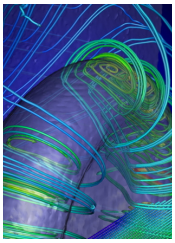
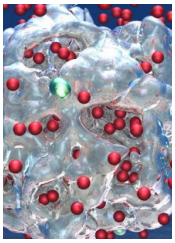
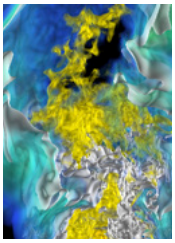
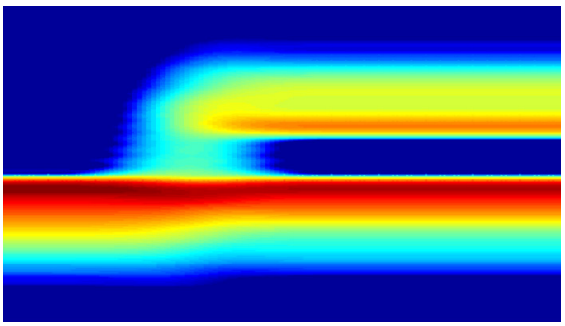
BES

BASIC ENERGY SCIENCES

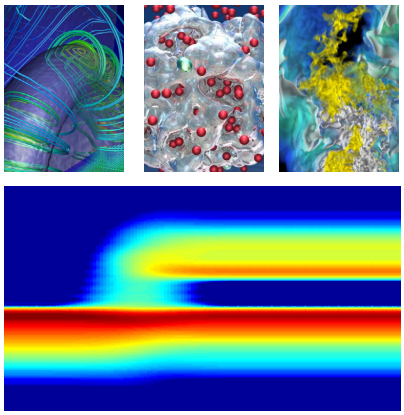
EXASCALE REQUIREMENTS REVIEW

An Office of Science review sponsored jointly by
Advanced Scientific Computing Research and Basic Energy Sciences

NOVEMBER 3–5, 2015

ROCKVILLE, MARYLAND

U.S. DEPARTMENT OF ENERGY



On the cover:

Top left: Billion-atom reactive molecular dynamics simulation of cavitation bubble collapse in water [Shekhar et al. 2013].

Top middle: A 16,611-atom quantum molecular dynamics simulation of H_2 production from water using a LiAl-alloy particle. The valance electron density (silver isosurface) is centered around Al atoms, whereas some of the Li atoms (red spheres) are dissolved into water. Produced H_2 molecules are represented by green ellipsoids. Water molecules are not shown for clarity [Shimamura et al. 2014].

Top right: Direct Numerical Simulation of a Turbulent Di-methyl Ether Lifted Jet Flame with Multistage Ignition in the Negative Temperature Coefficient Regime at 5 atm. The cool-flame marker is a mass fraction of $CH_3OCH_2O_2$ (yellow-red), and the flame marker is a mass fraction of OH (blue). Extending this to high pressure (40 atm) with n-dodecane and including cool-flame ignition and dynamics of lifted flames and soot generation will require exascale resources, chemical model development and reduction strategies guided by uncertainty quantification, and numerical algorithmic development [Minamoto and Chen 2016].

Bottom: Nonequilibrium "melting" of a charge density wave insulator via an ultrafast laser pulse [Shen et al. 2014].

Disclaimer

This report was prepared as an account of a meeting sponsored by the U.S. Department of Energy. Neither the United States Government nor any agency thereof, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Copyrights to portions of this report (including graphics) are reserved by original copyright holders or their assignees, and are used by the Government's license and by permission. Requests to use any images must be made to the provider identified in the image credits.

This report is available at https://science.energy.gov/bes/community-resources/reports/BES-EXA_rpt.pdf. This report, along with all of the DOE Office of Science Exascale Requirement Review reports, can also be accessed at exascaleage.org.

BASIC ENERGY SCIENCES

U.S. DEPARTMENT OF ENERGY

An Office of Science review sponsored jointly by
Advanced Scientific Computing Research and Basic Energy Sciences

Meeting Co-Chairs
Theresa Windus, Ames Laboratory, Iowa State University
Thomas Devereaux, SLAC National Accelerator Laboratory & Stanford University
Michael Banda, Advanced Light Source, Lawrence Berkeley National Laboratory

Meeting Organizer
Julia C. White, Oak Ridge National Laboratory

Report Section Leads
We want to thank the meeting attendees, who all contributed to the writing and reviewing of this report. In addition, these individuals took leading roles on the following sections.

