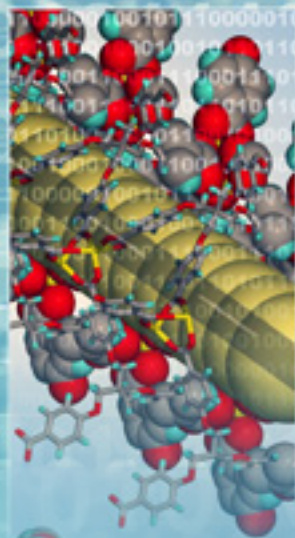


Harnessing the Department of Energy's High-Performance Computing Expertise to Strengthen the U.S. Chemical Enterprise

A workshop hosted by the Council for Chemical Research and the U.S. Department of Energy



U.S. DEPARTMENT OF ENERGY



THE COUNCIL FOR CHEMICAL RESEARCH

Harnessing the
Department of Energy's
High-Performance Computing Expertise
to Strengthen
the U.S. Chemical Enterprise

March 10 – 11, 2011
Rockville, Maryland

*Sponsored by the U.S. Department of Energy Office of Science's
Office of Basic Energy Sciences, Office of Advanced Scientific
Computing Research, and the Council for Chemical Research*

Acknowledgments

Harnessing the Department of Energy's High-Performance Computing Expertise to Strengthen the U.S. Chemical Enterprise

A Workshop hosted by the Council for Chemical Research and the U.S. Department of Energy

Organizing Committee

Dr. Steven Ashby

Pacific Northwest National Laboratory

Dr. Peter Cummings

Vanderbilt University / Oak Ridge National Laboratory

Dr. Larry Curtiss

Argonne National Laboratory

Dr. Krish Doraiswamy

E.I. Du Pont de Nemours & Co.

Dr. Susan Fitzwater

The Dow Chemical Company

Dr. Nikola Juhasz

Arkema, Inc.

Mr. Paul Mendez

Council for Chemical Research

Dr. Mark Pederson

DOE Office of Basic Energy Sciences

Dr. Walt Polansky

DOE Office of Advanced Scientific Computing Research

Dr. Eric Rohlifing

DOE Office of Basic Energy Sciences

Dr. Matthew Tarr (Chair)

University of New Orleans / Council for Chemical Research

Members of the Report Writing Committee

David Dixon

The University of Alabama

Michel Dupuis

Pacific Northwest National Laboratory

Bruce Garrett

Pacific Northwest National Laboratory

Joseph Golab

INEOS Technologies

Jeffrey Neaton

Lawrence Berkeley National Laboratory

Charity Plata

Pacific Northwest National Laboratory

Matthew Tarr

University of New Orleans / Council for Chemical Research

Jean-Francois Tomb

E.I. DuPont de Nemours & Co.

Executive Summary

For decades, the United States (U.S.) has been a world leader in many areas of science and technology. In recent years, increased economic and technological pressure on a global scale has threatened U.S. economic and technological superiority, as well as compromised national security issues. While the U.S. maintains many strengths derived from abundant natural and human resources, excellent higher education systems, and strong federal laboratories, growing international competition requires this country to develop well-focused strategies for maintaining world leadership in economic, technological, and security sectors. Without such focused approaches, the U.S. risks falling behind international competitors and suffering from severe economic and security consequences.

A key strategy for maintaining U.S. economic and technological strength involves improving innovation—the process of translating new ideas and basic knowledge into real products and services. Historically, the U.S. has made and continues to make a large investment in basic research and development in numerous science, engineering, and technology fields. This investment has resulted in robust institutions of higher education and excellent national laboratory facilities. The combination of these resources provides the U.S. with a distinct advantage over many international competitors. However, barriers that inhibit the translation of these resources into improved private sector competitiveness still remain. To bolster U.S. competitiveness and simultaneously create new jobs, improved transfer of academic and government sector expertise into the private domain is both necessary and paramount.

Energy and energy technology, including alternative energy technologies, are rapidly growing economic sectors. Not only is there tremendous capacity for economic growth in these areas, but there also is a direct connection between energy resources and U.S. national security and prosperity. The U.S. has the largest per capita energy consumption in the world, and dependence on foreign energy is a major vulnerability. Consequently, strategies for energy independence are urgently needed. One important strategy is to use existing Department of Energy (DOE) tools and expertise to stimulate innovation that will lead to economic growth, creation of new jobs, and improved energy independence. High-performance computing (HPC) is one area where the DOE has developed extensive expertise and capability. However, this expertise currently is not optimally shared with or used by the private sector to speed product development, enable industry to move rapidly into new areas, improve product quality, or minimize energy consumption. Such utilization would lead to substantial competitive advantages in global markets and yield important economic returns for the U.S., including increased gross domestic product (GDP) and job growth.

To stimulate the dissemination of DOE's HPC expertise, the Council for Chemical Research (CCR) and the DOE jointly held a workshop on this topic. As a starting point, four important energy

topic areas were chosen as the focus of the meeting: Biomass/Bioenergy, Catalytic Materials, Energy Storage, and Photovoltaics. Academic, industrial, and government experts in these topic areas participated in the workshop to identify industry needs, evaluate the current state of expertise, offer proposed actions and strategies, and forecast the expected benefits of implementing those strategies.

Overall findings of the workshop indicate there is little current translation of DOE HPC expertise into the private sector. Furthermore, a number of limitations in HPC tools exist. Areas where additional development is needed include:

1. Creation and maintenance of databases, including retrieval strategies
2. Model development addressing aging/decomposition of materials
3. Model development explicitly for new materials design
4. Computing approaches that couple multiple time and length scales
5. Model development for charge carrier transport
6. Methods to handle uncertainty prediction and quantification.

Additional interaction is needed with industry to transfer expertise gained by government sponsorship into the private sector. By addressing these HPC limitations and implementing strategies to improve transfer to the private sector, opportunities for U.S. economic growth will be augmented, while dependence on foreign energy will be decreased.

The workshop also provided a forum to expound key findings within each topic area. Within the biomass/bioenergy arena, tremendous potential exists for renewable and domestic energy resources. However, considerable challenges exist in realizing this potential. Flexible model systems with high temporal and spatial resolution for assessing risk, variability, vulnerability, and resilience in the biofuels systems are needed. Design and development of reaction systems to produce fuels from biomass, both by enzymatic and non-enzymatic pathways, require substantial fundamental knowledge that can be gained from HPC. In addition, HPC can be applied to solve complicated land-use and supply-and-demand questions that will afford the most successful use of domestic resources. Computational methods can dramatically aid in sustainable development because they offer clearer pictures of the long-term impacts of industrial and societal decisions.

Products synthesized from catalytic processes account for 20–30% of the U.S. GDP, while 90–95% of all chemical production depends on catalysis. The energy sector is heavily dependent on catalysts (considered as “catalytic materials” at the workshop). Several constraints require

careful development of new catalysts: 1) the cost of replacing a catalyst in an industrial process is high, even when the same catalyst is used to replace aging materials; 2) common elements used in catalysts are rare in the Earth's surface, and world demand for these rare elements (often mined outside of the U.S.) is rapidly increasing; and 3) development of new catalysts typically is an Edisonian process, which relies on trial and error guided by intuition and insight and can be tedious and time consuming. Computational tools for catalyst development must include mechanisms that can generate, manage, store, and analyze large computational data sets in conjunction with experimental data, as well as quantify uncertainty in models and compute uncertainties in computational data. Computationally guided design will greatly accelerate this development process and enable deployment of new catalytic processes in a much shorter time than now possible.

Energy storage is a crucial element to meeting U.S. energy needs, especially because many alternate energy sources are intermittent. Batteries of the future will require higher energy/power densities, lower cost, longer life, lower footprints, and a high degree of safety. Applying computational methods to energy storage requires multiple models that function across time and size scales, reflecting the multiscale nature of behavior in these systems. HPC methods applied to energy storage also must include improved abilities to deal with interfacial systems and materials degradation.

Because the amount of solar energy reaching the Earth's surface is tremendous—one hour of sunlight contains enough energy to meet the world's energy usage for an entire year—photovoltaic applications offer tremendous potential for delivering future energy sources. Successful utilization of solar energy requires the discovery of novel, high-performance materials; transformational advances in understanding the energy conversion process; control of material degradation processes; and a systems approach to managing, integrating, and manufacturing all solar cell components. Currently available HPC tools have not been effectively used to address these issues. Application of current tools, as well as developing new ones directly applicable to photovoltaics, will result in products that can compete globally, spurring new job and economic growth and bolstering U.S. energy independence.

To capitalize on current resources, HPC expertise must be disseminated rapidly to the private sector. Additional focused workshops in each of the four individual topic areas should be held in the immediate future. These workshops should involve DOE experts, leading academic authorities, and industrial leaders from both large and small firms. Stakeholders in energy-related businesses and representatives from software and hardware companies also should be involved. The process for developing government-industrial cooperative agreements needs to be streamlined, and viable models for managing intellectual property must be developed. To that end, legal representatives from industrial and government sectors also should be involved in any future workshops.

Acronyms and Abbreviations

ANL	Argonne National Laboratory
ASCR	Advanced Scientific Computing Research
BioKDF	Bioenergy Knowledge Discovery Framework
CEES	Center for Electrical Energy Storage
CFD	computational fluid dynamics
CIC	Center for Innovation through Computations
DFT	density functional theory
DOE	U.S. Department of Energy
<i>E. coli</i>	<i>Escherichia coli</i>
EFRC	Energy Frontier Research Center(s)
EPIC	Environmental Policy Integrated Climate model
HPC	high-performance computing
IP	intellectual property/properties
GDP	gross domestic product
GPU	graphical processing unit(s)
Li-ion	lithium-ion
MD	molecular dynamics
MM	molecular mechanics
PV	photovoltaic(s)
NIST	The National Institute of Standards and Technology
OCP	open circuit potential
ORNL	Oak Ridge National Laboratory
ORR	oxygen reduction reaction(s)
QM	quantum mechanics/mechanical
SEI	solid electrolyte interphase
SWAT	Soil and Water Assessment Tool
TS	transition state
TST	transition state theory
U.S.	United States

Optimizing the Bioenergy Supply Chain through the Application of High-Performance Computing

Biomass and bioenergy will play a key role in the production of renewable fuels and chemicals. Capitalizing on this energy source requires major science and engineering developments that span scales from electronic to the biorefinery to the ecosystem.

A holistic approach is necessary that accounts for the following:

- Feedstock availability, variability, and handling
- The science of biomass processing
- Integration of products into the existing fuels and chemicals infrastructure
- End users
- Economic, societal, and ecological impacts of extensive biomass utilization.

Among renewable energy technologies, biomass illustrates the complexity and conflicts associated with sustainability (Robertson et al. 2008 and Kline et al. 2009). No other renewable energy technology has received the scrutiny and analysis devoted to the debate over its sustainability. Still, researchers continue to struggle with a lack of answers to satisfy the questions raised by both private sector decision makers and government policymakers. Biofuels face substantial challenges on various fronts: potential land and water degradation, carbon dioxide (CO₂) emissions, and shocks to food supply and food security. This turmoil is not surprising given the exceedingly complex bioenergy supply chain and associated challenges, including:

- Raw feedstocks with different compositions and properties
- Pre-processing methods
- Storage effects
- Temporal, local, and regional variations
- Biomass conversion processes
- Biorefinery configurations
- Biocatalysts optimization for rates, titers, and yields of various biofuels
- Transporting logistics for raw and pre-processed feedstocks and intermediate and final products
- Sensitivity to petroleum industry price fluctuations and agricultural and energy policy.

The environmental, economic, and energy security implications of the bioenergy supply chain are necessarily linked to all of these systemic features. Optimized solutions will be configurations of all these elements.

INDUSTRY NEEDS

Sustainability and security of the bioenergy supply chain

Much well-executed, but disparate, modeling work has been conducted to evaluate the sustainability and security of bioenergy and bioproducts (Sheehan 2009). Most studies have addressed overall life cycle energy and energy security impacts (the proverbial, low-hanging analytic fruit), while others have addressed greenhouse gas impacts with mixed success. A few researchers even have dabbled in coupling social and economic systems with natural systems.

A comprehensive understanding of the energy security aspect and how it relates to technologies used, infrastructure, and practice has not been systematically researched and modeled for bioenergy (Leiby and Martinez 2011 and Leiby 2007).

The creation and application of flexible model systems with high temporal and spatial resolution for assessing risk, variability, vulnerability, and resilience in the biofuels system are needed. Because such knowledge can inform the economics, location, and configuration of biorefineries, it is an important benefit to the chemical industry. These flexible model systems also will illuminate system-wide pros and cons. Finally, they could provide insight for strategies to manage the risk and volatility of this emerging industry for all participants in the supply chain.

Biomass conversion to generate consumable feed and platform molecules

Biomass derived from corn stover, cobs, switchgrass, and wood residue comprises cellulose, hemicellulose, lignin, and pectin. These high-energy content polymers can provide sustainable raw materials for liquid fuels, high-value chemicals, and materials. Unlike fossil fuels, biomass contains significantly higher oxygen content, and the contents are contained within a highly heterogeneous solid

material. Thus, the two main problems facing the industry today in making renewable biofuels and bioproducts that can compete with fossil-fuel-based products are broadly characterized as 1) biomass recalcitrance and 2) reduction. Recalcitrance is related to the difficulty in breaking down a heterogeneous solid into small molecules that can be reduced even further to lower the overall oxygen content, the latter of which is especially important for the energy content of fuels. The conversion of biomass polymers to consumable intermediates and monomers can be accomplished to make fuels and products in many ways, including low-temperature routes that use a relatively mild thermochemical "pretreatment" step with solvent engineering and subsequent biochemical pathways, such as enzymes and fermentative organisms. Other routes include thermochemical conversion steps, such as pyrolysis, liquefaction, gasification, bio-oil upgrade, and selective transformation of biomass derivatives to value-added chemicals. In each of these areas, computational sciences have and will continue to play a significant role in understanding the fundamental chemistry, physics, and biology of these processing steps, as well as generating improvements to make these processes cost-competitive with fossil fuels.

