse. These national polytechnics have 19 engineering schools and 75 research laboratories. These public engineering schools award the "national diploma of engineer" accredited by the French Ministry of Education. INP Toulouse includes ENSAT, ENSEEIHT, (N7) and ENSIACET (A7), and an affiliated engineering school, ENIT in Tarbes. Research at INPT is concentrated in 15 laboratories in collaboration with CNRS (French National Center for Scientific Research), INRA (National Institute for Agricultural Research), and other universities in Toulouse.

GRID-TLSE PROJECT

The goal of the Grid-TLSE (Test for Large Systems of Equations) project is to design an expert site that provides a user-friendly test environment for expert and nonexpert users of sparse linear algebra software. Sparse linear algebra software provides sophisticated algorithms for pre/post processing of the matrices. The selection of the most efficient solver depends on several problem and computational parameters, such as ordering, amount of memory, computer architecture, libraries available, etc. Hence, there is a need for a decision support system to aid users in solving this multiparametric problem to select the most appropriate solver for their problem and computational environment.

This is a multiyear project that started in January 2003. It was originally funded under the ACI GRID Program by the French Ministry of Research. Currently it receives funding from ANR (Agence Nationale de la Recherche) under the SOLSTICE and LEGO projects, and from CNRS under the REDIMPS program. The following French research laboratories are involved in this project: ENSEEIHT, CERFACS-Toulouse, IRIT-Toulouse, LaBRI-INRIA, and LIP ENS-Lyon/INRIA. This project also involves the following industrial partners: CEA-CESTA, CNES, EADS, EDF, and IFP.

The Grid-TLSE project aims to develop a Web portal that will provide easy access to tools allowing comparative analysis of solvers available for sparse matrix computations. Users can submit their own problems or choose from a library of matrices, which include domain matrix collections, such as the Rutherford-Boeing and University of Florida sparse matrix collections.

The Grid-TLSE architecture consists of four levels. The user describes the problem to be solved using a Web portal, WebSolve. The encoded problem description is used by the second level, Weaver, to generate corresponding candidate solver/problem instance list. This list is converted to XML format by the Gridcom level. This specification can be sent to direct solvers on grid resources (Grid 5000 in France) using the DIET middleware, which is a simpler version of Globus developed in GRAAL at Ecole Normale Supérieure de Lyon. Direct solvers are launched onto a grid of remote servers using CORBA based DIET middleware. Currently, MUMPS (INRIA Bordeaux), UMFPACK (Florida), and SuperLU (Berkeley) solvers are being considered for

implementation. This architecture collects statistics on the performance of these solvers for different parameters and operating constraints. Details of the Grid-TLSE project are available at <http://gridtlse.org>.

CONCLUSIONS

At the time of the WTEC visit, the Grid-TLSE project was still in the design phase. Project team members expect it to go into production mode in a year's time (early 2009). They anticipate that this utility will lead to significant savings in researcher time in selecting the best solver for problems at hand—from weeks to a day or less. The researchers also plan to extend the current framework for domains beyond linear algebra.

A demonstration of the GRID-TLSE platform will take place during the November 2008 Supercomputing Conference in Austin, Texas, in the framework of the LEGO project.

REPRESENTATIVE PAPERS

- Amestoy, P., M. Daydé, R. Guivarch, C. Hamerling, and M. Pantel. 2007. Use of scenarios for generating dynamic execution workflows over the grid within the Grid-TLSE project. In *Proceedings, International E-Conference on Computer Science, Electronic Conference, 10/07/06-14/07/06*, Vol. 8, T. Simos and G. Psihoyios, eds., 1–11. Leiden, the Netherlands: BRILL (Lecture Series on Computer and Computational Science).
- Daydé, M., A. Hurault, and M. Pantel. 2006. Semantic-based service trading: Application to linear algebra. In *Proceedings, VECPAR'06 - Workshop on Computational Grids and Clusters (WCGC 2006)*, Rio de Janeiro, Brésil, 10/07/06-13/07/06, 622–633. Berlin: Springer-Verlag, LNCS 4395.
- Amestoy, P., M. Daydé, C. Hamerling, M. Pantel, and C. Puglisi. 2007. Management of services based on a semantic description within the GRID-TLSE project. In *High-Performance Computing for Computational Science - VECPAR 2006*, 634–643. Berlin/Heidelberg: Springer.
- Daydé, M., A. Hurault, and M. Pantel. 2005. Gridification of scientific application using software components: The Grid-TLSE project as an illustration. In *Proceedings, CSIT 2005, Yerevan, Armenia, 19/09/05-23/09/05*, 419–427. Yerevan, Armenia: National Academy of Sciences of Armenia.
- Amestoy, P., F. Camillo, M. Daydé, L. Giraud, R. Guivarch, V.M. Lamiel, M. Pantel, and C. Puglisi. 2007. “Goal and status of the TLSE platform.” Presentation at the First REDIMPS Workshop, Tokyo (Japan), 29/05/2007 (unpublished).
- Amestoy, P., I.S. Duff, and J.-Y. L'Excellent. 2004. GRID-TLSE: A website for experimenting with sparse direct solvers on a computational grid. In *Proceedings, SIAM conference on Parallel Processing for Scientific Computing*, San Francisco, USA, February 2004.

Site: Paris Simulation Network*Participating Universities***École Nationale Supérieure de Chimie de Paris (ENSCP; Hosts)**

11, rue Pierre et Marie Curie
75231 Paris, Cedex 05, France
<http://www.enscp.fr/>

Ecole Normale Supérieure (ENS)

45, rue d'Ulm
F-75230 Paris Cedex 05, France
<http://www.ens.fr>

Université Pierre et Marie Curie (UPMC)

4 Place Jussieu
75005 Paris, France
<http://www.upmc.fr>

l'Université Paris-Sud 11 (UPS)

Bât. 300
91405 Orsay Cedex, France
<http://www.u-psud.fr>

Université d'Evry-Val-d'Essonne (UEVE)

Boulevard François Mitterrand
91025 Evry Cedex, France
<http://www.univ-evry.fr/>

Date Visited: February 28, 2008

WTEC Attendees: P. Cummings (report author), K. Chong, M. Head-Gordon, S. Kim

Hosts:

Professor Alain Fuchs, Director, ENSCP

Email: directeur@enscp.fr

Carlo Adamo, Laboratoire d'Électrochimie et Chimie Analytique, ENSCP

Email: carlo-adamo@enscp.fr

Anne Boutin, Le Laboratoire de Chimie Physique, UPS

Email: anne.boutin@lcp.u-psud.fr

Damien Laage, Département de Chimie, ENS

Email: damien.laage@ens.fr

Rodolphe Vuilleumier, Laboratoire de Physique Théorique de la Matière Condensée, UPMC. Email: Rodolphe.vuilleumier@lptmc.jussieu.fr

BACKGROUND

The Paris Simulation Network is a loose association of researchers within Paris engaged in electronic, atomistic, and coarse-grained simulations of chemical, material, and biological systems. It involves researchers Alain Fuchs and Carlo Adamo from the Ecole Nationale Supérieure de Chimie de Paris, (ENSCP); Daniel Borgis, J.T. (Casey) Hynes, Damien Laage, and Rodolphe Vuilleumier from the Ecole Normale Supérieure (ENS); Bertrand Guillot and Pierre Turq from the Université Pierre et Marie Curie (Paris 6, UPMC); Anne Boutin and Bernard Rousseau from the Université de Paris-Sud (Orsay, UPS); and Marie-Pierre Gageot and Riccardo Spezia from the Université d'Evry-Val-d'Essonne (UEVE). Several of these eminent researchers hosted the WTEC team (see list of hosts, above).

The participating institutions are quite varied. ENSCP has just 300 undergraduates majoring in the chemical sciences and chemical engineering and 100 doctoral and post-doctoral candidates, with 59 teaching and research staff members. ENS is likewise small, enrolling students in both the humanities and sciences after two years of study at another institution. It is essentially a graduate-only school—students do not receive baccalaureate degrees from the ENS, only graduate degrees. UPMC and UPS are both large institutions, with approximately 30,000 students each. Pierre & Marie Curie University (UPMC) is one of the largest universities teaching science and medicine in France, and indeed in Europe, with 4000 researchers and teaching academics/researchers, 180 laboratories, and 8000 of its 30,000 students in graduate studies. UPS has 1800 teaching staff, 1300 engineers, technicians, administrative staff and maintenance staff, as well as 1200 research scientists and 900 technical and administrative staff of the national research organizations (CNRS, INSERM, INRA, CEA). UEVE is a new university, established in 1991, and is mid-sized, with just over 10,000 students.

R&D ACTIVITIES

The members of the Paris Simulation Network (PSN) cover a wide range of research areas, from first-principles molecular dynamics (Gageot, Laage, Spezia, and Vuilleumier) and quantum chemistry (Adamo), to atomistic simulation (Boutin, Fuchs, including Gibbs ensemble Monte Carlo simulation for fluid phase equilibria), to coarse-grained techniques such as dissipative particle dynamics (Rousseau), forcefield development (Guillot, particularly for water), chemical reaction and solution dynamics (Bougis and Hynes), and theory of electrolyte solutions (Turq). There are quite a number of researchers focused on using simulation methods to understand spectroscopy (IR, NMR, UV-Vis, EPR, RX absorption and diffraction) for both gas and condensed phases (including biological systems). The WTEC team's meeting with the PSN scientists did not focus on the research activities of the members *per se*, but rather on the questions and issues raised in the questionnaire sent earlier to the participants by WTEC. For specifics of the research activities of PSN members, please refer to the websites given in the references section (some of which are a little dated, but nevertheless give an accurate flavor of the research of each individual). The main research targets of PSN members can be summarized as follows:

- Multiscale simulations
- Realistic and direct calculation/prediction of experimental observables
- Development/implementation of theoretical models
- Quantum-classical simulations
 - Forcefield matching
 - Mixed quantum-classical simulation, with application to
 - Chemical reactions (liquid phase and/or biological systems)
 - Radiolytic processes
 - Catalysis (homogeneous and inhomogeneous)
- Micro-mesoscopic simulations, with applications to
 - Polymer systems
 - Colloids

Particular mention was made of the fact that there are considerable research funds available for research related to nuclear waste treatment. This is not surprising, since, as a result of a long-standing policy based on energy security, France derives 78% of its electricity from nuclear energy. It is the world's largest net exporter of electricity (earning €3 billion/year) due to its very low cost of generation. Consequently, France has been very active in developing nuclear technology, and nuclear-related equipment and services are a major export.

The PSN focuses that are related to energy are in materials design for carbon dioxide capture and storage, radioactive waste treatment at the atomistic level, enzymatic catalysis, and photovoltaic cells. The PSN members pointed to opportunities raised by the new European community regulation on chemicals and their safe use (EC 1907/2006), known by the acronym REACH (Registration, Evaluation, Authorization and restriction of CHemical substances, http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm). The REACH regulation

requires specific data on every chemical produced and/or transported into and out of the EU. It has apparently been established that data derived from quantum mechanical and atomistic simulations are permissible in the REACH database, thus establishing the opportunity for modelers to replace costly experiments with accurate computationally derived properties.

The software codes used within the PSN are 50% developed in-house. These codes include Gibbs (a general Monte Carlo code); Newton and MDVRY (classical molecular dynamics codes); a mesoscopic simulation code (for dissipative particle dynamics and Brownian dynamics); QCMD (a mixed quantum/classical molecular dynamics code); a classical DFT code; and many post-processing codes (e.g., to analyze Car-Parrinello molecular dynamics, CPMD, output files). About 30% of the codes used are of the open-source/noncommercial variety: DL_POLY, CPMD, PWSF, CP2K, GROMACS, GAMESS, and NW-CHEM. Some PSN researchers are involved in the development of extensions of CPMD and PWSCF. The remaining 20% of the codes used are commercial: GAUSSIAN, ADF, CRYSTAL, AMBER and CHARMM. Again, one PSN member (Adamo) is active in development for GAUSSIAN. For the in-house codes, at each release the code is validated on a standard set of test problems. Beyond this, the PSN researchers do not perform validation beyond what many researchers do—that is, compare with experiment whenever possible, test against known results on test problems for a given algorithm, etc.

In the area of big data, PSN researchers will be participating in a new CECAM initiative to develop a database of *ab initio* molecular dynamics trajectories. For PSN researchers, the breakdown of computing resources used is as follows: local clusters account for 80%, European computer centers (CINECA, CEA, Edinburgh, IFP, ...) constitute 10% (often made available through collaborations), and the CNRS national center (IDRIS, <http://www.idris.fr/>) accounts for the remaining 10%. Codes are typically used in parallel, particularly on the external resources. In the future, they would like to see order-N algorithms for first-principles calculations, improvements in processor-to-processor communication and memory access speeds.

In education and training, the WTEC team's hosts viewed the role of universities to be teaching theory rather than computer science. They view training in simulation methods as being achieved primarily through participation in the training programs of CECAM and the UK CCPs (Collaborative Computational Projects, <http://www.ccp.ac.uk/>). The PhD students at PSN institutions are essentially all French; post-doctoral researchers are primarily from Europe (Italy, UK, Germany, Spain, and others). Of the graduates of the PSN groups, 20% end up as teachers at undergraduate (nonresearch) universities, 60% end up in academia, and 20% in industry.

Funding in the areas relevant to the PSN comes from the federal government (France's Ministry of Research, "Agence Nationale de la Recherche" [ANR], CNRS, etc), Europe (the European Science Foundation, <http://www.esf.org>, and other EU sources), CECAM, and industry. Generally, the trend in funding for research in SBES is positive.

CONCLUSIONS

Researchers in the PSN undertake a wide range of research activities in molecular modeling, ranging from quantum methods to mesoscale simulations. The group was generally upbeat about the state of support for their research from French and European funding sources.

REFERENCES

- | | | |
|-------|-------------|---|
| | Borgis | http://www.univ-evry.fr/labos/lmsmc/membres/dborgis.html |
| ENS | Hynes | http://www.chimie.ens.fr/w3hynes/casey.php |
| | Laage | http://www.chimie.ens.fr/w3hynes/damien.php |
| | Vuilleumier | http://www.lptl.jussieu.fr/users/vuilleum/ |
| ENSCP | Adamo | http://www.enscp.fr/labos/LECA/Research/site_msc/index.htm |
| | Fuchs | http://www.enscp.fr/spip.php?article139 |

UEVE	Gaigeot	http://www.lambe.univ-evry.fr/spip.php?rubrique54&lang=en
	Spezia	http://www.lambe.univ-evry.fr/spip.php?rubrique57&lang=en
UPMC	Guillot	http://www.lptl.jussieu.fr/users/guillot/Welcome.html
	Turq	http://www.li2c.upmc.fr/-Turq-Pierre-?lang=en
UPS	Boutin	http://pagesperso.lcp.u-psud.fr/boutin/
	Rousseau	http://pagesperso.lcp.u-psud.fr/rousseau/

Site: Science & Technology Facilities Council (STFC) Daresbury Laboratory
 Computational Science and Engineering Department (CSED)
 Synchrotron Radiation Source (SRS)
 Daresbury Science and Innovation Campus
 Daresbury, Warrington, WA4 4AD, UK
 STFC Daresbury: <http://www.stfc.ac.uk/About/Find/DL/Introduction.aspx>
 HECToR: <http://www.hector.ac.uk>

Date Visited: February 25, 2008

WTEC Attendees: S Kim (report author), P. Cummings, K. Chong

Hosts: Prof. Nicholas M. Harrison Computational Science & Engineering Dept, DL and Imperial College; Email: n.m.harrison@stfc.ac.uk
 Dr. Richard Blake, Acting Director, Computational Science & Engineering Dept. Email: r.j.blake@dl.ac.uk
 Dr. Keith Refson, STFC Rutherford Appleton Laboratory, CSED
 Email: k.refson@rl.ac.uk
 Dr. Mike Ashworth, Advanced Research Computing Group, & Computational Engineering Group; Email: m.ashworth@dl.ac.uk
 Dr. Martyn Winn, Team Leader, Computational Biology Team, SED
 Email: m.d.winn@dl.ac.uk

BACKGROUND

The Daresbury Laboratory (DL) began operations in 1962 and officially opened in 1967 as the Daresbury Nuclear Physics Laboratory. Today it is one of the major national laboratories of the UK with a staff of 500. The nuclear legacy today takes the form of the Synchrotron Radiation Source (SRS) which is one of the major facilities of the Science & Technology Facilities Council (STFC), and accelerators ALICE and EMMA. A strong core of activities in SBES have evolved from DL's historical roots and is enabled by a multi-institutional supercomputing consortium, and by HPCx and DL's active participation in the UK's Collaborative Computational Project (CCP) model for supporting community-scale software development. Furthermore, SBES activities, most notably an initiative to create the Hartree Centre (a CSE institute), figure prominently in the future of the Daresbury Laboratory.

COMPUTING FACILITIES

The WTEC visiting team's hosts described HPC hardware resources physically at the Daresbury site and also those more broadly in the STFC framework but having DL as the STFC-responsible entity. Located at Daresbury, the HPCx supercomputer is an IBM System P575 high-performance server, maintained by a consortium led by the University of Edinburgh, UoE HPCX Ltd., and funded by the UK's Engineering and Physical Sciences Research Council (EPSRC).

More recently (2007), the STFC/DL joined a consortium (along with University of Edinburgh and NAG, Ltd.) that runs the High End Computing Terascale Resources (HECToR) at the Edinburgh supercomputer center (EPCC). In the first phase, HECToR featured 60 TFLOPs, but a subsequent phase (2009) will upgrade this to 250 TFLOPs. A third phase is planned for 2011. At the start of operations, the system featured 60 Cray XT4 cabinets connected to 576 TBs of storage with the UNICOS/lc operating system. The next upgrade is the addition of a Cray Black Widow supercomputer. For software, the discussion centered on the UK's experiences with large-scale efforts in community code development in the CCPs. (CCPs are discussed in greater detail in the first presentation below, and the list of all CCP projects is provided at the end of this report.)

PRESENTATIONS

The WTEC team's hosts at DL prepared a most informative lineup of presentations that mapped well to the main themes of our SBES report and that addressed our list of questions. After a brief discussion of the morning's agenda by our primary host, Prof. Nicholas Harrison, we listened to five presentations (PowerPoint files provided for each presentation), followed by a general discussion. We also had a poster session over the lunch break before departing.

Dr. Richard Blake: *The Hartree Centre: A New CS&E Institute on the Daresbury Science and Innovation Campus*

Dr. Blake, Acting Director of the CSED (Computational Science and Engineering Department), presented plans for a new CSE institute, the Hartree Centre (named in honor of the prominent computational chemist). His talk is reported in detail because it provides a helpful guide to the organizational structure of the UK's government funding bodies (the seven Research Councils of the UK's funding body, the Office of Science and Innovation), as well as the strategic context for SBES at DL.

- For funding of SBES activities, the three most relevant councils are the Science & Technology Facilities Council (STFC), the Biotechnology & Biological Sciences Research Council (BBSRC), and the Engineering & Physical Sciences Research Council (EPSRC).
- The STFC supports large-scale experimental facilities for the UK research community, including neutron sources; lasers/light source science; HPC, CSE, and e-science; engineering instruments; and accelerators; it also funds the facilities for particle physics and astronomy. The STFC is a recent creation formed by the merger of the Council for the Central Laboratory of the Research Councils (CCLRC) and Particle Physics and Astronomy Research Council (PPARC) in 2007 and operates from Swindon, Rutherford, Daresbury, Edinburgh, and Chilton with an annual budget of UK£735 million.
- DL benefits from strategic investments that resulted from a 2004 study that supported the creation of a "science and innovation in the campus model" for the Daresbury and Harwell locations. The plans for the Hartree Centre can thus be placed in the context of the 2004 plan. The recent growth of CSE (headcount of 90 full-time employees, operating budget of UK£7 million) would receive the additional impetus of UK£50 million of capital and UK£16 million operating costs and a machine room (10,000 sq. ft. and 10 MW power with UK£10 million machines refreshed on a two-year cycle).
- Dr. Blake provided an overview of the CCP (Collaborative Computational Project) model and the history of CCPs. There have been 13 to date, with the CCP1 for the electronic structure of molecules. The 13 CCPs encompass activities of 370 groups and 53 HPC consortia involving 150 groups from the UK and 50 groups from the EU network. The CCP model advances CSE with "flagship" code development; maintenance and distribution of codes; periodic meetings and workshops; facilitating collaborations with overseas visitors; and issuance of regular newsletters. The governance is by the CCP Steering Committee, currently chaired by Prof. Peter Coveney, and membership is composed of the CCP chairs, the director of the European Center for Atomic and Molecular Simulation (CECAM), and international members.
- Dr. Blake concluded his presentation with one-slide overviews of the main themes, each covered as subsequent presentations.

Dr. Keith Refson: *First-Principles Modeling*

This was a presentation on the past, present, and future of the computational chemistry (quantum level) capabilities at DL and its integration with the UK and European frameworks for computational research and applications in industrial technology transfer. The presentation of the history of the CCP3 development of CASTEP (plane wave DFT code) and CRYSTAL (periodic solid state code) provided useful context illustrating the importance of the CCPs. CRYSTAL is used by 450 groups worldwide (46 in the UK), and there have been 1530 downloads of the CRYSTAL06 demo. CASTEP is distributed worldwide by Accelrys and had 314 citations in 2007. Thanks to CCP3, the high quality of the software engineering expertise invested in these codes is reflected in their flexibility and modular architecture. New mathematical/numerical methods, parallelism, and

most importantly new science (NMR, phonons, hybrid XC functionals, PIMD, GA, E-field response, etc.) can all be readily incorporated into these codes.

An example of large-scale modeling of the simulation of hydrated polypeptide involving 300 atoms + 310 water as a CASTEP PW calculation (ultrasoft potentials) in 8 hours on 512 processors on HECToR will become a “routine calculation.” The talk also had a nice illustration of the complementary synergy between experiments and *ab initio* modeling: X-ray diffraction, XAS methods, IR and Raman spectroscopy, and Neutron spectroscopy: all of these experimental methods provide complete information; the gaps in the structure-property relationships are filled in by *ab initio* simulations. This is essential to advances in materials research and design of next-generation catalysts.

Prof. Nicholas Harrison: *Energy – The Materials Challenge*

The presentation focused on the impact of SBES materials research in the energy field, e.g., fuel cells, hydrogen storage, nuclear containment, photovoltaics, novel electronics/spintronics, and solid state battery electrodes. The overarching and unifying theme is that the design space for the new materials in each of these applications is so large that experiments alone are unlikely to lead to the optimal (or even adequate) materials design, and that with recent and expected advances in algorithms and HPC hardware, the SBES approach is entering the “sweet spot” with respect to applicability and relevance for bridging the length and time scales encountered in foundational scientific predictions for structure & composition, thermodynamics of phases, reaction kinetics & dynamics, and electronic structure/correlation. The toolkit of codes—DL-POLY, CASTEP, CRYSTAL, kppw—provided a nice integration with the presentations on software development and CCP. The presentation on SBES in materials design for nuclear (fuels/waste) containment was especially memorable and timely in the context of solving the looming energy challenges in the post-fossil-fuel era. (*Note:* Prof. Harrison is also on the faculty of Imperial College and was on our Friday 2/29/2008 schedule at that campus.)

Dr. Michael Ashworth: *Continuum Modeling and Engineering*

This presentation was on continuum scale modeling (CFD) with emphasis on geophysical fluid dynamics (ocean modeling). The presentation also included a discussion of consortia for combustion and aerodynamics. Pending advances in SBES resources and algorithms increase the resolution scale for ocean models. Most notable in terms of new science is the ability to incorporate continental shelves (shelf seas) in the model and the consequential impact on ecological and environmental sciences as we incorporate the most important oceanic habitats into the models. For CFD, in turbulence research such as higher resolution of eddy scales, the highlight is the impact on applications to combustion and aerodynamic simulations. In microscale flows, trends in SBES suggest closing of the gap between molecular-scale and continuum-scale models.

Dr. Martyn Winn: *Computational Biology in STFC*

This overview was for the entire network of STFC, including activities at DL and a review of the importance of the CCP (and in this case, CCP4) in advancing the computational biology capabilities of the STFC. This talk also gave insights into the increasing presence of the BBRC in funding SBES activities in the biological sciences. CCP4 on X-ray crystallography revolved around software development for structure solution and refinement. But the STFC also realized the importance of associated developments in workflow: automation of structure determination, e-HPTX (management of workflow), DNA data collection, and PIMS (laboratory information management systems). *This experience is especially relevant to our report chapter on “big data”* and much of this was funded under the UK’s well known e-Science program. In terms of applications and impact on society, Dr. Winn gave a brief overview of the role of protein structure science in: (1) probing the causes of motor neuron disease (MND); and (2) target informatics in modern drug discovery research, e.g., the targets for cancer research (the Human EGFR family of receptors is the target for the oncology therapeutic Herceptin). The talk also gave a nice illustration of the multiscale approach for linking QM to QM/MM. He concluded his presentation with an overview of the increased scale of activity in computational biology that would be achieved in the context of a fully implemented vision of the Hartree Centre for CSE.

CONCLUSIONS

There were three overarching themes that registered with this author. First of all, the UK's Science and Technology Facilities Council and its network of laboratories and facilities is of great interest as an organizational structure to sustain the deployment of cyber- as well as classical infrastructure. In comparison, it is noted that the United States has the mission-oriented national laboratories (e.g., DOE, NIH, etc.) and some examples in fundamental science (e.g., the National Center for Atmospheric Research, NCAR), but it has nothing on the systematic scale of the STFC. Especially in the context of software development for scientific research, the STFC structure in combination with the Collaborative Computational Projects allows the UK to unify and leverage resources to design, develop, and sustain community codes with a high degree of professional software engineering standards, as exhibited in the projects led by our hosts.

Secondly, our host department, the CSED of the STFC (Daresbury Laboratory and Rutherford Appleton Laboratory), has significant activities in SBES materials research from quantum to continuum scale. These are thoughtfully directed as some of the most promising thrusts to impact the present and future landscape of energy technologies. The presentation on optimal materials design for nuclear containment was an especially impressive illustration of the timeliness of SBES.

Third, the ambitious scope of the plans at Daresbury for the Hartree Centre, an institute for CSE, is an indication of the great strategic value placed on the future role of SBES. If this proposal moves forward, given the track record of achievements, the Hartree Centre is likely to become a magnet for global collaborations, including SBES researchers from the United States.

Listing of the UK's Collaborative Computational Projects (CCP) and Project Heads²⁰

CCP1: Electronic structure of molecules (P. Knowles)

CCP2: Continuum states of atoms and molecules (E. Armour)

CCP3: Computational studies of surfaces (S. Crampin)

CCP4: Protein crystallography (J. Naismith)

CCP5: Computer simulation of condensed phases (M. Rodger)

CCP6: Molecular quantum dynamics (S. Althorpe)

CCP7: Astronomical spectra (D. Flower)

CCP9: Electronic structure of solids (J. Annett)

CCP11: Biosequences and function (D. Gilbert)

CCP12: High performance computing in engineering (S. Cant)

CCP14: Powder diffraction (J. Cockcroft)

CCPN: NMR in structural biology (E. Laue)

CCPB: Bio-molecular simulation (C. Laughton)

CCPP: Plasma physics CCP (tba)

²⁰ In subsequent discussions at other sites in the UK, WTEC panelists learned that the scope of the CCP model has been scaled back in some areas to focus primarily on code maintenance rather than new code development. However, they were informed by their hosts at Daresbury Labs that the CCP programme as a whole is expanding into new areas, and some of the well-established CCPs also continue to take a leading role in developing new theoretical frameworks and associated codes.

Site: **Stuttgart Research Center of Simulation Technology**
Universität Stuttgart
Pfaffenwaldring 7a
70569 Stuttgart, Germany
<http://www.simtech.uni-stuttgart.de>

Date Visited: February 26, 2008

WTEC Attendees: C. Sagui (report author), G. Karniadakis, A. Deshmukh, G. Lewison, P. Westmoreland

Hosts: Prof. Dr.-Ing. Wolfgang Ehlers, Civil Engineering
 Email: ehlers@mechbau.uni-stuttgart.de

Prof. Dr.-Ing. Rainer Helmig, Civil Engineering
 Email: Rainer.helmig@iws.uni-stuttgart.de

Prof. Dr. Barbara Wohlmuth, Mathematics and Physics
 Email: wohlmuth@ians.uni-stuttgart.de

Prof. Dr. Thomas Ertl, Computer Science
 Email: Thomas.Ertl@vis.uni-stuttgart.de

Prof. Dr.-Ing. habil. Jadran Vrabec, Mechanical Engineering
 Email: vrabec@itt.uni-stuttgart.de;
 (Now at the University of Paderborn, Germany
 Email: jadran.vrabec@uni-paderborn.de)

Prof. Dr.-Ing. Hans Hasse, Mechanical Engineering
 Email: hasse@itt.uni-stuttgart.de
 (Now at the University of Kaiserslautern, Germany
 Email: hans.hasse@mv.uni-kl.de)

BACKGROUND

The University of Stuttgart (Universität Stuttgart) looks back at a 175-year history. It became a full university in 1967; its present rector is Prof. Dr.-Ing. W. Ressel. The university has 20,000 students (approximately 5,000 are international) distributed in 10 faculties (departments), and 5000 staff. It is a research university with a focus on engineering and the natural sciences. Relevant faculties (colleges or departments) in these areas are Civil and Environmental Engineering; Chemistry; Energy Technology, Process Engineering, and Biological Engineering; Computer Science, Electrical Engineering, and Information Technology; Aerospace Engineering and Geodesy; Engineering Design, Production Engineering, and Automotive Engineering; and Mathematics and Physics. Key research areas are modeling and simulation, complex systems, communications, materials, technology concepts and assessment, energy and the environment, mobility, construction and living, and integrated products and product design.

The university has an international range of study programs, including seven master's courses taught in English. In addition, it has a number of collaborative programs. It has an interfaculty (interdepartmental) research structure, with the different faculties linked through transfer and research centers interacting with national institutions (Max Planck Institutes, Fraunhofer Institutes, German Aerospace Center), international institutions, and industry. The university sits in one of Europe's strongest economic regions: Bosch, Fischer, Daimler, HP, IBM, Festo, Pilz Deutschland, Porsche, Trumpf, Stihl, Züblin, and BASF all have facilities in the area.

Stuttgart has a number of superlative technical and computing facilities and projects. These include the facilities of the High Performance Computing Center Stuttgart (HLRS) with a computing power of about 220 Tflops; *Gauss Center for Supercomputing* (Europe's most powerful high-performance computing alliance among the universities of Jülich, Munich, and Stuttgart); *VISUS* (Visualization Research Center, one of the leading facilities in Europe); *ASCS* (Automotive Simulation Centre Stuttgart, in close cooperation with industry since 2007); *Archi-Neering* (Bangkok's new airport is the result of close cooperation between Civil Engineering and

Architecture); *Baden-Württemberg Astronautics Center* (opening in 2010); *IZKT* (International Center for Cultural and Technological Studies); *SOFIA* (joint U.S.-German project, a Boeing 747SP, equipped with a high-performance mirror telescope); *wind tunnel* (tests of aerodynamic and aeroacoustic properties of vehicles, up to 265 km/hour); and *VEGAS* (Research Facility for Subsurface Remediation and for the simulation of contamination processes).

Rankings and Funding

The University of Stuttgart ranks number three in Germany in total amount of external funding (Aachen is first), and number one in Germany in external funding per professor (average of €400,000 per professor). In 2006 it was one of the top grant university recipients in grant ranking by DFG (Deutsche Forschungsgemeinschaft, the German Research Foundation); in the top three for a range of Engineering study programs (ranked by CHE, Spiegel, Focus 2007); number five in the CHE 2006 research ranking; most successful German University in the 6th EU Framework Program, particularly in the fields of simulation technology, energy, and e-health; and number two in Germany for research visits by international scholarship holders and award-winners of the Humboldt foundation, 2005.

GERMAN RESEARCH FOUNDATION (DFG) SUPPORT

DFG has provided support for collaborative research centers (SFB), transregional collaborative research centers (TRR), transfer units (TFB), research units (FOR), priority programs (SPP), and “Excellence Initiatives.” Many of these (listed below) are entirely based on or have major components in simulation-based engineering and science:

- *DFG Collaborative Research Centers (SFB) and Transregional Collaborative Research Centers (TRR)*: Spatial World Models for Mobile Context-aware Applications; Selective Catalytic Oxidation of C-H Bonds with Molecular Oxygen; Dynamic Simulation of Systems with Large Particle Numbers; Incremental Specification in Context; Control of Quantum Correlations in Tailored Matter (transregional project involving Stuttgart/Tübingen/Ulm). The budget of these centers is about €2–3 million/year for twelve years, in three 4-year terms.
- *DFG Transfer Units (TFB)*: Simulation and Active Control of Hydroacoustics in Flexible Piping Systems; Development of a Regenerative Reactor System for Autothermal Operation of Endothermic High-Temperature Syntheses; Transformability in Multivariant Serial Production; Rapid Prototyping; Computer Aided Modeling and Simulation for Analysis; Synthesis and Operation in Process Engineering; and Thermodynamic Properties for Process Engineering Applications. These units execute transfer from university to industry.
- *DFG Research Units (FOR)*: Nondestructive Evaluation of Concrete Structures Using Acoustic and Electromagnetic Echo Methods; Development of Concepts and Methods for the Determination of Reliability of Mechatronics Systems in Early Stages of Development; Noise Generation in Turbulent Flow; Multiscale Methods in Computational Mechanics; Specific Predictive Maintenance of Machine Tools by Automated Condition Monitoring; Positioning of Single Nanostructures—Single Quantum Devices. These units are smaller than the centers, generally consisting of 4–5 researchers.
- *DFG Priority Programs*: Molecular Modeling and Simulation in Process Engineering; Nanowires and Nanotubes: From Controlled Synthesis to Function. These programs bring €1.2–2 million/year (spread over 10–20 projects). Ten new priority programs are opened per year. The call involves preproposals, out of which 80 are invited for full proposals to compete for 10 awards. Reviewers are academics and convene in panels. Generally, the PI meets the reviewers.
- *Excellence Initiative*: The goal of the Excellence Initiative, planned for an initial period of 5 years, is to foster excellence in science and research and to raise the profile of top performers in the academic and research community by means of three lines of funding: strategies for the future; excellence clusters, and graduate schools. The University of Stuttgart has been successful in two of the Excellence Initiative’s three lines of funding. Funding has been granted to the Advanced Manufacturing Engineering Graduate School and the **Simulation Technology Excellence Cluster**.

SIMULATION TECHNOLOGY EXCELLENCE CLUSTER

The Simulation Technology Excellence Cluster is a University of Stuttgart DFG Excellence Initiative award. The SimTech Excellence Cluster is coordinated by Dr.-Ing. Wolfgang Ehlers, and the initiative's vision is to progress from isolated numerical approaches to integrative systems science. Interestingly, in order to obtain funding, the group gave arguments partly based on the U.S. NSF Blue Ribbon Panel Report of February 2006: "... [C]hallenges in SBES ... involve ... *multiscale and multiphysics modeling, real-time integration of simulation methods with measurement systems, model validation and verification, handling large data, and visualization*. ... [O]ne of those challenges is *education of the next generation of engineers and scientists* in the theory and practices of SBES" (Oden et al. 2006). DFG has agreed wholeheartedly with this report and provided funding accordingly. The SimTech Excellence cluster brings €7 million/year for 5 years.