Novel Quantum Materials and Chemicals
Aurora Clark, Washington State University
Andy Millis, Columbia University

Catalysis, Photosynthesis and Light Harvesting, and Combustion
Laura Gagliardi, University of Minnesota
Thanos Panagiotopoulos, Princeton University

Materials and Chemical Discovery
Ilja Siepmann, University of Minnesota
Chris Wolverton, Northwestern University

Soft Matter
Priya Vashishta, University of Southern California
Ilja Siepmann, University of Minnesota
Mark Stevens, Sandia National Laboratories

Advances in Algorithms for Quantum Systems
Mark Gordon, Iowa State University
Paul Kent, Oak Ridge National Laboratory

Computing and Data Challenges @ BES Facilities
Kerstin Kleese van Dam, Brookhaven National Laboratory
Thomas Proffen, Oak Ridge National Laboratory
Craig Tull, Lawrence Berkeley National Laboratory

Mathematics and Computer Science Transforming BES Science: Mathematics and Computer Science
Lori Diachin, Lawrence Livermore National Laboratory
Jamie Sethian, University of California-Berkeley

Mathematics and Computer Science Transforming BES Science: Programming Models
Anouar Benali, Argonne National Laboratory
Jackie Chen, Sandia National Laboratories

Path Forward
Katie Antypas, National Energy Research Scientific Computing Center
Richard Gerber, National Energy Research Scientific Computing Center
Katherine Riley, Argonne Leadership Computing Facility
Tjerk Straatsma, Oak Ridge Leadership Computing Facility

Sponsors and Representatives

James Davenport, DOE/Office of Basic Energy Sciences
Matthias Graf, DOE/Office of Basic Energy Sciences
Carolyn Lauzon, DOE/Office of Advanced Scientific Computing Research
Eliane Lessner, DOE/Office of Basic Energy Sciences
Mark Pederson, DOE/Office of Basic Energy Sciences

Additional Contributors

Argonne National Laboratory Document Design and Development Team
Mary Fitzpatrick, Andrea Manning, Michele Nelson, and Vicki Skonicki, Communications, Education, and Public Affairs
Beth Cerny and Richard Coffey, Argonne Leadership Computing Facility

TABLE OF CONTENTS

Executive Summary **1**
Abstract 1
ES.1 Summary and Key Findings 1
ES.2 Basic Energy Sciences Vision and Grand Challenges 3
ES.3 Priority Research Topics and Computing Needs 4
ES.3.1 Novel Quantum Materials and Chemicals 4
ES.3.2 Catalysis, Photosynthesis and Light Harvesting, and Combustion 4
ES.3.3 Materials and Chemical Discovery 5
ES.3.4 Soft Matter 5
ES.3.5 Advances in Algorithms for Quantum Systems 5
ES.3.6 Computing and Data Challenges at BES Facilities 5
ES.3.7 Mathematics and Computer Science Transforming BES Science 6
ES.3.8 Next-Generation Workforce 6
ES.4 Path Forward 6

1 Introduction **9**
1.1 Goal of the DOE Exascale Requirements Reviews 9
1.1.1 Previous DOE Requirements-Gathering Efforts: "Lead with the Science" 9
1.1.2 National Strategic Computing Initiative (NSCI) 9
1.2 BES Exascale Requirements Review 10
1.3 Report Organization 11

2 Basic Energy Sciences: Vision and Grand Challenges **13**
2.1 BES Vision 13
2.2 BES Grand Challenges 13
2.3 Mapping the BES Grand Challenges to the Computing Ecosystem and Exascale 16
2.3.1 Novel Quantum Materials and Chemicals 16
2.3.2 Catalysis, Photosynthesis and Light Harvesting, and Combustion 16
2.3.3 Materials and Chemical Discovery 17
2.3.4 Soft Matter 17
2.3.5 Advances in Algorithms for Quantum Systems 18
2.3.6 Computing and Data Challenges at BES Facilities 18

2.3.7 Mathematics and Computer Science Transforming BES Science 18
2.3.8 Next-Generation Workforce 19
2.4 Report Roadmap 19