Biocatalyst optimization

The engineering of microbial metabolism to meet bioprocess objectives, including volumetric productivity, titer, and yield, is a key component of bioprocess development. This process begins with computational searches of completed genomes, metagenomes of relevant microbial consortia, and enzyme and metabolic databases to identify and rank pathways based on yield, thermodynamic feasibility, and maximum achievable rate. Often, identification of most compatible enzymes and ideal pathways is hampered by poorly annotated databases, inadequate assemblies of massive numbers of sequence reads (resulting from second-generation microbial consortia sequencing), a limited number of known biochemical reactions, and a combinatorial explosion of possibilities when novel biochemistry is considered.

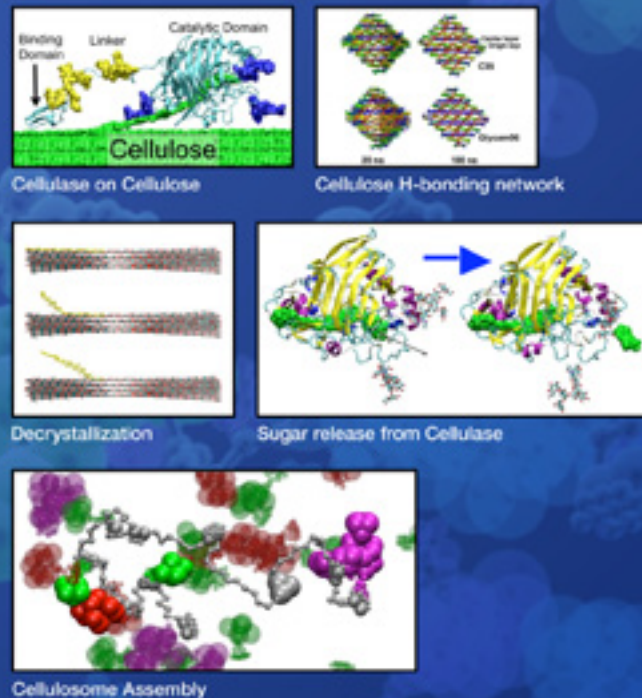
Because rate, titer, and yield objectives often are in conflict, methods for identifying microbial strain designs that consider these combined objectives are needed. Computational models of cellular metabolism and regulation can be used to identify cellular metabolic states, suggest metabolic bottlenecks to chemical production, and predict how genetic changes will affect biofuels production. Currently, there is extensive development of algorithms focused on improving yields using bi-level optimization algorithms that consider trade-offs between cell growth and product formation. These methods have been applied for the design of common industrial hosts, such as *Escherichia coli* (*E. coli*) and *Saccharomyces cerevisiae*.

Sophisticated data visualization tools that can integrate results of computational analyses along with genome-wide biological data from gene and protein expression, flux measurements, metabolite concentrations, and genetic mutations are needed for effective translation of simulation results into actionable steps.

STATE OF EXPERTISE IN COMPUTATION

Sustainability and security of the bioenergy supply chain

Qualitatively defining the environmental response surface of possible land-use changes due to biofeedstock production. High-performance computing (HPC) is just beginning to combine extensive high-resolution geospatial information on existing land use, soil, topography, and climate with possible biofeedstock production technologies and local-scale ecosystem models such as Environmental Policy Integrated Climate (EPIC) and Soil and Water Assessment Tool (SWAT) (Nichols et al. 2010 and Baskaran et al. 2009). HPC can enable the hundreds of thousands of runs needed to explore the response surface of specific land-climate configurations to land-use management for biofeedstocks. These response surfaces can be used in optimization



Cellulose, Cellulases, and Cellulosomes— Making Biofuels Cheaper

Computational experiments have been dedicated to dissecting enzyme mechanisms, as well as the structure and dynamics of natural and pretreated cellulose and discovering ways to improve performance of enzymes and cellulosome complexes. Using the largest currently available high-performance computers, details of cellulose microfibrils have been elucidated. The energy to decrystallize individual sugar polymer chains out of cell wall microfibrils also has been determined. Several mutations that can ease release of sugars from the enzyme were discovered using HPC modeling. These mutations may speed up the enzyme, reducing the need for other enzymes and potentially lowering the cost of the biofuels produced.

Image credit: Michael Crowley, Ph.D.
Used with permission by NREL.

modeling to balance environmental and economic feedstock production trade-offs. Likewise, HPC is being explored to optimize siting of potential biorefineries in the context of feedstock supply location, transportation costs, and biofuel demand location.

Quantitatively addressing energy security. The security problem associated with the bioenergy supply chain can be posed as a large combinatorial optimization problem with dynamic and stochastic elements. To date, multiple processors have been used to solve single instances of a large combinatorial system (Hartman-Baker et al. 2009). However, the ability to solve multiple instances of the problem over a range of parameter values is needed to develop trade-off surfaces for high-level models (Shinano et al. 2010).

Biomass monitoring in the field at large spatial scales. Developing biomass-monitoring capabilities at large spatial scales is important not only to estimate existing feedstocks accurately in the field, but to assess their impacts on competing agricultural production and indirect repercussions on climate change through shifts in surface albedo and evaporative losses. Mining techniques developed to address land-use/land-cover monitoring are computationally prohibitive at large spatial scales. Handling a large covariance matrix and solving a large system of equations are primary bottlenecks (Chandola and Vatsavai 2010).

Knowledge Discovery. In an effort to democratize access to bioenergy resources, Oak Ridge National Laboratory (ORNL) developed the Bioenergy Knowledge Discovery Framework (BioKDF), an online collaborative platform. BioKDF serves as a one-stop shop for resources related to biofuels and as a decision support tool for efficient planning, development, and management of bioenergy infrastructure. BioKDF stakeholders are researchers and private industry. For these stakeholders, the available online resources are not limited to geospatial data sets and relevant papers, but also include models (e.g., transportation routing and refinery siting) to run simulations. However, these models are being scaled down in computational complexity to provide simulation results instantaneously to stakeholders.

By eliminating some of the complexity, it dramatically reduces the scenario parameter space and accuracy that stakeholders can manipulate.

Biomass conversion to generate consumable feed and platform molecules

Thermochemical conversion

Several methods for the thermochemical conversion of biomass have been reported, including: gasification to syngas, pyrolysis to bio-oil, torrefaction to bio-char, cellulose to 5-hydroxymethyl furfural, and hemicellulose to furfural. Computational fluid dynamics (CFD) is recognized as a powerful design and development tool in many industrial areas. Application of numerical methods and computational resources to biomass thermochemical conversion systems reduces the risk associated with reactor scale-up and streamlines the development-to-deployment process. They are efficient alternatives to empirical approaches.

In their current state of development, CFD models for reactive multiphase flows do not provide the accuracy needed to answer some of the major challenges in thermochemical conversion, such as tar generation during gasification. In conducting CFD studies for multiphase flow modeling, for example of gasifiers, using the current state of the art is not practical. Three main topics can be identified for which advances are much needed: 1) kinetic modeling, 2) gas-solid system modeling, and 3) computational efficiency. Chemistry is a key element (Gaston et al. 2011 and Jarvis et al. 2010). However, detailed kinetic mechanisms describing biomass volatilization are still missing. Also, the current numerical description of dense particulate flows involves many assumptions and closure models that often lack thorough validation (Van der Hoef et al. 2008). Reactive simulations are difficult to perform due to lack of appropriate numerical algorithms that can account for detailed reaction chemistry in complex flows in an efficient manner. Finally, no practical strategies exist to accurately account for the multiscale nature of the conversion process from intra-particle heat and mass transfer to chemistry-hydrodynamics interactions at the reactor scale.

Biochemical conversion

To date, most strategies for improving cellulase enzymes have focused on thermal stability via directed evolution or targeted mutations since these typically are in the experimental efforts realm. However, focusing on improvements to specific cellulase activity and production of less-recalcitrant cellulose is likely to yield more dramatic cost savings in biofuels production. Applications of advanced computing to cellulase enzymes in the last 2 to 3 years have demonstrated that molecular dynamics (MD) simulations and associated statistical mechanics and quantum mechanical (QM)-based methods are able to yield quantitative insights into the structure and function of cellulases. These computational insights rely heavily on observations, insights, and molecular structures from experimental efforts and provide a complementary and powerful tool to elucidate critical features of enzyme action not readily obtained via experimental approaches.

The process steps that cellulase activity can be parsed into include:

- Binding of enzyme to cell walls or substrate
- Docking of the substrate in catalytic sites
- Chemical reaction (hydrolysis)
- Release of product.

The system elements are: the substrate (cellulose), binding modules, the linkers, and catalytic domains. To be thermodynamically meaningful, all parts of the problem must be modeled in a statistically relevant manner. Thus, this requires large amounts of sampling of the accessible configurational and energetic space.

Currently, the most useful computations produce statistical data on systems on the order of up to 1 million atoms and can run efficiently on up to 1,000 cores. Typical runs are on the order of 20 to 100 copies of approximately 100 cores. Usually, the statistics are gained by defining a path between two states of interest and sampling along that path to generate thermodynamic information. The path may include a chemical reaction or a physical process. As such, it often requires a QM description of, at least, part of the

system. The most effective way to generate statistical information is to run many instances of the system, often restraining the system to a particular region of the pathway of interest (i.e., umbrella sampling, free energy perturbation, and finite temperature strings) or enhancing the sampling across and in barriers by using temperature ladders, such as replica-exchange molecular dynamics (REMD) or metadynamics. In general, the current limitations of the software and hardware restrict the questions to be asked to smaller systems or subsystems of larger questions and usually require several months or up to a year to complete.

With rich simulation results at hand, the challenge is transforming simulation data into useful knowledge, i.e., mechanisms and rates of key molecular processes and difficult-to-measure thermophysical properties. For successful reduction to practice, the chemical industry will need easy access to this emerging knowledge base through fast and effective methods for data storage, annotation, and retrieval.

Novel abstract analysis methodologies also need to be developed to distill predicative descriptors for a given problem by exploring multiscale simulation data. An example of such a method is the fluctuogram analysis of all-atom MD for identifying mutation sites for protein engineering (Silvestre-Ryan et al. 2011).

Biocatalyst optimization

Biomass conversion generates carbohydrates, such as glucose, xylose, and mannose, which are metabolized by biocatalysts to produce a variety of products, including biofuels and specialty molecules. Biocatalyst optimization requires efficient uptake and metabolism of mixed sugar streams in the presence of inhibitors and against increasing amounts of toxic products under commercially viable process conditions. The search space for an optimum combination of bioprocess parameters is large. Furthermore, despite the tremendous increase in DNA reading and writing capacity and the identification of millions of new open reading frames, the ability to accurately ascribe functions to novel genes remains limited.

Two types of pathway search algorithms are in practice today—network- and constraint-based

(Ma and Zeng 2003, Price et al. 2002, and Schuster et al. 1999). In both approaches, an exhaustive enumeration of all possible pathways for large networks is computationally challenging. For exploring novel biochemistry, several methods are capable of automatically generating pathways of novel and known biosynthesis routes. The computational challenge arises from the combinatorial explosion in the number of possible reactions and products produced by each successive algorithm application.

Other computational tools are mathematical cellular models, constrained by metabolic parameters and interrogated by various simulation algorithms. Methods that focus on improving yield through the use of bi-level optimization algorithms are able to resolve the trade-off between cell growth and product formation and have been successfully applied to the design of common industrial hosts such as *E. coli* and *Saccharomyces cerevisiae*. The computational cost for identifying optimal strategies increases as more types of modifications are considered (e.g., addition of new pathways or altered expression) and additional properties are accounted for in the models (e.g., regulation, thermodynamics, and kinetics). In some cases, these methods result in large models containing non-linear equations and integer variables (Kim and Reed 2010, Ranganathan et al. 2010, Lun et al. 2009, Pharkya and Maranas 2006, Patil et al. 2005, Pharkya et al. 2004, and Burgard et al. 2003).

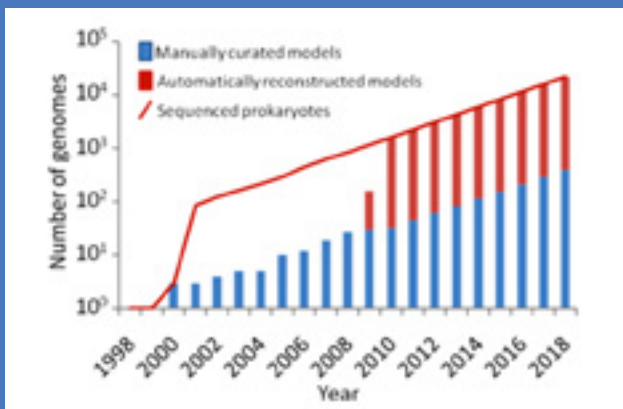
Using the ModelSEED platform, genome-scale models can now be developed in a high-throughput fashion for any sequenced prokaryote (Henry et al. 2010). This public resource uses annotated genomes to reconstruct metabolic pathways in an organism and then fills in potentially missing reactions by requiring the model be capable of predicting growth. To date, more than 2,000 metabolic models have been constructed using the ModelSEED platform. HPC is being applied in an ongoing effort to enhance the throughput of the ModelSEED; perform model refinements based on experimental evidence and flux estimates from ¹³C labeling experiments; and maintain the increasing number of completed genomes and improvements to gene models, gene annotations, and enzyme characterizations.