Long-Term Goals

The SimTech cluster ultimately aims at linking three types of interactions:

- the interactive handling of multiscale, multiphysics systems under consideration of uncertainty by simulation in order to gain a qualitative and quantitative understanding, to predict the system behavior, and to prepare decisions
- the interactive situation-adapted or context-adapted optimization of systems that are influenced by sensor data streams in real time
- the interaction between developers and systems for evaluating the consequences of system behavior for management, optimization, control, and automation

The long-term scientific vision is focused on advances that go

- from Empirical Material Description towards Computational Material Design
- towards Integrative Virtual Prototyping
- towards Interactive Environmental Engineering
- from Classical Biology to Systems Biology
- from Biomechanics towards the Overall Human Model

Research Areas

The only way to achieve this integrative, simulation-based vision is through a *long-term sustained research agenda* in the following research areas:

- A. Molecular and Particle Simulations, serving for nanoscale and microscale simulations, and bridging scales by coupling particle dynamics with continuum mechanical approaches
- B. Advanced Mechanics of Multiscale and Multifield Problems, exhibiting the basic scientific key for the description of complex problems in almost all branches of engineering simulation and design by bridging scales, coupling physics, and linking domains
- C. Systems Analysis and Inverse Problems, focusing on model validation, parameter identification and model reduction, as well as dynamical system analysis, control, aspects of autonomy, automation and hierarchical or networked structures of systems
- D. Numerical and Computational Mathematics, guaranteeing advances towards multiscale and multiphysics numerical models, including the quantification of uncertainty and a self-adaptive choice of scales and physics in order to simulate dynamic and coupled processes in complex real-world problems

- E. Integrated Data Management and Interactive Visualization, mastering the explosion of information in simulation technology through human-system interfaces for model setup, real-time simulation, control, and interactive visualization, and through design of sensor networks for real-time control simulations
- F. Hybrid High-Performance Computing Systems and Simulation Software Engineering, with the basic idea of harnessing the power of large-scale systems through advanced simulation software technology to solve grand challenge problems
- G. Integrative Platform of Reflection and Evaluation, which contributes to the above research areas regarding Theory of Science, Philosophy of Technology, Sociology of Technology, and Ethics

The linking of the above research areas will result in an integrative systems science that includes an interactive, computational approach to engineering and natural sciences. Molecular Simulations (A) and Advanced Mechanics (B) represent fundamental fields of engineering and science, where advances in Simulation Technology are most indispensable. Systems Analysis (C) and Computational Mathematics (D) constitute fundamentals of utmost importance that will take Simulation Technology to a higher level. Data Management and Visualization (E) as well as High Performance Computing (F) provide the necessary tools and framework. The Integrative Platform (G) acts as an overall bracket of reflection and evaluation, reflecting questions of social acceptance and supervising the research of A–G. Examples of complex modeling and simulation were given for the following:

- *Vaporisation of a droplet in an air stream, a fully 3D enhanced visualized phase transition problem (B, E)*
- *Salt water on fresh water in a porous medium, a multifield flow simulation obtained by parallel computing (B, F)*
- *Collision of methane and ethane droplets, a molecular dynamics nanoscale simulation with 3D point-based visualization (A, E)*
- *CO₂ sequestration in a geological formation, systems analysis based on a multiscale and multiphysics simulation (B–F)*
- *Flexion of a lumbar spine, a multifield and multiphysics biomechanical simulation (B–D)*

New Professorships and Research Positions

In order to support these efforts, new positions are being created. These include three new professorial positions:

- Mathematical Systems Theory
- Modeling of Uncertain Systems
- Human-System Interaction and Cognitive Systems

In addition, there are 13 new junior professorships and 7 new post-doctoral positions with up to 2 research associates each, tenure-track options for 4 of the junior professors, and a total of 72 scientific projects.

Structural Actions at the University and Long-Term Sustainability

Three *new* structural elements that provide long-term sustainability for the cluster have been founded at the university: *Research Centres*, *Transfer Centres*, and the *Stuttgart School of Science and Technology*. Compared to the traditional departments with their teaching-oriented "vertical" structure, the research centres are "horizontally" oriented, thus comprising researchers and their institutions from various departments under the common roof of a research goal.

Recognizing the enormous importance of simulation sciences, the university founded the **Stuttgart Research Centre of Simulation Technology (SRC SimTech)**. Opening on April 1, 2007, this research centre is the first one at the university and represents both a scientific research unit and a new structural element acting as a research department with its own organizational and administrative structure, including financial resources (€240,000/year) and personnel.

The **SimTech Transfer Unit** bundles all activities of the cluster that require uni- or bidirectional communication with external institutions and industrial enterprises, comprising Active Working Committees, a SimTech Industrial Consortium, and university-internal transfer to teaching and other departments. It will be embedded in the **Stuttgart Transfer Centre**, an overarching structure whose role is to transfer research results into application, bundle exchange activities with industrial partners, and provide a basis for all future fundraising activities of individual research centers.

To promote common research interests, the doctoral students supervised by their senior researchers and junior professors of the cluster are integrated in the cluster's **Graduate School of Simulation Technology**, which itself is part of the Stuttgart School of Science and Technology.

International Visibility

To attract scientists from all relevant fields of simulation sciences, the cluster will organize a series of international conferences on simulation technology at the University of Stuttgart. In addition, smaller thematic symposia, workshops, and meetings on specific topics of the cluster are planned.

A *guest program* is the most promising tool to obtain a fruitful exchange among researchers. It enables the participants of the cluster to invite international experts as visiting professors or visiting researchers to the university. They also plan to offer a Visiting Research Professorship of Simulation Technology for a period between half a year and a year.

Education and Promotion of Young Scientists

Highly qualified students and young scientists are being prepared both for research in industrial environments and for scientific careers. The cluster intends to establish new lines of elite education and promotion activities spanning the entire range from first studies to top-level scientific careers:

- Elite Study Programme in Simulation Technology
- SimTech Stipend and Qualification Fellowship
- Graduate School of Simulation Technology
- post-doctoral period
- independent young research groups
- a tenure-track program

At each of the different stages, a competitive selection procedure will allow both consecutive promotion and new admission of excellent candidates. The doctoral students of the graduate school essentially are the research associates funded by the cluster.

BSc/MSc Elite Study Programme

- ~20 ECTS lectures; ~10 ECTS research
- BSc and MSc theses in different research areas
- One term required abroad during MSc studies
- Flexible study regulation (more than 1 supervisor, and at least 1 from abroad)
- Teaching import/export within consistent 6 + 4 BSc/MSc framework
- E-learning, tele-teaching

Graduate School in Simulation Technology

- Stuttgart School of Science and Technology
- Concept similar to DFG/DAAD programs (ENWAT, NUPUS)

- Requirements: 9–12 ECTS, doctoral seminar
- Milestone presentation, international exchange program
- Optional: short course, summer schools, software skills program
- Joint internal/external and international supervision
- No more separation in Faculties (departments), new interdisciplinary flexibility

EXAMPLE OF DFG COLLABORATIVE RESEARCH CENTER: DYNAMICAL SIMULATIONS OF SYSTEMS WITH LARGE NUMBER OF PARTICLES

Directors: Rainer Trebin, Thomas Ertl

This is an interdisciplinary collaboration between Chemistry, Physics, Engineering, and Computer Science and the High Performance Computing Center, through 12 institutes. Applications include catalytic converters, enzymes, downstream processing, nanotechnology, materials, transport technology, etc. The methods include combining particle simulations with continuum simulations (multiscale simulations), and “simulations across scales,” where results on finer levels are transferred to coarser levels successively.

Briefly, the structure of this particular Collaborative Research Center is as follows.

Project Area A: Fluid Mechanics and Thermodynamics (PIs: Vrabec, Hasse, Harting)

- Molecular dynamics simulations of multiphase flow of real fluids in nanoscale channels
- Hybrid approaches for the simulation of microfluids
- Molecular simulations of hydrogels

Project Area B: Materials / Mechanics (PIs: Trebin, Roth, Schmauder, Seifried, Hilfer, Eberhard)

- Molecular dynamics of large systems with long-range forces
- Molecular dynamics simulations of fracture at metal/ceramics interfaces
- Breakdown and cracks in granular packings
- Granular processes and particle breakage
- Molecular dynamics simulations of laser ablation in metals

Project Area C: Biophysics / Biochemistry (PIs: Pleiss, Wachtrup)

- Influence of organic solvents on activity and stereo selectivity of Lipase
- Molecular modeling of inhibitor resistance β -Lactamase
- Translocation of proteins

Project Area D: Algorithms / Implementation (PIs: Bernreuther, Resch, Ertl)

- Software framework for scalable simulations of flow in nanoscale channels
- Performant and flexible particle simulation on different HPC architectures
- Visualization of systems with large number of particles
- Visualization of protein-solvent systems

EXAMPLE OF DFG PRIORITY PROGRAM: MOLECULAR MODELING AND SIMULATION IN PROCESS ENGINEERING (SPP 1155)

Director: Hans Hasse and Frerich Keil

- Priority Programs provide a collaborative framework for research in a defined area.

- Priority Programs coordinate knowledge and resources in their field to produce added scientific value.
- Within a Priority Program researchers conduct project work, they choose their topics, research plan, and methods.
- Evaluation of project proposals by a peer reviewer group. Interestingly, the PI proposes the priority area; if the DFG agency approves, a general call for a preproposal goes out. The PI meets the reviewer panel.
- Typical size: 20 projects, €1.2–2 million/year for 6 years.

The motivation for this DFG priority program was the following:

Modeling and Simulation in Process Engineering

- Key to progress
- Presently characterized by phenomenological methods
- Potential of phenomenological methods is largely exploited
- Potential of molecular simulations remains unexploited

Future Development of Process Engineering

- Driven by molecular simulation methods
- Need to foster molecular simulation methods in research and education
- The efforts in the United States motivate a focused effort in Germany

Main Features of the Program

- This is the first transnational DFG priority program
- It consists of 23 projects (4 with international participation); 11 programs belong to Engineering, 10 programs belong to Chemistry, and Physics and Computer Sciences contribute 1 program each
- Working areas
 - Predictive Fluid Property Models: (i) Development of Molecular Models of Real Fluid Process; (ii) Engineering Applications
 - Processes at Fluid-Solid Interfaces
 - Processes in Porous Media
- As an example of the interface between academia and industry, a workshop bringing together several academic (DFG, this priority program, and the Society for Chemical Engineering and Biotechnology) and industrial partners took place in March 2008 (*International Workshop on Molecular Modeling and Simulation in Applied Material Science*).

CONCLUSIONS

It is interesting to note that the group of people who came together under the different umbrellas of DFG support studied carefully two reports prepared for the government of the United States. One of them was the U.S. NSF Blue Ribbon Panel Report in February 2006, as quoted above. The other report was the July 2005 PITAC report to the President of the United States of America, *Computational Science: Ensuring America's Competitiveness*, which states that, "The multidisciplinary teams required to address computational science challenges represent what will be the most important mode of the 21st century science and engineering research and development."

Certainly neither the University of Stuttgart nor the German government is unaware of these computational challenges. On the contrary, they both took decisive action for meaningful changes. Thus, new academic structures were created such as the DFG-supported Collaborative Research Centers, Transfer Units, Priority Programs (e.g., *Molecular Modeling and Simulation in Process Engineering*), Excellence Initiatives (e.g., *Simulation Technology Excellence Cluster*), and new graduate programs (new educational model from BSc to PhD, Graduate School of Simulation Technology). All these programs are strongly supported economically, with

a few million Euros per year for several years (5–12 years). Both the academics and the government have understood that a **long-term sustained research agenda** is the only way this integrative, simulation-based vision can be realized. These initiatives have clearly strengthened the position of the University of Stuttgart as a major global player in simulation technology.

REFERENCES

Oden, J.T., T. Belytschko, T.J.R. Hughes, C. Johnson, D. Keyes, A. Laub, L. Petzold, D. Srolovitz, and S. Yip. 2006. *Revolutionizing engineering science through simulation: A report of the National Science Foundation blue ribbon panel on simulation-based engineering science*. Arlington, VA: National Science Foundation.

Selected Publications of the Stuttgart Research Center of Simulation Technology

1. Eberhard, P., and W. Schiehlen. 2006. Computational dynamics in multibody systems. *ASME Journal of Computational and Nonlinear Dynamics* 1:3–13.
2. Bischoff, M., W.A. Wall, K.-U. Bletzinger, and E. Ramm. 2004. Models and finite elements for thin-walled structures. In *Encyclopedia of computational mechanics*, T.J.R. Hughes et al., eds., 59–137. Chichester, West Sussex: John Wiley.
3. Kays, W.M., M.E. Crawford, and B. Weigand. 2004. *Convective heat and mass transfer*, 4th ed. New York: McGraw Hill.
4. Werner, H.-J., F.R. Manby, and P.J. Knowles. 2003. Fast linear scaling second-order Moller-Plesset perturbation theory (MP2). *Journal of Chemical Physics* 118:8149–8160.
5. Ehlers, W. 2002. Foundations of multiphase and porous materials. In *Porous media*, W. Ehlers and J. Bluhm, eds., 3–86. Berlin: Springer-Verlag.
6. Miehe, C., J. Schotte, and M. Lambrecht. 2002. Homogenization of inelastic solid materials at finite strains based on incremental minimization principles. *Journal of the Mechanics and Physics of Solids* 50:2123–2167.
7. Engel, K., M. Kraus and T. Ertl. 2001. High-quality pre-integrated volume rendering using hardware-accelerated pixel shading. *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Workshop on Graphics Hardware* 9–16.
8. Leymann, F. 2001. Web Service Flow Language (WSFL). IBM technical report. Available online: <http://www.ibm.com/software/solutions/webservices/pdf/WSFL.pdf>.
9. Schitter, G., P. Menold, H. Knapp, F. Allgöwer, and A. Stemmer. 2001. High performance feedback for fast scanning atomic force microscopes. *Review of Scientific Instruments* 72:3320–3327.
10. Vrabec, J., J. Stoll, and H. Hasse. 2001. A set of molecular models for symmetric quadrupolar fluids. *Journal of Physical Chemistry B* 105:12126–12133.
11. LeFloch, P.G., and C. Rohde. 2000. High-order schemes, entropy inequalities, and nonclassical shocks. *SIAM Journal on Numerical Analysis* 37:2023–2060.
12. Wohlmuth, B. 2000. A mortar finite element method using dual spaces for the Lagrange multiplier. *SIAM Journal on Numerical Analysis* 38:989–1012.
13. Gerstendörfer, S., and H.-J. Wunderlich. 2000. Minimized power consumption for scan-based BIST. *Journal of Electronic Testing, Theory and Applications* 3:203–212.
14. Westermann, R., and T. Ertl. 1998. Efficiently using graphics hardware in volume rendering applications. *Computer Graphics (SIGGRAPH '98)* 32:169–179.
15. Leymann, F., and D. Roller. 1999. *Production workflow: Concepts and techniques*. Upper Saddle River, N.J.: Prentice Hall PTR.
16. Gabriel, E., M. Resch, T. Beisel, and R. Keller. 1998. Distributed computing in a heterogeneous environment. In *Lecture notes in computer science 1497*, V. Alexandrov and J. Dongarra, eds., 180–187. Berlin: Springer-Verlag.
17. Baumann, J., F. Hohl, K. Rothermel, and M. Straßer. 1998. MOLE – Concepts of a mobile agent system. *World Wide Web* 1:123–137.
18. Ehlers, W., and W. Volk. 1998. On theoretical and numerical methods in the theory of porous media based on polar and non-polar elasto-plastic solid materials. *International Journal of Solids and Structures* 35:4597–4617.

19. Chen, H., and F. Allgöwer. 1998. A quasi-finite horizon nonlinear model predictive control scheme with guaranteed stability. *Automatica* 34:1205–1218.
20. Bastian, P., K. Birken, K. Johannsen, S. Lang, N. Neuß, H. Rentz-Reichert, and C. Wieners. 1997. *UG – A flexible software toolbox for solving partial differential equations*. Berlin: Springer-Verlag.
21. Helmig, R. 1997. *Multiphase flow and transport processes in the subsurface: A contribution to the modeling of hydrosystems*. Berlin: Springer-Verlag (also translated into Japanese, Springer, Tokyo 2004).
22. Stadler, J., R. Mikulla, and H.-R. Trebin. 1997. IMD: A software package for molecular dynamics studies on parallel computers. *International Journal of Modern Physics C* 8:1131–1140.
23. Mielke, A., and G. Schneider. 1995. Attractors for modulation equations on unbounded domains: Existence and comparison. *Nonlinearity* 8:743–768.
24. Simo, J.C., and C. Miehe. 1992. Associated coupled thermoplasticity at finite strains: Formulation, numerical analysis and implementation. *Computer Methods in Applied Mechanics and Engineering* 98:41–04.
25. Werner, H.-J., and P.J. Knowles. 1988. An efficient internally connected multiconfiguration reference configuration-interaction method. *Journal of Chemical Physics* 89:5803–5814.
26. Horsch, M., J. Vrabec, M. Bernreuther, S. Grottel, G. Reina, A. Wix, K. Schaber, and H. Hasse. 2008. Homogeneous nucleation in supersaturated vapors of methane, ethane, and carbon dioxide predicted by brute force molecular dynamics. *The Journal of Chemical Physics* 128:164510.
27. Grottel, S., G. Reina, J. Vrabec, and T. Ertl. 2007. Visual verification and analysis of cluster detection for molecular dynamics. *Proceedings of IEEE Visualization '07* 1624–1631.
28. Vrabec, J., and J. Gross. 2007. A molecular based approach to dipolar and quadrupolar fluids: vapor-liquid equilibria simulation and an equation of state contribution for dipole-quadrupole interactions. *Journal of Physical Chemistry B* 112:51–60.
29. Eckl, B., Y.-L. Huang, J. Vrabec, and H. Hasse. 2007. Vapor pressure of R227ea + Ethanol at 343.17 K by molecular simulation. *Fluid Phase Equilibria* 260:177–182.

Site: **Technical University of Denmark (DTU)**
Department of Chemical and Biochemical Engineering
Søltofts Plads Building 229
DK-2800 Kongens Lyngby, Denmark
<http://www.kt.dtu.dk/English.aspx>

Date Visited: February 27, 2008

WTEC Attendees: A. Deshmukh (report author), P. Westmoreland

Hosts: Professor Rafiqul Gani, Director, CAPEC
Computer-Aided Process Engineering Center
Tel: +45 4525 2882; Fax: +45 4593 2906
Email: rag@kt.dtu.dk

Professor Erling Stenby, Director, IVC-SEP, Center for Phase Equilibria and Separation Processes (Former Chair of the Engineering Research Council for Denmark)
Tel: +45 4525 2875; Fax: +45 4588 2258
Email: ehs@kt.dtu.dk

Professor Jorgen Mollerup, IVC-SEP
Tel: +45 4525 2866; Fax: +45 4588 2258
Email: jm@kt.dtu.dk

BACKGROUND

The Chemical Engineering Department at TU Denmark is divided into six main research areas: CAPEC (process engineering), IVC-SEP (petroleum and separation), CHEC (combustion), BIO-ENG (food and bioreactors), Polymer Center (polymer engineering and kinetics), and Aerosol and Reaction Kinetics. Modeling and simulation activities are conducted by most groups, and the department is thinking of establishing a virtual center on product engineering. The WTEC panel members met with representatives of two of the groups from the Chemical Engineering Department: CAPEC and IVC-SEP.

Computer Aided Process-Product Engineering Center (CAPEC)

CAPEC (Computer Aided Process-Product Engineering Center; <http://www.capec.kt.dtu.dk/>) was established in 1997 with the objective of developing and using a systems approach to solve and analyze problems related to Chemical and Biochemical Product-Process Modeling, Simulation, Synthesis, Design, Analysis, and Control/Operation for the Chemical, Petrochemical, Pharmaceutical, Agrochemical, Food, and Biochemical Industries. CAPEC has pioneered model-assisted process-product development methods that use simulations to identify the process-product parameters (design targets), then use predictive models to match the targets, and finally, use experiments (or data available in databases) to verify the predictions. Center researchers have produced state-of-the-art models and databases that allow users to correlate and estimate pure component and mixture properties of importance in the chemical, petrochemical-pharmaceutical, and biochemical industries. The researchers have implemented results from the above models and databases to develop phenomena-based and/or parameter-based models for process, product, and operations. CAPEC has developed computer-aided systems for process integration, product-process synthesis and design, hybrid separation, waste reduction, pollution control, batch operations, advanced process control, and process analytical technology.

CAPEC is funded by annual fees from member companies. Each company pays €7000 and has access to CAPEC-developed software exclusively. Currently, 32 member companies from Europe, North and South America, and Asia are members of CAPEC. The software developed under this arrangement is owned by CAPEC. It includes ICAS: Integrated Computer Aided System; CAPEC Database: Database of physical properties, solvents, specialty chemicals, reaction; UNIFAC-Utility: Groups, parameters, interface; Library: Solvers, MoT-based property models (UNIFAC, SRK, SAFT, PC-SAFT, CPA, Elec-UNIQUAC, GC-Flory), MoT-based process models (short-path evaporation, vacuum membrane distillation, pervaporation); property prediction packages

(ProPred and TML using the group-contribution-plus approach); Computer-Aided Molecular Design (CAMD) software, ProCAMD; Design/visualization: Synthesis & design tools based on driving forces and phase diagrams, reverse method; CTSM: Continuous time stochastic modeling tool; GoLM: Grid of Linear Models toolbox for data driven modeling. The center sees future growth areas in pharmacology, agribusiness, food, and aroma sectors. Tools such as ProPred (used for pure component property prediction for organic chemicals and polymers); ProCAMD (used for solvent search as well as chemical product design); MoT (used for quick and easy development of models without the user writing any programming code); and ICAS-utility (providing a number of useful design related calculations) are widely used among the member companies and close to 50 users from academia with special educational licenses.

Professor Gani, the director of CAPEC, is also involved in leading a discipline-wide effort on software integration. Computer-Aided Process Engineering Open Standards, or CAPE-OPEN, is a chemical-process-engineering community-based effort focusing on developing standards to facilitate integration of software and simulation packages. CAPE-OPEN is based on XML specifications, COM, and CORBA standards. This effort was initially funded by the EU. It has now evolved into Global CAPE-OPEN with academics, end user companies, and vendors from all across the world (U.S. researchers had to withdraw out of this effort due to lack of funding). Efforts to help users implement the CAPE-OPEN standards has resulted in the formation of CO-Lan, an organization that maintains and updates the CAPE-OPEN standards, runs workshops, and provides expert advice. CAPEC has linked its thermo and process models to external simulators through a special CAPE-OPEN compliant interface developed by ProSim (France). Even though applications highlighting interoperability between software from academia and commercial simulators are being developed continuously, much more effort is necessary to get the CAPE-OPEN standards established.

Center for Phase Equilibria and Separation Processes (IVC-SEP)

The IVC-SEP group (Center for Phase Equilibria and Separation Processes; <http://www.ivc-sep.kt.dtu.dk/>) focuses on developing state-of-the-art thermodynamic models and algorithms for implementation in simulation tools. These are applied in the chemical, bio-chemical, and petroleum industries. The faculty present at the meeting, Professor Stenby and Professor Mollerup, had research interests in phase behavior in petroleum and fine chemicals industries, protein purification, and modeling of chromatographic processes. This group is strong in generic algorithms for phase equilibrium calculations. They develop thermodynamic models and make them available for use in commercial process simulators. Furthermore they provide efficient routines for the so-called next generation reservoir simulators for oil and gas production. They have significant activities connecting extraction efficiency and sustainability. The group consists of 7 faculty members, 15 PhD students, 6 senior researchers, and 8 technical assistants. In particular, Professor Michael L. Michelsen is a world leading expert concerning efficiency in chemical engineering computations. This is true for phase equilibrium calculations, simultaneous chemical and phase equilibria, and implementation of complex thermodynamic models. The latter has been used to make thermodynamic models such as CPA and PC-SAFT available as robust and fast options for simulations.

Both models are available in the IVC-SEP CAPE-OPEN ThermoSystem, which has been successfully tested with HYSYS, PROII, and ASPEN Plus. IVC-SEP is a very active member of the Co-LAN. The IVC-SEP thermodynamic software package, SPECS, includes the above-mentioned models as well as Extended UNIQUAC for electrolyte systems, developed by Associate Professor Kaj Thomsen from the IVC-SEP, and a range of models for polymer solutions systems. SPECS also has several models implemented for characterization of petroleum reservoir fluids and simulation of relevant processes such as gas injection. Since 1982 the model and software development in IVC-SEP has led to 3 spin-off companies (Calsep, 1982; Tie-Line Technology, 2001; VLXE, 2003). The IVC-SEP center has collaborative research programs with the chemical and biochemical process industries, and petroleum technology and refining industries. The membership fee for industries to participate in the center's programs is €8000 per year. Although there are some related programs in geophysics and geology at other institutions, this is the only petroleum engineering research program in Denmark.

ISSUES RELATED TO MODELING AND SIMULATION

- *Computational Resources.* The research groups mainly use workstations for model building and computations. They have access to supercomputers but do not use them for current studies. The CAPEC group is starting activities on molecular simulations, which might need more significant HPC resources.
- *Algorithms.* The CAPEC director noted that they needed to implement their algorithms on the cutting-edge computational architectures to get the maximum payoff. The matching of algorithms and next-generation computational architecture was the key in scaling up, not necessarily more computational horsepower.
- *Uncertainty Quantification.* The consensus was that there were significant errors in the models developed for chemical and biochemical processes. It was especially important in the bio area, where model experimentation was necessary in order to develop models that would be suitable for model-assisted product development. Methods needed to be developed that would allow adaptive experimentation based on simulation models that are in turn refined based on partial experimental results.
- *Academic Environment.* The Denmark Technical University (DTU) faculty highlighted the difficulty of attracting young researchers to academia due to the structure of the Danish academic system, where a new researcher is initially hired for three years as an assistant professor, after which an associate professor position may be created in that area by the university. However this position is open for everyone to apply. This creates significant uncertainty for young faculty in the system. The associate professor position is equivalent to a tenured position in U.S. academic world. The full professor positions are limited; the total number is fixed by the Education Ministry and DTU. Most faculty members participate in research centers on a voluntary basis.
- *Education and Training.* The Chemical Engineering faculty noted that their PhD students are getting training in at least two of the following three areas: experiments, computations, and theory. The students at DTU are learning computational skills in the Informatics Department. The WTEC team's hosts pointed out that the industry wants engineers well-trained in basics, not necessarily specialized in specific areas. They noted that the Danish students are very flexible and are being hired by the member companies in the consortia. Most students sponsored by member companies in centers go to work in the sponsoring company after graduation. The Chemical Engineering Department at DTU has collaboration with a French University (CS students), who come for 4 months to work on domain problems.
- *Funding Issues.* Both CAPEC and IVC-SEP have strong collaborations with industrial partners. 40% of their total funding comes from member companies. Companies do not get anything specific for their membership in centers. Member companies can sponsor specific projects, PhD students, and post-docs. Typically the total cost of a research project for industry is the sum of the project cost plus 20% overhead if DTU owns the intellectual property (IP). The overhead is charged at 180% if the industrial partner intends to keep the ownership of any IP resulting from the research collaboration. DTU does not accept liability for any of the products. DTU has liability insurance for projects.

CONCLUSIONS

The Chemical Engineering Department at TU Denmark is a premier research and education center in Denmark. Both CAPEC and IVC-SEP have a long history of collaboration with industry. The base funding for centers comes through membership fees that are used to develop tools, software, and databases of interest to the funding industries. Faculty members in this department are also leading international efforts in developing open standards for chemical engineering software and tools.

Site: **Technical University of Denmark (DTU) Wind Engineering**
Department of Mechanical Engineering (MEK)
Nils Koppels Alle, Building 403
DK-2800 Lyngby, Denmark
<http://www.mek.dtu.dk>
<http://www.risoe.dk/>

Date Visited February 27, 2008

WTEC Attendees: G. Karniadakis (report author), C. Sagui

Host: Jens N. Sorensen, Professor of Fluid Mechanics
Tel: +45 4525 4314; Fax: +45 4593 0663
E-mail: jns@mek.dtu.dk
Jens H. Walther, Associate Professor of Fluid Mechanics
E-mail: jhw@mek.dtu.dk

BACKGROUND – WIND ENERGY

The number of kW of wind power generated in Denmark per 1000 inhabitants exceeds 570 and is by far the highest in Europe and indeed the world. Spain and Germany follow with 340 and 270 (kW/1000 inhabitants), respectively. By the end of January 2007, there were 5,267 turbines installed in Denmark, with a total power of 3,135 MW; the total wind power in European Union countries during that period was 56,535 MW with Germany leading at 22,247 MW. The total installed power and the numbers of turbines in Denmark had a continually increase until 2002. From 2001 to 2003, a replacement agreement was carried out, where smaller and badly placed turbines were replaced with bigger turbines. In 2004 a new replacement agreement running to 2009 was introduced. The agreement involves replacing old 175 MW turbines with 350 MW turbines. Besides the replacement agreement, two offshore wind farms, each for 200 MW, have been approved (Horns Rev II and Rødsand II). These farms are expected to be connected to the grid in 2009–2010.

Denmark's electricity production from wind in 2006 was 6.108 GWh, which corresponded to 16.8 percent of the electricity consumption in Denmark or to the consumption in about 1.73 million Danish households. In the first half of 2007, the turbines produced 3.934 GWh, which corresponds to 21.7 percent of the electricity consumption in Denmark. When offshore turbines at Horns Rev II and Rødsand II are installed, it is expected that wind power will account for 25 percent of Denmark's total electricity consumption.

Danish manufactures of wind turbines have in recent years had a total global market share of 40 percent. The global market for wind power has grown dramatically, and such development is expected to continue. In 2006 Denmark installed about 15,000 MW new capacity, and by the end of 2006 its total installment reached 74,300 MW. The yearly installment of new capacity is expected to increase about 17% to 2011.

Systematic research and education in wind turbines in Denmark started more than 30 years ago, and simulation-based design of wind turbines started about 15 years ago. Totally, it is expected that there will be about €150 million per year devoted to research in energy and environmental issues in Denmark (population about 5.5 million).

BACKGROUND – DTU

Many advances in simulation-based design of wind turbines are associated with researchers at DTU and at Risø DTU National Laboratory for Sustainable Energy, which became part of the Technical University of Denmark (DTU) in January 2007. Examples are the development of popular aeroelastic code FLEX and the design of parts of the Nible and Tjaereborg turbines. Researchers at DTU participated in the Danish National Program on Wind Energy (1977–1990). There are approximately 20 researchers at DTU but there are up to 150 such researchers if combined with those of Risø.

The annual wind engineering budget at DTU is DKK 20 million (Danish Krone); the sponsors of the program are DTU, EFP (Ministry of Energy), UVE, PSO (Electric Utility), and the EU. Researchers at DTU and Risø work closely with researchers from the University of Aalborg (AaU) and the Danish Hydraulic Institute (DHI) supported by the Danish Research Consortium on Wind Energy. DTU researchers work closely with private companies on wind turbines, and VESTAS (one of the leading companies) has established scholarships for students and five-year full professorships in Wind Energy. Also, Siemens has an office in DTU, and an Indian company SUZLON (fifth-largest wind energy company worldwide) opened an office recently in Copenhagen.

RESEARCH ACTIVITIES AND SIMULATION TOOLS

Wind energy involves multiple disciplines, e.g., fluid mechanics, aeroelasticity, and electrical engineering. The work at DTU is focused on the following areas: design of optimum airfoils; dynamic stall, especially 3-D stall; tip flows and yaw; heavily loaded motors; and interference (wake and park) effects. The following are ongoing fully funded projects currently at DTU:

- A joint program with Risø for research in aeroelasticity, funded by EFP
- Research on noise generation and its suppression for wind turbines, funded by the Research Council
- Power and Noise optimization of Wind Turbines, funded by EFP
- Development of airfoil prediction codes
- Database on wind power, funded by IEA
- Simulation of “shadow” effects in wind farms, funded by PSO
- Simulation of wakes behind wind turbines, funded by EFP
- Simulation of wind turbines in complex terrains and analysis of vortex generators, funded by the Danish Research Council
- Rotor blade with flaps on reducing oscillating loads, jointly with Risø
- Experiments on rotor in wind tunnel

In many of these simulation-based research programs, verification and validation is performed using standard benchmarks for flow but also with experimental data collected at the Department’s scaled-down wind tunnel facilities but also in collaboration with a research institute that is near-by and has full-scale experimental facilities.

There are many important simulation tools developed at DTU in collaboration with Risø researchers. In particular, the FLEX code has undergone developments over many generations (currently FLEX5) and is used for designing wind turbines and analyzing loadings. Also, EllipSys is a CFD code for wind turbine aerodynamics simulations in both 2D and 3D. Another program is WPPT/WASP for predicting wind energy; unlike FLEX5, WASP uses the more advanced code EllipSys. Finally, DTU researchers are the keepers of IEA—an international database on wind measurements to determine loadings (including USA).

EDUCATION

DTU offers a unique two-year MSc program open to international students that is taught in English. There is a total of 20 students per year with about 3-4 students per year from the United States. The program entails one semester of required courses (wind turbine aerodynamics, wind aeroelasticity), two semesters of elective courses (aerodynamics and fluid mechanics, structural mechanics, construction and materials, power electronics and grid connection prediction, and optimization), and a last semester for a final project in collaboration with wind energy companies or related institutions. This unique program was featured in the 2005 special issue of *Wind Engineering* where several MSc theses were published as papers; the WTEC team’s host, Prof Sorensen, was the guest editor of the special issue.

Another wind engineering-specific education activity by DTU is the PhD program supported by the Danish Academy on Wind Energy (DAWE) with participating institutions (DTU, AaU, and DHI). About 40–50 PhD students are enrolled in the program. Also supported by the program are summer courses, seminars, and guest researchers.

CONCLUSIONS

Research on wind energy in Denmark has been taking place for over 30 years, and DTU researchers in collaboration with those of its partner institution Risø have been the leaders in this field. The research approaches employed at DTU are both of the fundamental type (e.g., large-eddy simulations and sophisticated grid techniques for moving meshes) but also of practical use, e.g., the popular aeroelastic code FLEX that uses lumped modeling for wind energy prediction. DTU also maintains a unique data bank of international meteorological and topographical data. In addition, the education activities at DTU are unique and impressive. It is one of the very few places in the world to offer both MSc and PhD degrees on wind engineering, and its international MSc program attracts students from around the world, including the United States, where there are no such programs currently.

Site: Technical University of Denmark Center for Biological Sequence Analysis
Systems Biology Department (BioCentrum-DTU)
Kemitorvet, Building 208
DK-2800 Lyngby, Denmark
<http://www.cbs.dtu.dk/>

Date Visited: February 27, 2008

WTEC Attendees: C. Sagui (report author), G. Karniadakis, A. Deshmukh, G. Lewison, P. Westmoreland

Hosts: Prof.-Dr. Søren Brunak, Director of Center for Biological Sequence Analysis
Tel: +45-45 25 24 77; Fax: +45-45 93 15 85
Email: brunak@cbs.dtu.dk
David Ussery, Microbial Genomics, Visualization
Email: dave@cbs.dtu.dk
Thomas S. Jensen, Integrative Systems Biology
Email: skot@cbs.dtu.dk
Petek F. Hallin, Microbial Genomics
Email: pfh@cbs.dtu.dk

CENTER HIGHLIGHTS

The Center for Biological Sequence Analysis (CBS) at the Technical University of Denmark (DTU) was formed in 1993 by coalescing in one center diverse bioinformatics activities dating back to the mid-1980s. It conducts basic research in the field of bioinformatics and system biology. It employs 100 people (80% research personnel and 20% staff) with 2:1 bio to non-bio backgrounds, and with a highly multidisciplinary profile (molecular biologists, biochemists, medical doctors, physicists, and computer scientists).

CBS is one of the largest bioinformatics centers in the European Union.²¹ It has a strong teaching component, with many courses; some are transmitted in real time over the Internet. CBS showcases a highly popular suite of WWW servers and codes, and it has a very strong publication and citation profile.

CBS is funded—in addition to a contribution from the Technical University of Denmark—by the Danish Research Foundation, the Danish Center for Scientific Computing, the Villum Kann Rasmussen Foundation, the Novo Nordisk Foundation, other institutions in the European Union, in industry, and the U.S. National Institutes of Health.