3 Research Directions and Computing Needs/Requirements 21
3.1 Quantum Materials and Chemistry 21
3.1.1 Scientific Challenge and Opportunities 21
3.1.2 Priority Research Directions 23
3.1.2.1 Emergent Properties of Quantum Materials: Superconductivity, Magnetism, Topological Protection, and Beyond 23
3.1.2.2 Heavy-Element Science 24
3.1.2.3 Advanced Spectroscopies 25
3.1.3 Cross-Cutting Research Directions 26
3.1.4 Computing Needs and Requirements 28
3.2 Catalysis, Photosynthesis and Light Harvesting, and Combustion 30
3.2.1 Scientific Challenges and Opportunities 30
3.2.1.1 Catalysis 30
3.2.1.2 Photosynthesis and Light Harvesting 31
3.2.1.3 Combustion 32
3.2.2 Priority Research Directions 33
3.2.2.1 Catalysis 33
3.2.2.2 Photosynthesis and Light Harvesting 34
3.2.2.3 Combustion 35
3.2.3 Cross-Cutting Research Directions 36
3.2.4 Computing Needs and Requirements 37
3.2.4.1 Catalysis 37
3.2.4.2 Photosynthesis and Light Harvesting 38
3.2.4.3 Combustion 38
3.2.4.4 Other Needs and Requirements 38
3.3 Materials and Chemical Discovery 41
3.3.1 Scientific Challenges and Opportunities 41
3.3.2 Priority Research Directions 43
3.3.3 Cross-Cutting Research Directions 46
3.3.3.1 Validation and Uncertainty Quantification in Predictive Simulations 46
3.3.4 Computing Needs and Requirements 47

3.4 Soft Matter 49
3.4.1 Scientific Challenges and Opportunities 49
3.4.2 Priority Research Directions 51
3.4.2.1 Alternative Separations 51
3.4.2.2 Multicomponent Fluid Mixtures 52
3.4.2.3 Composite Polymeric Materials 52
3.4.2.4 Rational Design of Polymer Dielectrics for Energy Storage 53
3.4.3 Cross-Cutting Research Directions 54
3.4.3.1 Developing Polarizable and Reactive Potentials 54
3.4.3.2 Developing Methods for Long-range Electrostatics for Exascale Architectures 55
3.4.3.3 Developing a Framework for Multiscale Science 55
3.4.4 Computing Needs and Requirements 56
3.5 Advances in Algorithms for Quantum Systems 58
3.5.1 Challenges and Opportunities 58
3.5.2 Priority Research Directions 58
3.5.3 Cross-Cutting Directions 60
3.5.4 Computing Needs and Requirements 60
3.6 Computing and Data Challenges @ BES Facilities 63
3.6.1 Challenges and Opportunities 63
3.6.2 Priority Directions in User Support 64
3.6.2.1 Streaming Analysis 64
3.6.2.2 Multimodal Analysis 68
3.6.2.3 Data Curation 69
3.6.2.4 Accelerator Simulation 71
3.7 Mathematics and Computer Science Transforming BES Science 75
3.7.1 Challenges and Opportunities 75
3.7.2 Goals 75
3.7.2.1 Simulation 76
3.7.2.2 Experiment 77
3.7.2.3 Software Development and Optimization for Extreme-Scale Computers 79
3.7.3 Cross-Cutting Issues 81
3.7.3.1 Use of Applied Math and Computer Science Software Libraries and Tools 81

3.8 Next-Generation Workforce 83
3.8.1 Meeting the Exascale Challenge Is Grounded in Workforce Development 83
3.8.2 Priority Directions to Promote the Next-Generation Workforce 83
3.8.3 Cross-Cutting Opportunities to Promote the Next-Generation Workforce 84
3.8.3.1 Collaboration Opportunities 84
3.8.3.2 At the Science and Computing Interface: The Application Engineer Position 85

4 Path Forward 87
4.1 Methods Development 87
4.2 Computational Environment 88
4.3 Data 88
4.4 Communication and Community Involvement 89
4.5 Conclusion 89