ACTION ITEMS AND STRATEGIES

Sustainability and security of the bioenergy supply chain

Sustaining and securing the bioenergy supply chain will require multiscale, multi-objective modeling, biomass and land-use monitoring, and HPC that supports system-wide informed decision making.

High-throughput Reconstruction of Genome-scale Metabolic Models



For more than a decade, genome-scale metabolic models have been applied in biology to produce global predictions of metabolic phenotypes based on genome sequence. Methods and infrastructure for model reconstruction also needed to scale up. Recently, the ModelSEED resource was developed, applying HPC to generate more than 2,000 new draft genome-scale metabolic models in less than 6 months.

Image credit: Dr. Christopher S. Henry, University of Chicago, Searle Chemistry Laboratory

Action items

- Apply HPC to logistical and temporal aspects of bioenergy systems planning
- Establish biomass monitoring in the field at large spatial scales, combining “social” data with satellite imaging data and closer-to-real-time monitoring
- Conduct comprehensive global land- and water-use modeling that incorporates all aspects of land-use change, including local social and economic drivers not related to demand
- Dynamic social/economical models with broader scope than the current Computable General Equilibrium (CGE) models of global agriculture
- Use HPC to explore the environmental and economic consequences of potential land-use decisions for bioenergy at multiple geospatial and temporal scales (as appropriate)
- Enable the entire multidisciplinary bioenergy stakeholder community to access knowledge and resources, such as HPC, via a common interface (e.g., BioKDF).

Strategy includes right scale, complexity, and framework, such as:

- Multiscale modeling
- Multiple levels of complexity
- Linking complex and simple models
- Geospatial modeling framework—all data and models plugged into geospatial data and modeling framework.

Leadership-scale supercomputing resources are a promising source of computational power for solving large combinatorial optimization problems with dynamic and stochastic elements. We propose their application to the logistical and temporal aspects of bioenergy systems planning at a sufficiently fine spatial resolution, accounting for uncertainty and technological change. The availability of scalable combinatorial optimization solvers and HPC resources makes this possible.

For biomass monitoring in the field at large spatial scales, handling a large covariance matrix and solving a prodigious system of equations are primary bottlenecks (Chandola and Vatsavai 2010). To use the underlying computational architecture effectively, the non-linear computer complexity of this problem requires both HPC and efficient parallelized versions of mining algorithms. Access to HPC resources would enable stakeholders within the BioKDF community to run simulations asynchronously with higher accuracies and test out combinations of large input parameters space.

Biomass conversion to generate consumable feed and platform molecules

This requires improved scaling of MD and QM codes, efficient parallel solver implementations, and more cycles.

Biochemical conversion

Currently, the need for better-scaling MD and QM codes is paramount. Also, more cycles dedicated to sampling methods with a significant amount of loosely coupled computation and machines designed to favor that kind of computation would be remarkably useful. Specifically:

- Increase the parallel efficiency of molecular mechanics (MM) and QM codes
- Seamless incorporation of QM into MM/MD approaches
- Thorough examination of data analysis—both on-the-fly and post-simulation (essential)
- New algorithms that maximize the ability to run thousands of simultaneous instances of a problem and loosely couple them
- A concerted effort to run multiple length- and time-scale simulations to couple atomic and macroscopic scales
- New approaches to chemical environment of systems of interest (implicit and explicit solvent, many-body interactions, polarizability, and reactive dynamics)

- Streamlined mechanisms for collaborative efforts between industry, national laboratories, and academics
- The efforts need to encompass all of lignocellulosic biomass, not just cellulose, which will require new force fields and innovative multiscale methods.

Thermochemical conversion

It is computationally unrealistic to consider fully resolved detailed simulations of commercial-scale thermochemical conversion systems. An approach based on a hierarchical suite of models and numerical tools with a decreasing level of details and cost—yet applicable to increasingly larger scales—is needed. It includes:

- Development of a detailed description of the biomass conversion kinetics based on specifically designed experiments, quantum chemistry calculations, and multiscale methods
- Development of intra-particle heat and mass transfer to describe single-particle evolution
- Detailed simulations of small-size configurations using an explicit representation of the particles (Lagrangian approach); these simulations are needed to gain a fundamental understanding of the reactor dynamics and coupling between chemistry and hydrodynamics.
- Computational techniques to design suitable materials for biomass conversion
- Detailed simulations will serve as a reference database to formulate and validate closure models for computationally cheaper approaches (e.g., Eulerian-type framework), which, in turn, can be used to investigate industry-relevant configurations.
- Strategies need to be developed to incorporate the knowledge gained from the most detailed simulations into the larger-scale, coarser models while preserving essential physics and predictive capability; in particular, kinetic mechanism size reduction will be essential for reactive simulations.

Due to the lack of detailed experimental data, a major issue is code validation. Error-control strategies must

be developed, and model validation needs to be conducted at each scale. This hierarchical approach will allow CFD to be used reliably at the industrial scale to assist in reactor design and optimization because of an adequate description of all relevant phenomena.

Biocatalyst optimization

This requires faster algorithms for large-scale optimization problems, exhaustive searches, and parallel processing.

Optimization of biocatalysts can benefit from focused development of the following resources:

- Develop algorithms to solve linear and non-linear problems involving 10^6 variables and constraints
- Develop methods and resources for multi-level optimization that will facilitate bioprocess design, including dynamic simulation tools
- Create methods and models for identifying process and strain design strategies for improving performance in the presence of inhibitors and toxic products
- Implement methods for analyzing and visualizing high-dimensional spaces and mapping two spaces, such as perturbation and response, to each other
- Develop parallel computing resources to facilitate novel pathway design algorithms and bioprocess design in the presence of uncertainty
- Develop a microbial physiology database linked to visualization tools.

BENEFITS

In terms of the questions they are asking, decision-makers considering the role of biomass in a sustainable future are further advanced than the modelers' abilities to address their concerns. There is risk of not only suffering the consequences of uninformed decisions, but also becoming irrelevant in the decision-making processes as political and

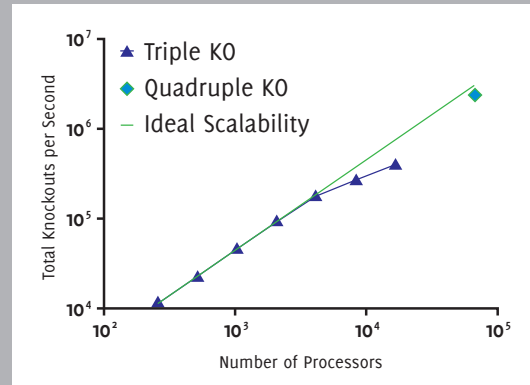
other concerns fill the technical and scientific vacuum that now exists. A more comprehensive and holistic modeling, analysis, and data collection infrastructure for global land use and the bioindustry will help break the logjam that now prevents responsible progress for the industry. This is a unique time that presents the ability to make societal choices embedded in the broader context of sustainable development. A responsive, nimble, and flexible modeling system that can engage at all levels of decision making can help steer the way to a sustainable role for biomass.

At the molecular level, several benefits are expected, including communication of questions, knowledge, and ideas that will produce new enzymatic cocktails, cellulose solvents, and treatments with much improved performance and optimized biocatalysts. Moreover, industry will benefit from the expertise of the national laboratories and universities in simulation and analysis methodologies.

Thorough investigations with fewer approximations can be routinely performed when simulation codes are optimized to perform orders of magnitude faster, from years or months to days or hours. When better-scaling QM can be coupled with improved-scaling MM, the reduction in approximations about the properties of an active site can allow us to predict behavior of enzyme cocktails at varying conditions in a time frame that will enable the biofuel industry to be competitive with fossil fuels.

The benefits of developing the proposed range of numerical tools in the field of biomass thermochemical conversion are outlined as follows:

- Development of more efficient technologies. While CFD is not expected to replace exploratory experimental work to develop novel concepts, it will facilitate fine tuning and optimization.
- Faster, cheaper deployment. CFD can prevent problems associated with scale-up from laboratory-scale reactors to commercial-size units.
- Reduced risks for investors due to reduced uncertainties on process efficiency.



Simulating Knockouts with Flux Balance Analysis

While individual Flux Balance Analysis (FBA) simulations are not computationally intensive, some applications require large numbers of simulations. Recently, 18 trillion gene knockouts were simulated in *B. subtilis* using FBA on Blue Gene/P. The total processing time scaled with the number of processors used, demonstrating near linear scalability up to 160,000 processors.

Image credit: Dr. Christopher S. Henry, University of Chicago, Searle Chemistry Laboratory

- Improved, more reliable models for techno-economic analysis can aid strategic planning at the highest level.
- Technology transfer (numerical algorithms, model-development strategies) to other types of conversion, such as biochemical.
- Development of inexpensive, active, selective, and stable catalytic materials that can operate in a complex, multiphase environment of biomass processing.
- An integrated multiscale design platform can lead to simultaneous optimization and effective characterization of production biocatalysts metabolism for three key components of bioprocesses—yield, titer, and productivity.

Catalytic Materials: Improving Chemical Transformations with High-Performance Computing

The ultimate goal of chemistry is to understand how chemical reactions—the making and breaking of chemical bonds—occur under realistic operating conditions so that one can control chemical reactions to efficiently and economically synthesize any molecule or material that is needed with the desired properties. Chemical transformations in industry take a cheap feedstock (often hydrocarbon-based) and convert it into a higher-value product by rearranging the atoms and adding functional groups to the compound. Catalysts are molecules or materials that facilitate chemical reactions without being consumed and play an essential role in the petroleum, chemical, and pharmaceutical industries, which rely heavily on catalysts. Products synthesized from catalytic processes account for 20–30% of the U.S. gross domestic product (GDP) and 90–95% of all chemical production depends on catalysis (mostly heterogeneous) (Panel/NRC 1992 and DOE 2008). The energy sector is totally dependent on catalysts (e.g., gasoline at the filling station has been processed by approximately eight catalysts). Development of new catalytic processes is necessary to meet challenges facing the U.S. today in creating alternative fuels to meet the energy needs of the nation and improve national security, cleaning up the environment and preventing future pollution, dealing with the causes of global warming, keeping populations safe from the release of toxic substances and infectious agents, and creating safer and more effective pharmaceuticals. Successfully addressing these challenges will advance industrial competitiveness, as well as provide innovative solutions to pressing societal issues.

For many catalytic processes, which usually involve a sequence of chemical reaction steps, it is still unclear how the catalyst works to increase the rates at which some chemical reactions occur and which reaction steps control selectivity. Building on an earlier Basic Energy Science Advisory Committee (BESAC) report (2003), a Department of Energy (DOE) Basic Research Needs report (2008) reached the conclusion that a key need is to “connect catalytic and photocatalytic reaction rates and selectivities to the kinetics, energetics, and dynamics of individual elementary steps and relate these to the structure and dynamics of the catalytic sites involved,” i.e., a molecular-level understanding of the detailed steps in chemical reaction mechanisms with knowledge of the energetics of the various steps. Computational methods hold the key to the fundamental understanding of catalytic processes, enabling catalyst design from first principles using the fundamental laws of physics. Catalysis is fundamentally about accelerating the rates of chemical reactions, and catalyst design will require quantitative information about reaction pathways (how reacting molecules interact with each other and the energies of these molecular interactions). Currently, information about critical points along these reaction pathways (e.g., transition states) in terms of their geometric structures and molecular interaction energies is only readily accessible via computational methods. As

products become more sophisticated, the need to quickly develop new catalysts grows rapidly in importance. New energy technologies are completely dependent on new catalytic processes and catalysts, e.g., artificial photosynthesis, battery electrode processes, and fuel cells. In many cases, no catalyst exists despite substantial amounts of trial-and-error experimentation. Moreover, the development of new catalytic processes typically takes 10 years or longer. Advances in computational approaches, many developed through support from the federal government, hold the promise of accelerating this development cycle and improving industry's ability to compete globally in these new technology markets.

The workshop identified a few key steps that must be acted upon to capitalize on the promise of using a computational approach to accelerate the design and deployment of new catalysts to greatly improve industrial competitiveness. First, it is crucial to develop sets of tools that can generate, manage, store, and analyze large computational data sets in conjunction with experimental data, which will provide new capabilities for discovering the underlying principles to design catalysts for specific applications. Second, methods must be developed to quantify uncertainty in models and compute uncertainties in computational data, which will provide reliability estimates in predictions of new catalytic systems. This also will require improvements in the theories and computational methods to improve their accuracy and predictive capabilities, especially for real catalysts under operating conditions. Finally, the collaboration between government laboratories, federally sponsored research programs at national laboratories and academic institutions, and industry must improve to enable the transfer of knowledge and understanding that will accelerate catalyst design, development, and ultimately deployment.