CBS RESEARCH

Information technology has become crucial for research in molecular biology, biotechnology, and pharmacology. Comprehensive public databases of DNA and protein sequences, macromolecular structure, gene and protein expression levels, pathway organization, and cell signaling have been created to facilitate the scientific manipulation of ever-increasing data within biology. Unlike many other groups in the field of biomolecular informatics, the Center for Biological Sequence Analysis directs its research *primarily towards topics related to the elucidation of the functional aspects of complex biological mechanisms.*

²¹ Other important centers are the European Bioinformatics Center near Cambridge, Bioinformatics in the Max Planck Institute in Munich, the European Molecular Biology Laboratory in Heidelberg, and the Centre for Genomic Regulation in Barcelona.

CBS currently has 10 research groups:

1. Integrative systems biology (Søren Brunak)
2. Systems biology of gene expression (Zoltan Szallasi)
3. Regulatory genomics (Chris Workman)
4. Protein post-translational modification (Henrik Nielsen)
5. Nutritional immunology and nutrigenomics (Hanne Frøkier)
6. Immunological bioinformatics (Ole Lund)
7. Comparative microbial genomics (David Ussery)
8. Metagenomics (Thomas Sicheritz-Ponten)
9. Molecular evolution (Anders Gorm Pedersen)
10. Chemoinformatics (Svava Jonsdottir)

The CBS publications appear in high-profile journals (*Science*, *Nature*, *Molecular Cell*, etc.) and command high citation records; even fairly recent papers easily surpass the 1,000 citations. Some of these papers appear in the top of the “Hot Biology” ISI list.

In the last decade, the Center for Biological Sequence Analysis has produced a large number of computational methods, which are offered to others via WWW servers.

- For NUCLEOTIDE SEQUENCES, these servers provide
 - Whole genome visualization and analysis
 - Gene finding and splice sites
 - Analysis of DNA microarray data
 - pH-dependent aqueous solubility of drug-like molecules
- For AMINO ACID sequences, these servers provide
 - Protein sorting
 - Post-translational modification of proteins
 - Immunological features
 - Protein function and structure
- Other bioinformatics tools
 - Development of neural network and weight matrix prediction methods for protein sequences
 - Combining protein sequence-based information with structural data from the Protein Data Bank
 - Multiple alignment of coding DNA using protein level information
 - Visualizing structural sequence constraints, etc.

These codes are highly popular, as evidenced by the number of page-views to the CBS WWW pages (over 2 million a month). The popularity of these servers is due to the fact that the computational models have higher fidelity (accuracy) than most of the experiments. For instance in protein sorting, the experimental precision is lower because things depend on where the protein is performing the function, how the cell directs the protein, and so forth.

As an example, the SignalP server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. The method incorporates a prediction of cleavage sites and a signal peptide/nonsignal peptide

prediction based on a combination of several artificial neural networks and hidden Markov models. The original paper (Nielsen et al. 1997) has over 3,300 citations to date. In general, the WWW sites provide only predictions, no data assessment of the algorithm prediction.

Most of the work is carried out on a variety of shared-memory SGI machines. One of the major challenges in the field is the integration of data. The amount of data generated by biology is just exploding. CBS has a relational data warehouse comprising 350+ different databases. In order to handle **data integration of 120 terabyte size**, CBS has developed its own integration tool (MySQL?).

At present there are 15 EU-funded projects dealing with European infrastructure for bioinformatics based on Web services. Parts of the infrastructure needs of Systems Biology are addressed by the European Commission and United States (EC-US) Task Force on Biotechnology Research (<http://ec.europa.eu/research/biotechnology/ec-us/>). The EC-US Task Force carries out numerous activities designed to bring European and U.S. researchers closer together. Its activities include sponsoring scientific workshops, short training courses, and short-term fellowships. In particular, WTEC sponsored the US-EC Workshop on Infrastructure Needs of Systems Biology (May 2007, Tufts University, Boston) under the auspices of the task force, which brought together 24 scientists from the European Union, the United States, and Canada. In this workshop only one session was dedicated to experiments, while the other three sessions were dedicated to databases, modeling applications, and software infrastructure (http://www.wtec.org/ec-us_sysbio_workshop).

An example of “messy” databank is given by the protein-protein interactions (PPI) databanks, which at present include 15 databases that are rather disorganized and chaotic, since there is no agreement on how to represent these interactions, on which format to choose, etc. CBS scientists view this challenge as a scientific window of opportunity, since the databases will simply be “exploding” with information. Data integration requires the analysis of data across different experimental platforms. The results can be used as proxy for experiments that are either impossible to carry out or that will only be feasible in the far future.

CBS’s goals go beyond one-gene biology. CBS aims at discovering novel specific functional aspects by exploiting systems-level principles that apply to one or more organisms, and also by finding new components. Ultimately, they aim to develop **data integration systems that can talk to one another**. This is clearly expressed in the *Nature* editorial “Let data speak to data” (2005): “Various sorts of data are increasingly being stored in formats that computers can understand and manipulate, allowing databases to talk to one another. This enables their users quickly to adapt technologies to extract and interpret data from different sources, and to create entirely new data products and services.... In biodiversity research, for example, rather than creating centralized monolithic databases, scientists could tap into existing databases wherever the data are held, weaving together all the relevant data on a species, from its taxonomy and genetic sequence to its geographical distribution. Such decentralization also helps to solve the problem that databases are often the fruits of individual or lab research projects that are vulnerable to the vagaries of funding, and to people and labs moving on to pastures new.”

As an example of the extremely complex and messy systems that CBS studies, the group studied the temporal aspects of biological networks (as opposed to static topological properties). Their integrative approach combined protein-protein interactions with information on the timing of the transcription of specific genes during the yeast cell cycle, obtained from DNA microarray time series. The resulting time-dependent interaction network places both periodically and constitutively expressed proteins in a temporal cell cycle context, thereby revealing previously unknown components and correlations. The dynamic proteins are generally expressed just before they are needed to carry out their function, generally referred to as **just-in-time synthesis**. However, they found that most complexes consist of both periodically and constitutively expressed subunits, which suggests that the former control complex activity by a mechanism of **just-in-time assembly or activation**. Transcriptional regulation influences almost all cell cycle complexes and thereby, indirectly, their static subunits. As a consequence, many cell cycle proteins cannot be identified through the analysis of any single type of experimental data but only through **integrative analysis** of several data types (de Lichtenberg et al. 2005).

Another interesting result the group found is that despite the fact the protein complexes involved in the transcription process are largely the same among all eukaryotes, their regulation has evolved considerably, and

thus the identity of the periodically expressed proteins differs significantly between organisms. In addition, these changes in transcriptional regulation have co-evolved with post-translational control independently in several lineages; loss or gain of cell-cycle-regulated transcription of specific genes is often mirrored by changes in phosphorylation of the proteins that they encode (Jensen et al. 2006). The assembly of the same molecular machines at the right time during the cell cycle has therefore evolved very differently, which can have dire consequences when, for instance, drugs designed for human consumption are tested in different organisms. Ultimately, the goal of these and future studies is to achieve cell life predictions and understanding.

At present the CBS group is moving towards other frontiers, the “disease interactomes.” This involves the use of text mining to relate which proteins relate to which diseases, or the identification of new disease gene candidates by the phenomic ranking of protein complexes. Briefly, the CBS scientists carried out “...a systematic, large-scale analysis of human protein complexes comprising gene products implicated in many different categories of human disease to create a phenome-interactome network. This was done by integrating quality-controlled interactions of human proteins with a validated, computationally derived phenotype similarity score, permitting identification of previously unknown complexes likely to be associated with disease. Using a phenomic ranking of protein complexes linked to human disease, we developed a Bayesian predictor that in 298 of 669 linkage intervals correctly ranks the known disease-causing protein as the top candidate, and in 870 intervals with no identified disease-causing gene, provides novel candidates implicated in disorders such as retinitis pigmentosa, epithelial ovarian cancer, inflammatory bowel disease, amyotrophic lateral sclerosis, Alzheimer disease, type 2 diabetes and coronary heart disease” (Lage et al. 2007). In this work, they define a “word vector” with the words describing the disease and then look at the relative angles between the vectors in “medical term space” to quantify the phenotypical overlap. The draft of 506 protein complexes associated with pathology is publicly available.

This new research leads to new types of “biobanks” with new computational challenges:

- Finding disease genes, and their “systemic” properties, in cases where the environment also plays a major role
- Extracting information from complex, messy “databases” and registries across countries

Nordic countries with long traditions for keeping medical records and biological samples (detailed from the 1950s, less detailed from the 1850s) provide ideal databases for this task. The combination of medical informatics with bioinformatics and systems biology represents a new trend in disease gene finding and phenotype association, which can also include social and behavioral levels. This also requires linking two 30–40 year old research traditions in systems biology: the data-poor (physics initiated) approach of the Kitano type, and the data-driven, model-fitted approach. Linking medical informatics with bioinformatics and systems biology requires bridging the gap between the molecular level and the phenotypic clinical levels, and linking two exponentially growing types of computer-accessible data: biomolecular databases and their clinical counterparts. The challenge ahead lies not only in linking the clinical level to the “parts list” of the human body (individual genes and proteins) but in understanding defects in biological mechanisms and disease aetiology in a network biology setting. This necessitates the development of a systems biology that includes a changing environment.

Finally, a lot of the information on the human genome input-output relation is and will be embedded in biobanks, and complex, messy databases, patient records, and registries in different countries. Text and other types of nontraditional “data types” contain information that can be used to reveal molecular disease mechanisms, finding disease genes and their systemic properties, in particular when a changing environment plays an important role. To achieve this, links to electronic medical records and real-time monitoring of living organisms is needed. This enterprise has over \$100 million in Danish funding, from Novo Nordisk and the Rasmussen Foundations. Interestingly, new biotechnology companies are bound to appear, like the recent “23andMe – Genetics Just Got Personal,” a privately held biotechnology company based in California that is developing new ways to help people make sense of their own genetic information. The company offers \$1000 tests for select, single-nucleotide polymorphisms.

Another big challenge is the **visualization** of large amounts of data (for instance, a run of a genome sequence generates 7 terabytes of data). CBS has an important effort in this area tackling different biological issues

(bacterial genomes, rRNAs, tRNAs, codon usage bias, mRNA, BLAST Atlases for proteome comparisons, etc). CBS has come up with the “DNA Structural Atlas,” which is a method of visualizing structural features within large regions of DNA. It was originally designed for analysis of complete genomes, but it can also be used quite readily for analysis of regions of DNA as small as a few thousand base-pairs in length. The basic idea is to plot the values for six different mechanical-structural properties of the DNA helix in a circle (or arc) representing the complete genome (or chromosome). At the level of whole genomes or chromosomes, large architecturally important regions can be seen. At the level of individual genes (e.g., usually around 10,000 to 20,000 bp for the entire plot), intergenic regions can be examined. The plots are created using the "GeneWiz" program, developed by Hans-Henrik Stærfeldt at CBS.

EDUCATION

- CBS offers an MSc degree in Bioinformatics and another MSc in Systems Biology that are not specifically linked to any department or school; the students can move around freely.
- PhDs are in Systems Biology and can combine courses and research done in CBS.
- A PhD costs about \$100 thousand a year; students get a salary of \$35 thousand.
- Through Macromedia Breeze CBS offers real-time Internet training
 - All lectures are transmitted in real time: 4 windows (teacher, PowerPoint, whiteboard, chat line)
 - All exercises are Web-compatible
 - All lectures are recorded for later view
 - All examinations use Breeze (examinations in cyberspace; Internet students act as presenters using microphones and webcam; poster parties)
 - Typically, half the students study onsite, half study via the Internet
 - The most popular 2-week PhD course is on Sequence Analysis (the course is oversubscribed)
 - The Internet courses cost about 20% more effort but bring in twice the revenues of onsite courses
- There are about 500 students enrolled in classes related to the center
- The teaching load per professor is 0-2 courses per year
- International exchanges are highly encouraged; students can take their salaries and move anywhere in the globe for half a year
- After the degree, most students go on to industry

CONCLUSION

CBS is a world-leading institution in Bioinformatics and Systems Biology. It has introduced substantial innovations in the field. The most recent developments such as those on data integration, the inclusion of a changing environment in systems biology, the integration with medical informatics, etc., open wide roads for research in the next decades.

REFERENCES

- Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* 10:1-6.
- Nature* editorial. 2005. “Let data speak to data.” 438:531 (Dec.).
- de Lichtenberg, U., L.J. Jensen, S. Brunak, and P. Bork. 2005. Dynamic complex formation during the yeast cell cycle. *Science* 307:724-727.
- Jensen, L.J., T.S. Jensen, U. de Lichtenberg, S. Bruanak, and P. Bork. 2006. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* 443:594-597.

Lage, K., E.O. Karlberg, Z.M. Storling, P.I. Olason, A.G. Pedersen, O. Rigina, A.M. Hinsby, Z. Tumer, F. Pociot, N. Tommerup, Y. Moreau, and S. Brunak. 2007. A human phenome–interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology* 25:309–316.

Site: Technische Universität München and
Leibniz Supercomputing Center (LRZ)
Boltzmannstrasse 1
D-85748 Garching, Germany
<http://portal.mytum.de/welcome>
<http://www.lrz-muenchen.de/wir/intro/en/>

Date Visited: February 27, 2008

WTEC Attendees: S. Glotzer (report author), L. Petzold, C. Cooper, J. Warren, V. Benokraitis

Hosts: Prof. Dr.rer.nat. Ernst Rank, Vice-President, TUM
Email: rank@bv.tum.de
Prof. Dr. Hans-Joachim Bungartz, Informatik V
Email: bungartz@in.tum.de
Dr. Matthias Brehm, Group Leader HPC
Email: brehm@lrz.de
Prof. Dr. Heinz-Gerd Hegering, Chairman of the Board, LRZ
Email: hegering@lrz.de
Dr.rer.nat. Ralf-Peter Mundani, CeSIM
Email: mundani@tum.de

BACKGROUND

The Bavarian Ministry of Sciences, Research, and the Arts is the umbrella organization that oversees both the Bavarian public universities such as Technische Universität München (TUM) and the Bavarian Academy of Sciences and Humanities, of which the Leibniz Supercomputing Center (Leibniz-Rechenzentrum, LRZ) is a part. These organizations reside in what is often referred to as the “Silicon Valley of Germany”, and they benefit from a large base of technically advanced industrial companies in their general vicinity, in areas such as life sciences, media, telecommunication, environmental engineering, aerospace, finance, software, automotive, and R&D.

The LRZ is the common computing center for the University of Munich, the Technische Universität München, the Bavarian Academy of Sciences and Humanities, the Munich University of Applied Sciences, and multiple other higher education institutions (50+) with 120,000 students and staff. Services provided include planning and running the Munich Scientific Network (MWN, networking services), providing and operating server clusters (storage, web, mail, etc.), providing backup and archiving capabilities and services, and maintaining special equipment such as VR, MM, VC, etc.

After extremely informative presentations and productive discussions with our hosts, the WTEC visiting team completed the visit with a tour of the world-class HPC facilities including HLRLB II, one of three of Germany’s national supercomputers. The panel was extremely impressed with the extensive and modern facilities, the programs, and the services provided by the LRZ and its partners. We considered this site a particular highlight of the week-long Europe trip.

SBE&S RESEARCH

The LRZ has a large portfolio of research areas, including IT-management (methods, architectures, tools), piloting new network technology, computational sciences (together with its partners), especially HPC, grid-computing, and long-term document archiving (digital library).

PRESENTATION SUMMARIES

The WTEC visiting team's hosts made presentations in the following areas (with free-flowing conversations among all participants).

The Role of the Leibniz Supercomputing Center (LRZ) in the Munich Area and in the National HPC Landscape

Prof. Dr. Heinz-Gerd Hegering, Leibniz-Rechenzentrum

In general, CFD was the original core application at the LRZ, but now, with the new SGI Altix, with an enormous 40 terabytes of shared memory, 10,000 cores and consuming more than 1 MW of power, the applications have become varied. Current jobs are CPU limited rather than memory limited. Application areas include cosmology & astrophysics, quantum chromodynamics, fluid dynamics, physics, chemistry, and material sciences, and others. In addition, the LRZ maintains substantial research in HPC itself. Areas of focus include IT management, virtualization, networking, archiving, grids, and HPC.

In Germany, there is a nationwide system for HPC, providing a pyramid of HPC infrastructure with three national centers for high-end computation at the top: NIC Jülich (IBM Hardware), LRZ München (SGI Hardware), and HLRS Stuttgart (NEC Hardware). These three have joined forces to form the Gauss Center for Supercomputing (GCS), which presents a unified face of German computation to the European Community.

Members of the LRZ have joint appointments with other university departments. LRZ provides a user support group to help researchers, and to investigate their code and adapt their code to the supercomputer.

The LRZ plans to continue to cycle its high-end hardware with a 5–6 year replacement time. Now also an EU project for a European HPC “ecosystem” aims at installing 3–4 European supercomputing centers, for which GCS will be an applicant (providing one voice towards the EU).

Computing and Computational Sciences Research & Education in Munich and Bavaria

Prof. Dr. Hans-Joachim Bungartz, Scientific Computing in Computer Science

Bavaria, and in particular the Munich and Garching campuses, provide a unique environment for HPC and Computational Sciences, encompassing Technische Universität München (TUM) and Ludwig-Maximilians University Munich (LMU) as two of nine elite universities in Germany, and a number of leading HPC centers, including LRZ Munich, the RZG Munich (MPI Supercomputing Center), the Regional Supercomputing Center Erlangen, and the Munich Computational Sciences Center. These institutions have supported a number of large-volume research programs since 1992 as core topics of Bavaria's political agenda (“Agenda 2020”), including

- FORTWIHR: Bavarian Consortium on HPC (1992–2001), largest and longest-running in Germany
- KONWIHR: Bavarian Competence Network on HPC (since 2002)

Of particular import are several “computational” programs at the bachelor's and master's levels, including BGCE, a Bavaria-wide honors program at the master's level, and IGSSE, TUM's postgraduate school in the German Excellence Initiative

Graduate and undergraduate programs allow specialization in CSE within the classical study programs at TUM, for example, an “Algorithms and Scientific Computing” in Informatics MSc. A full list includes master's programs in Mathematics for Engineering, Computational Physics, Computational Mechanics, and Computational Science and Engineering. At the PhD level, research training groups include the Centre for Simulation Technology in Engineering (CeSIM) and the Centre for Computational and Visual Data Exploration (exploraTUM).

For example, TUM has a CSE (international, i.e., in English) Master's program, with multidisciplinary and cross-departmental cooperation of 7 TUM departments (Maths, Informatics, Physics, Chemistry, Mechanical Engineering, Electrical Engineering, Civil Engineering). The master's program has a focus on simulation methodology, i.e., on advanced computing aspects. Examples include (in Engineering) Computational Fluid Dynamics, Computational Structural Mechanics, Computational Electrodynamics, and (in Science) Computational Physics, Computational Chemistry, Mathematics in Biology, and others. An industrial internship is the default and an external master's thesis is possible.

TUM also participates in the Elite Network of Bavaria (ENB), established in 2003 by the State of Bavaria with the objective of developing special programs for the most talented students. The funding for the program was developed by a remarkable initiative where increased working hours in public services were mandated in Bavaria (without salary compensation), and the net labor savings (>220 positions in total) were invested in ENB. The main realization of these savings was in two programs: (1) Elite graduate programs—study programs with their own master's degree, and (2) international doctorate programs. Currently there are more than 30 programs, with no topical restrictions. The expectation is that ultimately there will be < 1,000 students all over Bavaria participating.

In the same vein, TUM, in a joint venture with FAU Erlangen formed, in 2004, the Bavarian Graduate School of Computational Engineering (BGCE, <http://www.bgce.de>). The program provides an umbrella for three successful MSc programs in computational mechanics and CSE at TUM, and Computational Engineering (BSc and MSc) at FAU. These programs incentivize additional efforts by the students with additional rewards; in other words, “Do more, get more!” For example, if a student does 30 more credits during the remaining semesters while maintaining high average marks s/he is rewarded with individual guidance and a Master's degree “with honors.” Many students intend to move on to either a PhD (TUM, ETH, MPI Saarbrücken, St. Louis, ...) or to industry. Coursework is in appropriate areas such as Multigrid Methods, Lattice Boltzmann Methods, Stochastic Differential Equations in Finance, and others. Summer schools have been set up in this context at venues such as Ferienakademie (Sarntal, Italy), JASS (St. Petersburg, Russia), Indo-German Winter School (Kharagpur, India), Summer Academy (Montenegro), and others.

One of the more innovative efforts within the BGCE is the “Software Project,” which promotes the development of software for HPC/CSE as an educational goal. The program forms teams of 5–8 students who work together for 6–9 months and produce a complete software system—from the idea to the product. All aspects of this progression are represented, including customer, roles (project manager, experts), product specification, architectural design, implementation, documentation, presentation, fines for breach of contract, etc. Prior topics have included Computational Steering (2004), Molecular Dynamics (2005), Fluid-Structure Interaction (2006), Visualization and Finance (2007). The program has had about 15–17 students per year, with an international makeup, including participants from China, India, Pakistan, Saudi Arabia, Iran, USA, Costa Rica, Brazil, Germany, Belgium, Bulgaria, Turkey, Russia, Ukraine, and Serbia. Eleven students graduated in 2006.

International Graduate School of Science and Engineering: TUM's framework for CSE-Oriented Graduate Education

Prof. Dr.rer.nat. Ernst Rank, Computation in Engineering

The mission of the International Graduate School of Science and Engineering (IGSSE) is to bridge the two cultures of science and engineering through multidisciplinary, building on disciplinary excellence, the preparation of young scientists for cosmopolitan leadership, the promotion of international exchange and networking, a research training program, and the support of high-risk projects with the chance for high reward. The IGSSE encourages the formation of teams, with 2 PhD students funded by IGSSE and at least 2 PhD students funded by industry, as well as a postdoc (team leader). The IGSSE has a number of partners, including Stanford University, the University of Tokyo, Technical University of Denmark (DTU), Weizman Institute, ETH Zürich, Technical University of Vienna, University of New South Wales, National University of Singapore, and others, as well as DLR, NASA, Siemens, General Electric, and Fujitsu Labs.

The IGSSE is funded through sources within the university, industry and the Excellence Initiative. Selection of projects is by a review board at the TUM Graduate School. Current research areas include (i) Computational Science and Engineering, (ii) Biomedical Engineering, (iii) Nanotechnology and Advanced Materials, Energy, (iv) Geodynamics and the Environment. More than 100 PhD students team up with senior scientists, postdocs, and master's students in 30 interdisciplinary project teams. Many of these teams have their centre in HPC/CSE, most others have through their multidisciplinary approach at least a strong connection to CSE.

KEY QUESTIONS

The WTEC team's hosts provided the panel with extensive consideration of the questions of interest, and their responses to the panel's questionnaire are given here in full.

General

- What are the major needs, opportunities or directions in SBE&S research over the next 10 and 20 year time frames?
 - Establish a holistic approach to computing – widening the focus from HW/Moore's law and numerical analysis/O(h)-type considerations to other aspects of efficiency: hardware awareness, ubiquitous parallelism, data-driven computing, data exploration, software quality, ...; from the CPU-time focus to a development time focus
 - Establish a "trans-disciplinarity mainstreaming" – overcoming today's mindset that computational physics just needs physicists "with some additional courses"
 - Bring together excellence in computing infrastructure, computing methodology, and CSE applications
 - Doing the step from a mere number-driven computing to a data- and insight-driven one
- What are the national and/or regional funding opportunities that would support research to meet these needs, and/or take advantage of these opportunities? Are these funding opportunities expanding?
 - Computing: slightly expanding (program "HPC software" of the federal Dept. of Research (BMBF), EU project PRACE (Partnership for Advanced Computing in Europe) to prepare the installation of European supercomputing centres, ...)
 - CSE: expanding, but problematic (one problem being the main funding agencies such as German Research Foundation (DFG) still aligning everything with classical disciplines: a proposal on "Hardware-aware finite element computations for Tsunami simulation" was submitted to the math department, handed over internally to the informatics department (since the principal applicant came from informatics), and then reviewed by computer science experts the agent had on his/her list – obviously no CSE experts!)
 - This leads to the fact that there is a lot of "intra-discipline interdisciplinarity" in proposals and reviews, but not that much of a real trans-culture-cooperation
- **Material / Energy and Sustainability / Life Science and Medicine**
- What major breakthroughs in these fields will require SBE&S; and which are you and or your colleagues pursuing? Within your institution, region, or country, are there identified targets of opportunity for applications of simulation either for scientific research or for engineering applications in these fields?
 - Hardly any breakthroughs without simulation!
 - Activities in basically all of these fields (from a TUM-perspective (consider IGSSE's areas of research) to a Munich perspective (Max Planck institutes for plasma physics (ITER, tokamak; energy), physics (LHC-ATLAS; material), biochemistry (life science))
- Which problems could benefit most from a 1–2 order of magnitude increase in computational power?
 - Wrong question – most problems would benefit significantly; the question is how to ensure that the gain in computing power is really exploited to its full potential (a gain in computing power also helps those doing Gauss-Seidel iterations ...)

- What are examples of major SBE&S successes or failures in these fields?
 - Success story: What and where would astrophysics be without simulation?!
 - Problem field: weather forecasting, where many see fundamental limitations
 - Emerging general failure potential: lack of simulation software quality (example: Potsdam Institute for Climate Impact Research – very prominent, a lot of modeling and simulation, but so far hardly any concern about a systematic software approach)
- Do investigators, laboratories and institutions receive any financial compensation for patented inventions derived from their simulations?
 - Sure, but the problem frequently is the “hidden innovation phenomenon”: the Nobel Prize or the profit goes to the physicist or engineer, not to the simulation technologists behind the computational result enabling the scientific breakthrough
- Have any start-up companies spun-off based on simulation effort in your lab? If so describe them.
 - Such stories do exist: Tetaris, e.g., a start-up company doing computational finance, having got some venture capital support; two founders from TUM’s informatics and math dept.; or, on a Bavarian scale, INVENT Computing, a spin-off from the fluid mechanics chair at University of Erlangen-Nuremberg
 - However, far less activity than in other fields (engineering consulting, IT, ...)

Multiscale Simulation

- Describe efforts and advances within your institution to couple multiple simulation methods in order to bridge multiple length and/or time scales.
 - Multiscale modeling is a focus in TUM’s math department
 - Multiscale numerics (multigrid, ...) is a focus of TUM’s two scientific computing chairs; several co-operations with groups from the application side
 - A lot of research projects or consortia are run or coordinated by TUM: FOR 493 Fluid-Structure Interaction, FOR 507 Large-Eddy Simulation, SFB 438 Math modeling and simulation, ...
 - At least 6 “young Researcher’s Groups in TUM’s International Graduate School of Science and Engineering” work on multi-scale and multi-disciplinary problems
- Is the development of integrated Multiscale modeling environments a priority in research funding in your country or region?
 - The mathematics behind and the application are funded; for software development and PSE etc., it is much harder
 - Several initiatives to the German Science Foundation (DFG) are yet on the way
 - TUM itself focuses funding on this topic (see previous response)

Validation, Verification, and Quantifying Uncertainty

- Describe efforts and advances within your institution to validate and verify codes and to quantify uncertainty in simulation-based predictions.
 - Validation: a lot of “simulation-meets-experiment” projects (FSI, e.g.; experimental benchmarking)
 - Verification: difficult, due to the huge mental distance of CSE and software engineering; however, first successful steps (latest result @TUM: complete verification of a sparse grid hierarchical basis transform program with ISABELLE, a leading theorem prover)
 - Uncertainty: professorship in “Computational Statistics”, involvement in several MAC projects (with geophysics people, astrophysics, biologists, ...)

Simulation Software

- What percentage of code used in your group is developed in-house? What percentage of code is commercial? What percentage is open-source? What percentage has been developed by others (e.g., under

contract or by acquisition)? What are the biggest issues in using models/simulations developed by others? How easy/difficult is it to link codes to create a larger or multifaceted simulation environment?

- Depends on purpose, field and size of problem— hence no exact percentages available
 - Estimations by field for large scale projects:
 - CFD and Engineering: 90 % in-house /10% commercial
 - Chemistry: 40 % in-house / 30 % open source or academia software / 30 % commercial
 - Small scale projects: 10 % in-house / 30% open source/academia / 60 % commercial (Gaussian)
 - Physics / Astrophysics / Geophysics: 90 % in-house
 - Prototyping: Matlab & Simulink widespread
 - Industry co-operations in engineering: commercial software frequently used (legal aspects)
 - Quantum chemistry: Gaussian, and ...
 - But also a lot of “home-made” simulation codes: CFD, CSD, MD, ...
 - Especially in some simulation based engineering research groups: A significant number of large own software development (O(1 M lines of code))
 - Medium issue: groups tend to either commercialise it (via a spin-off, e.g.) or not to provide it at all (for fear of competitors)
 - Big issue: old-fashioned license models, in particular for parallel computing (example: coupling library MpCCI needs one license per core (!))
 - Linking is in most cases quite complex and costly as simulation codes often have proprietary standards, use different discretisation schemes, are transparent (i.e., black box) concerning implementation details, don't provide source code access, etc.
 - Example for one of our activities: FOR 493 FSI, where the coupling PSE FSIce has been developed
- Who owns the intellectual property rights (IP) to the codes developed in your group?
 - University and/or person/group developing the code

Big Data and Visualization

- What type of data and visualization resources do you need (and have access to) for SBE&S research?
 - Need
 - Large storage resources accessible locally or via LRZ/grid services for long/huge simulation runs; capable for tackling huge data advent arriving in single peaks (e.g., LHC)
 - Fast and efficient visualisation resources for handling huge data amount, preferably in real time, for interactive computational steering applications (as an example)
 - Access
 - Many small- and medium-size data and visualisation resources are owned by single groups/chairs
 - Huge data and visualisation resources (e.g., storage systems, holobench, visualisation cluster) are offered by LRZ

Engineering Design

- What type of models/codes do you use, develop or conduct basic research on pertaining to different phases of engineering system design (conceptual, parametric optimization, operational/control)?
 - DFG's Priority Programme 1103 “Network-based co-operative planning processes in structural engineering”, e.g., different approaches to support cooperation in (distributed) design processes
 - Process modeling (formal methods, relational model description, Petri networks, ...)
 - Distributed building information models (algebra of sets, model transformation, ...)

- Distributed simulation (computational steering, 4D simulation, ...)
- Multi-agent systems
- FORBAU - Virtual Building Site: ERP (Enterprise Resource Planning) systems are coupled with a construction and simulation model and the spatiotemporal dependences of construction proc. in a holistic construction-information-model
- What are the data requirements for a comprehensive life-cycle model of an engineered system? Do you have repositories in place that provide the data?
 - SFB 768 “Zyklusmanagement von Innovationsprozessen“, e.g., deals with related aspects such as the examination of technical, competitive, and social based cycles that essentially influence and hamper the development and introduction of innovative products and services; group consisting of several disciplines (engineering, computer science, sociology, marketing); just started in 2008, hence, no results so far
- How do you couple output of the models/simulations to the decision making process (including quantification of uncertainty/error in the predictions)?
 - Actually just emerging as a research topic; also in industry more or less hands-on strategies
- What is the curriculum for training doctoral students in all aspects of designing engineered systems using simulation/modeling tools?
 - There is no particular curriculum for doctoral students concerning this aspect. Yet, in the typical German engineering Ph.D. model a lot of competence is gained especially via industry cooperation, which is very often conducted in connection with the research work.
- Are there efforts within your institution to couple physics-based models with macroscopic logic or econometric models?
 - Micro-macro: yes, especially in an engineering context (example: former SFB 411 on Biological Wastewater Treatment; coupling of CFD to biofilm growth, coupling of spatial simulations to system simulators)
 - Micro-econo: first steps especially in the earth science context (e.g., LMU’s geophysics group and Munich Re insurance company)

Next-Generation Algorithms and High-Performance Computing

- Would you characterize the SBE&S research in your lab as needing and/or using primarily desktop, teraflop, or petaflop computing resources?
 - Garching campus as a whole: usage of all kinds/scales of computing resources
 - Desktop for code development
 - Small clusters for optimisation/testing of parallel code
 - Teraflop/petaflop for production runs
 - Max Planck institutes: clear petaflop needs
 - TUM
 - Esp. engineering research: also many scenarios, where the model is still in the fore (materials, optimization, ...) and where computational power is not the bottleneck
 - Algorithm development of math and informatics is more desktop- or cluster-related, the big machines primarily for scalability studies etc.
- What are the computational bottlenecks in your simulation problems? What solutions exist now and in the near future, either from a hardware perspective, algorithm perspective, or both?

Problems

- All facets of efficiency (such as data locality, cache usage etc.)

- Scalability: multi-core and the resulting massive parallelism (examples: hardly any robust multigrid scheme scaling well beyond 1,000 cores; many eigenvalue solvers not scaling beyond 10,000 cores)
- Communication/synchronisation for MPP, new parallel programming paradigms
- Data handling, analysis, exploration, rendering
- Still many open modeling issues (i.e., the “pre-computing stage”)
- Software (maintenance/re-engineering of older codes, version zoo, lack of reliability due to missing testing culture and verification)

Solutions

- Some ideas in the concept of MAC
- Hierarchy etc. as general paradigms
- More abstraction levels in software
- What is the state-of-the-art in your community in discrete event simulations (model size, parallelisation or thread limits, special techniques – time warp, rollback etc.)?
 - DES is used (by computer engineering groups, for traffic simulation—cf. Ph.D. thesis Srihari Narasimhan), but, as far as we know, no DES-centered research at present

Education and Training

- Is there a formal scientific computing or computational science and engineering graduate program at your institution? Or, if you are in a company, are their educational institutions/programs in your country or region that you find prepare students in formal scientific computing or computational science and graduate engineering effectively?
 - TUM: two international master’s programs Computational Mechanics and Computational Science & Engineering
 - TUM & Bavaria: Bavarian Graduate School of Computational Engineering (BGCE) as a Bavaria-wide honours program (“elite program”)
 - TUM: International Graduate School of Science & Engineering (IGSSE) on PhD level
 - Germany and neighbours: Working Group of CSE programs in Germany/Austria/Switzerland (München, Stuttgart, Darmstadt, Frankfurt, Aachen, Braunschweig, Rostock, Erlangen, Bochum, Bremen, Hannover, Dresden, Zürich, Basel, Graz) established 2005, will get the legal status of an “eingetragener Verein” in 2008
 - Undergraduate + graduate programs (Darmstadt, Erlangen)
 - Meer graduate programs (most)
 - Doctoral programs (Aachen, Darmstadt, Erlangen, München)
- What level of computing expertise do your incoming graduate students or employees arrive with? Do you feel that they have the necessary background to conduct research in SBE&S or do they need extensive preparation after arriving before they can be productive researchers?
 - Very heterogeneous (in particular to the international programs – even a CSE background does not at all ensure a decent computing background)
- What kind of training is available in your organisation for simulation and high performance computing? Is it adequate? Do you believe it will be adequate to address computing on multi-core or petascale architectures? What plans are in place for training programs in next-generation computing?
 - Tailored courses within specialized CSE programs and mainstreaming programs (informatics, mathematics, engineering, ...): all kind of lectures, a broad offer of lab courses (CFD lab, scientific computing lab, performance optimized computing lab, visual computing lab, augmented reality lab)
 - Additional (specialized) courses at Leibniz Supercomputing Centre (LRZ): parallelization, ...