5 References 91

6 Acronyms and Abbreviations 97

Appendices: Meeting Materials 101
Appendix A: Basic Energy Sciences Organizing Committee and Meeting Participants A-1
Appendix B: Basic Energy Sciences Meeting Agenda B-1
Appendix C: Basic Energy Sciences White Papers C-1
Appendix D: Basic Energy Sciences Case Studies D-1

FIGURES

Figure 2-1. Opportunities at the Frontiers of Discovery Science for Matter and Energy. 15

Figure 3-1. Metal and Metal-Oxide Catalyst of Atomic Precision Anchored at a Support for Natural Gas Conversion Catalysis. 34

Figure 3-2. Direct Numerical Simulation of a Turbulent Di-methyl Ether Lifted Jet Flame with Multistage Ignition in the Negative Temperature Coefficient Regime at 5 atm. 36

Figure 3-3.. Nanoparticles Entering a Solution Phase as Part of a Complex Synthesis Pathway. 43

Figure 3-4. The Ubiquitous Role of Heterogeneous Systems. 44

Figure 3-5. Historical Rates of Materials Discoveries. 45

Figure 3-6. Decision Tree Diagram of Structure Determination and Computing. 46

Figure 3-7. Photosystem II. 50

Figure 3-8. Prediction of Many Soft Materials Properties from Chemical Structure/State Consists of Multiple Steps. 51

Figure 3-9. The Primary Steps Involved in a Rational Collaborative Design (or Co-design) Approach. 53

Figure 3-10. Observed Standard Deviation for the Sum of Sets of *n* Random Numbers. 56

Figure 3-11. Chemical Composition Mapping with Nanometer Resolution by Soft X-ray Microscopy. 66

Figure 3-12. Common Publications. 68

Figure 3-13. Mathematics and Computer Science Embedded in Advanced Scientific Discovery. 75

EXECUTIVE SUMMARY

Abstract

Computers have revolutionized every aspect of our lives. Yet in science, the most tantalizing applications of computing lie just beyond our reach. The current quest to build an exascale computer with one thousand times the capability of today's fastest machines (and more than a million times that of a laptop) will take researchers over the next horizon. The field of materials, chemical reactions, and compounds is inherently complex. Imagine millions of new materials with new functionalities waiting to be discovered — while researchers also seek to extend those materials that are known to a dizzying number of new forms. We could translate massive amounts of data from high precision experiments into new understanding through data mining and analysis. We could have at our disposal the ability to predict the properties of these materials, to follow their transformations during reactions on an atom-by-atom basis, and to discover completely new chemical pathways or physical states of matter. Extending these predictions from the nanoscale to the mesoscale, from the ultrafast world of reactions to long-time simulations to predict the lifetime performance of materials, and to the discovery of new materials and processes will have a profound impact on energy technology. In addition, discovery of new materials is vital to move computing beyond Moore's law. To realize this vision, more than hardware is needed. New algorithms to take advantage of the increase in computing power, new programming paradigms, and new ways of mining massive data sets are needed as well. This report summarizes the opportunities and the requisite computing ecosystem needed to realize the potential before us.

In addition to pursuing new and more complete physical models and theoretical frameworks, this review found that the following broadly grouped areas relevant to the U.S. Department of Energy (DOE) Office of Advanced Scientific Computing Research (ASCR) would directly affect the Basic Energy Sciences (BES) mission need.

- **Simulation, visualization, and data analysis are crucial for advances in energy science and technology.**
- **Revolutionary mathematical, software, and algorithm developments are required in all areas of BES science to take advantage of exascale computing architectures and to meet data analysis, management, and workflow needs.**
- **In partnership with ASCR, BES has an emerging and pressing need to develop new and disruptive capabilities in data science.**
- **More capable and larger high-performance computing (HPC) and data ecosystems are required to support priority research in BES.**
- **Continued success in BES research requires developing the next-generation workforce through education and training and by providing sustained career opportunities.**

ES.1 Summary and Key Findings

The findings presented throughout this report are the result of a joint requirements review by BES scientists and DOE ASCR facilities teams. The mission of BES is to support fundamental research to understand, predict, and ultimately control matter and energy at the electronic, atomic, and molecular levels in order to provide the foundations for new energy technologies. The BES scientists focused on current scientific challenges in the areas of novel quantum materials and chemicals; catalysis, photosynthesis, and light harvesting; combustion; materials and chemical

discovery; soft matter; advances in algorithms for quantum systems; and computing and data challenges at BES user facilities (each described in greater detail in the following sections). These scientific challenges were then evaluated in the context of the computational science and computing ecosystems required to achieve the BES mission goals.