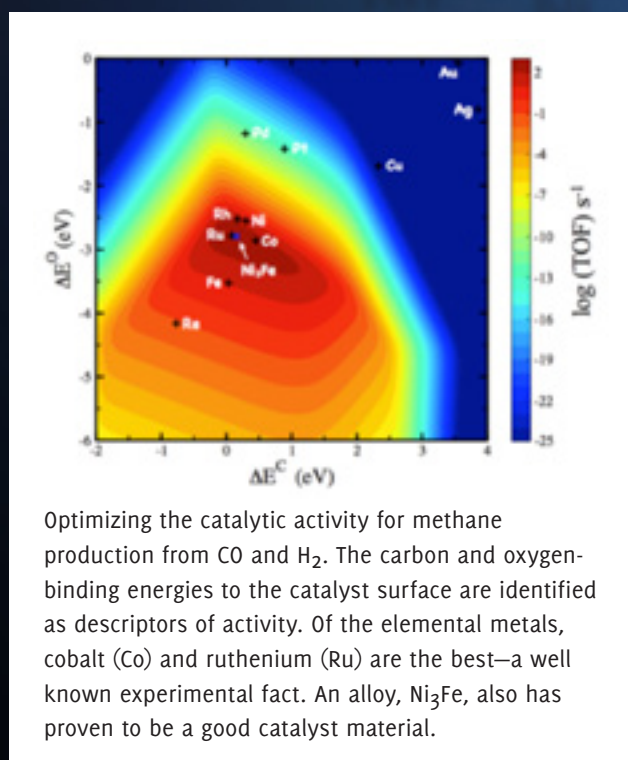
INDUSTRY NEEDS

Catalysts, both homogeneous and heterogeneous, facilitate the control of chemical reactions by improving the rates at which chemical bonds are formed or broken for specific steps. An effective catalyst can lead to improved selectivity and control over the formation of unwanted by-products, decreased energy consumption, and reduction of the waste stream for the process. These factors lead to minimization of the environmental impact of chemical production. The complexity and diversity of the chemical events embodied in catalysis demand a revolution in the way catalysts are designed and used. This revolution can become a reality only by the application of new methods in synthesis and characterization of molecular and extended material systems and computational methods, as well as an intimate coupling of experiment, theory, and simulation. New opportunities to understand and

predict how catalysts work at the atomic/molecular scale have been made possible by substantial breakthroughs during the past decade in synthesis, computation, and measurement techniques, such as advanced imaging methods that are providing new paradigms for catalyst design and eventual use. An additional set of challenges is that catalysts need to be designed for use in real environments in a chemical reactor for long periods of time, often in harsh conditions.

Industrial needs for catalyst development are driven by four primary issues: 1) time, 2) cost, 3) the increasing diversity of commercial applications, and 4) protection of the environment. Increasing the efficiency of the development and/or improvement cycle for a catalyst or catalytic process can occur by minimizing the typical trial-and-error approach. The use of reliable and robust predictions and correlations that are validated can be used to

Catalytic Materials by Design



Methods for the computer-aided design of heterogeneous catalytic materials involve several steps. The first is the identification of catalytically relevant reaction steps on the active surfaces. Then, key reaction steps are taken to the next level by identifying important trends for classes of catalysts across the periodic table. On this basis, it is possible, in many cases, to identify one or more descriptors of catalytic function. Identifying these descriptors is the main challenge for basic research. If these descriptors can be calculated efficiently or exist in a database, the basis for a computer-aided design strategy can be established.

One example of this strategy was the identification of two independent descriptors for the methanation reaction, hydrogenation of carbon monoxide (CO) to methane (see figure).

Image credit: J.K. Nørskov, SLAC National Accelerator Laboratory. Reproduced with Permission.

work together in a concerted way to solve practical problems in catalyst prediction and design. DFT is currently the most commonly used computational method for calculations of catalytic processes, and advances in DFT are enabling new understanding of catalytic mechanisms and the design of advanced heterogeneous catalysis.

DFT methods with most current exchange-correlation functionals are not sufficiently accurate for all applications, especially those with weak interactions and for calculations of energies along reaction pathways. There is an essential need for new, more accurate exchange-correlation functionals. To enable this development, more accurate predictions of reaction energetics are also needed. This also requires high-level methods, such as those based on coupled-cluster theory and its multi-reference extensions. The computational costs of these methods are much higher than DFT and significantly limit the size of the system that can be treated to only tens of atoms, yet they can provide unique insights into catalytic reaction processes and reliable energetics to ± 2 kcal/mol or better (Craciun 2010 and Vasiliu 2011). Recent advances in electronic structure methods, including those exploiting coupled-cluster ideas that either partition the system into smaller polyatomic fragments or use the local nature of interactions among electrons, enable the development of low-order scaling algorithms that can take full advantage of existing multi-node, multi-core computer architectures (Li 2010 and Piecuch 2010). With at least hundreds of non-hydrogen atoms and DFT methods capable of working with very large systems, these approaches show great promise for extending high-level methods to much larger systems. These emerging, reduced-scaling approaches will be valuable in providing benchmark calculations to validate less-reliable methods, such as DFT, as well as data needed for improving the parametrizations of DFT or for developing analytical representations of molecular interactions (e.g., force fields or molecular mechanics potentials). Such improvements are especially needed for studying catalytic processes at interfaces or in solution.

Although calculations of electronic energies for fixed molecular configurations are essential, they are insufficient for catalysis. Calculations of

thermodynamic and kinetic properties require determination of free energies that average electronic energies and vibrational frequencies over very large numbers of configurations. Methods for calculating the free energies (inclusion of the entropy component) of stable structures have become routine and can be applied to systems of practical interest in catalysis. To make reliable predictions for larger molecules, hybrid methods are commonly employed in which a typically small number of atoms in the system is treated using a high-level QM electronic structure method with the rest of the atoms being treated using MM or a much lower-level QM method. Advances in these QM/MM and QM/QM methods allow for efficient calculations of free energies for reaction pathways, where bond breaking and making is important. These methods promise to be a powerful approach for understanding the important pathways for catalytic processes on complex surfaces and in liquids. As with the electronic structure methods, the accuracy of these approaches, where part of the system is treated using more approximate MM approaches, is an issue that needs to be addressed.

The identification of reaction pathways for individual reaction steps becomes difficult when the reacting species and catalysts become larger. Rare-event discovery tools, such as accelerated dynamics methods (Voter 2002), could provide approaches to locating reaction pathways in these cases, although they are computationally intensive and have been applied only to relatively simple problems. This area would benefit from further research.

Once reaction pathways are identified, the rates at which individual chemical reactions proceed along these pathways can be calculated. Transition state theory (TST)-based methods are the most common approaches for using electronic structure information directly in calculations of rate constants for chemical reactions and are routinely used in catalysis studies (Truhlar 1998). In recent years, these approaches have been extended to liquid-phase, interfacial, and enzymatic reactions, where the effects of the molecular environment around the reacting components can have a significant influence on rate constants. Free-energy methods can be employed to account for these effects. Although TST approaches are approximate, they provide a

framework for systematically improving the accuracy of the computed rate constants. The larger errors are those introduced by the accuracy of the energetic information used in these calculations. At the lowest level, simple bond additivity or scaling arguments can be used to provide useful information about how rate constants change when the reacting molecules are altered. Highly accurate calculations of absolute rate constants require high-level calculations of electronic energies and accurate treatment of the variations of these energies with atomic coordinates. The treatment of quantum effects on nuclear motion, such as zero-point energy and tunneling effects, can be large for reactions involving light atoms (i.e., hydrogen atoms) and including these effects presents a challenge for condensed phase reactions. Rates of full reaction networks usually found in catalysis often depend less critically on the accuracy of individual rate constants. It is important to develop an understanding of these phenomena. In particular, it will be crucially important to develop systematic methods for assessing the accuracy of electronic structure calculations, as well as accuracy of predicted absolute rates and, most importantly, trends.

Many reaction steps (hundreds or more) typically make up a catalytic process. For example, conversion of cellulosic biomass into hydrocarbon fuels will require a multitude of bond breaking and making processes. Tools have been developed for the automated generation of the network of coupled reaction steps that make up catalytic reaction mechanisms (Neurock 1991). Although these approaches have been primarily applied to conversions of fossil fuels, they are appropriate for studies of emerging feedstocks such as biomass.

Industrially important catalytic processes occur in large reactors where the reaction mechanisms are coupled to heat and mass transport processes occurring in reactors. Accurate modeling of these catalytic reactors requires coupling reactions occurring at active sites (< 1 nm) that are influenced by the interactions of reactants and products across microregions (< 10 nm) with transport of reactant molecules to active sites of the catalyst that occur over length scales up to meters. In some cases, it is adequate to pass parameters for processes at one scale up to the next scale. For example, kinetic

parameters often can be used in computational fluid dynamics codes that model catalytic reactors. However, in other cases, changes at different scales can strongly affect phenomena at adjacent scales. In these cases, multiscale modeling approaches are required that seamlessly couple processes at the different scales. Advances in approaches that bridge scales, such as kinetic Monte Carlo, reaction networks, and coupling of discrete and continuum approaches, provide the basis for future development of the multiscale methods needed for more accurate modeling of catalytic processes.

Informatic approaches are beginning to be successfully applied to design new catalysts and hold great promise for accelerating the development of new catalytic processes. These approaches use chemical descriptors that can be calculated easily to predict catalytic behavior, e.g., "volcano" plots in which binding energies of atoms/molecules to surfaces are used as the descriptor to enable the rapid identification of active and selective catalysts.

ACTION ITEMS AND STRATEGIES

To improve catalyst modeling, there are a number of continuing and new developments needed in computational chemistry approaches. There is an obvious need for reliable predictions. As such, accuracy of computational results is important. Electronic structure methods are at the heart of catalyst design. Computationally efficient electronic structure methods that exceed current DFT capabilities in terms of accuracy and advanced correlated electronic structure methods in terms of system size are needed for energy, structure, and frequency calculations. New theories and computational approaches must deal with predicting the entropy of large systems and find rare events (e.g., finding extrema-like transition states on reaction pathways). Tools for the prediction of much longer time-period dynamics for large numbers of atoms are needed, coupled with the ability to treat the making and breaking of chemical bonds during dynamics simulations without developing specific, individual force fields. An important issue for heterogeneous catalysis is to study levels of

doping consistent with real doping levels, which are often at the 1% level. Such simulations require large structures, leading to additional computational complexity. Developing theories and tools for the reliable prediction of reaction rates in solution and at the solid/liquid interface is also necessary. This is an unsolved problem and another example of the need for a capability that can be used to predict entropy in complex systems. Photocatalysis will require developing new tools for the reliable prediction of excited state energies and dynamics for large systems that exceed current time-dependent density functional theory (TD-DFT) technologies. This will enable the prediction of ultraviolet (UV) spectra and reliable band gaps. A substantial breakthrough would involve reliable methods for the solid state that are equivalent in accuracy to current coupled-cluster calculations extrapolated to the complete basis set limit for small-to-moderate molecular systems. In addition, coupled-cluster methods for geometry optimization and frequency calculations, as well as other property calculations for larger molecules, should be further developed and deployed to establish benchmark data for homogeneous and heterogeneous catalysts.

Because of the complexity of a real catalytic reactor system, multiple time and length scales need to be coupled in a catalytic reactor simulation. To enable the seamless transition between scales, substantial improvements are needed in theory and computational tools. For example, consider what is needed to model a reactor containing 2- to 3-nm catalytic particles on a support under operating conditions in a fluidized bed reactor. Multiscale approaches with seamless coupling across scales are required from the sub-nanometer molecular level up to continuum fluid flow on the meter level over time scales from femtoseconds to months.

A consequence of continually evolving computer architectures, driven more and more by consumer technologies such as video games, smart phones, and iPads, is that software has to be continually updated for new central processor technologies, such as graphical processing units (GPUs). GPUs may be especially fast, but they do not offer all the functionality of a standard processor. Software conversion to new architectures is a costly, people-

intensive process. There must be continued investment in the development of new software systems, porting of legacy software functionality and capability to new computer architectures, and software maintenance for the general computational community.

It is crucial to develop sets of tools that can generate, manage, store, and analyze large datasets. Specifically, access to computer systems to generate large datasets, especially for mechanism development, is essential. This is a parallel problem, which does not require system-wide fast interconnects, but needs substantial central processing unit (CPU) resources. Statistical tools for mechanism searching and development must advance. Tools for building catalytic model systems need to progress, for example, through automated literature searching of experimental and computational results. Software is needed to manage large datasets generated from different codes and methods. To readily share data and leverage software tools developed under DOE/Advanced Scientific Computing Research (ASCR) efforts (such as cloud computing tools), standardized software incorporating open-source guidelines needs to be developed. In turn, this potentially will lead to standardized data formats as these software tools will enable users in three domains—academia, industry, and national laboratories—to more readily interact. This effort needs to be sponsored by the government with data centers located at the National Institute of Standards and Technology (NIST) and/or DOE national laboratories for long-term stability and maintenance. In addition, there is a tangible need to cultivate tools and capacity for long-term storage and maintenance of data—the so-called “scientific data life cycle” data management issue. These tools must address the fact that it is easier to deal with the pedigree of computational data than experimental data, especially if the original output files are available. Tools to search and analyze large data sets to enable discovery of new systems and generate new ideas also would be helpful. These tools could annotate the data and generate metadata for future search algorithms. Some points to consider include:

- Is there an equivalent of the IBM Watson architecture for chemical data analysis, especially for catalysis (Ferrucci et al. 2010)?