- A very solid basis – definitely the most elaborate in Germany, permanently (and currently) updated to meet future requirements (e.g., activities within the Munich Multi-Core Centre, of a lecture series on future architectural trends co-organized by TUM and IBM)
- Summer school courses (Simulation – from Models to Software; Simulation Technology; ...), European Advanced Courses (ATHENS program: Parallel Numerical Simulation)
- Also a lot of TUM activities on the international scale: establishing Computational Engineering in Belgrade, establishing Computational Science in Tashkent/Uzbekistan, establishing Applied and Computational Physics in St. Petersburg/Russia, planned joint CSE program with National University of Singapore
- What fraction of students in your institution study/work in the area of SBE&S, and how has this fraction changed over the past 5–7 years?
 - No general answer possible: a significant part in maths, a minority in informatics, in engineering the majority on a “use some simulation tool” basis, but a minority really doing CSE work
 - Increasing numbers
- What fraction of graduate students/postdocs in SBE&S comes from abroad? From which country do these researchers originate? How many students in your country would you estimate go abroad to do their PhDs in SBE&S?
 - Hard to give special numbers for simulation; however, simulation is a very internationally oriented field, many of TUM’s partnerships have their origin here
 - Incoming: About 80% from abroad in CSE/COME master’s programs (mainly from Near East, Asia, Eastern Europe, and Middle/South America)
 - TUM in general: about 20% students from abroad; higher-level staff (faculty, professors): still rather small, but increasing numbers
 - Outgoing: Increasing numbers for graduate students, since there are a lot of partnership programs (bilateral ones, ERASMUS, ...)
 - Long-term goal of the informatics dept., e.g., offer all students a one-semester-abroad option
 - Increasing numbers for Ph.D. students, since more recent programs establish this as a rule (e.g., IGSSE)
- After completing their PhD in an SBE&S related field, what is the route your students take? Do they take postdoctoral positions in your country or abroad? What fraction eventually achieve permanent jobs in SBE&S? Is this a desired path for students? Do they perceive many job opportunities related to their training in SBE&S? How does industry view students with training in SBE&S?
 - International students: all variants (returning for a career at home, aiming at an academic career here, trying to get a job in industry)
 - German students: main part heading for industry, small part staying at university

Funding, Organization, and Collaboration

- What are the roles of academic, government and industrial laboratories in SBE&S research in your country?
 - Academic: education; research depending on university’s focus
 - Government: topical research centers (FZJ, FZK, DESY, DKRZ, GSI, ...) – most of them with are smaller or larger simulation branch; organised in several associations (Helmholtz, Leibniz, Max Planck)
 - Supercomputers: operated by all types of institutions (HLRS Stuttgart: university; Jülich: government lab; LRZ München: Bavarian Academy of Sciences; RZG Garching: Max Planck Society; ...)
 - Industry: not that much research in companies; industry-funded labs at universities on special topics (examples: TUM & Siemens – Center for Knowledge Interchange; TUM & Audi: Ingolstadt Institut iniTUM)

- Who pays for, and/or is responsible for, the development and sustainability of SBE&S infrastructure (includes long term data storage costs, code maintenance, open-source software etc.)?
 - Funding mainly from state/government (hardware)
 - Code maintenance via student/postdoc work – actually hardly happening in a professionally organised way at present (one reason: models and algorithms are funded, simulation software is not)
 - Responsibility according to IT infrastructure plan: DFG requires an IT strategy (cf. DFG Committee on IT Infrastructure, publishing “Recommendations” every 5 years and evaluating all IT-related purchases by public universities in Germany) and recommends a CIO; depending on this strategy, responsibility is centralised (Computing / IT Services Center) or local
- What is the funding situation for SBE&S in your country? E.g., what are the major sources of funding? E.g., over the past 5 years, has funding of SBE&S increased, decreased or remained roughly constant relative to funding of all of science and engineering research? E.g., is most SBE&S research funded as single investigator grants, small term grants, or large teams and/or initiatives? E.g., what is the typical duration over which proposals in SBE&S are funded?
 - Computer systems: a national plan (by National Science Council / Wissenschaftsrat) with a “pyramid” of tiers (0: future European HPC centers; 1: national HPC centers (Stuttgart, Jülich, München); 2: regional; 3: departmental); funding as for all university buildings and equipment as a joint venture of the federal government and the state; situation is rather good in Germany
 - Network infrastructure: German Research Network (Deutsches Forschungsnetz, DFN) organising the science network nation-wide, with public funds; situation rather good
 - Projects on all levels (from individual projects to consortia); the bigger ones are primarily application driven (a research consortium “Fast linear solvers” has close-to-zero chances, a consortium on turbulence modeling and simulation has excellent ones; situation is clearly sub-optimal for cross-sectional topics such as simulation)
 - Terms: not topic-dependent, but depending on the funding program (for example DFG: individual grants 3 years, research units 6 years, SFB 12 years, priority programmes 6 years)
 - Amount for funding: not that much dedicated money, but an increasing part of the funds for engineering, for example, are simulation-related.

Site: Unilever Centre for Molecular Informatics
University of Cambridge
Department of Chemistry
Lensfield Road
Cambridge CB2 1EW, U.K.
<http://www-ucc.ch.cam.ac.uk/>

Date Visited: February 27, 2008

WTEC Attendees: M. Head-Gordon (report author), K. Chong, P. Cummings, S. Kim

Hosts: Prof. Robert Glen, Unilever Centre
Email: rcg28@cam.ac.uk
Dr. Jonathan Goodman, Senior Lecturer, Unilever Centre
Email: jmg11@cam.ac.uk
Dr. Dmitry Nerukh, Senior Research Associate, Unilever Center
Email: dn232@cam.ac.uk
Dr. Maxim Fedorov, Postdoctoral
Email: mvf22@cam.ac.uk
Dr. Volker Thome, Postdoctoral
Email: vt228@cam.ac.uk
Dr. Hamsa Y. Mussa, Postdoctoral
Email: hym21@cam.ac.uk
Mr. James Bell, PhD student
Email: jcb56@cam.ac.uk

BACKGROUND

The Unilever Centre for Molecular Informatics was established in 2000 by a large grant from Unilever that also supported the construction of a new building and provided up to 15 years of running costs. The Centre has 4 academic staff members, including the Director, Prof. Robert Glen, and roughly 40 additional students, postdocs, and visitors. The overall mission of the Unilever Centre is to develop new ways of extracting insight into molecular processes such as drug binding from the rapidly increasing amount of data of all types that is becoming available.

RESEARCH

All research at the center is academic in nature, with the goal of seeking alignment between academic interests and issues of interest to Unilever. Prof. Glen leads a large group in the general area of molecular informatics, with present interests in molecular similarity and docking, complexity analysis, aspects of drug design, new molecular property calculations, and data analysis. This work spans a wide range of computational methods from data mining to physics-based approaches, and it also includes some wet lab activities. Dr. Jonathan Goodman leads research into understanding organic reactivity and structure using both computational methods (electronic structure calculations) and synthetic approaches. Prof. Murray-Rust leads a research group that is also focused on molecular informatics, with a strong emphasis on data management through the development of the Chemical Markup Language, and other activities revolving around open data and data mining directly from the chemical literature. Dr. Mitchell performs research on enzyme-ligand interactions using both knowledge-based methods and physically based methods.

COMPUTING HARDWARE

The Unilever Centre exploits high-performance computing resources at the University of Cambridge (recently upgraded to over 2000 processors), and it employs distributed computing for some projects where many independent calculations are involved.

DISCUSSION

During the course of the meeting, the following additional points arose in connection with the operation of the Unilever Centre and its relationship to general issues concerning the future of simulation-based engineering and science:

- *Funding.* Roughly one-third of the Centre funding is provided by Unilever under the operating contract, which is currently in its second 5-year period. Roughly equal fractions also come from government research grants, and contracts from other companies that are not direct competitors of Unilever. The group is limited in size more by available space and facilities than by available funding.
- *Intellectual property.* Under the first 5-year contract, all intellectual property developed in the Centre was owned by the University of Cambridge. This was modified in the second (present) 5-year contract so that IP resulting from research that is fully funded by Unilever is owned by Unilever.
- *Research roadblocks for adoption of high-performance computation in the pharmaceutical industry.* It was discussed that accuracy attainable with molecular dynamics and free-energy simulations using existing molecular mechanics force fields is simply not high enough to be considered predictive. This lack of reliability is a key factor discouraging the pharmaceutical industry from investing in HPC.
- *Other fundamental questions of importance* to the chemical and pharmaceutical industry that cannot at present be satisfactorily addressed include the prediction of solubility, the polymorph into which a molecule crystallizes, and the reliable prediction of drug-binding affinities, already mentioned above.
- *Employment outcomes for graduates of the Unilever Centre:* There are good employment prospects for Centre graduates; at the moment approximately 50% go on to industry and 50% remain in universities.

REPRESENTATIVE PUBLICATIONS OF THE UNILEVER CENTRE

- Cannon, E.O., F. Nigsch, and J.B.O. Mitchell. 2008. Novel hybrid ultrafast shape descriptor method for use in virtual screening. *Chemistry Central Journal* 2:3; doi: 10.1186/1752-153X-2-3.
- Fedorov, M.V., J.M. Goodman, and S. Schumm. 2007. Solvent effects and hydration of a tripeptide in sodium halide aqueous solutions: An in silico study. *Phys. Chem. Chem. Phys.* 9:5423–5435; doi: 10.1039/b706564g.
- Holliday, G.L., D.E. Almonacid, J.B.O. Mitchell, and J.M. Thornton. 2007. The chemistry of protein catalysis. *Journal of Molecular Biology* 372:1261–1277; doi: 10.1016/j.jmb.2007.07.034.
- Hughes, L.D., D.S. Palmer, F. Nigsch, and J.B.O. Mitchell. 2008. Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and log P. *Journal of Chemical Information and Modeling* 48:220–232; doi: 10.1021/ci700307p.
- Llinas, A., and J.M. Goodman. 2008. Polymorph control: Past, present and future. *Drug Discovery Today* 13:198–210; doi: 10.1016/j.drudis.2007.11.006.
- Nigsch, F., and J.B.O. Mitchell. 2008. How to winnow actives from inactives: Introducing molecular orthogonal sparse bigrams (MOSBs) and multiclass winnow. *Journal of Chemical Information and Modelling* 48:306–318; doi: 10.1021/ci700350n.
- O'Boyle, N.M., G.L. Holliday, D.E. Almonacid, and J.B.O. Mitchell. 2007. Using reaction mechanism to measure enzyme similarity. *Journal of Molecular Biology* 368:1484–1499; doi: 10.1016/j.jmb.2007.02.065.
- Palmer, D.S., A. Llinas, I. Morao, G. M. Day, J.M. Goodman, R.C. Glen, and J.B.O. Mitchell. 2008. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Molecular Pharmaceutics* 5:266–279; doi: 10.1021/mp7000878.
- Paton, R.S., and J.M. Goodman. 2007. Exploration of the accessible chemical space of acyclic alkanes. *J. Chem. Inf. Model.* 47:2124–2132; doi: 10.1021/ci700246b.

- Paton, R.S., and J.M. Goodman. 2008. 1,5-Anti stereocontrol in the boron-mediated aldol reactions of β -alkoxy methyl ketones: The role of the formyl hydrogen bond. *J. Org. Chem.* 73:1253–1263; doi: 10.1021/jo701849x.
- Simon, L., and J.M. Goodman. 2007. The mechanism of TBD-catalyzed ring-opening polymerization of cyclic esters. *J. Org. Chem.* 72:9656–9662; doi: 10.1021/jo702088c.
- Smith, S.G., R.S. Paton, J.W. Burton, and J.M. Goodman. 2008. Stereostructure assignment of flexible five-membered rings by GIAO ^{13}C NMR calculations: Prediction of the stereochemistry of elatenyne. *ASAP J. Org. Chem.* 73; doi: 10.1021/jo8003138.
- Torrance, J.W., G.L. Holliday, J.B.O. Mitchell, and J.M. Thornton. 2007. The geometry of interactions between catalytic residues and their substrates. *Journal of Molecular Biology* 369:1140–1152; doi: 10.1016/j.jmb.2007.03.055.

Site: Unilever R&D Port Sunlight
Quarry Road East
Bebington
Wirral CH63 3JW, UK
<http://www.unilever.co.uk>
http://www.unilever.co.uk/ourvalues/sciandtech/hpc_randd

Date Visited: February 25, 2008

WTEC Attendees: P. Cummings (report author), M. Head-Gordon, S. Kim, K. Chong

Hosts: Dr. Dominic Tildesley, Vice President, One-Unilever Platform Leader, Structured Materials and Process Science; Email: dominic.tildesley@unilever.com
Dr. Janette Jones, Discovery Platform Director for Systems Biology
Email: janette.jones@unilever.com
Dr. Massimo Noro, Manager, Physical and Chemical Insights
Email: massimo.noro@unilever.com
Dr. Peter Olmsted, Professor of Physics, School of Physics and Astronomy, Polymers and Complex Fluids Group, University of Leeds (university-based collaborator)
Email: p.d.olmsted@leeds.ac.uk
Dr. Ian Stott, Lead Scientist for Informatics; Physical and Chemical Insights
Email: ian.stott@unilever.com
Dr. Patrick Warren, Scientist, Physical and Chemical Insights
Email: patrick.warren@unilever.com
Dr. Simon Watson, Scientist, Physical and Chemical Insights
Dr. Jerry Winter, Manager, Physical and Chemical Insights

BACKGROUND

Unilever is one of the world's most successful consumer goods companies, with focuses in three businesses: food, home care, and personal care products. Unilever leads the home care market in much of the world, which includes cleansing and hygiene products, with brand names such as Cif, Comfort, Domestos, Persil, and Comfort. Within the personal care market, Unilever is among the global leaders in products for skin cleansing, deodorants, and antiperspirants. It has annual sales of over €40 billion and has 174,000 employees in close to 100 countries.

Unilever devotes approximately 2.5% of its yearly turnover to research and development, making it one of Europe's biggest R&D investors in household care. The research center at Port Sunlight employs over 700 scientists researching home and personal care products. Each year research conducted at Port Sunlight results in over 100 patent filings and approximately 140 peer-reviewed papers and conference presentations. In addition to collaborations with other Unilever R&D centers located throughout the world, researchers at Port Sunlight have collaborations with universities and consultants, both within and outside the UK. In fact, Unilever has been successful in leveraging UK and regional government funding, as well as EU funding, to advance its research goals.

The visit was hosted by Dominic Tildesley, a former chaired professor of chemistry at Southampton (until 1996) and Imperial College (1996-98), who joined Unilever as Head of the Physical Sciences Group in 1998. Tildesley is a world-renowned expert on molecular simulation and coauthor with Mike Allen of the University of Warwick of the seminal textbook in the field (Allen and Tildesley 1994). The Unilever group is widely admired as one of the most successful modeling groups in industry. This same group was visited in 1999 as part of the WTEC study on molecular modeling, and it is interesting to note the evolution in the effort between 1999 and 2008. Although the size of the group is comparable, the focus has shifted somewhat, with less emphasis on traditional molecular modeling and an increased emphasis on systems biology and chemical informatics. The group's shifting interests reflect a growing interest in the development of targeted products that reflect customer characteristics. For

example, Unilever has announced a collaboration with the genome scale metabolic modeling company Genomatica to “to take advantage of Genomatica’s unique metabolic modeling and simulation technologies to accelerate the development of novel ingredients to improve the effectiveness of Unilever products” (<http://www.genomatica.com/news.20060809.html>).

R&D ACTIVITIES

The WTEC visiting team viewed presentations by members of the modeling group: (1) Molecular Dynamics of Lipids (Massimo Noro and Peter Olmsted), (2) Systems Biology (Janette Jones and Patrick Warren), (3) Finite Elements (Simon Watson), and (4) Polymer Informatics (Jerry Winter and Ian Stott). Afterwards, the WTEC team and our hosts discussed both technical issues and funding mechanisms.

Unilever researchers are leaders in industrial application of molecular modeling. The WTEC team very much appreciated the openness of our hosts in our discussions; the insights we gained from visiting Unilever were many. However, because much of the research we discussed had proprietary elements, we respect Unilever’s wishes that we not publish details of our discussions.

CONCLUSIONS

The WTEC visit to Unilever left the WTEC team with the strong impression that the Unilever R&D modeling group is evolving in line with the directions of the parent company and in response to the availability of new tools (particularly informatics- and informomics-based tools) relevant to the core businesses. The research efforts reflected maximum leverage of regional, national, and European funding sources, and collaborations with UK and European universities. The leadership of Dominic Tildesley, coming out of an academic environment with excellent connections to the UK academic institutions and national funding councils, may be a factor in Unilever R&D’s success in making external collaborations work for the company.

In contrasting the UK university collaborations with similar collaborations that might take place with U.S. universities, issues of negotiating intellectual property arrangements and contract terms with North American universities were cited as hurdles in starting and maintaining collaborations with leading U.S. groups.

REFERENCES

- Allen, M.P., and D.J. Tildesley. 1994. *Computer simulation of liquids*, 2nd ed. Oxford: Clarendon Press.
- Warren, P.B., and J.L. Jones. 2007. Duality, thermodynamics, and the linear programming problem in constraint-based models of metabolism. *Physical Review Letters* 99, Art. no. 108101.

Site: University College London
Gower Street
London WC1E 6BT, UK
<http://www.ucl.ac.uk/>

Date Visited: February 29, 2008

WTEC Attendees: P. Cummings (report author), M. Head-Gordon, S. Kim, K. Chong

Hosts: Prof. C. R. A. (Richard) Catlow, Dean, Mathematics and Physical Sciences Faculty Email: c.r.a.catlow@ucl.ac.uk; <http://www.chem.ucl.ac.uk/people/catlow/>
Prof. Peter V. Coveney, Director, Centre for Computational Science
Email: p.v.coveney@ucl.ac.uk; <http://www.chem.ucl.ac.uk/people/coveney/>
Prof. Michael Gillan, Director, Materials Simulation Laboratory, UCL
Email: m.gillan@ucl.ac.uk; <http://www.cmmp.ucl.ac.uk/~mjg/>
Dr. Paul Kellam, Reader, Division of Infection and Immunity, Medical School, UCL
Email: p.kellam@ucl.ac.uk; <http://www.ucl.ac.uk/medicalschoo/infection-immunity/research/group-leaders/pkellam.htm>
Dr. Maziar Nekovee, Royal Society Industrial Research Fellow (UCL) & BT Research
Email: maziar.nekovee@bt.com; <http://nekovee.info>
Prof. G. David Price, Vice Provost for Research, UCL
Email: d.price@ucl.ac.uk; <http://www.es.ucl.ac.uk/people/d-price/index.htm>
Prof. Sally L. Price, Professor of Theoretical & Computational Chemistry
Email: s.l.price@ucl.ac.uk; <http://www.chem.ucl.ac.uk/people/slprice/>

BACKGROUND

University College London (UCL) was established in 1826, the first university to be established in England after Oxford and Cambridge. UCL is regarded as one of the four top-ranked universities in the United Kingdom (along with Oxford, Cambridge, and Imperial College). It has more than 2,000 faculty members (including over 600 professors in established or personal chairs) in 72 departments, 19,000 students (more than a third whom are in graduate programs, with half of these pursuing research degrees), and 4000 post-doctoral researchers. Almost a third of the UCL students come from outside the United Kingdom (more than 140 countries). UCL is the largest medical school in Europe, and research in the medical and biological sciences constitutes a major institutional focus. UCL has considerable ties with industry, including a research and graduate training centre at BT's Adastral Park, as well as a £4.6 million program to equip UCL's scientists and engineers with enterprise and business skills.

Institutionally, UCL has made a major capital investment (£3.6 million) in a new 2560-core cluster with Infiniband interconnect rated at 27TF peak. Information about the cluster, called Legion, is available at the UCL Research Computing website, <http://www.ucl.ac.uk/research-computing/information/services/cluster>. With the transition to full costing of research in the UK, it is important to note that UCL is applying full costing upstream of the user, so that use of Legion is free to users. (In other UK universities, as a result of the transition to full costing of research, researchers are typically being charged for usage by the core-hour or node-hour consumed.) UCL is committing £1.5 million/year for research computing, including technical refreshing of the Legion cluster.

UCL also has a Centre for Computational Sciences, <http://ccs.chem.ucl.ac.uk/>, headed by Peter Coveney, which provides a bridge to external computing facilities, such as the UK CSAR and HPCx. CSAR (Computer Services for Academic Research) is a national high-performance computing (HPC) service for the UK run on behalf of the Research Councils by the Computation for Science (CfS) consortium and located at Manchester University. HPCx is the capability computing center for the UK (<http://www.hpcx.ac.uk>), located at the University of Edinburgh. Of the cycles at HPCx, 25–30% are associated with consortia led by or containing UCL faculty.

Much of the high-performance computing at UCL is related to research in the biological and medical sciences and research at the life sciences/medical interface, such as bioinformatics. An interesting aspect of this is that value-added tax (essentially a national sales tax of 17.5%) is exempt on biomedical-research-related purchases. This contrasts with the United States, where sales tax at the state level is generally not levied on any purchases by nonprofit educational institutions.

R&D ACTIVITIES

The WTEC visiting team heard presentations from each of the hosts, as summarized below.

Peter Coveney: Overview of Simulation-Based Engineering and Science Research at UCL

Peter Coveney began with the computing environment, describing both local resources (Legion and future anticipated systems at UCL, including 48-core SMP nodes) and connections with external resources. Researchers at UCL are connected to the UK National Grid Service (NGS, <http://www.grid-support.ac.uk>), the Distributed European Infrastructure for Supercomputing Applications (DEISA, <http://www.deisa.org>), and the TeraGrid in the U.S. (<http://www.teragrid.org>). The grid middleware enabling grid-based applications is GridSAM (<http://gridsam.sourceforge.net>), Globus (<http://www.globus.org>) and UniCORE (Uniform Interface to Computing Resources, <http://www.unicore.eu>). In view of the role of medicine at UCL, the data-driven “omics”—transcriptomics, proteomics, metabolomics, physiomics, phenomics, and pathomics—play a prominent role. The European data center for core biomolecular data is located in the Wellcome Trust Genome Campus south of Cambridge, and is called the European Bioinformatics Institute within the European Molecular Biology Laboratory (EMBL-EBI, <http://www.ebi.ac.uk>). UCL researchers are marrying “omics” data to dynamical models for signaling networks, protein dynamics, and molecular dynamics to understand diseased states. The goal is to develop grid-enabled patient-specific medical simulation. As part of their efforts, UCL researchers are leading the European-wide effort EU FP7 Virtual Physiological Human (http://ec.europa.eu/information_society/activities/health/research/fp7vph/index_en.htm) that began in 2007 and runs to 2013, with first-year funding of €72 million. UCL researchers, with Peter Coveney as PI, also participate in the GENIUS (Grid Enabled Neurosurgical Imaging Using Simulation) project that aims to use grid computing technology to provide real-time simulation tools for neurosurgeons to dry-run a surgery before performing it on a real patient (<http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=EP/F00561X/1>).

Coveney also gave the UCL perspective on questions raised by the WTEC study. He reported on multiscale simulation methodologies being developed at UCL (Delgado-Buscalioni et al. 2005) to combine particle-based methods (such as molecular dynamics for liquids or direct simulation Monte Carlo for rarified gases) and continuum methods (computational fluid mechanics) to derive a multiscale description of flow phenomena. On the question of validation and verification, he pointed out that this is an especially acute problem for grid-based computing – does an application produce the same results in grid-enabled form as it produces on a serial machine or tightly coupled parallel machine. Results for the grid-enabled version must be tested against serial and/or purely parallel implementations. Another aspect of validation is ensuring that model-based simulations of biomedical and clinical scenarios are “reliable” for use in medical decision-making. With regard to software, 50% of the software used by UCL researchers is developed in-house. Examples include HYPO4D (high Reynolds number fluid flow), LB3D (mesoscale 3-D complex fluid simulations based on lattice Boltzmann methods), and HemeLB (blood flow in the intracranial vasculature). The remainder of the codes used are external, and are a combination of open-source, community/academic and commercial codes. For example, for molecular dynamics, LAMMPS is used for materials simulations with 10M+ atoms), and NAMD, AMBER, CHARMM for biomolecular simulations. The applications housing environment at UCL is that developed under the GNU license by the UK-funded Open Middleware Infrastructure Institute (OMII-UK) project (<http://www.omii.ac.uk>).

In reference to big data and visualization, the UK’s education and research network is JANET (<http://www.ja.net>). Within the overall network (having capacity of 40Gbit/s) there are so-called Lightpaths, 1Gbit/s dedicated links to various resources, such as the TeraGrid and HPCx. The preference and trend at UCL is for in situ visualization (such as in HemeLB, http://wiki.realitygrid.org/wiki/HemeLB_Deployment), combined

with computational steering, rather than post-processing visualization. There are plans for UCL-wide visualization facility, implemented on the 10 Gbit/s UCL-wide infrastructure, to facilitate real-time steering and visualization from Legion. In terms of next generation algorithms and HPC, UCL researchers are, and will continue to be, at the forefront of exploitation of grid computing to solve problems; Coveney is an international leader in this area, having won awards (including a 2005 Supercomputing Conference HPC Challenge Analytics Award and the 2006 International Supercomputing Conference Award in the Life Sciences) for achievements of the grid-enabled application SPICE (Simulated Pore Interactive Computing Environment) in advancing the understanding DNA translocation across membrane-bound protein channel pores in biological cells [<http://www.realitygrid.org/Spice>].

Funding for simulation-based sciences at UCL comes from the Engineering and Physical Sciences Research Council [EPSRC, <http://www.epsrc.ac.uk>], the Biotechnology and Biomedical Sciences Research Council [BBSRC, <http://www.bbsrc.ac.uk>], the Technology Strategy Board [TSB, <http://www.innovateuk.org>], the European Union (EU) and the US NSF. As noted above, one particularly large EU project under UCL leadership is EU FP7: Virtual Physiological Human, with a budget of around €300M over 7 years. At UCL, the impression was that the current funding environment for simulation-based research, particularly in the biological and biomedical sciences, was positive.

In training and education, UCL researchers find that few starting PhD students are adequately prepared for research and careers in computational sciences, and have little or no knowledge of good software engineering practice, etc. UCL participates in a Boston University NSF-funded IGERT grant supporting interdisciplinary training of graduate students in computational science as one of the overseas centre in IGERT participants can choose to work for a 3-month period during their PhD degrees. Problems cited by Coveney were the difficulty of developing and maintaining software beyond the lifetime of project funding (one attempt to do this is the materials database at <http://db.foxd.org>) and the ongoing challenges endemic to grid computing of security, confidentiality, privacy, authentication and authorization balanced against speed and efficiency.

Richard Catlow: UK HPC Materials Chemistry Consortium (HPC-MCC)

In his presentation, Richard Catlow focused on the UK HPC Materials Chemistry Consortium (HPC-MCC). The HPC-MCC consists of 23 faculty members across the UK and their group members, for a total of almost 100 participants. HPC-MCC members employ the latest developments in HPC technologies, in a wide-ranging program of development, optimization, and applications studies aimed at modeling and predicting the structures, properties and reactivities of functional materials, including catalysts, ceramics, minerals and molecular materials. The techniques include large-scale forcefield-based simulations (molecular dynamics and Monte Carlo methods) as well as electronic structure techniques employing density functional theory, Hartree-Fock and hybrid techniques. Computational chemists and chemical engineers find joining the HPC-MCC is a quicker and more efficient route to gaining access to large-scale HPC resources than trying to access those resources directly. This is similar to the process available to materials/chemical/biological simulators in the United States through the Department of Energy (DOE) nanoscience centers. For example, the users of the Nanomaterials Theory Institute (NTI) in the Center for Nanophase Materials Sciences (CNMS) at Oak Ridge National Laboratory (ORNL) have access to the large allocations the NTI-CNMS has on the DOE's capacity and capability HPC platforms. NTI researchers work with users to convert their problems into codes that will run efficiently on the HPC platforms, allowing users to get access quickly to large levels of state-of-the-art computational capabilities. Likewise, new members of the HPC-MCC can draw on the combined experience and expertise of HPC-MCC members in order to use HPC facilities efficiently. Several examples of HPC-MCC members research accomplishments were given, including a study of chlorine adsorption on Ag(111) surfaces (de Leeuw and Nelson 2004) and a study of the active sites in titanium silicate (To et al. 2007). Information about the HPC-MCC, its meetings and its membership is available at its website, <http://www.dfri.ucl.ac.uk/mcc>.

Mike Gillam: History of Computational and Simulation-Based Materials Sciences in the UK

Mike Gillam highlighted the Daresbury Laboratory-based Collaborative Computational Projects (CCPs, <http://www.ccp.ac.uk>) beginning in the 1970's, the national HPC consortia (including the UK Car-Parrinello

consortium, the Materials Chemistry consortium covered in detail by Richard Catlow, and the Mineral Physics consortium), and the national-level HPC resources from the 1990 onwards, culminating in HECToR (<http://www.hector.ac.uk>, scheduled to reach 250 Tflop by October, 2009). Focusing specifically on UCL, he noted that UCL has more than 30 research groups engaged in materials modeling (in the departments of Physics and Astronomy, Earth Sciences, Chemistry, Mechanical Engineering, etc, and the London Centre for Nanotechnology, <http://www.london-nano.com>), coordinated under the Materials Simulation Laboratory led by Gillam. UCL researchers account for ~30% of national computational resource use, and have strong involvement in international grid projects (particularly through the activities of Peter Coveney). Several internationally used codes (such as GUESS, Gaussians used for embedded system studies (Sushko et al. 2000) and CONQUEST, <http://www.conquest.ucl.ac.uk>, a linear scaling density-functional-theory-based electronic structure code) were developed at UCL. Some of the technical issues he outlined that impact current and next-generation HPC algorithms and hardware are *development time*—during the time taken to develop, implement, validate, optimize the algorithms, computer power may have increased by a factor 1000—and *the trend of increasing numbers of processors and number of cores rather than more powerful processors*—this is useful for increasing the size of problems that can be addressed, but does not impact the time-scale problem, which in many ways is the most fundamental and will require new theory. As educational and training challenges, he cited formal education in computational science as being the only way to redress the trends of inefficient HPC codes and the growing “black-box” mentality to simulation codes. At present, he offered the opinion that these issues are not well recognized in the UK funding agencies.

Sally Price: Computational Prediction of Organic Crystal Structures and Polymorphism, and the CPOSS Project

Sally Price described her research on the computational prediction of organic crystal structures and polymorphism and the closely related control and prediction of the organic solid state (CPOSS) project (<http://www.cposs.org.uk>) that she leads. The goal of her work, reviewed recently (Price 2008), is to develop a computational method to predict the crystal structure of a molecule without experimental input, and to predict all (practically important) polymorphs and properties, as aid to experimental discovery and understanding crystallization. One of the important practical applications of polymorph prediction is in the pharmaceutical industry, where variations in crystal polymorph can alter dramatically the bio-availability of a drug. Her methodology is to search systematically for plausible crystal structures (considering 3000-100000 possible crystal structures), using quantum mechanics to predict molecular structure and represent the charge distribution within the molecule, high-quality forcefield models to minimize the lattice energy of each crystal structure, and analyze the structures with the lowest energies (calculate properties and store in database to compare with experiment). The prediction of progesterone crystal structures was given as an example of the methodology and its ability to understand puzzling experimental results. The results are placed in an on-line searchable database. Although the work is scientifically interesting and industrially valuable, challenges include the use of 12 substantial codes (all published, licensed and documented), of which only 2 are in-house, and problems of continuity and stability in funding, computational infrastructure and personnel. Right now the project depends significantly on the goodwill of several contributors, some of whom are retired and contribute out of personal interest.

Maziar Nekovee: Ad Hoc Wi-Fi Networks

The final speaker, Maziar Nekovee, is a senior scientist at British Telecom (BT) and an Industry Fellow of the Royal Society. He is currently on a secondment at UCL's Centre for Computational Sciences, working with Peter Coveney and others. Nekovee is interested in *ad hoc* wi-fi networks, created on the fly by, e.g., laptops interacting with hot spots and each other, and vehicles interacting with each other and traffic control systems. Given that a single wi-fi device is already a complex stochastic state-machine, which is difficult to model, putting these devices together to form networks is creating extremely complex engineering systems that are very large scale (over a million BT hubs, many millions of wi-fi-enabled smart phones, are highly distributed (unlike cellular systems), and may show unpredictable behavior. Hence, simulation-based planning and evaluation is becoming essential, to address issues such as security (e.g. in response to cyber attacks), predictability (reliable performance), mobility (pedestrians, buses, cars, trains) and scalability (interference/radio spectrum). One recent

application, the simulation of worm propagation in *ad hoc* wi-fi networks (Noekovee 2007), was described. Nekovee enumerated some of the computational challenges of these kinds of simulations, including efficient updating of large-scale mobile wireless communication graphs, coupled simulations of vehicular traffic and high-fidelity wireless communications, and parallel simulation and synchronization of discrete event wireless simulations. Many of these challenges have analogues in molecular/materials simulation. E.g., borrowing from the ideas of neighbor lists in molecular dynamics simulation has led to advances in updating large-scale mobile wireless communication graphs; and the coupling of vehicular traffic and wi-fi communications is a multi-timescale problem, with the typical timestep for traffic simulations being 1 second, while the typical timescale of wi-fi wireless simulations is about 1 microsecond. Another challenge is for the combination of algorithmic efficiency and HPC capability to reach the point where simulations run faster than real time, allowing the possibility of control; right now, even for modest-sized systems (640 node wi-fi network), simulations run between one and two orders of magnitude slower than real time.

CONCLUSIONS

Researchers at UCL are engaged in a wide variety of simulation-based scientific and engineering research projects. UCL's high level of simulation-based research activity, its institutional commitments to provide, now and into the future, state-of-the-art capacity computational capabilities to its users with zero cost at the point of use, and its leadership in applications of grid computing, make it an exemplar of simulation-based engineering and science (SBES).

Clearly, UCL's commitment to support SBES research is paying dividends in recruitment of faculty in this area. One example is Richard Catlow, who left the prestigious Royal Institution (RI) in 2007 after 18 years to move permanently to UCL (<http://education.guardian.co.uk/higher/news/story/0,,2063727,00.html>), taking a very large group (30–40 scientists) with him.

The WTEC team was left with the impression that UCL is aggressively positioning itself as a leader in SBES within the UK and internationally, and that the connection to the medical applications of SBES is a significant driving force in the strategic planning at UCL.

REFERENCES

- de Leeuw, N.H., C.J. Nelson, C.R.A. Catlow, P. Sautet, and W. Dong. 2004. Density-functional theory calculations of the adsorption of Cl at perfect and defective Ag(111) surfaces. *Phys. Rev. B*, 69:045419.
- Delgado-Buscalioni, R., P.V. Coveney, G.D. Riley, and R.W. Ford. 2005. Hybrid molecular-continuum fluid models: Implementation within a general coupling framework. *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* 363:1975-1985.
- Lancaster, R.W., P.G. Karamertzanis, A.T. Hulme, D.A. Tocher, D.F. Covey, and S.L. Price. 2006. Racemic progesterone: predicted in silico and produced in the solid state. *Chem Commun*, 4921-4923.
- Lancaster, R.W., P.G. Karamertzanis, A.T. Hulme, D.A. Tocher, T.C. Lewis, and S.L. Price. 2007. The polymorphism of progesterone: Stabilization of a 'disappearing' polymorph by co-crystallization. *Journal of Pharmaceutical Sciences* 96:3419-3431.
- Nekovee, M. 2007. Worm epidemics in wireless *ad hoc* networks. *New J Phys*, 9:189.
- Price, S.L. 2008. From crystal structure prediction to polymorph prediction: Interpreting the crystal energy landscape. *Phys. Chem. Chem. Phys.*, DOI: 10.1039/b719351c [in press].
- Sushko, P.V., A.L. Shluger, and C.R.A. Catlow. 2000. Relative energies of surface and defect states: *Ab initio* calculations for the MgO(001) surface. *Surface Science* 450:153-170.
- To, J., A.A. Sokol, S.A. French, and C.R.A. Catlow. 2007. Formation of active sites in TS-1 by hydrolysis and inversion. *J. Phys. Chem. C* 111:14720-14731.

Site: University of Cambridge Centre for Computational Chemistry
Department of Chemistry
Lensfield Road
Cambridge CB2 1EW, UK
<http://www-theor.ch.cam.ac.uk>

Date Visited: February 27, 2008

WTEC Attendees: M. Head-Gordon (report author), K. Chong, P. Cummings, S. Kim

Hosts: Prof. Daan Frenkel
Email: df246@cam.ac.uk
Prof. Michiel Sprik
Email: ms284@cam.ac.uk
Dr. David Wales
Email: dw34@cam.ac.uk
Dr. Ali Alavi
Email: asa10@cam.ac.uk

BACKGROUND

Theoretical Chemistry at Cambridge traces its origins back to the appointment of Prof. Lennard-Jones in 1933. One of his students, John Pople, was the 1998 Nobel Prize winner in chemistry for developments in electronic structure theory. The present head of the centre, Prof. Frenkel, has just taken over from Prof. Hansen, who had been leader of the sector for the past decade. The Centre presently has 6 academic staff, 2 Royal Society Fellows, a number of distinguished emeritus faculty, and approximately 30 graduate students and postdoctoral fellows. It is housed in contiguous recently renovated space within the main chemistry building.