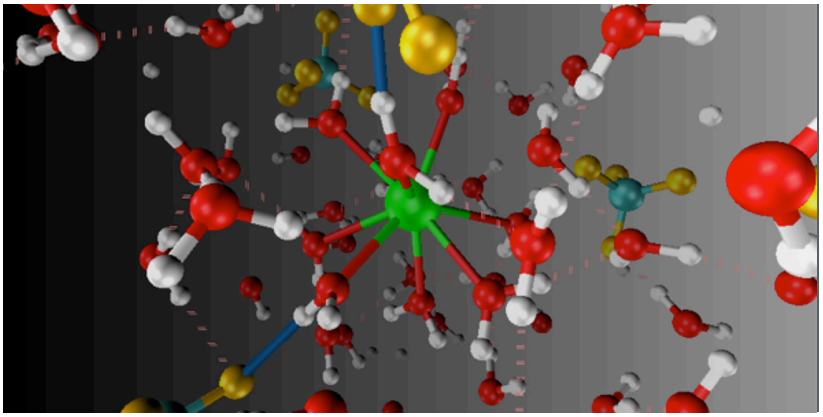
In order for BES science to take advantage of exascale computing architectures, software and algorithm development is required in all areas. Researchers need a new suite of sustainable and performant software tools, programming models, and applications to enable effective use of exascale systems. New algorithms are needed to enable codes to run efficiently on upcoming HPC architectures, allowing scientists to model larger materials and chemical systems with greater fidelity and predictive power. New mathematics are required to reduce the computational scaling of current algorithms and to develop new physical models. New software tools are needed to analyze and manage data produced by experiments, including tools that can drive complex workflows. New programming models are required to enable facile development of software that performs across multiple, distinct computing platforms. Attendees agreed that there is an opportunity for mathematics and computer science to have a transformational impact on BES science.

BES and the ASCR facilities are experiencing a pressing need to mature their capabilities in data science. Improvements and new capabilities at BES facilities are creating challenges that the community is not prepared to address. These include unprecedented growth in data volume, complexity, and access requirements; the need for curation of the massive amounts of data that are retained; and integration of diverse datasets from different experiments to enable new scientific conclusions. Efficient and effective use of BES facilities requires real-time access to ASCR HPC facility-class resources to support streaming analysis and visualization to guide experimental decisions.

Last, and of equal importance, a strong vision for workforce development is crucial for scientific success within future computing ecosystems. BES science will require cross-cutting, multidisciplinary teams of domain scientists, applied mathematicians, and computer scientists who can work together to address the broad range of challenges. Apparent to both BES and ASCR participants is a need to vastly improve the current training strategies and opportunities to develop the next generation of scientists and engineers.

Participants in the BES Exascale Requirements Review were asked to articulate the BES vision and grand challenges, identify priority research topics and computing needs, and outline a path forward for BES and ASCR. The following subsections summarize each of these topics in detail.

A VIEW OF THE SCIENCE



Thorium in Water with Perchlorate Counterions. Figure courtesy of Wibe DeJong, Lawrence Berkeley National Laboratory.

ES.2 Basic Energy Sciences Vision and Grand Challenges

BES supports fundamental research to understand, predict, and ultimately control matter and energy across scales from atoms to molecules to materials. This research probes the electronic and magnetic properties of matter and energy — an effort that is foundational to discovery of future energy solutions for the nation. Theory, computation, and the user facilities have played key roles in the BES portfolio to provide these solutions and have been recognized as essential for the future of the program. In 2015, the DOE's Office of Science charged the BES Advisory Committee (BESAC) to develop a series of "grand challenges" that would inspire and guide BES research. That tasking resulted in a report entitled *Challenges at the Frontiers of Matter and Energy: Transformative Opportunities for Discovery Science* (DOE-BESAC 2015), which identified five new transformative opportunities. If realized, these would lead to the control of matter and energy at the molecular, atomic, and quantum levels and could spark revolutionary changes in technologies to help us meet some of humanity's most pressing needs, including the need for renewable, clean, and affordable energy. The transformative opportunities are expressed in the BESAC report as three themes:

- Mastering Hierarchical Architectures and Beyond-Equilibrium Matter
- Beyond Ideal Materials and Systems: Understanding the Critical Roles of Heterogeneity, Interfaces, and Disorder
- Harnessing Coherence in Light and Matter

and two cross-cutting opportunities:

- Exploiting Transformative Advances in Imaging Capabilities across Multiple Scales
- Achieving Revolutionary Advances in Models, Mathematics, Algorithms, Data, and Computing

The energy systems of the future — whether they tap sunlight, store electricity, or make fuel from splitting water or reducing carbon dioxide — will revolve around materials and chemical changes that convert energy from one form to another. These advanced materials and chemical processes are not found in nature; they must be designed and fabricated to exacting standards using principles revealed by basic science. Acquiring the ability to reveal this fundamental knowledge has largely come about as a result of making extremely rapid progress in our ability to analyze and simulate experimental data and systems.

Not surprisingly, one of the transformative opportunities is to develop revolutionary advances in models, mathematics, algorithms, data, and computing. Today we are on the verge of standing up 100-petaflop machines. Furthermore, the Office of Science envisions reaching the exascale level of computing within the next decade (with machines capable of performing a million trillion floating-point calculations per second). This enormous growth in computational power, coupled with major advances in other fields, such as the development of coherent light sources and increased imaging resolution, is accelerating the pace of materials and chemical sciences research that will enable basic understanding and control at the quantum level.

The BES program also operates major scientific user facilities to serve researchers from universities, national laboratories, and private institutions. The availability of state-of-the-art BES user facilities has opened new avenues of research that were not available just a few years ago. The planned and potential upgrades of several BES facilities in the next five to ten years will support more detailed and expanded experiments, producing an exponential growth of data. To manage and extract useful scientific information from those experimental data will require computational resources beyond the capabilities of current resources.

In addition, a strong connection is required between new computing landscapes and experimental and theoretical toolsets to realize transformational opportunities. This review articulated the requirements of BES research and user facilities for models, mathematics, algorithms, data, computing, and workforce development. Implicit in these requirements is the need to increase engagement with ASCR facilities and resources with the goal of developing sustained partnerships to make full use of the ASCR resources to advance the BES science mission.

ES.3 Priority Research Topics and Computing Needs

Review participants focused on eight areas in which advancement in the transformative opportunities can be achieved through key and sustained efforts in computation, simulation, and advanced tool design.

ES.3.1 Novel Quantum Materials and Chemicals

Technologies of the future will rely upon the specific combinations of elements, compounds, materials, and phases whose behaviors are "emergent": synergistically enhanced and not predictable from studying the components in isolation. The range of technological needs is vast, from alternatives to silicon electronics, including spin- and Mott-tronics, optical and neuromorphic device components that take computing past the end of the "Moore's Law" barrier; to high-transition-temperature superconductors that move electricity with no loss of energy; to better thermoelectrics that capture and utilize waste heat; to new, strong, earth-abundant magnets for turbine engines; to extractants for the separation of heavy elements in waste mixes; and to better photovoltaics. The need for new and improved functionalities can be met by new generations of "quantum materials and chemicals," with greatly enhanced responses emerging from the "quantum chemistry" of strongly interacting electrons. Critical to the design of chemicals and materials with the properties we need is the development of new predictive theories, efficient and adaptive software, and exploitation of the full capabilities of new and future computing architectures at ASCR facilities, coupled synergistically with advanced experimental facilities at BES nanocenters and photon and neutron sources.

ES.3.2 Catalysis, Photosynthesis and Light Harvesting, and Combustion

Catalysis is the essential technology for accelerating and directing chemical transformation and is a key to efficient fuel production and industrial processes. To realize the full potential of catalysis for energy applications, scientists must develop a profound understanding of catalytic transformations so that they can design and build effective catalysts with atom-by-atom precision and can convert reactants to products with molecular precision, requiring a fundamental understanding of catalytic processes occurring in multiscale, multiphase environments. Likewise light harvesting and photosynthetic pathways hold the promise of efficient, inexpensive power sources if the processes can be understood and manipulated. Until these alternative energy sources are economically available, combustion will continue to be a dominant mode of energy conversion for transportation, power generation, and industrial thermal processes. High-performance computing, in particular exascale computing, is playing and will play a central role in providing the insight needed to design these energy-efficient transformations that involve processes at multiple length and time scales under real-world — rather than the idealized — conditions that constitute current simulation and modeling scenarios. However, methodologies that deterministically describe the coupling between different scales must be solved to make full use of future HPC platforms.