Developing Catalysts for Novel Materials

INEOS Group Limited (hereafter Ineos) investigated a family of single-site, dual-state polymerization catalysts based on work done in Robert Waymouth's group at Stanford University.¹ These are unbridged metallocene catalysts with substituted indene ring systems at the red positions (see Figure 1). Because the catalysts are unbridged, the rings are able to rotate back and forth between geometries. One geometry, the *rac*, produces isotactic polypropylene, while the other, the *meso*, produces atactic polymer within the same polymer strand.

In 1957, Giulio Natta first isolated elastomeric homopolypropylene, or EHPP, as a minor component extracted from reaction products using traditional multi-site Ziegler-Natta catalysts. The elastomeric polymer properties come from the microstructure—a balance between regions of crystalline isotactic and elastic atactic polymers. Experiments demonstrate that these properties come from the stereoblock structure formed as the catalyst rotates between geometries, alternating between isotactic and atactic insertion. **Mixing isotactic and atactic polymers does not produce these properties.**

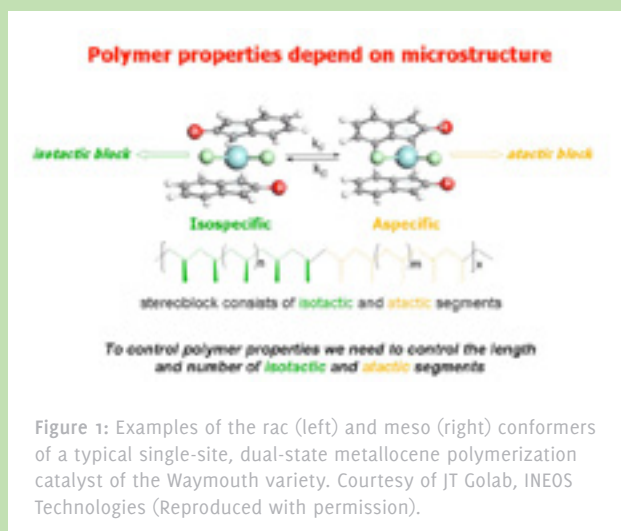


Figure 1: Examples of the *rac* (left) and *meso* (right) conformers of a typical single-site, dual-state metallocene polymerization catalyst of the Waymouth variety. Courtesy of JT Golab, INEOS Technologies (Reproduced with permission).

To control the polymer properties, the length and number of isotactic and atactic segments need to be controlled. Clearly, a detailed understanding of the underlying, enabling catalyst technology of this process would lead to fine-tuned control of the product polymers properties.

To investigate the mechanism of this catalysis and influences of catalyst structure on the mechanism, Ineos modeled dichloride coordinated to metallocene catalyst systems, as well as systems coordinated to a methyl group and C₇ polymer chain, to simulate the effect of polymer growth. It was assumed that the equilibrium constant for the stereoisomer interconversion in the metallocene catalyst system is directly proportional to the Gibb's free energy of reaction. Of course, considering only the energy difference of the *rac* and *meso* isomers may be an oversimplification of the overall mechanism. However, reasonable results from chemistry modeling needed to be obtained quickly to use modeling techniques as an effective tool for directing experimental work. The hypothesis was, by comparing the *rac* and *meso* energy differences as a function of phenyl substituents, correlations to polymer tacticity could be made.

The assumption in relating this energy difference to the overall equilibrium between the stereoisomeric states of the metallocene enabled Ineos to correlate the calculated energy difference data to observed polymer microstructure. These systems have zirconium (Zr) coordinated to a methyl group and C₇ polymer chain to simulate the effect that polymer growth would have on the energy difference.

¹ The heritage of the work described belonged to Amoco Chemical and BP Chemicals. Both entities now are part of INEOS Group Limited (<http://www.ineos.com>).

For example, an energy difference greater than 3 kcal/mol in favor of the rac isomer, which produces isotactic crystalline polypropylene, was predicted to create polymers of high m_4 (m_4 is a measure of the relative tacticity of four consecutive insertions in a polymer chain and can be determined using ^{13}C nuclear magnetic resonance spectroscopy) because the meso isomer would be accessed rarely, if at all. In fact, polymers created from this catalytic system are 75% m_4 . Highly isotactic polymers have correspondingly high m_4 values. Similarly, the energy difference for a different ligand system predicts that the meso isomer is more stable than the rac by more than 1.5 kcal/mol. The meso metallocene isomer produces atactic amorphous polypropylene, so the prediction is this catalytic system would create polymers of low m_4 . In fact, polymers created from this catalytic system are 12% m_4 . If the rac isomer is more stable in the range of 0 to 3 kcal/mol, it is predicted to produce polymers that would have medium m_4 between 30% and 40%.

Due to the large molecular size of these systems, *ab initio* calculations are often prohibitive. As computer technology advances, the problems that chemistry modeling can address broadens, while the answers such modeling provides are more detailed and accurate. This problem continues to be studied using quantum chemical approaches, such as those in DMol³, NWChem, and GAMESS (the latter two program suites were developed with DOE support), to perform full optimizations on the rac and meso conformers, as well as the transition state (see Figure 2). Encouragingly, these results agree with initial “fast sweep” molecular mechanics results.²

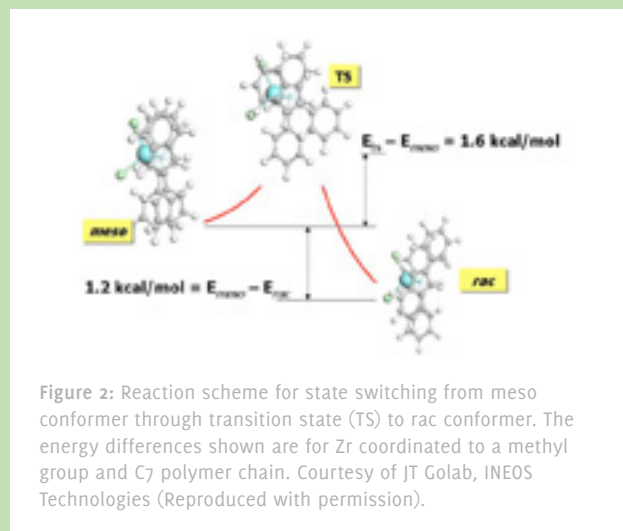


Figure 2: Reaction scheme for state switching from meso conformer through transition state (TS) to rac conformer. The energy differences shown are for Zr coordinated to a methyl group and C7 polymer chain. Courtesy of JT Golab, INEOS Technologies (Reproduced with permission).

In summary, trends in the relative energies of the rac and meso isomers of various single-site, dual-state metallocene catalytic systems were studied using molecular chemistry modeling. The energy difference was calculated using several diverse computational chemistry techniques, including molecular mechanics, density functional theory, and molecular orbital theory. Assuming this energy difference is directly proportional to the equilibrium constant for state switching, the modeling and calculations revealed predictors of polymer microstructure as a function of ligand type. The computational results uniformly compare very well with experiment, further supporting this fact. Outliers have led to new interpretations of the interplay between the state-switching equilibrium and properties of the final polymer.

Significant, rapid progress was made in this project because both modeling and experiment were conducted simultaneously in the early phases. Without fast computers and modeling algorithms, these trends would not have been readily discerned and, in turn, would not be able to provide more accurate insights for future experiments.

² Golab JT. 1998. “Making Industrial Decisions with Computational Chemistry.” *Chemtech* 28(4):17-23 and Maiti A, M Sierka, J Andzelm, J Golab, and J Sauer. 2000. “Combined Quantum Mechanics: Interatomic Potential Function Investigation of rac-meso Configurational Stability and Rotational Transition in Zirconocene-Based Ziegler-Natta Catalysts.” *The Journal of Physical Chemistry A* 104(46):10932-10938.

- How can computing be used for rapid data validation, e.g., the active thermochemical tables (ATcT) approach developed at Argonne National Laboratory (ANL) for thermochemistry (Ruscic et al. 2004)? There already is a tool for reaction barriers and energies for simple reactions on metals developed at Stanford University.

Theories and software tools also must be developed to quantify uncertainty in models and provide uncertainties for computational data. For example, there is a need to determine which component data (e.g., elementary reactions in complex molecular processes involving a competition of multiple elementary reactions) have to be targeted at the highest levels of theory to obtain reliable predictions. Ideally, this would be automated and extend beyond Monte Carlo sampling of parameter space or *ad hoc* choices of processes or properties to study. One idea is to consider incorporating or adapting some of the new work in applied mathematics being used, for example, in reactive fluid flow and transport, such as sensitivity analysis methods exploiting high-dimensional model representation and variance decomposition techniques. This approach already is being used to predict which elementary reactions (of many) need to be studied theoretically and/or experimentally to obtain precise chemical kinetics models of complex combustion processes.

The collaboration between government laboratories, federally sponsored research programs at national laboratories and universities, and industry must improve. There is a need for the government and national laboratories to use HPC to generate basic science data of interest to industry, which then can be used for proprietary industry work. For example, consider NIST, which is a government laboratory with a mission to collaborate with industry. NIST has broadly addressed intellectual property (IP) issues, while its reorganization into facilities has made collaboration with industry easier. There are high-quality NIST databases, but NIST needs broader outreach and to improve integration of its databases. NIST can substantially benefit by input from industry (e.g., what data does industry need?).

There was ample agreement that annual workshops on computational technologies relevant to catalysis would be useful and beneficial. To foster collaborations with industry, these workshops would focus on a particular class of computational (electronic structure, plane waves, etc.) methods and codes, provide a forum to disseminate new developments in terms of methods and benchmarks, offer hints on how to best use codes, and host hands-on training. In addition to continuing a dialogue with industrial practitioners, federally funded application developers would obtain valuable user feedback on how a specific code is to be applied, leading to better interfaces and new methods and benchmarks required to address specific industry problems. These workshops would facilitate knowledge transfer from basic federally funded research and identify important problems from industry. To begin this process, a virtual network via the web that can provide updates on new developments in methods, codes, and benchmarks should be established.

BENEFITS

The rational design of catalysts and catalytic processes for a specific application is a valuable goal of modern chemistry and chemical engineering. Recent advances in computational approaches, tools, and hardware hold great promise for making this objective a reality. Computationally guided design will greatly accelerate this development process and enable deployment of new catalytic processes in a much shorter time than now possible. Accelerating the design of new catalysts from first principles of chemistry and engineering would have a dramatic impact on important societal issues, as well as competitiveness of industries engaged in addressing new technologies that are critical to U.S. economic security. Although much progress can be made with existing tools, continued research is required to reach the goal of rational catalysis design as the fundamental rules governing how to control the kinetics of making and breaking chemical bonds continues to expand and evolve.

Maximizing Energy Storage Using High-Performance Computing Capabilities

In recent years, energy storage technologies have emerged as key enablers of a fossil-fuels-reduced future economy that will rely extensively on electrical energy storage and the inter-conversion of electrical and chemical energy. There is urgency in developing and deploying ecologically friendly energy supplies—an endeavor whose societal importance cannot be over emphasized. Before these technologies can become economically widely adopted (Yang et al. 2011), scientific and technical challenges remain, and their resolutions will require a close cooperation between government, academia, and industry (DOE 2010a). Much fundamental scientific research and knowledge about these technologies is being pursued in governmental institutions¹, academia, and industry, using state-of-the-art synthesis, characterization, measurement, and simulation tools.

In the last several years, tremendous progress has taken place in theory and simulations, greatly facilitated by advances in HPC that allow the use of more complex and realistic models of energy storage materials and processes. Expertise in HPC is widely available in the DOE national laboratory system and already is applied successfully to important problems relevant to new energy technologies, including those analyzed in the various panels of this workshop, namely catalysis, photovoltaics, biomass, and energy storage.

The energy storage science and technology domain is ripe for a partnership between DOE and industry, a concept strongly supported by this workshop panel. Establishing a partnership in the form of a Center for Innovation through Computation (CIC) on Energy Storage Science and Technology would immensely strengthen the potential for industry innovation through HPC. The CIC's primary goal would be to make current HPC capabilities available to assist in designing improved materials and properties for next-generation energy storage technologies while also providing a forum for defining further required HPC advances in theories and simulation tools. This would be achieved through close interactions between industrial and DOE partners. The HPC knowledge transfer would take place through industry and DOE/academia scientists conducting joint research. The partnership's distinguishing trait—its HPC-centric *raison d'être*—would be augmented by the ability to leverage experimental data from industry and other DOE-sponsored energy storage efforts (Energy Frontier Research Center [EFRC] and Hubs) for verification and validation of computational approaches and findings. Projects would be agreed upon by all partners to assure relevance while protecting the individual partners' IP and competitive advantages.

¹ Center for Electrical Energy Storage (CEES) Energy Frontier Research Center at Argonne National Laboratory, see <http://www.anl.gov/energy-storage-science/index.html>.

INDUSTRY NEEDS

The sustainable use of energy from solar, wind, hydro, and geothermal sources requires storing this energy directly as electricity in electrical energy storage devices (batteries and capacitors) and/or inter-converting electricity on demand into chemical energy (fuel cells and electrocatalytic systems) as these energy sources are either intermittent, not evenly distributed, or not locally available. Advances in the efficiency of these technologies touch upon several scales of length and time, from a materials scale to a device scale to a system scale. As such, they involve a full range of underlying scientific and engineering issues, all of which can benefit greatly from computer-aided design and HPC (Mukherjee et al. 2011).

The present panel focused its discussions and analyses to batteries—in particular lithium-ion (Li-ion) batteries as prototypes of issues that are found in batteries—and fuel cells. Much information can be found about the differences between types of energy storage systems in a number of reviews (Yang et al. 2011 and Mukherjee et al. 2011) and DOE reports (2007).