RESEARCH

The research activities of the academic staff span most of modern theoretical chemistry, with significant overlap into areas of condensed matter physics and biology. Prof. Frenkel is a leading developer and applier of condensed phase molecular simulation methods, with present focus on bio-inspired materials and nucleation phenomena. Prof. Sprik performs condensed phase simulations using *ab initio* molecular dynamics, with present focus on modeling redox chemistry. Dr. Wales is interested broadly in the characterization and exploration of potential energy surfaces, including applications to protein folding and glass formation. Dr. Alavi works on fundamental problems in electronic structure theory, including a novel formulation of quantum Monte Carlo in determinant space. Dr. Althorpe works on quantum dynamical methods and applications, and Dr. Vendruscolo works on computational biology with a focus on protein folding and misfolding.

COMPUTING HARDWARE

Cambridge University has a large computer cluster, with more than 2300 processors, which are available on a recharge basis (\$0.14 per cpu hour). As a result of the recharge system, most researchers within the center utilize their own clusters (some of which have over 200 processors). Some projects are not computer-intensive and just make use of desktop computing resources.

DISCUSSION

The WTEC visit to the Centre was very short, but during the discussion, the following issues concerning present trends in simulation-based engineering and science were raised:

- The Centre is able to recruit first-rate students and postdocs, which is crucial for its continued success. The WTEC team's hosts emphasized that they are far more often limited by brain power in their research than by computer power.
- For those aspects of research that depend upon computing, the team's hosts felt that one area in which they were particularly fortunate at present is having a first-rate computer officer within the center to provide dedicated support.
- There are significant funding challenges at present. To run a larger group requires a "patchwork quilt" of multiple grants and a strong focus on grant writing. The direct support of graduate students by the department has diminished greatly (about a factor of 4 less today than it was a decade ago) to now amount to only about 0.1 student per faculty member.
- Several sector members work on or use large codes, whose size is beyond the ability of small groups to develop or even maintain. This means that close associations are needed with either an active open source development effort or an active commercial effort. The number of British "CCPs" (computational chemistry projects), which have been one mechanism for developing community codes in the UK, seems to be diminishing.
- Some groups produce large amounts of data in their research. At present they take the view that such data is disposable, on at least some timeframe, and therefore do not make any special efforts to manage or maintain it. It can be far more cheaply regenerated in the future if that should be required.

Site: University of Cambridge Dept. of Applied Mathematics and Theoretical Physics (DAMTP)
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WA, UK
<http://www.damtp.cam.ac.uk/>

Date Visited: February 27, 2008.

WTEC Attendees: S. Kim (report author), P. Cummings, M. Head-Gordon, K. Chong

Hosts: Prof. Peter Haynes, Head of Department
Email: phh@damtp.cam.ac.uk
Prof. Raymond Goldstein, Schlumberger Chair in Complex Physical Systems
Email: R.E.Goldstein@damtp.cam.ac.uk
Dr. Stephen Eglén
Email: S.J.Eglen@damtp.cam.ac.uk
Prof. Ron Horgan
Email: R.R.Horgan@damtp.cam.ac.uk
Dr. Paul Shellard
Email: E.P.S.Shellard@damtp.cam.ac.uk
Prof. Arie Iserles
Email: A.Iserles@damtp.cam.ac.uk

BACKGROUND

The Department of Applied Mathematics and Theoretical Physics (DAMTP) at the University of Cambridge was founded as a department in 1959 by George Batchelor, but its lineage from the 17th century remarkably forms the summary of the history of applied mathematics and theoretical physics: Newton, Stokes, Clerk, Maxwell, Rayleigh, Babbage, Eddington, Dirac, G.I. Taylor, Jeffreys, and Lighthill. The first Lucasian Professor was Sir Isaac Newton, and this oldest professorship in mathematics (established 1663) is currently held by Prof. Stephen Hawking. Today, the DAMTP research areas include astrophysics, geophysics, fluid and solid mechanics, mathematical biology, quantum information, high-energy physics, relativity, and cosmology. The SBES theme runs through a common core of activities in applied and computational analysis. These groups are loosely defined; many staff contribute to more than one area as well as participating in a number of collaborative/multidisciplinary activities at the Center for Atmospheric Science, Center for Micromechanics, Institute for Theoretical Geophysics, Institute for Aviation and the Environment, Center for Quantum Computation, Cambridge Computational Biology Institute, COSMOS Supercomputer, BP Institute, Millennium Mathematics Project, and the Cambridge eScience Center. DAMTP staff members are also active users of the High-Performance Computing (HPC) Facility.

The department has 50 academic staff (faculty), 70 research staff, 30 support staff, and 100 PhD students. At the undergraduate level (years 1–4), 900 students are enrolled in the Mathematics Tripos (shared with pure math). DAMTP is also responsible for the mathematics curriculum of 600 students enrolled in the natural sciences. The computational biology program (1-year M.Phil. degree) has 25 students. The department is housed in a spacious grounds and a new set of buildings (completed in 2002) at the Centre for Mathematical Sciences, along with the physical sciences library, the Department of Pure Mathematics and Mathematical Statistics, and the Isaac Newton Institute.

COMPUTING HARDWARE

The operation of the HPC facility is noted in the report on the Cambridge University Centre for Computational Chemistry site visit. But the WTEC visiting team's visit to DAMTP was an opportunity to learn about the history

of the decision-making of the faculty in the cosmology group when they opted to purchase a high-end SGI/Altix SMP (shared memory) machine. The SMP/shared memory architecture has proven to be very popular with users, and time on this machine (COSMOS) is in very high demand.

PRESENTATIONS

After a succinct overview of the department by Prof. Haynes, the bulk of the WTEC team's visit to DAMTP was devoted by presentations representing the SBES activities of several selected but representative research groups. Brief summaries of and comments on those presentations follow.

Prof. R. E. Goldstein: *Research in Complex and Biological Systems*

These research activities continue the department's strong tradition in fluid mechanics of biological systems, most notably the legacy of Sir James Lighthill. The WTEC team saw experimental, computational, and theoretical elastohydrodynamic studies of microorganism swimming, but also learned of new efforts in elucidating the evolutionary transition to multicellularity. Free-boundary problems in the precipitative growth of stalactites represented current research in multiscale physics. The use of a state-of-the art experimental apparatus in microscopy and micromanipulation represented a significant but welcome break from the well-known DAMTP tradition of fluid mechanics experiments on a shoestring budget.

Prof. S. J. Eglén: *Mathematical Biology*

This presentation discussed the current research themes of this group: computational molecular biology, cancer biology, gene expression (data generation, storage, analysis), disease dynamics, and computational neuroscience, as well as a new education initiative, the M.Phil. in computational biology. The most important quest of the "post genomic" era ("sequencing of the human genome was the easy part") is understanding the organizational principles or how the biological parts work together. The presentation showed the visualization of microarray data (provided by Prof. Simon Tavaré and illustrating the multidisciplinary collaboration with his Cambridge Computational Biology Institute), e.g., to examine molecular signals for cancer. In computational neuroscience, Prof. Eglén described the Blue Brain project, a joint effort between IBM and Brain Mind Institute. This project hopes to elucidate the workings of the brain with computer models running on a BlueGene/L. The immediate task is the simulation of the workings of one neocortical column, which involves simulating 100,000 neurons. A million-fold increase in computational power can be applied towards the buildup of a model of the entire brain. The presentation then turned to some activities of the EPSRC's "life sciences initiative" and the pilot e-Science grant CARMEN (code analysis, repository, and modeling for e-Neurosciences, at UK£4.5 million). The UK also contributes UK£500,000 over three years to the International Neuroinformatics Coordinating Facility (INCF). Prof. Eglén concluded with a list of major issues facing SBES in biology: the diversity and (un)reliability of experimental data, choice of model organisms, difficulty and cost of code maintenance for large scale simulations, and the need for improved visualization techniques.

Prof. R. K. Horgan: (High-Energy Physics) *Computing Lattice Gauge Field Theory*

(This report author makes the following observations on the topic.) Lattice-QCD computations in particle physics involve multidimensional integrals after inversion of poorly conditioned matrices, and slowly converging sums. The computational resolutions of physical interest take on the order of 1000 years at Tera-FLOP speeds, and thus to no surprise, members of this community show up on the "top ten users" lists for HPC facilities. The pending transition to petascale computing is of great importance to this community: results obtained on the order of one year for current targets in lattice resolution scales. The experiences of this community are of importance to our SBES study as the numerical methods share common algorithmic roots with molecular simulations for material science research.

Dr. E. P. S. Shellard: *Relativity and Cosmology*

(This report author makes the following observations on the topic.) A major goal of cosmology is to explain the present state of the universe in terms of the processes that happened early in its history. The “Big Bang” model, well known to the general public thanks to science outreach efforts, makes a number of predictions that can be tested with a combination of astronomical observations and computer simulations. Although these SBES efforts are far removed from the three primary thrusts of the WTEC SBES study, significant advances in HPC hardware and algorithms are motivated and generated by this community. A good part of the presentation and subsequent discussion focused on recent experiences (at Cambridge and beyond) with shared memory SMP in the form of a significant investment in the SGI Altix (COSMOS). The COSMOS facility was founded in 1997 and evolved through several generations of supercomputers, with the most recent upgrade being to the Altix in 2003. These experiences are relevant to possible future directions for the “Track 2” and “Track 3” machines of the NSF Office of Cyberinfrastructure HPC roadmap.

Prof. A. Iserles: *Applied and Computational Analysis*

This research group complements the strong tradition in DAMTP of constructing solution methods for specific computational problems that arise from applications of mathematics with the more general machinery of numerical analysis, i.e., investigation of algorithms for fundamental mathematical calculations that are common across many applications and theoretical work on mathematical issues that are vital to the success of algorithms. In the discussion on emerging trends relevant to SBES, Prof. Iserles noted the increasing importance of understanding highly oscillatory solutions to PDEs.

CONCLUSIONS

Historically, DAMTP has possessed a singular tradition of excellence in applied mathematics that extends to its SBES applications. DAMTP today has not lost any of its traditional strengths but has gained new capabilities in the form of multidisciplinary collaborations with new institutes in the mathematical, physical, and biological sciences on the University of Cambridge campus, strong partnerships with leading multinational companies (BP, Rolls Royce, Schlumberger, and others), and leadership roles in EU framework projects. At the risk of wearing out the term “multiscale,” it must be mentioned that DAMTP has state-of-the-art facilities at multiple scales: from the new building and grounds to experimental apparatus and instruments in the laboratory. From our past observations and expectations that algorithmic advances are perhaps even more important than hardware advances in the push to multicore petascale computing, DAMTP can be expected to play an important role in shaping the SBES landscape in the petascale era.

REFERENCES

- Department of Applied Mathematics and Theoretical Physics. 2005. *Department Guide*. Cambridge: DAMTP.
- Cosmos website. N.d. <http://www.damtp.cam.ac.uk/cosmos>.
- Markram, H. 2006. The Blue Brain project. *Nature Reviews Neuroscience* 7:153-160.

Site: University of Cambridge Theory of Condensed Matter Group
Cavendish Laboratory
J.J. Thomson Avenue
Cambridge CB3 0HE, UK
<http://www.tcm.phy.cam.ac.uk/>

Date Visited: February 26, 2008

WTEC Attendees: P. Cummings (report author), K. Chong, M. Head-Gordon, S. Kim

Hosts: Professor Mike C. Payne, Head, Theory of Condensed Matter Group
Email: mcp1@cam.ac.uk
Professor Richard J. Needs, Theory of Condensed Matter Group
Email: rn11@cam.ac.uk
Website: <http://www.tcm.phy.cam.ac.uk/~rn11/>

BACKGROUND

The Theory of Condensed Matter Group is housed in the Cavendish Laboratory, home of the Physics Department, at the University of Cambridge. The focus of the group is to predict the properties of materials from first principles—i.e., in one form or another, solving the Schrödinger equation. The group members refine and develop new calculational tools and apply them to problems in physics, chemistry, materials science, and biology. Several widely used codes, including the first-principles total energy pseudopotential code CASTEP (Cambridge Sequential Total Energy Package, <http://www.castep.org>), have originated within this group.

CASTEP is an interesting case in point with respect to code development in the European environment. Because it is developed within the UK by academic and government laboratory employees (at Daresbury Laboratory), it is available free of charge to UK academics. For other users, CASTEP is licensed to Accelrys Software, Inc. (<http://www.accelrys.com>), which sells CASTEP as part of Materials Studio. This means that CASTEP, which is free to academic and government laboratory users in the UK, costs real money for U.S. academic researchers.

The WTEC delegation met with two members of the Theory of Condensed Matter group, Professors Mike Payne and Richard Needs. Mike Payne is currently Head of Theory of the Condensed Matter Group and has worked on first-principles total energy calculations since 1985; he is the principle author of CASTEP. He was awarded the 1996 Maxwell Medal and Prize by the Institute of Physics and gave the 1998 Mott Lecture. He is responsible for many of the technical developments that have led to the widespread adoption of the total energy pseudopotential technique and has pioneered the application of this technique to a wide range of scientific problems in physics, chemistry, materials science, earth sciences and, most recently, biology.

Richard Needs has worked on a wide range of complex systems such as surfaces, interfaces, defects, and clusters, mainly studying structural properties, including phase transitions and excitation energies. He has used a variety of computational techniques, including density functional theory methods, many-body perturbation theory, and quantum Monte Carlo methods. In recent years he has been developing continuum fermion quantum Monte Carlo methods and applying them to problems in condensed matter. He and his group have developed the CASINO quantum Monte Carlo code that is now used in a number of groups around the world.

R&D ACTIVITIES

Richard Needs described his main research focuses as solving the many-body Schrödinger equation using statistical methods, predicting and understanding the optoelectronic and physical properties of nanoparticles and solids, applications to ultracold atoms, and development of the CASINO quantum Monte Carlo code (<http://www.tcm.phy.cam.ac.uk/~mdt26/casino2.html>). Quantum Monte Carlo methods are stochastic wave-function-based approaches that include direct treatment of quantum many-body effects. They are computationally more demanding than density-functional theory (DFT) approaches; however, they serve as benchmarks against

which the accuracy of other techniques may be compared (Foulkes et al. 2001). CASINO is available free of charge to academic and nonprofit users. One of the recent focuses of the Needs group, highlighted in the presentation to the WTEC team, has been using DFT methods and random search to discover new structures for materials at high pressure, including silane (SiH_4) and hydrogen (Pickard and Needs 2006, 2007). Some of these structures have now been found experimentally.

Mike Payne focused on the program ONETEP (Order-N Electronic Total Energy Package) (Haynes et al. 2006) (<http://www.onetep.soton.ac.uk>), which he is developing with coworkers Peter Haynes, Chris-Kriton Skylaris, and Arash A. Mostofi. ONETEP achieves order-N scaling (i.e., computational cost that scales linearly, $O(N^1)$ in the number of atoms) through the use of non-orthogonal generalized Wannier functions to represent electron density instead of orthogonal extended wavefunctions used in molecular orbital methods. DFT methods typically scale as the cube of the system size, $O(N^3)$, and molecular orbital methods as $O(N^5)$ – $O(N^7)$. An example of calculating the ground state structure of DNA exhibited linear scaling in the number of atoms in the DNA; other examples are provided by Skylaris et al. (2008).

The other research area highlighted by Payne was multiscale modeling of dynamics of materials, specifically crack propagation in graphine sheets. The technique used is called learn on the fly (LOTF) (Csanyi et al. 2004, 2005). Rather than construct a new hybrid Hamiltonian that combines different models (as is frequently done in other hybrid approaches that combine first-principles and atomistic molecular dynamics simulation), the LOTF method uses a unique short-ranged classical potential parameters whose parameters are continuously tuned to reproduce the atomic trajectories at the prescribed level of accuracy throughout the system.

CONCLUSIONS

The Theory of Condensed Matter Group is clearly among the forefront groups in materials simulation, both within the UK and internationally, with several signature codes to its credit (CASTEP, ONETEP, and CASINO).

When asked what were the conditions that have made it possible in the UK (and Europe) to develop these codes while this has not happened in the United States, the surprising opinion offered was that the traditional scarcity of resources within the UK and Europe has made it almost obligatory for groups to collaborate on code development. Collaboration of this nature is endemic in the European system and is actively funded. By contrast, in the U.S. materials modeling community, there is much pressure within promotion and tenure processes to actively discourage collaboration until a faculty member reaches a senior level, by which time the U.S. faculty member has to consider the future careers of his/her students and post-doctoral researchers, again mitigating against collaboration.

Looking to the future, Payne and Needs did not view the number of people coming into the field as big an issue of concern as the declining programming skills of incoming graduate students, being addressed to some degree by the educational programs at centers for scientific computing established at the universities of Warwick (<http://www2.warwick.ac.uk>) and Edinburgh (<http://www.epcc.ed.ac.uk>), and by workshops hosted by the Collaborative Computational Projects (CCPs, <http://www.ccp.ac.uk>) at Daresbury Laboratory. Their view is that the Engineering and Physical Sciences Research Council (<http://www.epsrc.ac.uk>) is no longer funding code development; in particular, the CCPs are now funded only for code maintenance and training, not for code development. A long-term problem for the development of materials modeling codes is that the expectation of funding agencies and university administrators is for “sound-bite science,” publishable in the highest-profile journals. Software engineering is incremental by nature, and these “sound-bite science” expectations mitigate against it. Finally, it is too early to tell whether the reorganization of the UK science laboratories under the new Science and Technology Facilities Council (<http://www.scitech.ac.uk/>) will be beneficial for the development of materials modeling codes.

In summary, Payne and Needs see future challenges for continuing efforts to build large codes, unless there is broad recognition of the value that these codes bring to the scientific enterprise.

REFERENCES

- Csanyi, G., T. Albaret, M.C. Payne, and A. De Vita. 2004. "Learn on the fly": A hybrid classical and quantum-mechanical molecular dynamics simulation. *Physical Review Letters* 93:4.
- Csanyi, G., T. Albaret, G. Moras, M.C. Payne, and A. De Vita. 2005. Multiscale hybrid simulation methods for material systems. *Journal of Physics-Condensed Matter* 17:R691-R703.
- Foulkes, W., L. Mitas, R. Needs, and G. Rajagopal. 2001. Quantum Monte Carlo simulations of solids. *Reviews of Modern Physics* 73:33-83.
- Haynes, P.D., C. Skylaris, A.A. Mostofi, and M.C. Payne. 2006. ONETEP: linear-scaling density-functional theory with local orbitals and plane waves. *Phys Status Solidi B* 243:2489-2499.
- Pickard, C.J., and R.J. Needs. 2006. High-pressure phases of silane. *Physical Review Letters* 97 Art. No. 045504.
- . 2007. Structure of phase III of solid hydrogen. *Nature Physics* 3:473-476.
- Skylaris, C., P.D. Haynes, A.A. Mostofi, and M.C. Payne. 2008. Recent progress in linear-scaling density functional calculations with plane waves and pseudopotentials: the ONETEP code. *Journal of Physics-Condensed Matter* 20:064209.

- Site:** Universität Karlsruhe (TH) and
Forschungszentrum Karlsruhe (Karlsruhe Research Center),
Karlsruhe Institute of Technology (KIT), and other affiliated institutes
Kaiserstrasse 12
D-76131 Karlsruhe, Germany
<http://www.kit.edu>
<http://www.fzk.de/fzk/ideplg?IdcService=FZK&lang=en>
<http://www.ipc.uni-karlsruhe.de/english/>
<http://www.mathematik.uni-karlsruhe.de/ianm/de>
<http://www.rz.uni-karlsruhe.de/rz/scc/>
- Date Visited:** February 25, 2008
- WTEC Attendees:** M. Head-Gordon (report author), C. Sagui, P. Westmoreland
- Hosts:** Prof. Dr. Wim Klopper, Institute of Physical Chemistry, Universität Karlsruhe (TH)
Email: willem.klopper@kit.edu
Prof. Dr. Reinhart Ahlrichs, Institute of Physical Chemistry, Universität Karlsruhe (TH);
Email: reinhart.ahlrichs@chemie.uni-karlsruhe.de
Prof. Dr. Matthias Olzmann, Institute of Physical Chemistry, Universität Karlsruhe (TH);
Email: matthias.olzmann@chemie.uni-karlsruhe.de
Prof. Dr. Henning Bockhorn, Institute for Chemical Technology and Polymer Chemistry,
and Engler Bunte Institute, Universität Karlsruhe (TH)
Email: bockhorn@ict.uni-karlsruhe.de
Priv.-Doz. Dr. Wolfgang Wenzel, Institute of Nanotechnology, Forschungszentrum
Karlsruhe; Email: wolfgang.wenzel@int.fzk.de
Prof. Dr. Christian Wieners, Institute for Applied and Numerical Mathematics, Universität
Karlsruhe (TH); Email: wieners@math.uni-karlsruhe.de
Prof. Dr. Willy Dorfler, Institute for Applied and Numerical Mathematics, Universität
Karlsruhe (TH); Email: willy.doerfler@math.uka.de
Prof. Dr. Vincent Heuveline, Institute for Applied and Numerical Mathematics, Universität
Karlsruhe (TH), and Steinbuch Center for Computing
Email: vincent.heuveline@rz.uni-karlsruhe.de

BACKGROUND

Universität Karlsruhe (TH)²² is one of Germany's leading technical universities, a position recognized by its being awarded one of the first three "Excellence Initiative" awards in 2006. These prestigious awards are designed to permit a small number of leading German universities to more effectively compete against leading international rivals. A key part of reorganizing for the Excellence Initiative is the merging of the Universität Karlsruhe (TH) with the nearby national research laboratory, Forschungszentrum Karlsruhe, to create the Karlsruhe Institute of Technology (KIT). The combination of their large computer centers has created the largest high-performance computing center in Europe, the Steinbuch Center for Computing (SCC), with over 200 staff. New research opportunities are being explored in areas of science and engineering that are of industrial and societal importance, including nanoscience, energy research, and biology, including positions that are jointly funded by private industry. This is a time of significant change at Karlsruhe.

²² TH means "Technische Hochschule" (technical university); it is part of the proper name of Universität Karlsruhe.

RESEARCH

During a 5-hour minisymposium on simulation-based science and engineering, the visiting WTEC panel heard seven technical presentations on relevant research at Karlsruhe, as well as a short overview of the new organization of KIT from Prof. Klopfer. Short summaries of these presentations are given below.

- *Accurate simulation of electronic structure.* Prof. Klopfer described new developments in highly accurate electronic structure methods, using bielectronic basis functions in addition to the usual 1-particle expansions. He also discussed how these accurate but computationally expensive approaches could be integrated with simpler and more approximate models to treat multiscale problems such as the interaction of methanol with carbon nanotubes, and solid C₅₈ films.
- *Prediction of gas phase rate coefficients from first principles.* Prof. Olzmann reviewed the present state of the art in unimolecular, bimolecular, and complex-forming bimolecular reaction rate theory. He also discussed future needs such as further improvements in statistical rate theories, advances in the treatment of nonadiabatic effects on kinetics, the treatment of very high temperatures, and the development of an integrated universal kinetics program suite with standard interfaces to electronic structure programs for potential energy surface information.
- *Numerical simulation on multiple scales in chemical engineering and combustion.* Prof. Bockhorn discussed problems that involve interacting length scales ranging from 10⁻¹⁰ m to 10² m, such as propagation of reaction zones involving the interaction of diffusive transport with radiation and chemistry, the functioning of a catalytic converter, and the modeling of a technical burner in gas turbines. These are all cases where true multiscale approaches are essential to obtaining physically realistic modeling.
- *Simulation of nanoscale pattern formation.* Priv.-Doz Wenzel discussed problems in which long time scales are important, ranging from protein folding, *in silico* screening for promising drug leads, modeling of new light-emitting devices based on amorphous thin films, and the pursuit of a transistor that operates at the atomic level.
- *Mathematical modeling and scientific computing.* Prof. Wieners discussed activities underway in both research and education at the Institute for Scientific Computing and Mathematical Modeling, which seeks to bring students with mathematical training towards applications in the biological and physical computational sciences. Examples included dynamics of the electrochemical double layer for colloidal particles, discrete methods for cellular reactions in systems biology, and porous media models for biological soft tissues, such as the human spine.
- *Analysis, simulation, and design of nanotechnological processes.* Prof. Dörfler described research and training that bridges mathematics and nanoscience and technology, including issues such as optimization of the band gap in photonic crystals, and he gave an overview of the mathematical modeling used.
- *Research projects in high-performance computing (HPC).* Prof. Heuveline briefly introduced a wide range of high-performance computing projects underway at the Steinbuch Center for Computing (SCC) (see below). These range from shape optimization (ship hull, keel, etc.), to weather forecasting, to respiratory tract simulation. Some of these projects have specific software packages as outcomes, which are generally open source codes.

COMPUTING HARDWARE

High-performance computing at Karlsruhe is very strong, as the Steinbuch Center for Computing (SCC) is Europe's largest with more than 200 personnel. The SCC was formed by the merger of the Universität Karlsruhe (TH) and Forschungszentrum computer centers, as part of the overall merger to form the Karlsruhe Institute of Technology. The SCC aims to provide both research and development (under the direction of four professors representing numerical methods, networks, parallel and distributed computing, and computer systems and information processing), and information technology services under one roof. Their hardware resources are currently ranked third in Germany, with a 15.6 TFlop machine containing 3000 CPUs and 12 TB of memory. SCC is a Tier 1 center for repository of data generated by the high-energy physics experiments performed at CERN.

DISCUSSION

During the talks, as well as at lunch where we were joined by the President of the Universität Karlsruhe (TH), Prof. Horst Hippler, there were some opportunities for discussion of both scientific issues directly related to the content of the talks, and also of more general issues surrounding simulation-based engineering and science. There was also a small amount of time for some general discussion at the end of the minisymposium. Some of the additional issues that came up included:

- Karlsruhe faculty members are enjoying very good resources and research opportunities due to the key achievement of Centre of Excellence status.
- Computation and theory in chemistry are also an integral part of a Cluster of Excellence on Functional Nanostructures, which involves a total of 49 professors and group leaders and more than 70 projects, mostly funded by the DFG (German Research Foundation). This cluster takes pure science towards practical applications with a focus on optical, electronic, and biological properties at the nanoscale.
- There are new mechanisms for interaction between industry and KIT. Karlsruhe has begun to appoint faculty members who are also employees of leading local industry. These appointments involve arrangements for periodic absences from the university to permit the management of dual responsibilities.
- Multimethod/multiscale applications at the nanoscale inevitably involve hierarchies of methods operating at differing levels of accuracy. Inexpensive methods such as empirical potentials or tight-binding molecular dynamics can be used for preliminary exploration of potential surfaces followed by refinement with density functional theory, and in turn, still more accurate explicitly correlated methods. This demands a strong emphasis on validation of the low-level results, which in turn requires considerable experience and chemical knowledge in order to be accomplished effectively.
- It was noted that within the Institute for Scientific Computing and Mathematical Modeling there are several issues involved in more effectively training students at the interface of mathematics and physical applications. It was felt that math students need more guidance in applications in the future, which might be provided by, for example, a competence center in modeling and simulation.
- The interplay between high-end supercomputing and mid-range cluster computing was discussed. The perceived benefits vary strongly between fields and even between groups. For example, speaking in favor of mid-range computing, one KIT professor commented that “supercomputers are important, but most of science gets along without them.” To illustrate one instance in slightly greater detail, the quantum chemistry group at KIT has roughly 300 high-end server CPUs in its own cluster. At the university, more than 450 CPUs in total are dedicated to this one field alone—24 hours a day.

Site: **University of Oxford Condensed Matter Theory Group**
Rudolf Peierls Centre for Theoretical Physics
1 Keble Road
Oxford OX1 3NP, UK
<http://www-thphys.physics.ox.ac.uk/user/CondensedMatter/index.php>

Date Visited: February 26, 2008

WTEC Attendees: P. Cummings (report author), M. Head-Gordon, S. Kim, K. Chong

Hosts: Prof. Julia Yeomans, Professor of Physics
 Pauline Chan Fellow and Tutor in Physics, St. Hilda's College
 Email: j.yeomans1@physics.ox.ac.uk
 Website: <http://www-thphys.physics.ox.ac.uk/user/JuliaYeomans/>
 Dr. Ard Louis, Royal Society University Research Fellow
 Email: ard.louis@physics.ox.ac.uk
 Website: <http://www-thphys.physics.ox.ac.uk/user/ArdLouis/>

BACKGROUND

Julia Yeomans and Ard Louis are both faculty members in the Condensed Matter Theory Group, which is part of the Rudolf Peierls Centre for Theoretical Physics (<http://www-thphys.physics.ox.ac.uk>), a subdepartment of the Oxford University Physics Department (<http://www.physics.ox.ac.uk>). The other departments in theoretical physics are theoretical astrophysics and elementary particle theory

Prof. Yeomans and Dr. Louis are both theorists and simulators working on soft condensed matter; the other members of the condensed matter theory group work on the statistical mechanics of lattice systems (Abraham) and spin glasses (Sherrington), and on strongly correlated electron systems (Essler, Chalker) and field theory in condensed matter (Cardy). The research that Yeomans and Louis perform is unusual. It involves coarse-grained modeling of biological, polymeric, and flowing systems, focusing on the fundamentals of self-assembly, rheology, and microfluidics.

In the WTEC visiting team's discussions with our hosts, questions arose about the funding of the work described by Yeomans and Louis. The biological work, in particular, seemed too "physics"-based and fundamental to be funded by the Biotechnology and Biological Sciences Research Council (BBSRC) or the Medical Research Council (MRC). Louis expressed the opinion that in general the funding for biophysics in the UK is not sufficient. However, it turns out that the students that work for Yeomans and Louis are funded primarily through the Theoretical Physics Department's Doctoral Training Account (DTA) (<http://www.epsrc.ac.uk/PostgraduateTraining/DoctoralTrainingAccounts/default.htm>). The DTA is a source of funding for graduate students allocated annually to each department by the Engineering and Physical Sciences Research Council (EPSRC) in proportion to the external funding received by that department from the EPSRC.

Some faculty members in the Theoretical Physics department prefer primarily to work with post-docs, and grant income is currently buoyant; as a result, there is a fairly large pool of DTA funding available for graduate students—enough, in the case of Yeomans and Louis, to be able to take up one new student per year on DTA funds. This contrasted with most departments the WTEC team visited, where generally the DTA funds were dismissed as being insufficient to have any real impact on the funding of research—that perhaps once every five years a researcher might receive DTA funding for a student. Hence, it is clear that the unusual situation in Theoretical Physics at Oxford is enabling Yeomans and Louis to pursue lines of research that they might otherwise not be able to follow.

R&D ACTIVITIES

Yeomans and Louis both gave presentations about their research activities. Louis described research focused on biological self-assembly, biological networks, and dynamics of complex fluids. He is interested in minimal models that exhibit self-assembly similar to biological systems, specifically virus capsids and DNA tetrahedra. The model for virus capsid self-assembly is a minimal model, similar to those studied previously by Louis and co-workers (Wilber et al. 2007). Louis is additionally working on a minimal coarse-grained model for single strands of DNA that exhibits self-assembly into double-stranded DNA. Louis also indicated an interest in understanding biological networks and their evolution. Finally, he showed various simulations of complex fluid phenomena, including the impact of hydrodynamic forces on the sedimentation of Brownian particles, and modeling of peristaltic pumps created by colloidal spheres (Terray et al. 2002). Louis holds a Royal Society University Research Fellowship, which means that he has a prefaculty position funded by the Royal Society. Royal Society University Research Fellowships provide partial salary support and research support for five years (with a possible extension for another three years (and possibly two more), with the expectation that the fellows will be strong candidates for permanent posts in the host universities at the end of their fellowships. (Louis has a permanent lectureship post in Physics at the end of his fellowship.) Approximately 30 such fellowships are awarded per year.

Yeomans presented simulation-based research on four topics: (1) drops on patterned surfaces (e.g., see Kusumaatmaja and Yeomans 2007), (2) liquid crystal hydrodynamics (e.g., Marenduzzo et al. 2003), (3) low Reynolds number swimmers (Pooley et al 2008), and (4) polymer hydrodynamics (e.g., Ali and Yeomans 2005). In the first research area, for example, her interest is in the shape and motion of liquid water drops on patterned hydrophobic surfaces using a combination of Cahn-Hilliard-type free energy models and Navier-Stokes equations solved using the lattice Boltzmann method. Both analytic theory and simulations are used. In the third area of research, inspired by bacteria (which are the ultimate in low-Reynolds-numbers swimmers), she is investigating minimal models of swimmers that achieve motion by exploiting hydrodynamic forces. In all of the research presented, an elegant combination of minimal model(s), macroscopic/hydrodynamic theory, and mesoscale simulation methods are employed.

CONCLUSIONS

The two researchers Yeomans and Louis presented a wide range of simulation-based biophysics and soft materials research. On the issue of computing, their view was that the most important resource was to have local, powerful computers. All of the codes used in their respective groups are written within the groups; no community-based or commercial codes are used in either group.

REFERENCES

- Ali, I., D. Marenduzzo, and J.M. Yeomans. 2006. Polymer packaging and ejection in viral capsids: Shape matters, I. *Phys. Rev. Lett.* 96:208102.
- Kusumaatmaja, H., and J.M. Yeomans. 2007. Modeling contact angle hysteresis on chemically patterned and superhydrophobic surfaces. *Langmuir* 23:6019-6032.
- Marenduzzo, D., E. Orlandini, and J.M. Yeomans. 2003. Rheology of distorted nematic liquid crystals. *Europhysics Letters* 64:406-412.
- Pooley, C.M., G.P. Alexander, and J.M. Yeomans. 2008. Hydrodynamic interaction between two swimmers at low Reynolds number. *Phys. Rev. Lett.* 99: 228103.
- Terray, A., J. Oakey, and D.W.M. Marr. 2002. Microfluidic control using colloidal devices. *Science* 296:1841-1844.
- Wilber, A.W., J.P.K. Doye, A.A. Louis, E.G. Noya, M.A. Miller, and P. Wong. 2007. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *Journal of Chemical Physics* 127, Art. no. 085106.

Site: University of Oxford Department of Engineering Science
Parks Road
Oxford, OX1 3PJ, UK
<http://www.eng.ox.ac.uk/>

Date Visited: February 26, 2008

WTEC Attendees: K. Chong (report author), P. Cummings, M. Head-Gordon, S. Kim, P. Roe

Hosts: Prof. Alistair Borthwick, Environmental Flow Simulation
Email: alistair.borthwick@eng.ox.ac.uk
Prof. Rodney Eatock Taylor, Offshore Engineering, Structures, Marine CFD
Email: r.eatocktaylor@eng.ox.ac.uk
Dr. Janet Smart, Manufacturing Systems, Major Programmes, Networks
Email: janet.smart@sbs.ox.ac.uk
Dr. Viannis Ventikos, Micro-, Nano-, Bio- Transport Phenomena
Prof. Paul Taylor, Offshore and Coastal Engineering, Reliability/Risk Analysis
Email: paul.taylor@eng.ox.ac.uk
Prof. Li He, Turbomachinery, CFD
Email: li.he@eng.ox.ac.uk

BACKGROUND

The Department of Engineering Science at Oxford is the only unified department in the UK that offers accredited courses in all the major branches of engineering. Its students develop a broad view of the subject much appreciated by employers, but they can also choose from a very wide range of specialist options.

Every year the Department of Engineering Science, one of the largest departments in the university, produces around 160 new engineering graduates. They go off to a huge variety of occupations: designing cars, building roads and bridges, developing new electronic devices, manufacturing pharmaceuticals, healthcare and aerospace, into further study for higher degrees, and in many other directions. Some graduates also develop their managerial, financial, or entrepreneurial skills and go into commerce, financial services, or start their own companies. Each year 60 to 70 students take higher degrees, either MSc or PhD by research, and since October 2006, a number take an MSc course in Biomedical Engineering.