ES.3.3 Materials and Chemical Discovery

In an ideal world, researchers could create a new material or chemical with exacting properties to meet the need at hand, thus saving extensive costs associated with experimental trial and error. However, predictive modeling of properties, tailored synthesis of chemicals, and control of materials require advances in modeling capabilities and hardware resources and computational interpretive software for experimental techniques — to enable the characterization of spatial and temporal fluctuations and of short-lived intermediates along synthesis and degradation pathways. New computational tools will enable (1) computational discovery of novel materials and chemicals with target properties (including hierarchical structures with multiple functionalities); (2) prediction of pathways to the synthesis of these materials and chemicals (with consideration of sustainability and green chemistry principles); and (3) prediction of their kinetic or thermodynamic stability and degradation pathways.

ES.3.4 Soft Matter

Soft matter provides unique and critical materials behavior in a wide range of industrial products. Polymers, surfactants, electrolytes, and microheterogeneous fluids have long been key components in a multitude of applications, including energy storage (e.g., batteries and capacitors) and energy production (e.g., photosystems), chemical separations, enhanced oil recovery, food packaging, chip manufacturing, and health care products. Soft materials composed of molecular and/or modular building blocks can provide the hierarchical complexity and tunability for making paradigm-shifting materials that can accomplish multiple tasks. The complexity of soft materials presents scientific as well as computational challenges that make exascale computing a pivotal resource in achieving the goal of designing functional matter, which requires not only orders-of-magnitude greater scalability in both dimensional and time scales, but also seamless integration with exabyte big data analytics and mining so as to extract maximal scientific knowledge.

ES.3.5 Advances in Algorithms for Quantum Systems

Implicit in many of the BES mission phenomena is the need to develop truly multiscale methods that can span multiple time and length scales in a seamless and self-consistent manner. However, realistic simulations of complex materials and chemical problems remain out of reach due to their high cost. Truly predictive simulations will require development of robust hierarchical theories and algorithms to treat electron correlations across all relevant length scales. To take advantage of exascale systems, it is necessary to develop highly parallel, low-scaling algorithms for each scale of the system (quantum and classical) and for multiscale methods. In particular, because computer configurations have ever-increasing numbers of computational elements on each node, the newly designed algorithms must be able to take advantage of multilevel parallelism with multiple layers of memory and communication hierarchies.

ES.3.6 Computing and Data Challenges at BES Facilities

Computing and data challenges may be characterized as streaming analysis and steering of experiments, multimodal analysis of results from different instruments, and long-term data curation. The growing complexity of the analysis process (mixing fast data analysis and numerical modeling) will require capabilities beyond the petascale-level capabilities that the ASCR facilities currently offer. Addressing these challenges by providing for the analysis, management, and storage of user data signals a fundamental change in the operation and responsibility of BES user facilities. In addition, the computational capabilities of the future will provide a platform for real-time modeling and simulation so that experiments can be augmented and understood as they are in progress. This coupling, along with the need to manage data and the ability to steer and make decisions during the experiment to optimize the scientific outcomes, will require significant “on-demand” exascale types of computational resources to deliver the necessary feedback and insights in real time as the experimental process unfolds.

ES.3.7 Mathematics and Computer Science Transforming BES Science

Bridging the gap between BES scientific goals and ASCR computing capabilities will fundamentally depend on the collective abilities of our science domains and facilities to deliver transformative breakthroughs in mathematics and computer science — an objective that will require investing in and capitalizing on evolving, state-of-the-art mathematics and computer science. Mathematics will need to be developed that enables order-of-magnitude improvements in speed and accuracy in predictive materials and chemistry modeling. Mathematical algorithms and unified software environments must be delivered to allow fast, multimodal analysis of experimental data across different imaging modalities and DOE facilities. Software tools must be built that will make efficient programming of tomorrow's machines as straightforward as programming today's laptops. Finally, these three advances must be tied together to significantly advance our understanding in scientific domains.