Fuel cells

Briefly, fuel cells contain electrochemical cells consisting of two electrodes separated by an ion-conducting medium or membrane that conducts protons or ions, depending on the fuel, temperature of operation, and specific application of the technology. The associated science and technology issues include:

- The catalysis of fuel oxidation at the anode and oxygen reduction reactions (ORR) at the cathode with the goal of finding alternate catalysts based on non-precious metals with low over-potential and increased tolerance to poisoning
- The transport of protons or other ions through the membrane, a complex molecular environment operating under varied temperature and hydration levels, which must retain material stability and impermeability to fuel

- HPC studies involving catalysis of electrochemical reactions, in particular the ORR, with potential to lead to great progress in the performance and stability of fuel cells (DOE 2010b and 2008).

HPC issues discussed in this workshop's catalysis materials panel apply to this technology. HPC studies also can help develop novel, stable, inexpensive, and transport-efficient membranes for fuel cells. Such studies can form the basis for a rational design of electrodes and membrane assemblies with controlled, reliable properties for improved performance to enable widespread adoption of the technology. Already, HPC membrane studies exist that could greatly benefit from a partnership with industry to advance the rational design of membrane materials.

Batteries as an energy storage paradigm

Several technologies are being considered for widespread deployment of electrical energy storage. Due to their varied performance and cost characteristics, these technologies vary with the intended applications (Yang et al. 2011). Technologies include ubiquitous Li-ion batteries, Li-air batteries, redox flow batteries, and sodium-based batteries. At the lowest level, energy storage systems are made of electrical cells that are combined into modules, as well as modules combined into packs and systems. Loss of performance is associated with all levels of assembly. In terms of understanding the limitations and deficiencies of their physical-chemical properties and performances, cells have been the primary focus with the goal of devising solutions that improve efficiency and global performance.

Batteries of the future will require higher energy/power densities, lower cost, longer life, and lower footprint—all while operating safely. These improvements will emerge from the discovery of new materials, device architectures, and processing techniques. In fact, with electrodes packing ever more energy in closer proximity to each other, safety is a major concern when it comes to

increasing performance. There are many factors that need to be considered, including:

- Safety and thermal stability of the combined electrode-electrolyte system
- Electrode-electrolyte interfaces
- Reversibility of the electrochemical reaction
- Electronic conductivity and ionic mobility of the constituents
- Operating temperature range
- Component manufacturability and availability
- Raw materials costs.

Over the years, a wide range of cell chemistries and additives has been developed to optimize cell performance for different applications. The material substitution has been targeted to improving a

variety of performance characteristics: energy/power density, reduction in impedance, diminution of self discharge, increase in voltage, reduction in cost, increase in life cycle, corrosion and leakage prevention, control of polarization, and increase in safety. In addition, varying device architectures, both at the cell- and pack-level, have been considered.

STATE OF EXPERTISE IN COMPUTATION

Table 1 provides a categorization of the science and technology challenges associated with cell design along with a broad categorization of the applicable HPC approaches. The interrelated challenges are in the areas of interfacial processes, electrode materials, aging, and safety. For example, the safety of contemporary transportation batteries

Process and Issue	Methods		
	Quantum	Atomistic / Mesoscopic	Continuum
Interfacial Process			
Anode solid electrolyte interphase (SEI) effect on resistance and open circuit potential (OCP)	●	●	●
Explore new electrode/electrolyte combinations	●	●	●
Electrode Materials			
Phase transitions as a function of state of charge (SOC) and effect on OCP	●	●	
Explore new high capacity and performance electrode materials	●	●	●
Aging			
Effect of SEI on OCP, resistance, and capacity	●	●	
Determination of SEI growth as a function of temperature and pressure	●	●	●
Safety			
Factors controlling the stability of SEI layer	●	●	●
Role of heterogeneities at the battery pack level on thermal runaway			●

Table 1. Scientific Challenges in Li-ion Battery Technology and Computational Approaches (courtesy of S. Pannala, ORNL)

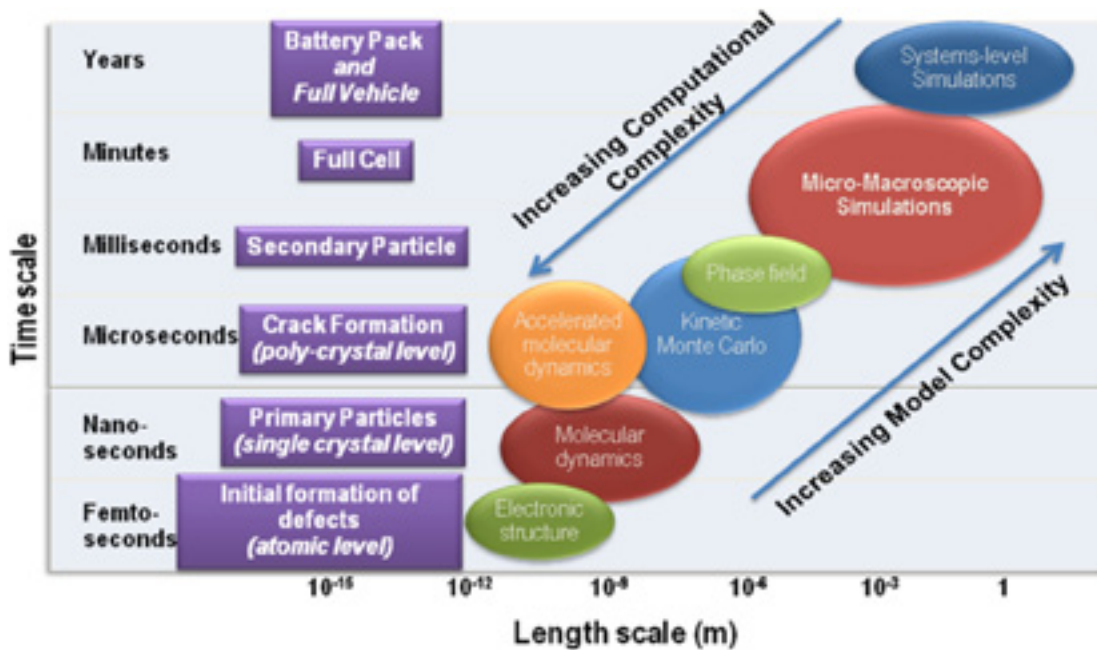


Figure 3.1: Multiscale and Multiphysics Models for Batteries and Electrochemical Systems (courtesy of S. Pannala, ORNL)

A partnership in the form of a CIC on Energy Storage Science and Technology would be a perfect forum to develop and showcase HPC-based research toward industry and manufacturers' specific needs, which include understanding, predicting, and most importantly discovering improved materials and systems for the next generation of energy storage technologies, batteries, capacitors, and fuel cells. In addition to demonstrating HPC capabilities in this field, these challenging applications are sure to uncover new advances in theories and tools that will be required to critically impact further industry needs. Experimental validation of the computational approaches and findings, achieved in part by leveraging experimental research in this field in industry and DOE-sponsored projects, would be a required and integral part of the partnership.

The partnership would involve chemical industry and manufacturers vested in the development and production of materials, components, or systems for energy storage, along with appropriate DOE offices motivated by HPC applications, e.g., Basic Energy Sciences (BES), Energy Efficiency and Renewable

Energy (EERE), and ASCR. To be successful, appropriate resources would be made available for an initial duration (5 years as with current EFRCs) that allows realistic demonstrations of the potential impact of HPC for innovation in these technologies. Such a partnership would implement elements of a broad strategy toward the advancement of energy storage science and technology through HPC, such as:

- Definition and prioritization of goals, milestones, and metrics of tactical scientific challenges, from materials science to cells to packs and systems that are amenable to HPC investigations in the short- and long-term.
- Selection of prototypical model-benchmark systems that the joint partners would endorse as representative of real science and technology challenges in the field without compromising partner IPs or associated competitive advantages.
- Selection and development of simulation strategies and tools relevant to each of the defined domains—interfacial processes, electrode

materials, aging, and safety; capturing the multiple scales, phases, and time scales; and multiphysics (ionic and electronic transport, thermal structural mechanics, and electrochemistry).

- Deployment and feasibility demonstration of HPC models and computations, along with validation by leveraging domain experiments toward understanding performance, aging, and safety factors in energy storage technology.

Action Item No. 1—Define and deploy a comprehensive approach to the multiscale, multiphysics investigation of interfacial processes as they pertain to energy storage. Assess current HPC capabilities and gaps to tackle modeling and simulations of realistic models of interfacial processes.

Action Item No. 2—Define and deploy a comprehensive approach to the multiscale, multiphysics investigation of electrode materials as they pertain to energy storage. Assess current HPC capabilities and gaps to tackle modeling and simulations of realistic models of the physics and chemistry of electrode materials.

Action Item No. 3—Define and deploy a comprehensive approach to the multiscale, multiphysics investigation of aging as it pertains to energy storage. Assess current HPC capabilities and gaps to tackle characterization and understanding of aging.

Action Item No. 4—Define and deploy a comprehensive approach to the multiscale, multiphysics investigation of safety as it pertains to energy storage. Assess current HPC capabilities and gaps to tackle physical and chemical processes that lead to breakdown in safety associated with the use of energy storage devices.

HPC capabilities and gaps

While significant advances have taken place among the HPC tools in use to date, there are many challenges that remain to take HPC to a

higher level of capability for characterization, understanding, and prediction. The preceding description is consistent with priority research directions identified in the context of extreme-scale computing (Mukherjee et al. 2011). In brief, those are:

- Advance simulation techniques to control chemical interactions at the molecular levels, including advances in fundamental theories, phase space sampling techniques, free energy determination, chemical rate calculations, and rare events characterization. These tools are central to the modeling of a variety of phenomena in the materials science and chemistry associated with energy storage technology.
- Develop simulation techniques for matter transport in heterogeneous environments. Such efforts would deal with the dynamics of electrons, protons, ions, and molecules at the nanoscale and overall dynamics at the macro- and field-scale. The focus on matter transport requires long time simulations of the coupled dynamics of these species in heterogeneous environments at the nanoscale and beyond.
- Develop general and robust multiscale modeling techniques for the description of integrated, complex materials. Multiscale modeling over a range of length and time scales is critical in dealing with multiphase descriptions of the energy storage science and, in some cases, toward real system characterization.
- Implement data analytics for combinatorial explorations and knowledge discovery, particularly in the context of energy storage materials and processes. Data analytics focuses on inference, or the process of drawing a conclusion based on knowledge obtained through computational investigations. Knowledge discovery techniques rely on machine learning to connect materials properties with the details of materials structure and composition using descriptor-based screening of many candidates for new and improved materials for energy storage applications with improved stability, safety, and energy densities.

BENEFITS

The type of partnership described would mutually benefit industry and DOE partners. HPC capabilities have progressed to the point that they are usefully applied to selected scientific issues of energy storage. In this context, applications of HPC methods often are challenging and require expertise in designing and executing the computations and simulations. Additional expertise and extensive experience is required to translate science questions about device design and performance into models and simulation approaches. Close cooperation between industry and HPC partners would provide a perfect environment to advance both the domain science and HPC technology.

Information sharing

Beyond the partnership, information sharing about HPC research approaches, methodologies, and technology would be broadly applicable to industrial developments. This information transfer would take place as industry scientists participate in HPC application projects in cooperation with DOE scientists. Outcomes specific to energy storage would include:

- Advances in battery cell performance (e.g., safety, life cycle, cost, energy power density) and fuel cell performance (membrane for proton transport under low hydration levels and no fuel crossover) that will increase U.S. competitiveness in a world market
- For batteries, anticipated success stories might include: design of new cathode materials, new redox shuttles for over-charge protection, and electrolyte additives for SEI formation
- In the area of low-temperature fuel cells, successes may include: design of new membrane materials for selective proton transport under low humidity and design of new catalytic electrode materials.

Test bed for HPC advances

In the face of practical scientific challenges to be solved in the context of the requirements of industrial research, vivid demonstrations would emerge showcasing current limitations of existing theories and methods in their ability to solve pressing technology problems—no matter how much computer time and power is aimed at some problems. Close interactions in the partnership would provide an ideal test bed for implementing the most advanced simulation approaches and analyzing their successes and limitations. Ideas for new methods and more complex, realistic simulations would undoubtedly emerge, and their implementation would be initiated, advancing the HPC enterprise.

Establishing a long-term relationship

Joint participation in a successful partnership would serve as the foundation for a long-term relationship between industry and DOE science offices, clearing a path for the effective use of DOE's HPC resources and expertise and strengthening the U.S. chemical enterprise via innovation through HPC, including transitioning from simulations for science to computer-aided technologies design.

Material Changes

HPC optimization enhances energy storage design options

High-performance computing holds significant promise for design and optimization of new materials for electrodes and electrolytes used in electrical energy storage and in simulating these materials at larger and longer scales for improved system performance and safety. Recently, Kang and Ceder (2009) reported a striking example¹. Using high-throughput computations, they discovered a new material for Li-ion energy storage that combines both high energy density (typical of batteries with fast charge) and ultrahigh discharge rate.

Electrical energy storage with both high energy density and high power density is critical for future mobile applications and balancing intermittent energy sources. Typically, electrochemical systems with high energy densities used in battery technology do not have the high charge and discharge rates characteristic of supercapacitors. The fundamental difference is how these systems store electric charge. In battery materials, electric charge is stored in the bulk. In supercapacitor materials, the charge is stored at the interfaces by forming an electrical double layer.