Oxford's Department of Engineering Science has a substantial research portfolio, including much that is directly supported by industry. In the department there are no barriers between the different branches of engineering, and they are involved in a great deal of multidisciplinary research, collaborating with groups in other departments from Archaeology to Zoology.

This broad view of engineering, based on a scientific approach to the fundamentals, is part of the tradition that started with the department's foundation in 1908—one hundred years of educating great engineers and researching at the cutting edge.

There are opportunities in the Department of Engineering Science for postgraduate study and research over a wide range of engineering disciplines, including civil, electrical, mechanical, process, biomedical, and information engineering. This breadth of expertise makes it possible to apply a multidisciplinary approach to the solution of engineering problems, as appropriate.

At present there are approximately 220 postgraduate students, about 70 academic staff, and over 100 post-doctoral research workers, supported by 70 technicians. The department works in close collaboration with many

government and industrial laboratories, and 40 percent of the Department's research funding comes from industry.

RESEARCH

An indication of Department of Engineering Science research areas is provided in the list of its graduate degrees below; further details are available online at <http://www.eng.ox.ac.uk> and <http://www.admin.ox.ac.uk/postgraduate/caz/engi.shtml>.

- Chemical and Process Engineering and Biotechnology
 - Bioprocessing and Tissue Engineering
 - Environmental Biotechnology
 - Sustainable Development
 - Liquid Surfaces and Foam
 - Multiphase Flow and Boiling
 - Combustion in Internal Combustion Engines
 - Cryogenic Engineering and Vacuum Technology
- Civil Engineering
 - *In situ* Testing
 - Soft Soils and Environmental Applications
 - Unsaturated and Gassy Soils
 - Offshore Foundations and Sea-bed Soil Mechanics
 - Reinforced Soil
 - Dynamics of Civil Engineering Structures
 - Tunnelling and Settlement Damage
 - Theoretical Modelling of Soils
 - Structural Glass
 - Deployable Structures
- Electrical Engineering
 - Communications and Sensors
 - Functional Materials
 - Imaging and Displays
 - Pulsed Power Technology, Plasmas and Machines
 - Neural Networks
 - Analogue Microelectronics
- Information Engineering
 - Control and System Dynamics
 - Advanced Instrumentation
 - Robotics and Sensor Systems
 - Medical Image Understanding
 - Machine Vision, Inspection and Visualisation
 - Production Engineering; Complexity in Manufacturing

- Mechanical Engineering
 - Turbomachinery
 - Hypersonics and Low Density Flows
 - Aerodynamics, Heat Transfer and Measurement
 - Ocean Engineering
 - Coastal Engineering
 - Water Resources Engineering
 - Vibrations and Condition Monitoring
 - Contact Stresses, Fatigue, Cracking and Plasticity
 - Structural Strength under Impact
 - Theory of Structures and Structural Integrity
 - Advanced Materials Engineering
- Biomedical Engineering
 - Gene and Drug Delivery Research
 - Orthopaedic Engineering
 - Ophthalmic Engineering
 - Medical Signal Processing
 - Medical Image Understanding

R&D ACTIVITIES AND DISCUSSIONS

Prof. Alistair Borthwick led the presentations to the WTEC visiting team. He indicated that Oxford's Department of Engineering Science teaches computational methods to undergraduates. It usually takes about 6 months to train graduate students to use simulation and modeling software (typically by their supervisors). They have excellent computer lab facilities; students take the lab course (e.g., MATLAB) for one term, about 90 minutes per week. The department needs more next-generation computing capabilities. About two-thirds of the students are trained in computer imaging, especially in the bio-med areas. Overseas students are mostly from China and India.

The following is a summary of the presentations and discussions.

Alistair Borthwick

- River eco-sustainability: low flow and floods in Yellow River
- Simulation of dyke break, solitary waves; some collaboration with Peking University and Delft
- Urban flooding—Thamesmead
- Chaotic advection; lead-zinc mines in Canada
- Tsunami simulation
- Waves up the beach
- Ice mechanics

Rodney Eatock Taylor

- Interactions with offshore structures (Question: Ekofisk offshore design in North Sea)
- Wave diffraction: develop software – DIFFRACT
- Resonances in gaps between vessels, liquid natural gas tanker mechanics.
- Wave energy

- Fluid structure interaction (BP-funded)

Paul Taylor

- Coastal engineering
- Wave hitting a structure; can go up to 35 meters
- Offshore platforms (with BP)
- Extreme waves on deep water
- Raleigh tail study
- Energy and physics studies
- Boussinesq wave modeling

Janet Smart

Simulation and analysis of various kinds of network systems:

- Power grid
- NY garment industry
- Data by UNITE, 7000 firms to 300 firms
- Defects spread, affected by 40% of network
- Simulated 1000 times for random input; same result
- Engineering networks simulations: computers/telecoms, electricity, etc; fungal networks vs. minimum spanning tree
- Manufacturing systems simulation: reconfigurable, make to order, cycle time, etc.

Li He

- Computational aerodynamics and heat transfer (funded by Rolls Royce)
- Fluid-structure interaction (flutter and forced vibration), (funded by Siemens, Alstom, EC)
- Adjoint shape optimization (funded by Siemens)
- Computational aero-acoustics
- Linear/Nonlinear CFD method development

Yiannis Ventikos

- Biomechanics
- Clinical driven collaboration with Princeton, Harvard, and others in the United States
- Patient-specific aneurysms; modeling and curing
- Micro- and nano-fluid mechanics and technologies

REPRESENTATIVE PUBLICATIONS

A.G.L. Borthwick

Rogers, B.D., A.G.L. Borthwick, and P.H. Taylor. 2003. Mathematical balancing of flux gradient and source terms prior to using Roe's approximate Riemann solver. *J. Computational Physics* 192(2):422–451.

Liang, Q., A.G.L. Borthwick, and G. Stelling. 2004. Simulation of dam- and dyke-break hydrodynamics on dynamically adaptive quadtree grids. *Int. J. for Num. Methods in Fluids*, 46:127–162.

- Borthwick, A.G.L., M. Ford, B.P. Weston, P.H. Taylor, and P.K. Stansby. 2006. Solitary wave transformation, breaking and run-up at a beach. *J. Maritime Engineering* 159(3):97–105.
- Ni, J.R., X.X. Li, and A.G.L. Borthwick. 2008. Soil erosion assessment based on minimum polygons in the Yellow River basin, China. *Geomorphology* 93(4/3):233–252.
- Liang, Q., and A.G.L. Borthwick. 2008. Adaptive quadtree simulation of shallow flows with wet-dry fronts over complex topography. *Computers & Fluids* (accepted manuscript; available online).

Li He

- He, L. 2006. Fourier modelling of nonaxisymmetrical steady and unsteady flows. *Journal of Propulsion and Power* 22(1): 197–201.
- Moffatt, S., and L. He. 2005. On decoupled and fully-coupled methods for blade forced response prediction. *Journal of Fluids and Structures* 20(2): 217–234.
- Li, H.D., and L. He. 2005. Towards intra-row gap optimization for 1&1/2 stage transonic compressor. *ASME Journal of Turbomachinery* 127(3): 589–598.
- He, L., V. Menshikova, and B.R. Haller. 2007. Effect of hot-streak counts on turbine blade heat load and forcing. *Journal of Propulsion and Power* 23(6):1235–1241.
- He, L. 2008. Harmonic solution of unsteady flow around blade with separation. *AIAA Journal* 46(6): 1299–1307.

R. Eatock Taylor

- Smith, C.W., J. Zang, and R. Eatock Taylor. 2008. Wavelet-based adaptive grids as applied to hydrodynamics. *International Journal for Numerical Methods in Fluids*. <http://dx.doi.org/doi:10.1002/fld.1657>.
- Eatock Taylor, R. 2007. On modelling the diffraction of water waves. 28th Georg Weinblum Memorial Lecture. *Journal of Ship Technology Research* 54:54–80.
- Bai, W., and R. Eatock Taylor. 2007. Numerical simulation of fully nonlinear regular and focused wave diffraction around a vertical cylinder using domain decomposition. *Applied Ocean Research* 29:55–71.
- Buldakov, E.V., P.H. Taylor, and R. Eatock Taylor. 2006. New asymptotic description of nonlinear water waves in Lagrangian coordinates. *Journal of Fluid Mechanics* 562:431–444.
- Chern, M.J., A.G.L. Borthwick, and R. Eatock Taylor. 2005. Pseudospectral element model for free surface viscous flows. *International Journal of Numerical Methods for Heat and Fluid Flow* 15:517–554.

P.H. Taylor

- Walker, D.A.G., R. Eatock Taylor, P.H. Taylor, and J. Zang. 2008. Wave diffraction and near-trapping by a multi-column gravity based structure. *Ocean Engineering* 35:201–229.
- Liang, Q., P.H. Taylor, and A.G.L. Borthwick. 2007. Particle mixing and reactive front motion in unsteady open shallow flow – modelled using singular value decomposition. *Computers & Fluids* 36(2):248–258.
- Liang, Q., J. Zang, A.G.L. Borthwick, and P.H. Taylor. 2007. Shallow flow simulation on dynamically adaptive cut-cell quadtree grids. *International Journal for Numerical Methods in Fluids* 53(12):1777–1799.
- Borthwick, A.G.L., A.C. Hunt, T. Feng, P.H. Taylor, and P.K. Stansby. 2006. Flow kinematics of focused wave groups on a plane beach in the U.K. Coastal Research Facility. *Coastal Engineering* 53(12):1033–1044.
- Gibbs, R.H., and P.H. Taylor. 2005. Formation of walls of water in fully nonlinear simulations. *Applied Ocean Research* 27:142–157.

Y. Ventikos

- Arcidiacono, S., D. Poulikakos, and Y. Ventikos. 2004. Oscillatory behavior of nanodroplets. *Phys. Rev. E* 70(1):011505(1–7).
- Zeng, D., A. Ferrari, J. Ulmer, A. Veligodskiy, P. Fischer, J. Spatz, Y. Ventikos, D. Poulikakos, and R. Kroschewski. 2006. 3D modeling of mechanical forces in the extra-cellular matrix during cystogenesis. *Biophys. J.* 90:4380–4391.

- Kurtcuoglu, V., M. Soellinger, P. Summers, K. Boomsma, D. Poulikakos, P. Boesiger, and Y. Ventikos. 2007. Computational investigation of subject-specific cerebrospinal fluid flow in the third ventricle and aqueduct of Sylvius. *J. Biomech.* 40(6):1235–1245.
- Moyle, K.R., and Y. Ventikos. 2008. Local remeshing for large amplitude mesh deformations. *J. Comp. Phys.* 227:2781–2793.
- Mitsos, A.P., N.M.P. Kakalis, Y. Ventikos, and J.V. Byrne. 2008. Haemodynamic simulation of aneurysm coiling in an anatomically accurate Computational Fluid Dynamics model. *Neuroradiology* 50:341–347.

Site: University of Oxford, Structural Bioinformatics
and Computational Biochemistry Group
Department of Biochemistry
South Parks Road
Oxford OX1 3QU, UK
<http://www.ox.ac.uk/>
<http://bioch.ox.ac.uk>
<http://sbc.bioch.ox.ac.uk/>

Date Visited: February 26, 2008

WTEC Attendees: S. Kim (report author), P. Cummings, M. Head-Gordon, K. Chong

Hosts: Prof. Mark Sansom, Overall Coordinator, and P.I., Membrane Proteins
Email: mark.sansom@bioch.ox.ac.uk
Dr. Philip Biggin, P.I., Computational Studies of Receptors
Email: philip.biggin@bioch.ox.ac.uk

BACKGROUND

For the past fifty years, the Department of Biochemistry at the University of Oxford has been one of the preeminent institutions in the SBES field with pioneering contributions in metabolism, immunology, protein structure and function, microbiology, cell biology, and genetics. In 2008, the department will move to a new, state-of-the-art building. Within the department, three faculty members are the most active in simulation-based research. Of the three, Prof. Mark Sansom is the most senior, and his research group (Structural Bioinformatics and Computational Biochemistry Unit), composed of 26 researchers (13 postdocs and 13 graduate students), is the largest. Dr. Philip Biggin and Dr. Bela Novak are more recent arrivals to the department, and their newer research groups (each circa 5 researchers), are expected to grow over time.

COMPUTING HARDWARE

As noted during our earlier visit to the University of Oxford, the university is in the process of moving its supercomputer facility to a new site that will accommodate a \$6 million, 15 TFlop machine. Larger-scale national facilities (HECToR) are also important for computational simulations of biological macromolecules.

OPEN-FORMAT DISCUSSION

In lieu of formal presentations, our two hosts engaged the WTEC visiting team in a free-format discussion on issues pertaining to SBES activities (computational biochemistry) in their research groups, in their department, in their interdepartmental collaborations, and in the biochemistry landscape in the UK. In the first part of this discussion and after our overview of the goals of the WTEC-SBES study, we were briefed on the research interests of two of our hosts, Prof. Sansom and Dr. Biggin (Dr. Novak was on travel and not on campus during our visit).

Prof. Sansom's group is working on topics ranging from the dynamics of water in nanopores to large-scale MD simulations of bacterial membranes. The group's computational studies of membrane proteins range from molecular simulations of channels and transporters, to computational bionanoscience and membrane protein folding and stability. In his research description, Prof. Sansom notes that an estimated 50% of potential new drug targets are protein membranes (see the group's research website <http://sbc.bioch.ox.ac.uk/index.php>). Also of relevance to our report's life sciences chapter, Dr. Biggin's simulation-based research activities are centered around receptor dynamics and ligand binding. Dr. Bela Novak's interests are in computational and mathematical modeling of regulatory networks.

In the second part of our discussion, our hosts considered some of the broader SBES perspectives aligned with our list of questions. The following points were particularly memorable:

- Prof. Sansom sees the growth of computational biology along two axes: one axis is the connection with experiments and closer collaborations with other disciplines; and the other is the build-up of the systems biology framework.
- High-quality graduate students and postdocs drawn from the global pool are readily available and interested in joining their research groups (the strong reputation of Oxford and its effect on student recruitment was a recurring theme throughout our visit to the four departments at Oxford). For large-scale simulations, e.g., of protein membranes, the students do not develop new codes but use the standard MD packages on the major computational facilities.
- Consistent with the WTEC team's findings throughout the week, the biologically oriented research groups at Oxford biochemistry are well funded, and scarcity of research funding does not appear to be an issue. In addition to the BBSRC (Biotechnology & Biological Sciences Research Council, <http://www.bbsrc.ac.uk>) the funding stream emanating from the Wellcome Trust (the world's largest medical research foundation funding research on human and animal health, <http://www.wellcome.ac.uk>) provides a major boost for high-quality research in the biological sciences, including SBES activities therein.
- The near completion of the new biochemistry building and the pending move to these state-of-the-art facilities is a major boost to the research efforts of the department, including those of our hosts.

Site: University of Oxford Theoretical Chemistry Group
Physical and Theoretical Chemistry Laboratory
South Parks Road
Oxford, OX1 3QZ, UK
<http://www.chem.ox.ac.uk/tcg/>

Date Visited: February 26, 2008

WTEC Attendees: M. Head-Gordon (report author), P. Cummings, S. Kim, K. Chong

Hosts: Prof. David Logan
Email: david.logan@chem.ox.ac.uk
Prof. Graham Richards
Email: graham.richards@chem.ox.ac.uk
Dr. William Barford
Email: william.barford@chem.ox.ac.uk
Dr. Jonathan Doye
Email: jonathan.doye@chem.ox.ac.uk
Dr. Peter Grout
Email: peter.grout@chem.ox.ac.uk
Prof. David Manolopoulos
Email: david.manolopoulos@chem.ox.ac.uk
Dr. Mark Wilson
Email: mark.wilson@chem.ox.ac.uk

BACKGROUND

The University of Oxford has a long tradition in theoretical chemistry, going back to the 1956 Nobel Prize-winning work of Hinshelwood on chemical kinetics and the work of Coulson's group on the theory of valence. Today the Theoretical Chemistry Group has 8 academic staff and approximately 30 students and postdocs studying with them. Prof. David Logan is the head of the group, and serves as the Coulson Professor.

RESEARCH

During the WTEC visiting team's two-hour visit to the Oxford Theory Group, our hosts first briefly introduced their research interests. Prof. Logan focuses on the study of strongly correlated electrons in condensed matter; Dr. Grout studies defects in solids. Prof. Richards is a widely recognized pioneer in biological simulations. Dr. Barford studies electronic properties of conjugated polymers. Dr. Doye works on simulations of soft condensed matter with a focus on biological systems, and Dr. Wilson also studies condensed matter systems including low-dimensional crystals, glasses and networks, and self-assembly. Finally, Prof. Manolopoulos is interested in dynamics and has recently developed tractable semiclassical methods for treating condensed matter systems.

COMPUTING HARDWARE

Different parts of the research depend to very different degrees on high-performance computing. The Barford, Doye, and Wilson groups are in the process of installing a large local cluster to meet a significant part of their computing needs. Oxford University itself is in the process of moving its supercomputer facility to a new site that will accommodate a 15 TFlop machine (at a cost of approximately \$6 million). Additionally, the Oxford campus as a whole is experimenting with grid computing through the creation of the OxGrid facility, which uses spare cycles from across the campus. The new Barford/Doye/Wilson cluster received local funding support that was conditional on it becoming a part of OxGrid.

DISCUSSION

The WTEC team exchanged views with our Oxford hosts on a variety of issues regarding opportunities and challenges in research and training in simulation-based engineering and science. Some of the perspectives expressed by the hosts included the following points:

- At Oxford, high-quality students and postdocs are readily available and comprise one of the most important strengths of the theory program. The hosts felt that members of their group were able to become adept quite quickly at developing appropriate tools, whether they be formal models or computational methodologies.
- Tie-ins to industry can be very strong, as exemplified by Prof. Richards' personal involvement in the founding of several companies. However this remains the exception rather than the rule. Most of the Group members operate in the traditional academic manner of pursuing curiosity-driven research.
- Towards the physical end of theoretical chemistry, funding is a challenge, and the funding models have changed in several important respects. First, the group no longer receives significant numbers of studentships directly to redistribute to its academic staff. Second, scientific grants to researchers are now funded based on full cost accounting; it is not yet clear whether this will lead to increased opportunities or not. Third, there is now a national Materials Modeling Consortia (started about 3 years ago) that brings together teams and provides some additional funding (overall at the rate of about \$2 million/year).
- Towards the biological end of theoretical chemistry, the funding situation is considerably better, due primarily to the availability of the Wellcome Trust, which sponsors research relevant to human and animal welfare.
- Funding for the new Oxford supercomputer facility is for hardware only.

Site: University of Zurich Physical Chemistry Institute
Computational Chemistry Group of Prof. Dr. Jürg Hutter
Winterthurerstrasse 190
8057 Zurich, Switzerland
<http://www.pci.uzh.ch/e/index.php>

Date Visited: February 29, 2008

WTEC Attendees: S. Glotzer (report author), L. Petzold, C. Cooper, J. Warren, V. Benokraitis

Hosts: Prof. Dr. Jürg Hutter, Head, Computational Chemistry Group
Email: hutter@pci.uzh.ch
Prof. Kim Baldridge, Organic Chemistry Institute
Email: kimb@oci.unizh.ch
Prof. Jay Siegel, Organic Chemistry Institute
Email: jss@oci.unizh.ch
Dr. Joost VandeVondele
Email: vondele@pci.uzh.ch

BACKGROUND

The Institute of Physical Chemistry is part of the Faculty of Mathematics and Science and the Department of Chemistry and Biochemistry at the University of Zurich. The institutes the Department of Chemistry and Biochemistry collaborate mostly in teaching but also partially share infrastructural facilities like workshop and service units. The Computational Chemistry Group of Prof. Dr. Jürg Hutter is part of a collaborative network, the Competence Center for Computational Chemistry “C4,” which is a collaboration between ETH Zurich, University of Zurich, and the IBM Research Laboratory. Activities include a seminar program, tutorials, and computational resources. C4 is the Zurich node for the European Centre of Atomic and Molecular Computations (CECAM).

SBE&S RESEARCH

Computational Science groups at the University of Zurich include

- Amedeo Caflisch, Biochemistry – Protein Folding/Aggregation, Docking
- Kim Baldridge, Organic Chemistry – Quantum Chemistry, Grid Computing
- Jürg Hutter, Physical Chemistry – *ab initio* Molecular Dynamics, Condensed Phase Chemistry/Physics
- Ben Moore, Astrophysics
- George Lake, Astrophysics
- Thomas Gehrmann, Particle Physics
- Andreas Wagner, Systems Biology
- Thomas von Mehring, Functional Genomics

The Computational Chemistry group at PCI includes

- Prof. Jürg Hutter, group leader
- Dr. Joost VandeVondele, senior postdoc
- Drs. Teodoro Laino, Valery Weber, and Urban Borstnik, postdocs
- Manuel Guidon, Florian Schiffmann, and Samuele Giani, PhD students

Professor Hutter's group recruits from the departments of Chemistry, Physics, Mathematics, and Computer Science. Funding from the university currently supports 4 postdocs and 2 PhD students. The main source of external funding is the Swiss National Science Foundation. Research activities of Prof Hutter's group include:

- *Car-Parrinello Molecular Dynamics*. Dynamical computer simulations of complex chemical systems are performed using accurate descriptions of the electronic structure. This includes (a) Method Development with a focus on the description of electronically excited states using density functional theory; and (b) Applications—calculation of the chemical and physical properties of complex molecular systems, molecules in solution, and molecular crystals. Special topics include hydrogen bonded systems and proton transfer in ground and excited states.
- *Methods for Large-Scale Density Functional Calculations*. Theoretical approaches are developed and tested for application of DFT to large molecules and condensed phase systems. New *algorithms for high-performance computer applications are designed and implemented in computer codes*.

Software Development

Prof. Hutter gave a presentation to the visiting WTEC team on the CP2K Software Development Project, a major open source code development in his group. CP2K provides a full spectrum of codes for computational chemistry. The project started in 1999 from scratch. The code is being developed as open source under the Gnu Public License. It uses object-oriented Fortran 95 and is maintained on a CVS server at <http://cp2k.berlios.de> (<http://www.berlios.de> is a German SourceForge site). Fifteen developers have write access to the code; they are located in Zurich (UZH, ETH, PSI), Berlin, PNNL, LLNL, and Minnesota. There is a user Google group for the CP2K community that contains 94 members. The code has increased steadily in size from 100,000 lines in 2001 to 500,000 lines in 2008.

Tools for CP2K software development include CVS source control system, code quality control, automatic regression testing, ca. 900 tests, serial, parallel, memory leaks, bug reporting, automatic documentation (using Doxygen), and other free Tools for Fortran programmers such as g95, gfortran, and valgrind. Methods and best practices for software development include “commit early, commit often”; “commit many tests”; force parallelization; write general code; extend and reuse code, don't duplicate code; code refactoring (constantly improve code); and minimize library dependencies.

Features of CP2K include *Force Methods* such as QM, MM, Semi-empirical, and QM/MM, and *Sampling Methods* such as MD, MC, Free energy, and nudged elastic band. Specialty features include job farming, CP2K shell. The CP2K module Quickstep is based on Kohn–Sham density functional theory, GGA, Hybrid functionals, Gaussian basis sets, (Non-)Periodic, and linear scaling. The CP2K module Fist is based on Classical Force Fields, Pair potentials, EAM, Tersoff potentials; General non-bonded interactions; SPME; and multiple force fields. The force methods in CP2K include empirical models such as DFTB, NDDO, frozen density embedding, and Kim-Gordon. Sampling methods in CP2K include molecular dynamics with NVE, NVT, and NPT ensembles; shock wave; Monte Carlo with NVT, NPT, and Gibbs ensembles; meta-dynamics; thermodynamic integration, and nudged elastic band.

Parallelization of CP2K is based on MPI and OpenMP; is multilevel on job level (farming), on multiple force calculation level (NEB, Hessian), on single force calculation level, and on loop level (OpenMP); and employs fast networks for parallel 3D-FFT and parallel linear algebra.

High-Performance Computing

Dr. VandeVondele presented a talk on high-performance computing to the WTEC panel. He defined high-performance computing as the lowest end of the Top 500 list of supercomputers. Today, the slowest machine on the top 500 list runs at roughly 5 teraflops. Analysis of the number of systems on the top 500 list relative to the number of millions of inhabitants in the country shows that the United States and Switzerland are equivalent, with US=0.94 (283/301) and CH=0.93 (7/7.5). Other ratios include UK = .78, Denmark = 0.37, France = 0.28, and Luxembourg 2.08!

Analysis of the top ten machines on the 2007 Top 500 list shows that these machines contain between 212,992 cores (#1, at LLNL) and 36,864 cores (# 10 at BNL). Both of these are IBM machines.

Next-generation chips will be multicore and involve graphical processing units. For example, an 8 teraflop NVIDIA machine can be acquired for \$2000. Intel's teraflop chip has 80 cores, requires only 64W of power. This group is now trying to port CP2K to gpu; it can do FFTs now on gpu, but only 25% of job, not the bottleneck. VandeVondele believes that a marked presence on the Top 500 list should be a goal of computational chemistry and computational materials science.

Using CP2K, very large and complex systems have been studied. An examples is a DNA crystal, fully solvated, using 12661 atoms, > 100,000 basis functions, <5 minutes per SCF step on 1024 CPUs of an XT3 (Weber et al. 2008). Another is QM simulations of finite temperature water.

CPMD scales better but is slower than CP2K, since CPMD uses plane wave basis sets and thus dense matrix algebra (mostly multiplication by zero). So CPME will be faster for smaller systems, while CP2K will be faster for larger systems. The group is looking to further improve times by distributing individual atoms among multiple CPUs, which involves many sophisticated parallelization schemes involving dense linear algebra, sparse linear algebra, FFTs, and grid kernel routines (this involves a collaboration with University College, London)

Solar cell applications of CP2K is one next application goal, requiring adding two-component relativistic with spin-orbit coupling to go from iodine I- to I2-. Another grand challenge is to include electron transport.

Overall, this group is carrying out state-of-the-art SBE&S for computational chemistry. Code development is at a very sophisticated level, and the group's code is being used to solve many hard problems.

SBE&S EDUCATION

Training of students in SBE&S is largely organic within the group. Half of their PhD students come from the CSE program at the Swiss Federal Institute of Technology (ETH) Zürich. For engineering applications, the group favors students coming from the CSE program, whereas for scientific applications, they favor specialization. Prof. Hutter is very happy with CSE students; "they are excellent." There is lots of code development here, though, with a distinct emphasis on algorithm development; students are less worried about publications than at U.S. institutions.

Professor Hutter teaches Quantum Chemistry (together with K. Baldridge) as a basic course with computer lab; Molecular Dynamics Simulations as a basic course with computer lab, advanced courses on specialized topics, and advanced topics in tutorials (within C4 and CECAM)

COMPUTING FACILITIES

At the Physical Chemistry Institute, resources include a 16-CPU shared memory (64 Gbytes) machine for code development, group owned and operated (used by half the group that does code development); a 100-CPU cluster (AMD 2.6 GHz, 4 CPU nodes, 8 Gbytes), group owned, operated by Astrophysics as part of a 500 CPU cluster; a share of UZH cluster Matterhorn, with 756 CPUs, 256 of which are connected via a fast network, operated by ID (Computer center); shares at CSCS (the Swiss national supercomputer center), project-based, competitive allocation (Cray-XT3, 3328 CPUs, 80, 000 h/month [3.3%], Cray-XT4, 896 CPUs, 0 h/month [0%] IBM-SP5, 768 CPUs, 50,000 h/month [9%])

Plans are underway to acquire a 4000-core machine funded from a combination of start-up funds, internal university funds for equipment replacement, and 1M Swiss Francs/year provided by the Vice Chancellor (this last is exceptional) for the Natural Science faculty.

ADDITIONAL INFORMATION

The WTEC team was fortunate to meet briefly with Professors Baldrige and Siegel of the Organic Chemistry Institute, both formally of University of California, San Diego, and the San Diego Supercomputer Center, respectively. Switzerland had virtually no grid computing presence before their arrival. Baldrige initiated many grid activities, including participation as a first European node of the PRAGMA International grid effort, was part of the core team that started the Swiss national grid initiative (SwiNG), and recently formulated a grid competence center (GC3) at the University of Zürich. In Switzerland, the grid is bottom-up, whereas HPC is top-down. In Switzerland, there is now a grid testbed at many of the key Universities participating in the SwiNG initiative in grid computing, contributing applications or development of middleware tools. As in her keynote presentation at the 2006 NSF Workshop on Petascale Computing in the Biological Sciences (Baldrige 2006), Prof. Baldrige stressed the need for people, middleware, and software, not just hardware.

Regarding students, both Baldrige and Siegel noted that it is difficult to find students with the programming capability and knowledge to do algorithmic development, analogous to that found in the United States. They additionally noted that they found shortcomings in the NSF funding in the United States in the limited to nonexistent funding of fundamental algorithmic development, something that is still very much needed.

REFERENCES

- Baldrige, K. 2006. "Strategies for enabling petascale computing in biology." Keynote address, NSF Workshop on Petascale Computing in the Biological Sciences, Arlington, VA, 29–30 August, 2006. More details are available at <http://www.sdsc.edu/PMaC/workshops/bio2006/index.html>.
- Weber, V., J. VandeVondele, J. Hutter, and A.M.N. Niklasson. 2008. Direct energy functional minimization under orthogonality constraints. *J. Chem. Phys.* 128:084113–084122.

Site: Vrije University Amsterdam
Dept. of Molecular Cell Physiology and
BioCentrum Amsterdam, Faculty of Biology
De Boelelaan 1085
NL-1081 HV Amsterdam, The Netherlands
<http://www.vuamsterdam.com/home/index.cfm>
<http://www.bio.vu.nl/vakgroepen/mcp/index.html>

Date Visited: February 25, 2008

WTEC Attendees: L. Petzold (report author), S. Glotzer, J. Warren, V. Benokraitis

Hosts: Prof. Hans Westerhoff, Department Head
Email: Hans.Westerhoff@manchester.ac.uk
Prof. Frank Bruggeman
Email: frank.bruggeman@falw.vu.nl
Prof. Barbara Bakker
Email: barbara.bakker@falw.vu.nl
Prof. J.C. de Munck
Email: jc.munck@vumc.nl
Prof. Ivo van Stokkum

BACKGROUND

Systems Biology is defined as the science that studies how properties that are important for biological function emerge from the interactions between components of biological systems. Systems Biology bridges molecular biology and physiology. It needs mathematical approaches to deal with the complexity of the nonlinear interactions. It also needs quantitative experimentation in actual living systems, or well-designed *in vitro* representations thereof. Professor Westerhoff is a driver behind the Silicon Cell program (<http://www.siliconcell.net>), which makes computer replicas of pathways in living cells available on the Web for *in silico* experimentation. He heads a transnational research group on Systems Biology that includes the Manchester Centre for Integrative Systems Biology (MCISB) in the Manchester Interdisciplinary BioCentre (MIB) and the BioCentrum Amsterdam. He is also the director of the UK's EPSRC (Engineering and Physical Sciences Research Council) Doctoral Training Centre in Integrative Systems Biology.

R&D ACTIVITIES

The systems biology effort at Vrije University (VU) Amsterdam is affiliated with Integrative BioInformatics VU, the Netherlands Institute for Systems Biology, TopMaster Systems Biology, The Trans North Sea Centre (Manchester Centre for Integrative Systems Biology; Manchester Doctoral Training Centre Systems Biology), and the BioSim EU Network of Excellence. The WTEC visiting team heard about efforts in signal transduction, metabolic engineering, hierarchical control and regulation, and digital human from Professor Westerhoff. A theme of the efforts at this center is that to cure a disease, one must cure the network. This requires the understanding of multifactorial relationships. A strong case was made of why simulation is an integral part of this process.

Professor Bruggeman described the Silicon Cell project. This is an initiative to construct computer replicas of parts of living cells from detailed experimental information. The Silicon Cells can then be used in analyzing functional consequences that are not directly evident from component data but that arise through contextual dependent interactions.

Professor Barbara Bakker described efforts in network-based drug design, showing the interplay between metabolism, gene-expression, and differentiation. The premise is that the recent decline in drug discovery may be

due to the focus on single molecules, neglecting cellular, organellar, and whole-body responses. Specifically, she outlined a research effort focusing on the treatment of African sleeping sickness using trypanosome glycolysis as a drug target. This entailed development of an ODE model of trypanosome glycolysis, and sensitivity analysis to identify those reactions that can best control the flux. The researchers found that the network behaviour was complicated by the fact that gene-expression response can either counteract or potentiate the primary inhibition; hence, it was necessary to model the regulation of gene expression. Some of the model parameters were obtained from the literature, and others were fitted from experiments.

Professor J.C. de Munck described efforts in brain imaging. EEG and fMRI are used simultaneously to localize pathological waves such as epileptic spikes. The project requires modelling, solution of both forward and inverse problems, signal/image processing, and visualization, all in close coordination with experiment.

Dr. Ivo van Stokkum described a problem-solving environment (PSE), to identify and quantitatively model complex biomolecular systems. The PSE takes data from multiple time-resolved spectroscopical experiments and *a priori* knowledge, and outputs the model structure and estimated physicochemical parameters. A partitioned variable projection method was developed for the (large-scale) parameter estimation problem. The PSE is used by many collaborating scientists to study problems including natural and artificial photosynthetic systems, photoreceptors, plant metabolism, and imaging of the corneal surface. The work is done in close coordination with experiment.

CONCLUSIONS

The systems biology and biophysics efforts in Amsterdam are world-class. There is an impressive integration of simulation, visualization, and experiment. They demonstrate both examples and a tremendous potential for how modeling and simulation can impact drug development and medical diagnosis and treatment. The WTEC team's hosts expressed some frustration in finding appropriately trained students.

SBES QUESTIONNAIRE

The panel's hosts graciously provided answers to our written questions, as noted below.

General

What are the major needs, opportunities or directions in SBE&S research over the next 10- and 20-year time frames?

Silicon/ digital human (90% complete in 15 years)

Construction, validation, analysis of large models of biological systems, most urgently cells. This should go hand in hand with technological advances for cellular measurements. The feasibility of large predictive models is still questionable given current technological approaches. This is anticipated to change significantly within the next 5-10 years making predictive biosimulation reality; allowing for rational engineering of biological systems for medical and industrial purposes.

What are the national and/or regional funding opportunities that would support research to meet these needs, and/or take advantage of these opportunities? Are these funding opportunities expanding?

National: left-overs from genomics plus small scale

Germany, UK, EU: large investments

Transnational

Funding of institutions (excellence centres) to carry out integrative biology in a setting of technology and simulation development. Such institutions should offer facilities to groups outside of the institutes. Institutes should be involved in strategic collaboration with other institutes worldwide.

Materials / Energy and sustainability / Life sciences and medicine

What major breakthroughs in these fields will require SBE&S; and which are you and or your colleagues pursuing? Within your institution, region, or country are there identified targets of opportunity for applications of simulation either for scientific research or for engineering applications in these fields?

Digital human

Self-sustaining systems (Systems Biology to green planet)

Quantitative analysis of signal and metabolic networks; theory and experiment

We are engaged in: 1. SMAD, MAPK signalling 2. Central carbon and energy metabolism in yeast, E coli and extremophiles, 3. Modelling/experimentation of eukaryotic transcription initiation (PPAR induced transcription of human PDK4 gene), 4. nucleocytoplasmic shuttling of signalling proteins, 5. Nuclear receptor signalling

Formulation of concepts/approaches to facilitate comparative systems biology (discovery of principles) related to control, regulation and adaptation of intracellular processes.

Single-cell stochastic simulation of transcription and signalling; we are currently setting involved in initiating experiments

Which problems could benefit most from a 1-2 order of magnitude increase in computational power?

Distributive digital human

Web services based linkage of groups each responsible for pathway/organ

Analysis of populations of models for experimental design, system identification, stochastic simulation of large networks

What are examples of major SBE&S successes or failures in these fields?

SUCCESSSES: Application of models in biotechnology and bioengineering, improved understanding/exploitation of control of metabolism; hybrid modelling approaches; applications of engineering concepts/approaches to biology; role of simulations in bioinformatics/molecular dynamics.

Network-based drug design

FAILURES: small number of predictive models (depend much on quantitative experimentation of enzyme kinetics (enzymology) which is out of fashion and has no high-throughput method at the moment); ab initio prediction of kinetic constants from molecular dynamics simulations; solutions to deal with multi-scale simulation of biological systems (how to cross levels?; formal methods to allow for coarse graining);

Do investigators, laboratories and institutions receive any financial compensation for patented inventions derived from their simulations?

No

Have any start-up companies spun-off based on simulation efforts in your lab? If so please describe them.

No; potential was there; but by nature of the projects too much interconnection required

Multiscale Simulation

Describe efforts and advances within your institution to couple multiple simulation methods in order to bridge multiple length and/or time scales.

Modular approaches for control and response analysis

Is the development of integrated multiscale modeling environments a priority in research funding in your country or region?

No?

Validation, Verification, and Quantifying Uncertainty

Describe efforts and advances within your institution to validate and verify codes and to quantify uncertainty in simulation-based predictions?

Silicon cell/JWS: models are recalculated before put onto live web site and collaborating journals

Simulation Software

What percentage of code used in your group is developed in-house?

30% Mathematica code

What percentage of code is commercial?

50%; Mathematica

What percentage is open-source?

50% Copasi, JDesigner

What percentage has been developed by others (e.g., under contract or by acquisition)?

0%

What are the biggest issues in using models/simulations developed by others?

Link to experimental data; judgement of biological quality

How easy/difficult is it to link codes to create a larger or multifaceted simulation environment?

Hard for large models; SBML is promising effort but not yet sufficient multiscale supportive; no good multiscale software available for novice

How do you deal with liability issues for products developed with codes from other sources?

Not relevant for us now

Engineering Design

What type of models/codes do you use, develop or conduct basic research on pertaining to different phases of engineered system design (conceptual, parametric optimization, operational/control)?

Computer replica of reality

Site: Vrije University Theoretical Chemistry Section
De Boelelaan 1105
1081 HV Amsterdam, The Netherlands
<http://www.chem.vu.nl/en/sec/tc/>

Date Visited: February 25, 2008

WTEC Attendees: S. Glotzer (report author), L. Petzold, C. Cooper, J. Warren, V. Benokraitis

Hosts: Prof. Evert Jan Baerends, Head, Theoretical Chemistry Section, Department of Chemistry and Pharmaceutical Sciences
Tel: +31-20-598 7621; Email: baerends@chem.vu.nl
Dr. Stan van Gisbergen, CEO, Scientific Computation and Modeling NV (SCM)
Email: vangisbergen@scm.com
Prof. Luuk Visscher

BACKGROUND

Prof. Evert Jan Baerends leads the Theoretical Chemistry group at Vrije University, which is comprised of 30–35 people, including 5 permanent members (Dr. Luuk Visscher, Dr. Matthias Bickelhaupt, Dr. Oleg Gritsenko, Dr. Drew McCormack, and Prof. Baerends) plus postdocs and graduate students. An additional five senior researchers (including Dr. Stan van Gisbergen, CEO of SCM) are employed by Scientific Computation and Modeling NV (SCM), a spin-off company that supports the Amsterdam Density Functional (ADF) simulation software developed by the Baerends group.

Research within the Baerends group deals primarily with development and application of density functional theory. The computational advantages offered by DFT are being exploited in computational method development. Further research emphases within the group include theoretical aspects of DFT (Kohn-Sham potentials, meaning and status of Kohn-Sham orbitals, orbital-dependent exchange-correlation functionals); MO analysis of bonding and spectroscopy, mostly in transition-metal (organometallic) complexes; molecule-surface interactions (scattering, dissociation, heterogeneous catalysis); spectroscopic (response) properties; and relativistic effects in heavy element compounds. This group is one of the leading theoretical chemistry groups in Europe; its ADF simulation software is highly popular and used around the world.

The Baerends Theoretical Chemistry group is part of a new center, the Amsterdam Center for Multiscale Modeling (ACMM), which includes this group, Computational Chemistry and Physics at the University of Amsterdam, and Theoretical Biophysics at AMOLF Amsterdam.²³ The mission of the ACMM is to model systems “from electrons to electricians,” spanning all of the scales needed to bridge the physics of electrons to the biology of living systems. Over the next 20 years, ACMM researchers aim to focus on building a bridge from quantum mechanics to thermodynamics.

Other groups within the Netherlands involved in molecular modeling and simulation include groups at Eindhoven, Groningen, Leiden, Nijmegen, Twente, and Utrecht. These groups all participate in activities of the European Centre of Atomic and Molecular Computations (CECAM).

SBE&S RESEARCH

Simulation-based science research in theoretical chemistry at Vrije University includes the following:

- Fundamental research and development of density functional (DFT) and density matrix functional theory (DMFT)

²³ AMOLF is the Fundamental Research on Matter laboratory of the Institute for Atomic and Molecular Physics.

- Code development

The group is well known for its development of the Amsterdam Density Functional program (ADF). Originally developed to treat homogeneous catalysis and distributed free to the research community, ADF was spun-off in 1995 to form a new company, SCM (<http://www.scm.com/>), for continued professional code development and maintenance. The cost of the code varies depending on whether purchased by an academic site (roughly \$1000/year), industrial site (roughly \$10,000/year), or government site. The profits from sales of the code are sufficient to support five full time researchers/code developers, and revenue is shared with the university.

ADF has several competitors, including NWChem, QChem, Turbomole, and Jaguar, with Gaussian being the largest. ADF was the first of these codes to include numerical integration. ADF strengths include its unique ability to accurately treat transition metals, heavy metals, and spectroscopic properties; its ease-of-use as an analysis tool; as well as the longevity and stability of the code. ADF is very popular around the world. SCM has sold several hundred licenses, which represents many more users. In industry, the primary users include oil, automotive, and electronics companies. The code is built on HP-MPI, which enables it to run efficiently within a multicore architecture (currently under Unix [Linux], Windows, and Apple OS), and it can operate on a single processor or many processors in a parallel environment. The development of ADF has benefited from a long-term collaboration with the Ziegler group at Calgary University of over 30+ years, e.g., they developed the popular NMR module for ADF.

Typical application and development areas for ADF are:

- Time dependent density functional theory (TDDFT), response theory
- Embedding (frozen density), QM+MM multiscale methods (Wesolowski, Neugebauer, Baerends, Jacob, and Visscher). One can now easily calculate important properties on systems containing up to 150 atoms. More complex systems, such as molecules in water and undergoing reactions, require multiscale methods that bridge quantum mechanics and classical mechanics. Dr. Visscher received a large grant from the Netherlands Science Foundation (NWO) to develop these methods.
- Structure, reactivity, and catalysis (homogeneous). Bickelhaupt has received a large grant for this research from NWO. There are collaborations with UvA (Reek), Leiden (Reedijk and Buda), and Eindhoven (van Santen, Jansen), as well as participation in the top research school "Catalysis by Design."
- Relativistic effects and their inclusion in density functional methods. There are collaborations with Gronigen (Broer and Filatov).
- Under the leadership of Dr. Visscher, the group develops two other codes complementary to ADF—MOLFDIR and Dirac—that deal with relativistic effects for heavy atoms. MOLFDIR is not open source per se, but the group will provide it (primarily to academics) upon request. Dirac is a collaborative, bottom-up effort involving five groups throughout Europe that currently operate with only a small amount of funding from the EU.
- Molecule-surface interactions (McCormack in VU group, and collaboration with former group members Kroes and Olsen, now at Leiden,).
- Theoretical biochemistry, molecular recognition, DNA (Bickelhaupt). Questions addressed, in particular within ACMM, include, what are the reactions that lead to the formation of DNA?

COMPUTING FACILITIES

Some universities within the Netherlands, including Vrije University, acquire and support the equivalent of Tier 3 computational resources locally to faculty, researchers, and students. These resources are paid for by the university with central funds. According to Baerends, the Universities of Groningen and Amsterdam are today at Tier 3, while Vrije University is at the low end of Tier 3. National computing facilities provide access to Tier 2 resources.

NWO (the Netherlands Science Foundation) supports large-scale computing in the Netherlands by funding the regular purchase of "supercomputers" (every 3–6 years), and by staffing and funding their section called National Computing Facilities (NCF). NWO (through NCF) has always chosen SARA, the computer center at Amsterdam, as the site to host the national supercomputer, in this way strongly supporting this computer center. Only part of SARA's revenues comes from hosting the national supercomputer; SARA serves and is financed by a mixture of commercial and academic computer users.

Faculty can apply for grants (proposals are peer reviewed) from NCF and HPCC programs to obtain computing time at NCF facilities (primarily the national supercomputer at SARA, but also some smaller machines). It is also possible to apply for personnel to develop and port simulation codes to the newest machines of NCF. These codes must then be freely shared with other NCF users. This sharing requirement occasionally discourages some Dutch groups from using NCF resources, since a particular code may represent a significant competitive edge for that research group. There is also a constraint that no single group can use more than roughly 10% of the computational resources at NCF. Grand-challenge-type problems requiring substantially more resources can sometimes be investigated at the very early installation stages of new supercomputers at NCF.

DISCUSSION/OTHER COMMENTS

Regarding the impact of SBE&S and looking to the future, many important problems involving small molecules can be solved today via SBE&S. As one example, gas molecules (N_2 , O_2 , CO_2 , H_2 , CO ,...) constantly hit surrounding surfaces, with large effects. Using codes like ADF, we can answer today's questions such as: Why does iron rust? Why do copper cupolas turn green? Why is gold a noble metal? Why is its neighbor, platinum, a good hydrogenation catalyst? Why do we need a three-way catalyst in our car? However, major challenges remain with larger molecules and complex systems, such as large molecules embedded in a solvent environment. There, much faster computers, such as petascale computers, are required, along with the ability to efficiently run codes on those new platforms. Petaflop computing is also required to treat dynamics, which is important, for example, in problems in biology, energy, and those involving nuclear waste (such as the separation of actinides from lanthanides by judicious design of selective ligands). Such problems will require the coupling of molecular dynamics and time-dependent density functional theory, which in turn will require significant method and algorithm development.

ADF has been implemented on quad-core chips. However, implementation of ADF by SCM on chips on multicore chips with substantially more than four cores will be a major undertaking and will require a cost-benefit analysis taking into account the likely number of massively multicore users.

Regarding human resource development, education, and training in SBE&S, deep, solid knowledge of a core discipline, as well as computational skills, will be required to solve "grand challenge" problems with computational chemistry. (Half of the PhD students in the Baerends group currently are physicists, and the other half are chemists.) Baerends noted that many users of computational chemistry codes are not very good at mathematics and have limited knowledge in theoretical chemistry. To be innovative, computational chemists need to be trained in both theoretical chemistry and computational methods. Students in his group generally take graduate courses in computational chemistry and programming. A computer center at Vrije University, supported by NWO, provides training to students in parallel programming. In addition, the national computer center offers computing courses supported by NWO.

Regardless of country, students pursuing a PhD must contribute original ideas to their field of study. However, in the Netherlands, Baerends noted that in theoretical chemistry, code development and application are not considered "new." As a result, students generally focus on theoretical developments and not code development. He contrasted this situation with that in Germany or the United States, where code development is viewed as original.

At Vrije University, students pursuing a bachelor's degree in chemistry take courses in computational chemistry, where they are exposed to codes like Gaussian and Spartan. The Pharmaceutical Chemistry program offers

modeling courses for master's students. All undergraduates take a course in computer programming. It requires 5 (3+2) years to obtain the master's degree, and then an additional 4 years for the PhD.

Regarding Dutch government support of SBE&S, Professors Baerends and Visscher noted that in the Netherlands, it is very difficult to obtain research funds to support code development and long-term maintenance and support, since the peer-review process for grant applications is based on science and not code. Even more generally, code maintenance languishes because there is no reward system for non-science activities. They noted that due to lack of coordination and support of code development, there is an enormous amount of duplication of effort within academia where individual groups develop their own codes or parts of codes, in order to avoid having to rely on other's codes that may not be supported after a few years. They further noted that the UK stands out as having paid attention to code maintenance through long-term investment.

The researchers noted that in the Netherlands, most grants cover the full period of a student's PhD, which enables continuity of code development and science projects. Young investigator grants (VICI, as awarded to Bickelhaupt and Visscher) support 5–6 students for the duration of the PhD thesis, and “top grants” to established investigators support three students for the duration of the PhD thesis, plus equipment. At Vrije University, similar to other universities within the Netherlands, students cost approximately €40,000/year.

SOFTWARE INVESTMENT AS CRITICAL INFRASTRUCTURE AND NATIONAL ECONOMIC SECURITY

The theoretical chemistry community within The Netherlands is close-knit and meets at annual meetings. When asked to compare computational chemistry code development in Europe with the rest of the world, the WTEC visiting team's hosts stated that Europe leads in the development of relativistic QM codes, whereas the United States leads in the development of QM codes based on couple-cluster methods. They felt that algorithm development perhaps receives somewhat more attention in Europe. They further noted that contributions to computational chemistry at the U.S. Pacific Northwest National Laboratory (PNNL), namely the package NWChem, was innovative not in the theoretical developments in the code, but rather in their implementation. Amber and Charmm are still strong U.S. programs. Europe has strong Gromacs.

Approximately 50% of Vrije University graduate students in the Science faculty are from outside the Netherlands, with many from Eastern Europe. Our hosts noted that, in theoretical chemistry, the strongest and most disciplined students come from Germany, with whom they have easy communication. Trade with Germany and China has increased in The Netherlands, while at the same time, the country is clamping down on immigration, making it harder to recruit students from abroad.

Our hosts noted many success stories for simulation in terms of impact on industry. ADF is used to calculate, for example, solubilities and phase diagrams. A large unnamed Dutch company uses ADF for this. Most companies use ADF for homogeneous and heterogeneous catalysis. AkzoNobel used ADF to scan the hyperpolarizabilities of hundreds of molecules to narrow down potential targets for synthesis, and in so doing discovered molecules with enhanced nonlinear optical properties. Another Dutch company used ADF to improve the properties of an existing catalyst.

In the pharmaceutical industry, there is increased interest today in virtual screening, including dynamics and finite temperature (for entropic effects) to predict binding affinities. Many HIV transcription inhibitors have important contributions from simulation, and one compound went in production based on only 31 iterations, with significant savings.

Many pharmaceutical companies within The Netherlands have modeling groups, but these groups do not necessarily contain expert simulators. In contrast, BASF has a wide range of simulation capabilities, including a modeling group very strong in QM and MD. AstraZeneca is now building up its modeling and simulation capabilities, with an eye toward computational methods development.

Site: Zuse Institute Berlin (ZIB)
(Konrad-Zuse-Zentrum für Informationstechnik)
Takustrasse 7
D-14195 Berlin-Dahlem, Germany
<http://www.zib.de/>

Date Visited: February 26, 2008

WTEC Attendees: L. Petzold (report author), S. Glotzer, J. Warren, C. Cooper, B. Benokraitis

Hosts: Prof. Dr. Dr. h.c. Peter Deuffhard, President, ZIB
Scientific Computing, Department of Numerical Analysis and Modelling
Email: deuffhard@zib.de

BACKGROUND

The Konrad-Zuse Zentrum für Informationstechnik, or Zuse Institute Berlin (ZIB), is a non-university research institute of the State of Berlin that operates in the fields of applied mathematics, and computer science. It provides solutions for complex problems in science, engineering, environment, and society. In close cooperation with partners from science, economy, and society, ZIB develops mathematical models and efficient algorithms.

ZIB was created in 1986, when Professor Peter Deuffhard was recruited to Berlin. Now the institute includes about 200 people, including 3 professors, 100 scientists, and 60 students. The scientists are supported on soft money, most frequently on competitive projects lasting 3–4 years. The goal of the institute has been to be a center of high-level scientific computing, emphasizing mathematics where it matters and where it is difficult.

R&D ACTIVITIES

ZIB collaborates with Free University Berlin, Humboldt University Berlin, the Technical University of Berlin, and the Weierstrauss Institute for Applied Analysis and Stochastics, in R&D activities of the DFG Research Center MATHEON (<http://www.matheon.de/index.asp>). (DFG is the German Research Foundation Deutsche Forschungsgemeinschaft, analogous to NSF, although it also supports research in medicine.)

DFG Research Center MATHEON

The DFG Research Center MATHEON was initiated in 2001 as a result of the Berlin consortium winning a major DFG funding initiative. The MATHEON effort currently involves 40–50 professors of mathematics. It is funded through May 2010 and then has a possibility of extension into 2014, after which it must by the rules stop. DFG contributes €5.6 million per year, and the 5 other participating institutions contribute €3 million per year. This supports 6 new full professor positions, 7 junior research groups, 65 researchers, and 21 research students. MATHEON funds projects and terminates roughly the bottom 10% each year. To receive funding, each project does a 15-minute presentation every two years.

Fields of study at MATHEON include optimization and discrete mathematics; numerical analysis and scientific computing; and applied and stochastic analysis. Organization is in terms of application areas: life sciences; logistics, traffic and telecommunication networks; production; circuit simulation and optoelectronic devices; finance; visualization; and education, outreach, and administration.

MATHEON professors visit high schools to talk to the students and show them interesting applications of mathematics. They also have a program for embedding high school teachers into a research group to better equip them to bring more modern topics into the schools

Teaching loads for faculty in the participating departments are very high, at 8–9 hours per week; however, research professors in MATHEON receive reduced teaching loads. There are few pure mathematicians in MATHEON, but those who are involved are well integrated.

ZIB and MATHEON have an excellent track record of both significant mathematical accomplishments and collaborations with industry on cutting-edge problems.

The ZIB group has been very successful in obtaining funding. The bottleneck is people; where will they come from? The WTEC visiting team's hosts reported difficulty in recruiting good people in applied mathematics. Very few ZIB students are non-German, although in the Berlin Mathematical School, which is a newly funded international graduate school, about 40% of the students are foreign.

CONCLUSIONS

The focus of ZIB is mathematics. The ZIB President, Prof. Peter Deuflhard, believes that the innovations of the future will have their roots in mathematics. ZIB is built around the philosophy that mathematical modeling is a source of innovation and understanding of problems, and that when mathematics gets translated into software, it makes an impact, and the real-world problems provide challenges, which in turn lead to advances in the mathematics. The problems ZIB researchers choose to focus on are very broad, often requiring substantial mathematical expertise. The ZIB has an excellent track record, substantial industrial collaboration that includes generation of successful spinoff companies, and amazing success in attracting funding. The only apparent limitation to its further growth is in obtaining appropriately trained people.

APPENDIX D. SURVEY QUESTIONNAIRE

The following questions were developed by the U.S. delegation to prepare for the visit. We hope that they will help you to understand some of our objectives during this tour of leading European sites involved in simulation-based engineering and science (SBE&S) research. We do not expect detailed answers to each of these questions. Therefore, please feel free to examine the list and determine which questions would be most appropriate for you and your organization. We would like to emphasize that we are not seeking proprietary information.

It is our goal that discussion of issues addressed by these questions will lead to a productive exchange of views that will benefit programs in SBE&S research in both of our countries. Of course, our delegation is prepared to share its perspectives on research activities in the US.

Our study has three primary areas for which we are assessing the current state and the future trends of SBE&S research. The primary thematic areas are:

- Materials
- Energy and sustainability, and
- Life sciences and medicine

We are also interested in the impact of infrastructure and other crosscutting issues on the advancement of the field of SBE&S. We have thus identified eight issues applicable to all of the thematic areas. These are:

- Multiscale simulation
- Validation, verification, and quantifying uncertainty
- Simulation software
- Big data and visualization
- Engineering design
- Next-generation algorithms and high performance computing
- Education and training
- Funding, organization, and collaboration

General

What are the major needs, opportunities or directions in SBE&S research over the next 10- and 20-year time frames?

What are the national and/or regional funding opportunities that would support research to meet these needs, and/or take advantage of these opportunities? Are these funding opportunities expanding?

Materials, energy and sustainability, life sciences and medicine

What major breakthroughs in these fields will require SBE&S; and which are you and or your colleagues pursuing? Within your institution, region, or country, are there identified targets of opportunity for applications of simulation either for scientific research or for engineering applications in these fields?

Which problems could benefit most from a 1-2 order-of-magnitude increase in computational power?

What are examples of major SBE&S successes or failures in these fields?

Do investigators, laboratories and institutions receive any financial compensation for patented inventions derived from their simulations?

Have any start-up companies spun-off based on simulation efforts in your lab? If so please describe them.

Multiscale simulation

Describe efforts and advances within your institution to couple multiple simulation methods in order to bridge multiple length and/or time scales.

Is the development of integrated multiscale modeling environments a priority in research funding in your country or region?

Validation, verification, and quantifying uncertainty

Describe efforts and advances within your institution to validate and verify codes and to quantify uncertainty in simulation-based predictions?

Simulation software

What percentage of code used in your group is developed in-house? What percentage of code is commercial? What percentage is open source? What percentage has been developed by others (e.g., under contract or by acquisition)? What are the biggest issues in using models/simulations developed by others? How easy/difficult is it to link codes to create a larger or multifaceted simulation environment?

Who owns the intellectual property rights (IP) to the codes developed in your group?

How do you deal with liability issues for products developed with codes from other sources?

Big data and visualization

What type of data and visualization resources do you need (and have access to) for SBE&S research?

Engineering design

What type of models/codes do you use, develop or conduct basic research on pertaining to different phases of engineered system design (conceptual, parametric optimization, operational/control)?

What are the data requirements for a comprehensive life-cycle model of an engineered system? Do you have repositories in place that provide the data?

How do you couple output of the models/simulations to the decision making process (including quantification of uncertainty/error in the predictions)?

What is the curriculum for training doctoral students in all aspects of designing engineered systems using simulation/modeling tools?

Are there efforts within your institution to couple physics-based models with macroscopic logic or econometric models?

Next-generation algorithms and high performance computing

Would you characterize the SBE&S research in your lab as needing and/or using primarily desktop, teraflop, or petaflop computing resources?

What are the computational bottlenecks in your simulation problems? What solutions exist now and in the near future, either from a hardware perspective, algorithm perspective, or both?

What is the state of the art in your community in discrete event simulations (model size, parallelization or thread limits, special techniques - time warp, rollback, etc.)?

Education and training

Is there a formal scientific computing or computational science and engineering graduate program at your institution? Or, if you are in a company, are there educational institutions/programs in your country or region that prepare students in formal scientific computing or computational science and graduate engineering effectively?

What level of computing expertise do your incoming graduate students or employees possess? Do you feel that they have the necessary background to conduct research in SBE&S or do they need extensive preparation after arriving before they can be productive researchers?

What kind of training is available in your organization for simulation and high performance computing? Is it adequate? Do you believe it will be adequate to address computing on multi-core and petascale architectures? What plans are in place for training programs in next-generation computing?

What fraction of students in your institution study/work in the area of SBE&S, and how has this fraction changed over the past 5-7 years?

What fraction of graduate students/postdocs in SBE&S comes from abroad? From which country do these researchers originate? How many students in your country would you estimate go abroad to earn PhDs in SBE&S?

After completing a PhD in an SBE&S-related field, what route do your students take? Do they accept postdoctoral positions in your country or abroad? What fraction eventually obtains permanent jobs in SBE&S? Is this a desired career path for students? Do they find many job opportunities related to their training in SBE&S? How does industry view students with training in SBE&S?

Funding, organization, and collaboration

What are the roles of academic, government and industrial laboratories in SBE&S research in your country?

Who pays for, and/or is responsible for, the development and sustainability of SBE&S infrastructure (include long term data storage costs, code maintenance, open-source software, etc.)?

What is the funding situation for SBE&S in your country? Example considerations: What are the major sources of funding? Over the past 5 years, has funding of SBE&S increased, decreased or remained roughly constant relative to funding of all of science and engineering research? Is most SBE&S research funded for single investigators, small teams, or large teams? What is the typical duration of SBE&S funding?

APPENDIX E. BIBLIOMETRIC ANALYSIS OF SIMULATION RESEARCH

Grant Lewison, *Evaluametrics Ltd, 50 Marksbury Avenue, Kew, Richmond, Surrey, TW9 4JF, UK*

INTRODUCTION

This analysis was carried out in order to assist the panel in two ways:

- to compare the outputs of the USA and other leading countries in simulation research, in terms of both volume and impact, so as to evaluate the position of the USA relative to that of other countries and groups of nations, and establish time trends; and
- to identify the institutions in Europe and the Far East that were active in simulation research, so as to inform the panel's selection of places to visit.

Bibliometrics is the quantitative study of publications, normally of papers in the peer-reviewed serial literature. It can provide objective evaluation data as a complement to the views of an expert panel, and is increasingly being used in this way on a national and institutional level, even sometimes on an individual level. However the conclusions drawn from bibliometrics analysis depend on the accuracy with which the subject area of interest (here, simulation research) is defined, and hence on the database of the bibliographic details of papers that has been created for analysis.

In this study, the original data were drawn from the Science Citation Index (SCI) on CD-ROM for the ten years 1996-2005; some additional data were taken from the Web of Science (WoS) version (which has a somewhat wider journal coverage) for more recent years to allow for any later changes to be seen. All the data were based on articles and reviews only because they alone embody substantive research findings.

For the first task, the comparison of the USA with other countries, data were obtained on both numbers of papers and also some measures of esteem or impact – percentage of reviews, potential citation impact (based on journal citation impact factors) and actual citation impact (citations to the individual papers). These are useful partial indicators of merit, and if they agree, one can have more confidence that the message they are conveying is reliable. For the second task, the identification of possible sites for the panel to visit, listings were made of the numbers of papers from different locations and an indication was given of what type of simulation was of primary interest (biological, chemical or physical).

METHODOLOGY

The first step was to write down a short definition of the subject area of simulation research. This was done by Professor Sharon Glotzer, the panel chairman, at a meeting with the author in Ann Arbor, MI, in April 2007, and it reads as follows:

Simulation involves the application of mathematical models using a computer to the study of the underlying physical and chemical processes, and prediction of the behaviour and properties of systems, including natural and artificial materials, flow in liquids and gases, energy at all scales including the cellular level, and biomedical *sequelae*.

This definition states both what is included and what is excluded, and served to define a “filter” that, when applied to the SCI would selectively identify relevant papers. The filter was developed by an interactive process between Professor Glotzer and the author. It consisted of three parts: specialist journals (Table E.1), specialist title words (Table E.2) and negative title words (Table E.3). Papers were selected if they were either in one of the named journals, or had one or more of the title words, or both, but were de-selected if they also contained one or more of the negative title words. The latter were mainly designed to exclude papers involving laboratory animals, but also experiments unless they were specifically designed to check simulations.

Table E.1. List of Specialist Simulation Journals (29-character max. abbreviated names)

<i>Calphad*</i>	<i>J Comput Chem</i>
<i>Combust Theory Model</i>	<i>J Comput Neurosci</i>
<i>Comput Biol Chem</i>	<i>J Comput Phys</i>
<i>Computation Mech</i>	<i>Macromol Theory Simul</i>
<i>Comput Method Appl Mech Eng</i>	<i>Math Model Method Appl Sci</i>
<i>Comput Phys Commun</i>	<i>Med Biol Eng Comput</i>
<i>Comput Fluids</i>	<i>Model Simul Mater Sci Eng</i>
<i>Int J Numer M*</i>	<i>Mol Simulat</i>
<i>J Comput Acoust</i>	<i>Theor Comput Fluid Dynamics</i>
<i>J Comput Appl Math</i>	

* = any character(s) or none

Table E.2. List of Positive Title Words Used to Define the Simulation Filter

1st principles	dynamic* and model*	Monte Carlo
<i>ab initio</i>	finite difference	multiscale and model*
Bayesian	finite element*	neural net*
boundary element*	finite volume	numeric* and model*
cellular automata	integral equation*	numerical calculation*
CFD	kinetic* and model*	numerical prediction
comput* and model*	lattice Boltzmann	particle in cell
computational	lattice gas	phase diagram*
continuous* optimiz*	Lennard Jones	phase equilibri*
density functional	Markov chain*	simulat*
discrete element*	mathemat* and model*	theoretical and computational
dissipative particle dynamics	MCMC	
dynamic optimization	molecular dynamics	

Table E.3. List of Negative Title Words Used to Disqualify a Paper for the Simulation File

experiment* not simulat*		pig	rat
mice	mouse	pigs	rats
monkey*	murine	rabbit*	

The filter was used to generate samples of papers that could be marked for relevance. This was done subsequently by three members of the panel, and the precision, p , was 0.86 while the recall, r , was only 0.45. These figures suggest that the filter could have been further improved by the addition of extra title words, or additional journals, in order to increase its recall.

The filter was applied to the SCI on CD-ROM, and all papers (articles and reviews) from the ten years, 1996-2005, had their bibliographic details (authors, title, document type, full source and addresses) downloaded to three MS Excel files, covering the years 1996-99, 2000-02 and 2003-05. Altogether, there were just over 149,000 papers in the three files.

The papers were now classified by major field, on the basis of the journals in which they were published, using a scheme devised by CHI Research Inc. (now The Patent Board, a subsidiary of IPIQ Inc.) for the US National Science Foundation. Nine fields were used (there are some others with small numbers of papers included in “others”), as listed in Table E.4, which also gives the percentage of papers in each.

Table E.4. Major Fields Used to Classify the Downloaded Simulation Papers, and % in Each.

<i>Field</i>	<i>%</i>	<i>Field</i>	<i>%</i>	<i>Field</i>	<i>%</i>
physics	29.6	mathematics	8.0	clin. medicine	5.7
engineering and techn	21.5	earth and space	7.9	biology	3.2
chemistry	16.2	biomed. research	7.1	other	0.8

The percentage of reviews is a new measure of research esteem and gives a simple indication of how a country’s (or an institution’s) senior researchers are seen by journal editors. It was calculated for papers from different countries. These percentages can be compared with the corresponding values for all science²⁴

The potential citation impact of each paper was determined from the five-year mean citation score of papers in the same journal and year. This is also a simple indicator of research esteem, and although it has been criticised as being inferior to counts of citations to actual papers, it actually measures something different. Values for the year 2002 for some leading journals are shown in Table E.5.

²⁴ Lewison G (2009) The percentage of reviews in research output: a simple measure of research esteem. *Research Evaluation*, in press.

Table E.5. Mean Five-year Citation Score (*i.e.*, cites in 2002-06 to papers published in 2002) for Papers in Some Journals Much Used by Simulation Researchers.

<i>Journal</i>	<i>PCI</i>
<i>Physical Review Letters</i>	24.1
<i>Journal of Physical Chemistry B</i>	15.4
<i>Journal of Chemical Physics</i>	12.3
<i>Journal of Computational Physics</i>	9.3
<i>International Journal for Numerical Methods in Engineering</i>	6.6
<i>International Journal for Numerical Methods in Fluids</i>	3.6
<i>Journal of Computational and Applied Mathematics</i>	2.6

The actual citation impact of the simulation papers was determined from another version of the SCI, the Web of Science (WoS). This has a rather wider journal coverage than the CD-ROMs, and this affects national outputs differently – generally boosting those from Far Eastern countries such as China more than those from European and North American ones. The simulation filter was first applied to the WoS in order to learn about any recent changes in national outputs (with an overlap to earlier years so that outputs could be calibrated against those from the CD-ROMs). The WoS also provided citation data on individual papers and groups (provided that they numbered fewer than 10,000). This allowed average actual citation scores for simulation papers from different countries to be determined, but only on an integer count basis. (A paper with one US address and two from France would count unity for each on integer counting, and 0.33 and 0.67 respectively on fractional counting.) However, the determination of potential citation impact was also done on a fractional count basis, which is technically more accurate because it allocates credit on a proportional basis..

Since one of the main purposes of the bibliometric analysis was to determine the relative position of the USA in simulation research, comparisons were made with selected countries in Europe and Asia, as listed in Table E.6. The digraph ISO codes are used hereinafter to designate the different countries: they are used as the last element in web domain names and so have become reasonably familiar to web users.

Table E.6. List of Countries Used for the Bibliometric Analysis.

<i>Code</i>	<i>Country</i>	<i>Code</i>	<i>Country</i>	<i>Code</i>	<i>Country</i>
AU	Australia	EUR12	(see note)	PL	Poland
BR	Brazil	FR	France*	RU	Russia
CA	Canada	IN	India	SE	Sweden*
CH	Switzerland*	IT	Italy*	TW	Taiwan
CN	China (Peoples Rep)	JP	Japan	UK	United Kingdom*
DE	Germany*	KR	South Korea	US	United States
ES	Spain*	NL	Netherlands*		

EUR12 = Eight countries marked * + Austria AT, Belgium BE, Denmark DK, Finland FI

The table lists eight of 12 western European countries that have been taken as a group for comparison with the USA, and five Asian countries that may be of interest to the panel.

RESULTS: NATIONAL COMPARISONS

Figure E.1 shows the numbers of simulation papers, year by year, with for comparison the number of papers in the SCI, both from the CD-ROMs. Because of late processing, the tally of 2005 simulation research papers is expected to be about 10% low, so an allowance has been made for this²⁵. Data for 2006 and 2007 have been taken from the WoS. However in 2004 and 2005, there were more papers in this version of the SCI, as shown in Table E.7. The data suggest that the coverage of simulation research was about 41% greater in these years, and so the WoS outputs have been reduced by this amount to give estimated values for the CD-ROMs (which were not available). Overall outputs were only about 34% higher.

Table E.7. Comparison of WoS and CD-ROM Outputs of Simulation Research and All SCI papers for 2004 and 2005. Note: CD-ROM data for 2005 increased by 10% to allow for late processing.

<i>Year</i>	<i>WoS</i>		<i>CD-ROM</i>		<i>Ratio</i>	
	<i>SCI</i>	<i>SIMUL</i>	<i>SCI</i>	<i>SIMUL</i>	<i>SCI</i>	<i>SIMUL</i>
2004	861651	25138	621998	17936	1.39	1.40
2005	913393	27742	703805	19563	1.30	1.42
2006	955610	30112	<i>707859</i>	<i>21356</i>		
2007	941543	27926	<i>697439</i>	<i>19806</i>		

Figures in *italics* are estimated on the basis of the mean ratio of WoS to CD-ROMs.

²⁵ The data for the simulation papers are for publication years, but for all science they were for database years and do not need correction for the late processing of some 2005 papers.

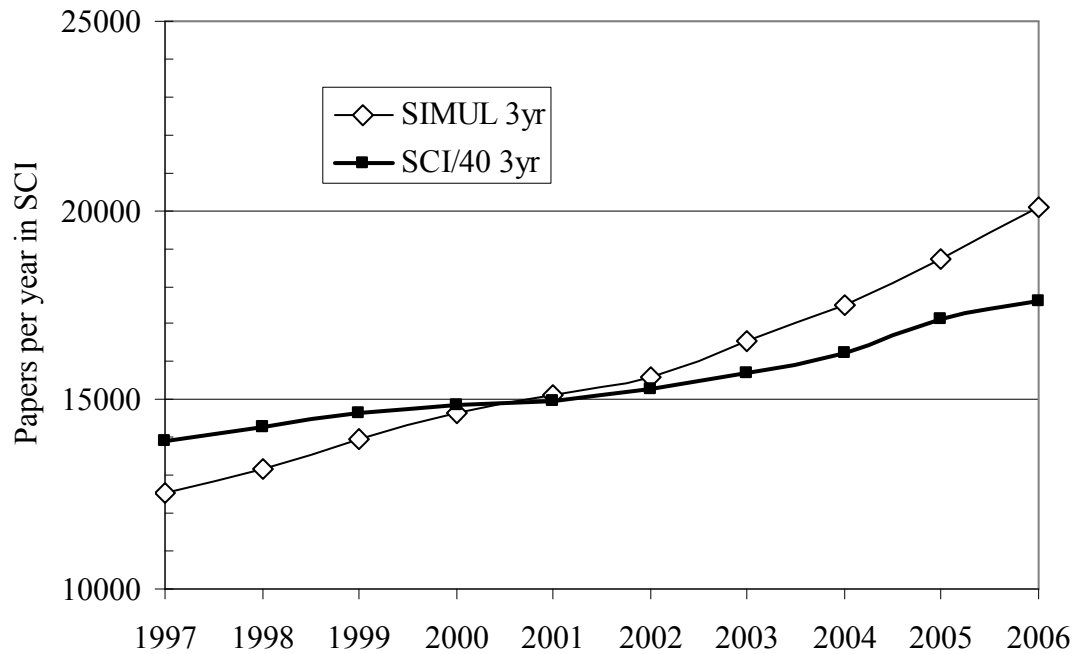


Figure E.1. Number of papers per year (three-year running means) in simulation research (open diamonds, light line) and in all science (/40; solid squares, heavy line).