Based on computational first-principles predictions of phase diagrams (Figure 1a), a new high-rate lithium iron phosphate (LiFePO_4)-based material was designed and synthesized. This new material offers the extremely attractive bulk property of high lithium mobility in nanoscale LiFePO_4 combined with a fast ionic transport in the lithium phosphate surface layer. The shaded area in the calculated phase diagram (Figure 1a) represents optimal coating compositions. The resulting material has a measured rate capability equivalent to an unprecedented full battery discharge in 10–20 seconds (Figure 1b).

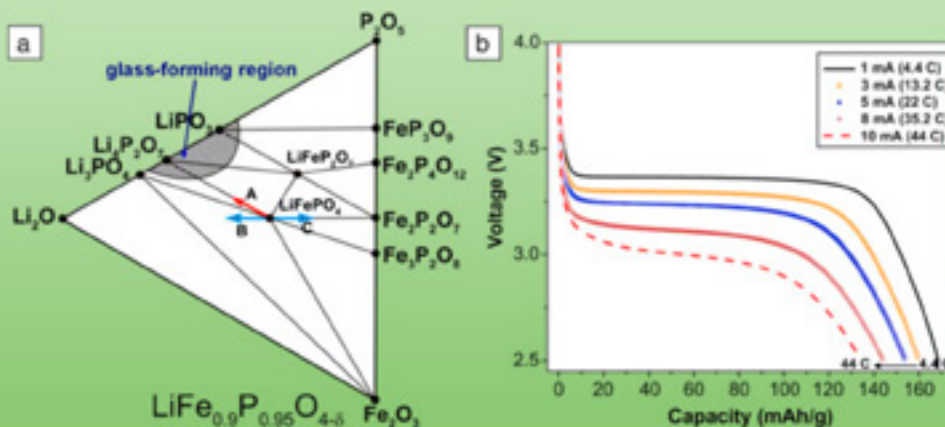


Figure 1: Design and synthesis of a new LiFePO_4 -based material. Reproduced with permission from G. Ceder, Massachusetts Institute of Technology².

This example illustrates how insights achieved through computations into charge and ion transport in an existing Li-ion battery cathode material have led to considerable improvement in rate capability. By applying HPC to the science of electrical energy storage, further breakthrough discoveries can be expected.

¹ Kang B and G Ceder. 2009. "Battery materials for ultrafast charging and discharging." *Nature* 458:190-193.

² Ceder G, G Hautier, A Jain, and SP Ong. 2011. "Recharging lithium battery research with first-principles methods." *MRS Bulletin* 36(3):185-191. DOI: 10.1557/mrs.2011.31.

Advancing Photovoltaic Efficiency and Capacity through High-Performance Computing

Enough sunlight reaches the earth's surface every hour to power the planet for an entire year. Harnessing even a small fraction of this sizable, renewable energy source using solar cells, or photovoltaics (PVs), would significantly alter our nation's energy balance and allow us to achieve energy independence. In addition, PVs are increasingly recognized as an essential part of future global energy production, and the PV industry is positioned to be a significant driver of new jobs in high-tech industry in parallel with microelectronics, display, and telecommunications industries.

For solar energy to compete with fossil fuels, significant problems with photovoltaic efficiency, durability, and cost must be overcome (Figure 4.1). In particular, cost-competitiveness requires breakthroughs in solar cell performance at nearly every component level and cannot be achieved through incremental reductions associated with economies of scale. Many fundamental mechanisms and basic properties of solar cells that govern efficiency and degradation are simply unknown, too difficult to measure, or only partially understood. Thus, realization of a globally competitive, economically viable renewable solar energy industry requires the discovery of game-changing materials and transformational leaps in understanding the energy conversion and degradation processes at work in component materials. A partnership between industry, academia, and DOE experts in HPC in materials and chemistry is essential in achieving these goals.

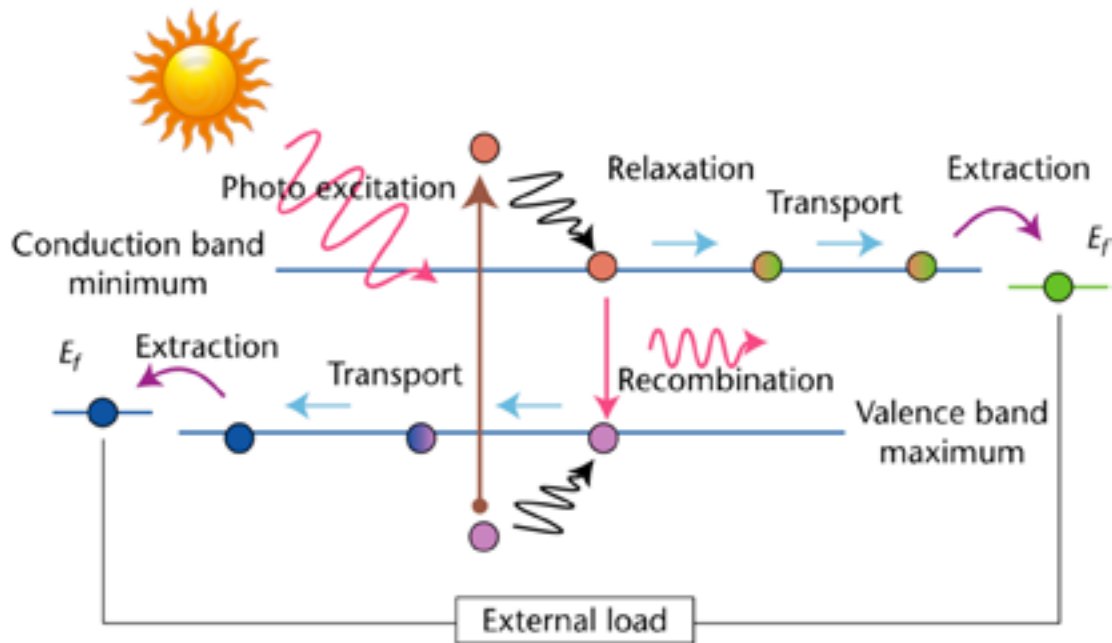


Figure 4.1: Fundamental Processing Contributing to Photovoltaic Efficiency (courtesy of J. Neaton, Lawrence Berkeley National Laboratory)

and manufacturing, and 4) understanding and controlling aging (the evolution and degradation of component materials over time).

STATE OF EXPERTISE IN COMPUTATION

The development of new, inexpensive solar cells with enhanced performance requires new breakthroughs in materials discovery and understanding. Present-day DOE high-performance computational tools—particularly recent innovations in software and algorithms—increasingly allow the prediction and understanding of the evolution of the atomic-scale structure, including defect and vacancy diffusion of complex materials on multiple time and length scales and, with some refinements, the efficiency at which they concentrate and absorb light and transport charge. As these methodologies truly represent the frontiers of DOE capabilities in computational materials science and chemistry, further developments of these computational tools and methods in the context of industrial needs for PVs will have a tremendous positive impact on both industry and computational researchers.

Solar energy conversion phenomena present additional challenges to the development of computational tools. For example, although computing electronic properties of periodic bulk materials is routine with contemporary DFT methods and calculating accurate electronic properties of individual molecules involves a standard set of quantum chemical methods that take advantage of their molecular size, neither approach is entirely sufficient to treat aperiodic systems, i.e., complex interfaces, point defects, and systems out of equilibrium with electrons entering and leaving the device through long leads held at fixed potentials. Only recently have DFT- and wave-function-based methods been formulated for treating nanostructures under finite bias. Further development, validation, and use of these software tools will be required for problems relevant to optical absorption, thermalization, charge separation, and carrier transport. Examples

of relevant capabilities include the solution of the time-dependent hybrid DFT coupled-cluster equations, Bethe-Salpeter equations, and so-called GW equations.

Understanding the manner in which these materials—in particular, complex interfaces and heterogeneous assemblies—evolve with time, initially during growth and later during cycling and associated underlying atomic-scale diffusion kinetics, can unlock mysteries of how aging and degradation proceeds. Recent advances in MD methods, including accelerated, rare-event approaches; reactive force-fields; kinetic Monte Carlo; and phase-field and continuum modeling, can be used to study both processes relevant to the long-time dynamics of materials at multiple length scales, from atoms to dislocations and cracks to microcrystalline domains.

Additional factors controlling their external quantum efficiency include anti-reflective packaging or surface treatments and photonics that further concentrate light into active regions. Electrodynamics simulations can be used to predict absorption of solar cells to design more effective device geometries for maximal light collection. Integration of photonic structures that allow wavelength shifting and optical field concentration can lead to substantial increases in efficiency.

Finally, as shown in recent work by Kang and Ceder (2009) in energy storage and Nørskov et al. (2011) in heterogeneous catalysis, it is now possible to screen materials for one or more properties over hundreds of thousands of possible combinations of elements across the entire periodic table, suggesting many new materials solutions that defy intuition. In principle, light-harvesting materials with new composition profiles could be discovered using this approach. However, recent advances in theory, algorithms, hardware, and materials and chemical sciences have not been available to the majority of the commercial sector—a situation that seriously threatens the realization of revolutionary breakthroughs in complex materials chemistry and materials design.

ACTION ITEMS AND STRATEGIES

Industry faces several materials-based and conceptual barriers to achieving the low-cost, high efficiency, robust solar cell technology that can lead to energy independence. Developments in high-performance computational tools, particularly emerging methods

in computational materials and chemistry, have the ability to address many of these roadblocks (pending some critical refinements) with implications beyond PVs. However, widespread use of more than any one of these methods by experts in computational chemistry and materials science is rare, and significant but well-defined efforts are required to translate these tools to industry.

Needs and Issues in System Performance Modeling

SunPower created its own proprietary simulation tool, called PVSim. The model is generally built from publicly available algorithms, such as irradiance translation algorithms and the Sandia module and inverter performance models. Some algorithms, such as those for shading and tracking, were developed by SunPower and are specific to the company’s products. SunPower has instrumented more than 650 systems and has used these data to validate and improve the accuracy of PVSim. This allows SunPower to provide customers with accurate estimates of power and energy production and gives SunPower the information they need to price their systems. As shown in Figure 4.3, on average, annual delivery is 1.2% greater than predicted.

SunPower cited soiling as the greatest source of uncertainty. While simple soiling models work well most of the time, some climates and regions are more difficult because of soiling composition, variable rainfall, ambient conditions, the surrounding environment, and avian migration patterns.

SunPower highlighted the need for evaluation and validation of performance models, including reconciliation of the models and third-party field test data. Testing standards supported by quality requirements and audits are needed and must account for baseline and evolving module ratings. Also, there needs to be a clear definition and understanding of metrics including delivered AC power, annual energy, yield (in kilowatt hours/kilowatt-peak [kWh/kWp], noting the importance of the watt-peak [Wp] rating), and Levelized Cost of Energy.

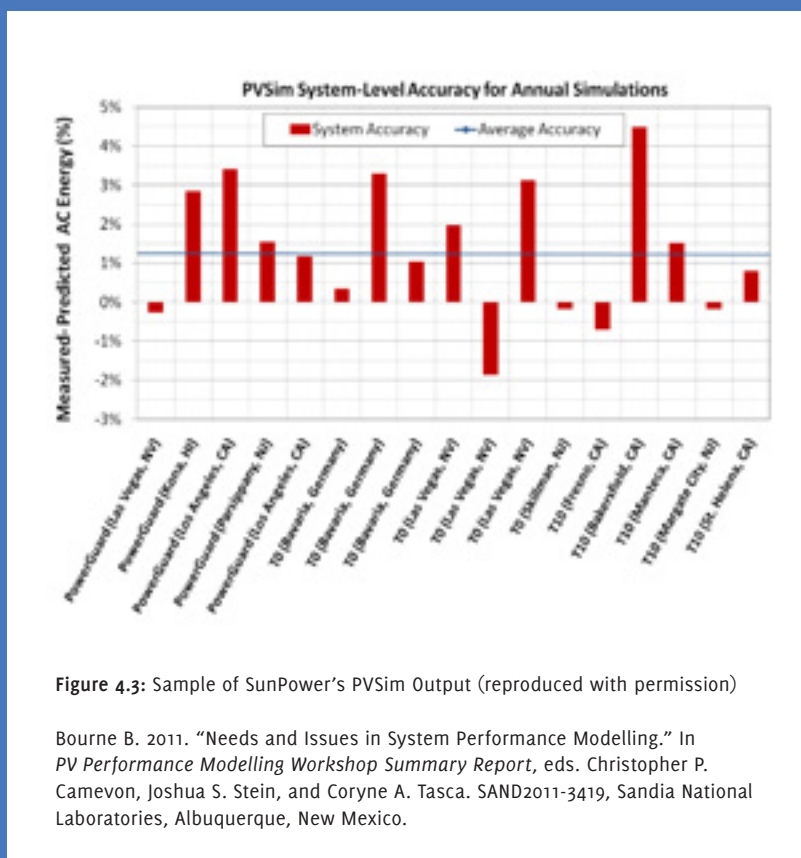


Figure 4.3: Sample of SunPower’s PVSim Output (reproduced with permission)

Bourne B. 2011. “Needs and Issues in System Performance Modelling.” In *PV Performance Modelling Workshop Summary Report*, eds. Christopher P. Camevon, Joshua S. Stein, and Coryne A. Tasca. SAND2011-3419, Sandia National Laboratories, Albuquerque, New Mexico.

One way to catalyze translation would be to incentivize their refinement through supported partnerships aimed at the most important problems facing industry in solar-to-electric energy conversion. Such partnerships must be developed with the notion that industry often is constrained by the need for fast turnarounds and flexibility, while effective legal negotiations between DOE laboratories and industry may need to be further developed. However, prior to the formation of such partnerships, communication between industry and DOE computational scientists must increase and resources be made available to unite DOE researchers in the community with specific industrial partners. Industry is largely unaware of the computational resources and capabilities available within the DOE—much less under development. New avenues for communication between industry and computational researchers in chemistry and materials are needed.

A curated website, or virtual hub, that organizes and summarizes DOE-funded activities within EFRCs, Nanoscience Research Centers, and other DOE programs with computational materials and chemistry components by sub-topic (photovoltaics, energy storage, etc.) would help industry identify specific DOE partners and assess the value they might add. A DOE virtual hub also could amass relevant data, such as optical properties and resistivity, defect energies, etc., on materials of importance to PVs from computation, experiments, or both. A virtual hub also would include additional information important for partnering with DOE laboratories when IP is involved (e.g., legal negotiations). Joint workshops that expose industrial leaders to DOE computational capabilities, and, simultaneously, computational scientists to industry was a shared idea among all session panels and would readily apply here.

BENEFITS

Photovoltaic technologies present an extremely important opportunity to help meet future energy needs of the United States and the world. Because sunlight is abundant, renewable, and free, its

conversion to electricity represents a huge potential for meeting this nation's energy needs. However, many barriers make such conversions economically challenging. Limited global quantities of necessary raw materials prevent certain photovoltaic technologies from being able to meet global energy demands. Technological limits of current technologies restrict the efficiency and lifetime of PV devices. All of these barriers and limits prevent photovoltaics from being a major source of renewable energy. Computational methods can dramatically improve the design, efficiency, cost, global capacity, and longevity of photovoltaics, as well as speed the time required to make technological advances. Furthermore, computational tools can improve the way these devices are operated to maximize overall efficiency and reduce costs. Utilization of computational tools can speed discovery and development of more efficient, less expensive, and longer-lasting materials to provide U.S. companies a competitive advantage in worldwide markets. Modeling tools also can improve production processes.

Specific areas where HPC can assist in the development of new and improved photovoltaic technologies include:

- Providing a better understanding of charge carrier transport in various cell designs
- Enhancing understanding of device losses
- Improving understanding of aging and degradation of each component of photovoltaic devices
- Speeding the time for development of new photovoltaic materials, especially those that do not include rare earth elements
- Minimizing material demands and production costs
- Improving design elements for PV devices, including affording improved ability to translate laboratory-scale designs to production-scale devices
- Maximizing the efficiency of large-scale arrays of devices to achieve greater efficiency.

Overarching Research Priorities and Action Items

While research priorities and action items have been presented for each of the four breakout topics, there are several overarching issues common to all of the breakouts, as well as to other areas where HPC can be applied. These key priorities are:

- **Databases**—There is a need for developing and maintaining databases that are accessible to the broad community in all sectors (academia, government, industry). Protocols for supporting long-term storage, data management, and retrieval are needed.
- **Design of Materials**—Computational tools are lacking to design new materials. As a result, the cost and time frame of new materials development leaves the U.S. at a disadvantage versus international competitors.
- **Stability and Aging**—In all areas of energy modeling, it is important to develop tools that can predict the stability, degradation, and aging of materials. Understanding these processes can result in new designs and strategies that can increase the lifetime of materials and substantially decrease costs.
- **Multiple Length and Time Scales**—Energy challenges require coupled modeling across multiple time and length scales. New approaches are needed that can effectively span very large time and length scales in an integrated model.
- **Electron and Charge Carrier Transport**—Most energy applications involve transfer and transport of electrons or other charge carriers. Better tools to model these transport properties in relevant materials are needed to improve performance in multiple sectors.
- **Entropy and Complexity**—All energy applications include important parameters related to disorder, and most systems have a high level of complexity. Modeling tools need to predict disorder and function in highly complex systems.
- **Uncertainty**—Methods are needed to accurately estimate the level of uncertainty in model predictions so investment risk can be evaluated.
- **Collaboration**—Improved collaboration between all stakeholders (academia, government, and industry) is needed to speed the transition from basic research to commercial application and improve U.S. competitiveness in the presence of increasing international competition. Furthermore, more effective mechanisms are needed for dealing with IP issues.

- Henry CS, M DeJongh, AA Best, PM Frybarger, B Lindsay, and RL Stevens. 2010. "High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models." *Nature Biotechnology* 28:977-982. DOI: 10.1038/nbt.1672.
- Jarvis MW, TJ Haas, BS Donohoe, JW Daily, KR Gaston, WJ Frederick, and MR Nimlos. 2010. "Elucidation of Biomass Pyrolysis Products Using Laminar Entrained Flow Reactor and Char Particle Imaging." *Energy & Fuels* 25(1):324-336. DOI: 10.1021/ef100832d.
- Kang B and G Ceder. 2009. "Battery materials for ultrafast charging and discharging." *Nature* 458:190-193.
- Kim J and JL Reed. 2010. "OptORF: Optimal Metabolic and Regulatory Perturbations for Metabolic Engineering of Microbial Strains." *BMC Systems Biology* 4:53. DOI: 10.1186/1752-0509-4-53.
- Kline KL, VH Dale, R Lee, and P Leiby. 2009. "In Defense of Biofuels, Done Right." *Issues in Science and Technology* 25(3):75-84.
- Leiby PN. 2007. *Estimating the Energy Security Benefits of Reduced U.S. Oil Imports*. ORNL/TM-2007/028, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- Leiby PN and RU Martinez. 2011. "Quantifying Energy Security Benefits of Biofuels." Presentation at the Transportation Research Board Annual Meeting, January 25, Washington, D.C. Session: *Transportation Energy: Climate Change and Energy Security—Role of Biofuels*.
- Li W and P Piecuch. 2010. Improved Design of Orbital Domains within the Cluster-in-Molecule Local Correlation Framework: Single-Environment Cluster-in-Molecule Ansatz and its Application to Local Coupled-Cluster Approach with Singles and Doubles." *Journal of Physical Chemistry A* 114:8644-8657.
- Lun DS, G Rockwell, NJ Guido, M Baym, JA Kelner, B Berger, JE Galagan, and GM Church. 2009. "Large-scale identification of genetic design strategies using local search." *Molecular Systems Biology* 5:296. DOI: 10.1038/msb.2009.57.
- Ma HW and AP Zeng. 2003. "Reconstruction of Metabolic Networks from Genome Data and Analysis of their Global Structure for Various Organisms." *Bioinformatics* 19(2):270-277.
- Mukherjee PP, S Pannala, and JA Turner. 2011. "Modeling and Simulation of Battery Systems." In *Handbook of Battery Materials*, eds. C Daniel and JO Besenhard, Wiley-VCH, Weinheim, Germany.
- Neurock M, LJ Broadbelt, HC Foley, and MT Klein. 1991. "Reaction Pathway Analysis: Global Molecular and Mechanistic Perspectives." *Selectivity in Catalysis*, ACS Symposium Series 517, pp. 290-315. American Chemical Society, Washington D.C.
- Nichols JA, S Kang, WM Post, X Zhang, D Manowitz, TO West, VP Bandaru, and RC Izaurralde. 2010. "Approaches for high resolution land use and bioenergy modeling." Presented at *Center for Bioenergy Sustainability Monthly Forum*, November 18, 2010, Oak Ridge National Laboratory, Oak Ridge, Tennessee. Accessed on March 28, 2011 at http://www.ornl.gov/sci/ees/cbes/forums/Slides_Nov10.pdf.
- Nørskov JK, F Abild-Pedersen, F Studt, and T Bligaard. 2011. "Surface Chemistry Special Feature: Density functional theory in surface chemistry and catalysis." *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 108(3):937-943.
- Panel on New Directions in Catalytic Science and Technology and National Research Council. 1992. *Catalysis Looks to the Future*, National Academies Press, Washington D.C.
- Patil KR, I Rocha, J Förster, and J Nielsen. 2005. "Evolutionary programming as a platform for *in silico* metabolic engineering." *BMC Bioinformatics* 6:308. DOI: 10.1186/1471-2105-6-308.
- Pharkya P and CD Maranas. 2006. "An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems." *Metabolic Engineering* 8(1):1-13.

Conference Agenda

Harnessing the Department of Energy's High-Performance Computing Expertise to Strengthen the U.S. Chemical Enterprise

A Workshop hosted by the Council for Chemical Research and the U.S. Department of Energy

Thursday, March 10

- 7:00 am – 8:00 am **Breakfast and Registration**
- 8:00 am – 8:10 am **Welcome and Opening Remarks**
Matt Tarr and Terry Ring
- 8:10 am – 8:40 am **Broad DOE Perspective**
Dr. Steven E. Koonin
U.S. Department of Energy
- 8:40 am – 9:10 am **Computational Tools: Overview of Tools Available**
Mark Gordon
Iowa State University / Ames Laboratory
- 9:10 am – 9:40 am **Computational Tools: Current Length and Time Scales**
Priya Vashishta
University of Southern California
- 9:40 am – 10:10 am **Computational Tools: Informatics**
Jens Norskov
Stanford University
- 10:10 am – 10:30 am **Break**
- 10:30 am – 11:00 am **Industrial Perspectives**
Tim Mueller
DuPont
- 11:00 am – 11:30 am **Industrial Perspectives**
Mark Schure
The Dow Chemical Company
- 11:30 am – 12:00 pm **Charge and Organization into Breakout Groups**
Matt Tarr and Terry Ring
- 12:00 am – 1:00 pm **Lunch**

Thursday, March 10 (Continued)

- 1:00 pm – 2:00 pm **Breakout Presentations**
- Presentation 1: Biomass and Bioenergy
 - Presentation 2: Catalytic Materials
 - Presentation 3: Energy Storage
 - Presentation 4: Photovoltaics
- 2:00 pm – 6:00 pm **Breakout Discussions**
- Breaks as needed

Friday, March 11

- 7:00 am – 8:00 am **Breakfast**
- 8:00 am – 10:00 am **Breakout Groups Wrap-up and Prepare Summary**
- 10:00 am – 10:30 am **Break**
- 10:30 am – 11:30 am **Breakout Groups Report to All Attendees on Progress**
15 minutes each
- 11:30 am – 12:00 pm **Meeting Summary and Future Goals**
Krish Doraiswamy

Support for this workshop was provided by the U.S. Department of Energy Office of Science through the Office of Basic Energy Sciences and Office of Advanced Scientific Computing Research.

Breakout Session Participants

Catalytic Materials

NAME	INSTITUTION	EMAIL ADDRESS
Dr. Thomas Allison	NIST	thomas.allison@nist.gov
Dr. Edoardo Apra	Oak Ridge National Laboratory	aprae@ornl.gov
Dr. David Dixon	The University of Alabama	dadixon@bama.ua.edu
Dr. Jack Douglas	NIST	jack.douglas@nist.gov
Dr. Marivi Fernandez-Serra	Stony Brook University	mvfernandezs@ms.cc.sunysb.edu
Dr. Bruce Garrett	Pacific Northwest National Laboratory	bruce.garrett@pnnl.gov
Dr. Joseph Golab	INEOS Technologies	joseph.golab@ineos.com
Dr. Shiv Khanna	Virginia Commonwealth University	snkhanna@vcu.edu
Dr. Istvam Lengyel	The Dow Chemical Company	ilengyel@dow.com
Dr. Timothy Mueller	DuPont Central Research & Development	timothy.e.mueller@usa.dupont.com
Dr. Jens Nørskov	SLAC National Accelerator Laboratory	norskov@stanford.edu
Dr. Mark Pederson	Department of Energy	Mark.Pederson@science.doe.gov
Dr. Piotr Piecuch	Michigan State University	piecuch@chemistry.msu.edu
Dr. Tom Pierce	The Dow Chemical Company	tpierce@dow.com
Dr. Sriraj Srinivasan	Arkema	sriraj.srinivasan@arkema.com

Biomass and Bioenergy

NAME	INSTITUTION	EMAIL ADDRESS
Dr. Jhih-Wei Chu	University of California at Berkeley	jwchu@berkeley.edu
Dr. Michael Crowley	National Renewable Energy Lab	michael.crowley@nrel.gov
Dr. Robin Graham	Oak Ridge National Laboratory	grahamrl@ornl.gov
Dr. Christopher Henry	Argonne National Laboratory	chenry@mcs.anl.gov
Dr. Wesley Jones	National Renewable Energy Lab	wesley.jones@nrel.gov
Dr. Donghei Mei	Pacific Northwest National Laboratory	donghai.mei@pnnl.gov
Dr. Perrine Pepiot	National Renewable Energy Lab	perrine.pepiot@nrel.gov
Dr. Marcellinus Pont	E.I. Du Pont de Nemours & Co.	marcellinus.pont@usa.dupont.com
Dr. Rahul Ramachandran	Oak Ridge National Laboratory	ramachandran@ornl.gov
Dr. Jennifer Reed	University of Wisconsin-Madison	reed@engr.wisc.edu
Dr. John Sheehan	University of Minnesota	sheeh179@umn.edu
Dr. Jean-Francois Tomb	E.I. Du Pont de Nemours & Co.	jean-francois.tomb@usa.dupont.com
Dr. Dionisios Vlachos	University of Delaware	vlachos@udel.edu