It is clear that simulation research is growing much faster than science overall, the annual average percentage growth rate being 5.0% compared with 2.5% for all science.

The outputs of papers from individual countries are shown as fractional counts in Table E.8 for the three periods, 1996-99, 2000-02 and 2003-05. Data are also given for the 12 European countries as a group (with fractional counts the numbers can be simply added) and for the five “Asian tigers” of China, India, Singapore, South Korea and Taiwan (AS5).

**Table E.8. Annual Outputs of Simulation Papers (fractional counts) from Leading Countries and Regions, 1996-99, 2000-02 and 2003-05 (SCI on CD-ROM, with correction for late processing of 2005 papers).
Countries are ranked by overall mean.**

<i>ISO</i>	<i>1996-99</i>	<i>%, 96-99</i>	<i>2000-02</i>	<i>%, 00-02</i>	<i>2003-05</i>	<i>%, 03-05</i>	<i>Mean</i>
Wld	12881	100.0	15123	100.0	17955	100.0	15076
EU12	4497	34.9	5123	33.9	5805	32.3	5077
US	3775	29.3	4086	27.0	4643	25.9	4129
AS5	1746	13.6	2556	16.9	3452	19.2	2501
JP	940	7.3	1150	7.6	1257	7.0	1098
DE	957	7.4	1055	7.0	1126	6.3	1037
UK	964	7.5	1041	6.9	1099	6.1	1028
FR	742	5.8	865	5.7	964	5.4	846
CN	396	3.1	750	5.0	1358	7.6	791
IT	496	3.9	611	4.0	719	4.0	598
CA	478	3.7	502	3.3	587	3.3	518
RU	486	3.8	476	3.1	432	2.4	467
ES	320	2.5	382	2.5	508	2.8	395
AU	255	2.0	284	1.9	330	1.8	286
KR	188	1.5	302	2.0	393	2.2	284
NL	253	2.0	281	1.9	326	1.8	283
IN	221	1.7	243	1.6	336	1.9	262
SE	192	1.5	231	1.5	260	1.4	224
TW	165	1.3	227	1.5	262	1.5	213
CH	165	1.3	188	1.2	236	1.3	193
PL	144	1.1	210	1.4	230	1.3	190
BR	118	0.9	187	1.2	252	1.4	179

Both the USA and Europe12 have reduced their share of world papers (by 12% and 8% of their 1996-99 presences in 2003-05), but their absolute numbers have increased in successive periods. The increase in share has gone to the Asian countries, particularly China, whose output increased from 308 papers in 1996 to 1691 in 2005 (after correction for late processing). Its output overtook that of Japan in 2004, and has been growing at a massive 18.5% per year, compared with 3.7% for Europe12 and only 2.8% for the USA. The other Asian countries' outputs are growing at intermediate rates: 4.7% for Japan but 11.5% for South Korea.

It is worthwhile to see how these outputs compare with the countries' overall production of scientific papers: the ratio of percentage presences is their relative commitment (RC) to simulation research. This is shown in chart form for the whole decade in Figure E.2, which demonstrates that China has much the largest RC (among leading nations), and that the USA (and Canada) do slightly less simulation research than their overall presence in science would suggest. Japan appears to be relatively inactive in this field. Over the decade, Switzerland and Sweden have increased their RC, but South Korea, China and Russia, and in Europe, Germany and the UK, have decreased it.

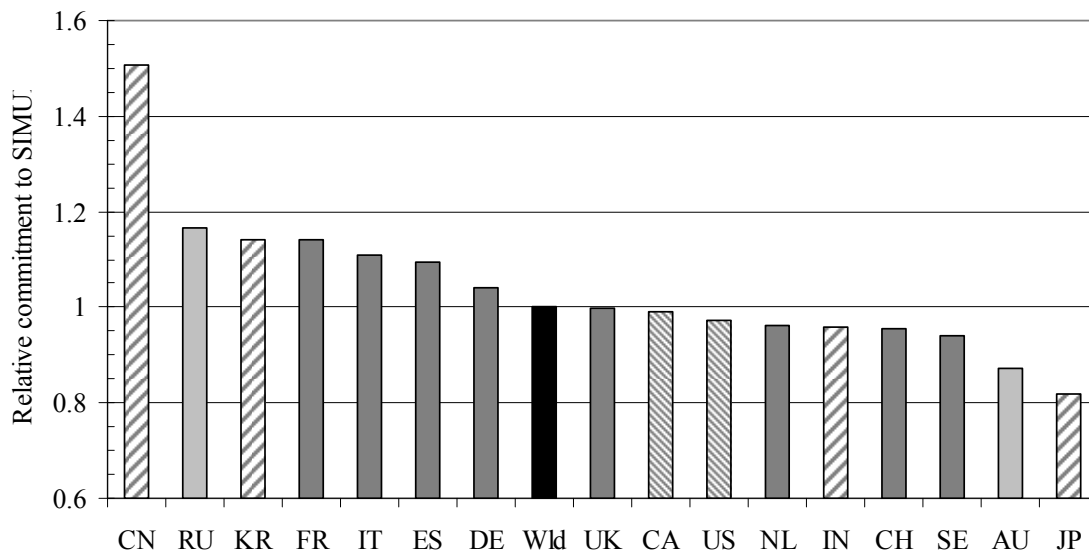


Figure E.2. Relative commitment of some leading countries to simulation research, 1996-2005 (integer counts). Shading shows geographical region of countries.

The percentage of reviews is a simple measure of the esteem in which a country's senior researchers are held, but it needs to be normalised with respect to the average value for all world papers to give a ratio. Within the total numbers of papers, the percentage of reviews has been steadily growing, as shown in Figure E.3. However it is clear that simulation has a much smaller percentage of reviews than the mean for all science, and that it has not been growing. This has to be taken into account when the performance of individual countries is considered.

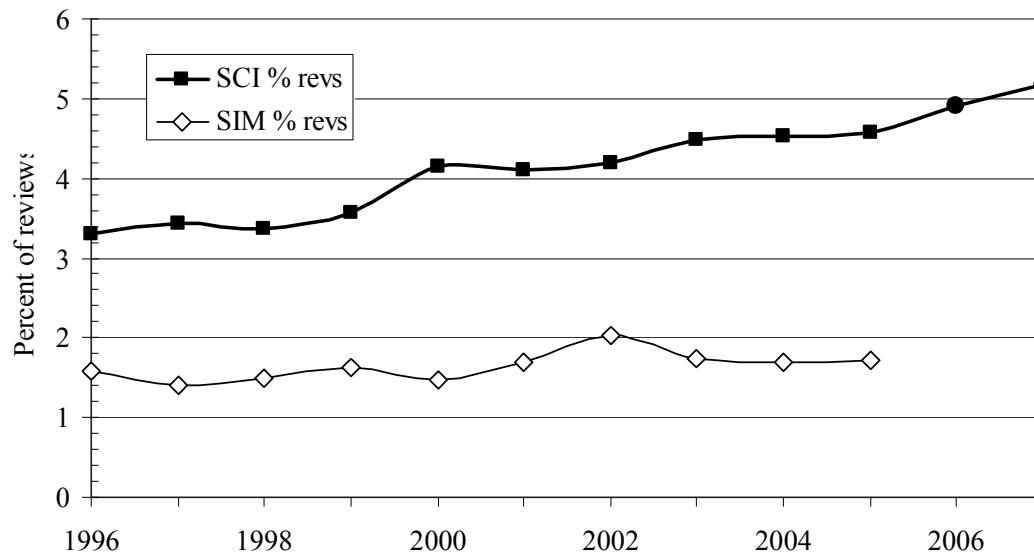


Figure E.3. Percentage of reviews (among articles + reviews) in simulation (open diamonds, light line) and in all science (solid squares, heavy line), 1996-2007. Values taken from CD-ROMs except for science in 2006, 2007 (solid circles).

Figure E.4 shows this ratio both for simulation (gray bars) and for all science (white bars) as the average ratios to the world values for the decade. The USA, closely followed by Germany, and also Spain and Russia, show to advantage in simulation, better than their positions in all science, whereas the UK's scientists are more highly esteemed in science overall than in simulation, as are the researchers in Australia. South Korea and China do not score well on this criterion.

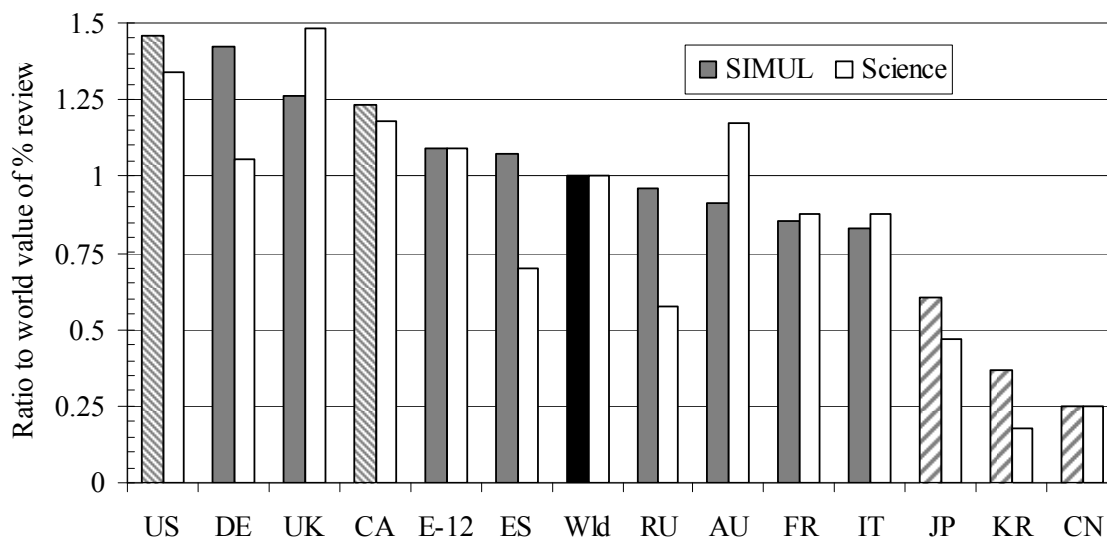


Figure E.4. Ratios of percentage of reviews for leading countries to world values, mean of values for 1996-99, 2000-02 and 2003-05, in simulation research (dark and shaded bars) and in all science (open bars). Based on integer counts.

A more conventional measure is the impact factor of the journals in which a country's papers are published. Figure E.5 shows the mean values for the last three years, 2003-05, based on both integer and fractional counts, which give a better indication of the potential impact of countries' research, particularly small countries, where international collaboration can mask their true performance on this measure.

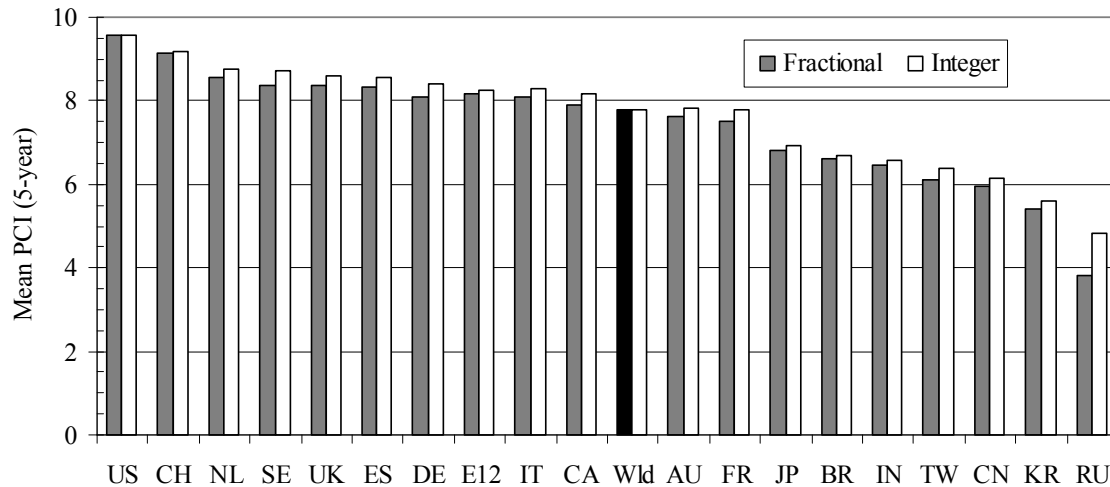


Figure E.5. Mean journal 5-year impact factor (PCI) for leading countries in simulation research, 2003-05, fractional (solid bars) and integer counts (white bars).

The USA leads also on this measure, but is closely followed by several European countries, notably Switzerland, the Netherlands and Sweden. There is a noticeable drop in journal impact factor from the level for the European countries to that for those outside Europe, notably countries in the Far East.

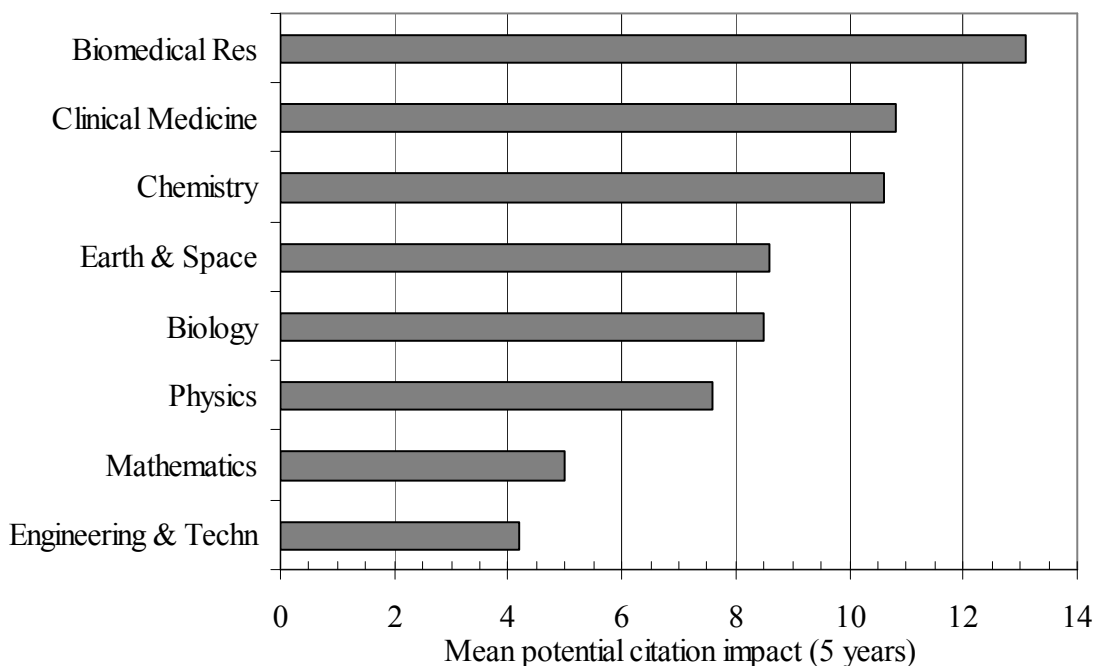


Figure E.6. Mean journal 5-year citation impact factor for simulation papers in different fields, 2003-05.

One of the reasons for the USA's superior performance may be that its simulation research is relatively more in those fields where papers are published in high impact journals. There are quite large differences between the fields in this respect, see Figure E.6, with biomedical research being published in journals having more than three times as many cites, on average, as work in engineering.

Some of the differences between countries in the PCI of their papers can be explained by their relative concentration on different aspects of simulation, because, for example, China and Japan do much more engineering and much less biomedical research. Table E.9 shows the effects of this differential field pattern, on the simplistic assumption that all countries' simulation research in each field had the world mean PCI.

Table E.9. Distribution of Simulation Research Between Fields for the World, the USA, EUR12, China and Japan, 2003-05, Percentages, and Estimates of the Corresponding Contribution (cont) of Each Field to the Overall PCI (shown in bold).

	<i>World</i>		<i>USA</i>		<i>EUR-12</i>		<i>China</i>		<i>Japan</i>	
	%	cont	%	cont	%	cont	%	cont	%	cont
Biology	3.1	0.26	3.7	0.32	3.1	0.27	1.5	0.13	1.4	0.12
Biomed. res	7.3	0.96	9.0	1.18	7.4	0.97	4.2	0.55	5.4	0.71
Chemistry	15.1	1.60	13.6	1.44	14.4	1.52	15.9	1.68	13.2	1.40
Clinical Med.	5.6	0.61	7.5	0.81	5.8	0.62	3.0	0.32	3.9	0.42
Earth & Space	7.9	0.68	10.7	0.92	8.4	0.72	4.7	0.41	7.8	0.67
Eng & Techn	21.9	0.92	19.5	0.82	18.8	0.79	25.6	1.08	25.3	1.06
Mathematics	8.5	0.42	8.1	0.41	10.2	0.51	9.1	0.45	3.4	0.17
Physics	29.9	2.27	26.9	2.04	31.2	2.37	35.8	2.72	39.3	2.99
Others	0.7	0.07	1.0	0.09	0.7	0.06	0.3	0.02	0.3	0.03
Total	100	7.79	100	8.03	100	7.84	100	7.36	100	7.56

The differences between the mean PCI values are not large, 0.67 between the USA and China. This only accounts for one fifth of the difference between the two countries seen in Figure E.5, but it does show that the effects of field distribution should not be neglected when national research evaluations are being made.

The actual citation impact was determined from the WoS for papers from the leading countries for the year 2004, with citations counted over the four years, 2004-07. This shows that the papers from some European countries, Switzerland, Denmark and the Netherlands, are actually better cited, on average, than those from the USA. However the analysis shows that papers from the four Asian countries (Japan, India, China and South Korea) are all rather poorly cited. The numbers of papers from the individual countries are not large, so this result should be treated with caution and would need to be confirmed for other publication years.

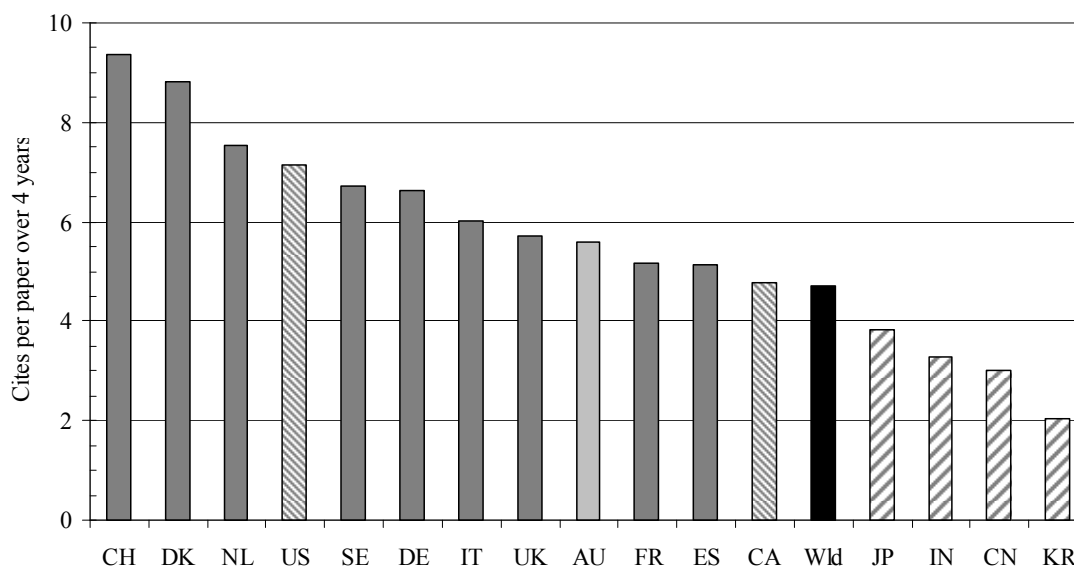


Figure E.7. Mean actual citation impact for 2004 simulation papers in four years (2004-07) from leading countries.

RESULTS: LEADING INSTITUTIONS

Because of the lack of unification of institution names, a different approach was used, and in each of the selected countries (which it was thought possible that the panel might visit) the leading cities represented in the addresses of the papers from the years 2003-05 were tabulated, on both fractional and integer counts using a special macro written by Dr Philip Roe, and then the leading institutions within these cities were identified (not listed here). The intention here was to give additional weight to cities where there were more than one active institution working in simulation research so as to make the panel's visit program more efficient. In addition, some indication was given of the relative balance of simulation research effort between the main fields, in the form of ratios of research output to the overall mean. The purpose here was to indicate to the panel which expertise should be available from within its membership (biology, chemistry, engineering or physics) so that there would be at least one person present to take part in the discussions from the discipline in which the proposed institution for the visit was particularly strong.

Table E.10 shows the results for 44 cities in 12 western European countries and Table E.11 shows the results for 20 cities in four Far Eastern countries. The cities listed are not necessarily the ones selected for visits by the panel because of other considerations, such as the presence of particular individuals known to panel members, or simple logistics.

Table E.10. Leading Cities in Western Europe, 2003-05, in Terms of Their Outputs of Papers on Simulation Research. Listing in descending order by fractional counts (*Frac*), with cities listed if they had more than 100 papers.

<i>ISO</i>	<i>City</i>	<i>Frac</i>	<i>Int</i>		<i>ISO</i>	<i>City</i>	<i>Frac</i>	<i>Int</i>
FR	Paris	368	661		DE	Garching	137	243
ES	Madrid	328	497		UK	Manchester	136	218
IT	Rome	283	441		SE	Uppsala	134	231
UK	Cambridge	279	475		SE	Gothenburg	133	197
CH	Zurich	270	425		NL	Eindhoven	129	185
UK	Oxford	249	411		DE	Karlsruhe	128	204
IT	Milan	236	367		IT	Trieste	127	214
DE	Berlin	227	379		IT	Turin	127	198
ES	Barcelona	217	343		DE	Dresden	123	194
AT	Vienna	200	326		UK	Edinburgh	118	190
DE	Stuttgart	197	311		UK	Leeds	116	172
UK	London SW	192	302		BE	Ghent	113	152
UK	London WC	189	324		FR	Orsay	112	221
SE	Stockholm	188	300		DE	Aachen	111	158
NL	Delft	183	262		FR	Marseille	110	190
FR	Toulouse	182	299		DE	Darmstadt	105	153
NL	Amsterdam	176	277		BE	Brussels	104	183
FI	Helsinki	149	220		DE	Munich	104	173
FR	Grenoble	146	269		FR	Villeurbanne	103	160
CH	Lausanne	143	234		FR	Montpellier	103	187
IT	Bologna	143	220		SE	Lund	103	155
BE	Louvain	138	223		IT	Pisa	101	163

Table E.11. Leading Cities in the Far East (China, Japan, South Korea and Taiwan), 2003-05, in Terms of Their Outputs of Papers on Simulation Research. Listing in descending order by fractional counts (*Frac*), with cities listed if they had more than 125 papers.

<i>ISO</i>	<i>City</i>	<i>Frac</i>	<i>Int</i>		<i>ISO</i>	<i>City</i>	<i>Frac</i>	<i>Int</i>
CN	Beijing	916	1273		JP	Aichi	239	376
JP	Tokyo	734	1137		KR	Taejon	232	333
JP	Ibaraki	428	691		TW	Taipei	226	306
KR	Seoul	413	584		CN	Nanjing	223	349
CN	Hong-Kong	377	591		JP	Fukuoka	143	211
CN	Shanghai	369	510		TW	Hsinchu	132	175
JP	Osaka	291	451		CN	Changchun	129	163
JP	Kanagawa	273	496		CN	Dalian	127	170
JP	Miyagi	267	395		CN	Xian	126	181
JP	Kyoto	259	403		CN	Wuhan	125	183

The listing is not perfect as it ignores the research efforts of commercial companies who may publish only occasional papers, and also those of major collaborative research centres such as CERN in Geneva, Switzerland, where the papers may have literally hundreds of addresses of which only a few will be local ones. London has been divided into several parts in Table E.10 on the basis of the postcode areas in the addresses: London SW connotes Imperial College and London WC connotes University College. If these were combined (with other postcode areas), London would easily top the list, but its output would still be much less than those of Beijing and Tokyo.

The bibliometric analysis also indicated to panel members the relative specialisation of the different cities, in terms of whether the numbers of their papers in the different fields were above or below what might have been expected on the basis of the overall distribution of papers between fields, see Figure E.6 and Table E.9. The results are shown in Table E.12 for ten leading European cities and six in the Far East.

Table E.12. Relative Concentration of Simulation Research in Six Fields (BIOL = biology, biomedical research and clinical medicine; CHEM = chemistry; EA&SP = earth and space; ENGR = engineering and technology; MATH = mathematics; PHYS = physics), 2003-05, for Ten European and Six Far East Cities. Values over 1.41 shown in bold; values below 0.71 shown in *italics*.

<i>City</i>	<i>ISO</i>	<i>BIOL</i>	<i>CHEM</i>	<i>EA&SP</i>	<i>ENGR</i>	<i>MATH</i>	<i>PHYS</i>
London	UK	1.70	0.73	1.06	0.84	1.06	0.81
Paris	FR	0.86	0.93	1.51	<i>0.63</i>	1.54	1.11
Cambridge	UK	1.22	0.99	1.53	<i>0.68</i>	0.74	1.06
Zurich	CH	1.25	1.28	1.36	<i>0.66</i>	1.11	0.86
Oxford	UK	2.00	0.74	0.91	<i>0.46</i>	0.93	1.04
Milan	IT	1.35	1.01	0.81	0.99	1.38	0.78
Berlin	DE	0.87	1.13	0.76	<i>0.46</i>	1.30	1.39
Barcelona	ES	0.75	1.70	1.05	<i>0.60</i>	1.23	1.00
Stuttgart	DE	<i>0.28</i>	0.87	<i>0.32</i>	1.31	1.63	1.25
Amsterdam	NL	1.52	1.49	0.90	<i>0.33</i>	1.05	0.94
Beijing	CN	<i>0.59</i>	0.76	1.01	1.21	1.03	1.19
Tokyo	JP	0.71	<i>0.65</i>	1.42	1.06	<i>0.47</i>	1.33
Ibaraki	JP	<i>0.53</i>	0.76	1.12	1.01	<i>0.12</i>	1.60
Seoul	KR	<i>0.62</i>	<i>0.69</i>	1.25	1.36	0.83	1.09
Hong Kong	CN	0.74	<i>0.50</i>	0.83	1.59	1.67	0.82
Shanghai	CN	<i>0.57</i>	0.93	<i>0.39</i>	1.42	1.09	1.10

Again, this listing is not necessarily exact because company outputs may well not be fully represented, and a city may contain several institutions whose outputs may differ from the average for the locality. But it does suggest what expertise should be available within sub-panels, *e.g.*, biology/medicine for Amsterdam, London and Oxford, chemistry for Amsterdam and Barcelona and physics for Ibaraki and Berlin.

CONCLUSIONS

The main lessons from this analysis are as follows:

- Simulation represents (after correction for the lack of recall of the filter) about 5% of the papers in the SCI, but as a subject it is growing relatively quickly, at about 5% per year compared with 2.5% for science as a whole (Figure E.1).
- Within simulation research, the country with the largest output of papers is the United States, with a fractional count of 26% of world papers in 2003-05. This is less than that of 12 leading western European countries combined (32%). Both outputs are declining as a percentage presence in the world, but are increasing in absolute amount (Table E.8).
- China's output was second highest, reaching 13% of the world total (integer counts) in 2007, followed by that of Germany and the UK (8%), France (7%) and Japan (6%).
- China also had the highest relative commitment to simulation research over the decade 1996-2005, with 50% more papers than expected on the basis of its overall output. The US published about 10% fewer papers than expected on this basis (Figure E.2).
- Three measures of esteem were determined: percentage of reviews, potential and actual citation impact. The USA was in first place on the first of these (Figure E.4), followed by Germany and the UK. It was also in first place for PCI (Figure E.5), followed by Switzerland, but only in fourth place for actual citations to 2004 papers (Figure E.7).
- The Far East countries all scored relatively poorly on each of these three indicators; this was partly because they concentrated their efforts in fields where the papers were in low impact journals and received few citations (such as mathematics and engineering, compared to biomedical research; Figure E.6).
- The tally of papers for 2003-05 enabled lists of the leading cities in western Europe and the Far East to be produced (Tables E.10, E.11) for the information of the panel members when they were deciding which sites to visit. The relative specialisation of these cities in different scientific fields was also determined (Table E.12) in order to help the panel to arrange appropriate sub-panel membership for its visit program.

APPENDIX F. GLOSSARY

211	trans-centenary key construction in higher education projects (P.R. China)
863	high-tech program/project (P.R. China)
985	projects focused on building world-class universities in the 21st century (P.R. China)
3D-RISM	three-dimensional reference interaction site model of Japan's IMS
ADME/Tox	absorption, distribution, metabolism, excretion, and toxicity (assay; screening)
AIAA	American Institute of Aeronautics and Astronautics
Algorithm	a sequence of instructions, often used for calculation and data processing;...a method in which a list of well-defined instructions for completing a task will, when given an initial state, proceed through a well-defined series of successive states, eventually terminating in an end-state.
ASCE	American Society of Civil Engineers
BABE	burn at both ends
CAE	computer-assisted engineering
CAE	Chinese Academy of Engineering (PRC)
CAPE	computer-aided process engineering
CAS	Chinese Academy of Sciences (PRC)
CBS	Center for Biological Sequence Analysis at the Technical University of Denmark (DTU)
CCP	collaborative computational projects (United Kingdom)
CFD	computational fluid dynamics
CFD/FEM	computational fluid dynamics/finite element method
CNOOC	China National Offshore Oil Corporation
COE	centers of excellence (Japan, MEXT program/University of Tokyo)
CoW	circle of Willis (a ring-like arterial structure sitting at the base of the brain whose main function is to evenly distribute oxygen-rich arterial blood to the cerebral mass)
CPU	central processing unit
CRDL	(Toyota) Central R&D Labs, Inc.
CREST	core research for evolutionary science and technology program of the Japan Science and Technology Agency, JST
CSL	Computational Science Laboratory (Mitsubishi)
CSE	computational science and engineering
CSTP	Council for the Science and Technology Policy (Japan)
CT	computed tomographic
DFG	German Research Foundation (Deutsche Forschungsgemeinschaft)
DFT-GGA	density functional theory- generalized gradient approximation
DG	distributed generation
DMFT	density matrix functional theory (also, dynamical mean field theory)
DPD	dynamic panel data

DSA-CT	digital subtraction angiography-computed tomography
EBI	Energy Biosciences Institute
EEG	electroencephalography
EMSL	Environmental Molecular Science Laboratory at the Department of Energy's (DOE's) Pacific Northwest National Laboratory (PNNL)
EPRI	Electric Power Research Institute
ESC	Earth Simulator Center of the Japan Agency for Marine-Earth Science and Technology
FE	finite element
FEAP	finite element analysis program
FEM	finite element method
flops	(or FLOPS or flop/s) floating point operations per second
FMO	fragment molecular orbital
fMRI	functional magnetic resonance imaging
FPGA	field-programmable gate array
FPMD	first principles molecular dynamics
GGA DFT	generalized gradient approximation density functional theory
GPU	graphic processing unit
GSIC	Global Scientific Information and Computing Center at the Tokyo Institute of Technology
gTOW	Genetic tug of war
Hamiltonian	In quantum mechanics, the Hamiltonian H is the observable corresponding to the total energy of the system; ...the set of possible outcomes when one measures the total energy of a system. (http://wikipedia.com)
HEP	high-energy physics
HPC	high-performance computing
HPRC	high-performance reconfigurable computing
ICME:	integrated computational materials engineering
IEEE	Institute of Electrical and Electronics Engineers
IP	intellectual property
IRMOF	isoreticular metal organic framework
IUPS	International Union of Physiological Sciences
IWM	Fraunhofer Institute for Mechanics of Materials (Institut Werkstoffmechanik)
JLPSM	Joint Laboratory of Polymer Science and Materials of the Institute of Chemistry, Chinese Academy of Sciences (ICCAS)
LANL	Los Alamos National Laboratory
LBM	lattice-Boltzmann method
LES	large eddy simulation
LIDAR	light detection and ranging
Linpack	A collection of FORTRAN subroutines that analyzes and solves linear equations and linear least-squares problems.

LLNL	Lawrence Livermore National Laboratory
MCC	Mitsubishi Chemical Corporation
MCHC	Mitsubishi Chemical Holdings Company
MCRC	Mitsubishi Chemical Group Science and Technology Research Center
MD	molecular dynamics
MEMS	microelectromechanical systems
METI	Ministry of Economy, Trade, and Industry of Japan
MEXT	Ministry of Education, Culture, Sports, Science, and Technology of Japan
MLWF	maximally localised Wannier function
MM	molecular mechanics
MOF	metal organic framework
MOST	Ministry of Science and Technology (P.R. China)
MOU	memorandum of understanding
MPI	message-passing interface parallel programming standard, the de facto industry standard
MRA	magnetic resonance angiography
NAREGI	National Research Grid Initiative (Japan)
NCAR	National Center of Atmospheric Research (Japan?)
NCSA	National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign
NEDO	New Energy Development Organization (Japan)
NEMS	nanoelectromechanical systems
NGSP	Next Generation Supercomputing Project (Japan/Riken/IMS)
NINS	National Institutes of Natural Sciences (Japan)
NIST	National Institute of Standards and Technology
NMR	nuclear magnetic resonance
NNSA	National Nuclear Security Administration
ODE	ordinary differential equation
OECD	Organisation for Economic Co-operation and Development
OGSA	open grid services architecture
OLED	organic light-emitting diode
ONIOM	our own N-layered integrated molecular orbital and molecular mechanics
PAYAO	Web 2.0 community tagging system
PDE	partial differential equations
PDF	probability density function (distribution)
PET	photoinduced electron transfer
PGAS	partitioned global address space programming languages
PHG	parallel hierarchical grid

PI	principal investigator
PMS	process modeling and simulation
PPD	parallel partitioning diagonal
PRACE	partnership for advanced computing in Europe
QCD	quantum chromodynamics
QM	quantum mechanical
RLC	(circuitry consisting of) resistive, inductive, and capacitive (elements)
RMB	ren min bi (the currency of the People's Republic of China)
RRPS	ratio of reduction in power supply
SAC	symmetry adapted cluster
SAC-CI	SAC-configuration interaction
SBGN	systems biology graphical notation
SBI	Systems Biology Institute (in the Department of Systems Biology at the Japan Foundation for Cancer Research in Tokyo, Japan)
SBML	systems biology markup language
SBW	systems biology workbench
SCF	self-consistent field
SC-PRISM	self-consistent polymer reference interaction site model
SIMS	secondary ion mass spectrometry
SNL	Sandia National Laboratory
SOC	self-organized criticality
SPDE	stochastic partial differential equation
Speedup	In parallel computing, speedup refers to how much a parallel algorithm is faster than a corresponding sequential algorithm. Linear speedup or ideal speedup is obtained when $S_p = p$ [p is the number of processors]; that is, ideally, when running an algorithm with linear speedup, doubling the number of processors doubles the speed; this is considered very good scalability.
super SINET	super science information network
TCRDL	Toyota Central R&D Labs, Inc
Tflops	one trillion floating point operations per second
TSUBAME	Tokyo tech supercomputer and ubiquitously accessible mass storage environment
UPC	unified parallel C (programming language)
UQ	uncertainty quantification
V&V	validation and verification
WSRF	Web services resource framework