ES.3.8 Next-Generation Workforce

The complexities and multiple layers of hierarchy of next-generation programming environments will include developing the physical and mathematical models, expressing the scientific workflow, developing the numerical algorithms, decomposing the algorithm to the optimal level of task granularity, expressing fine-grained parallelism in domain-specific languages (DSLs), and ensuring that all of the layers of the programming model and runtime have the right abstractions to enable flexibility and performance. Providing the kinds of education and training deliverables to prepare future computational and domain scientists and computational software developers for exascale computing presents a major challenge.

ES.4 Path Forward

The support and development of our evolving computing ecosystem relies on continued collaboration between BES and ASCR. Rooted in the discussions about the BES vision, research directions, and computing needs, four categories grew out of the review: methods development, computational environment, data, and cross-community engagement.

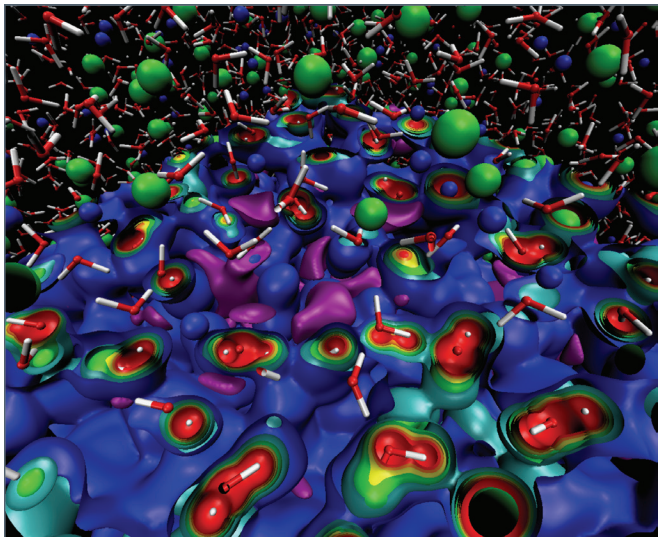
Methods development includes researching new, scalable algorithms; building better memory management techniques; exploring data-related methods; incorporating the full physical and chemical environment in a theoretically sound and yet affordable manner; providing parallel-in-time and new sampling approaches; and developing end-to-end analysis pipelines to solve workflow, visualization, and optimization concerns.

The **computational environment** addresses requirements for access, scheduling, and software environments. A greater demand for near real-time (or on-demand) computing and analysis is driven by HPC needs at experimental facilities during or shortly after experiments. At the same time, the HPC market appears to be delivering exascale and affordable petascale within the same time frame. This creates opportunities for improvement of tools not only for exascale systems but also for the petascale mid-range of tomorrow.

The **scale of data** generated these days creates challenges in performance, locality, analytics, access, and curation. Today, we find that with the increase of data outputs, analysis has moved from the workstation back to the supercomputing center. Analysis now includes machine learning and deep learning techniques. Curation comes into play because communities want the capability to provide access and knowledge of the location of the data in a push to the community, as well as the ability to combine data and analysis to discover new insights into experiment and simulations.

Finally, **cross-community engagement** brings together three fronts: workforce development, collection and feedback with standards, and the development of training materials and best practices. While there are existing efforts, there is an opportunity to create cohesive approaches on one or more of these fronts. The formation of new institutes to connect mathematicians and computer scientists and engineers with BES researchers and facilities is essential toward progress in addressing the challenges of exascale computing.

A VIEW OF THE SCIENCE



Recent Experiments Have Verified That Nucleation Can Emit and Be Induced by Electromagnetic Fields. We are developing a fundamental understanding of the quantum mechanics, statistical mechanics, and electrodynamics to control and exploit the synthesis of matter relevant to DOE's mission. Reproduced from B. Sellner and S.M. Kathmann, "A Matter of Quantum Voltages," JCP, 141, **18C534 (2014).**

1 INTRODUCTION

1.1 Goal of the DOE Exascale Requirements Reviews

During fiscal years (FYs) 2015 and 2016, the Exascale Requirements Reviews brought together key computational domain scientists, U.S. Department of Energy (DOE) planners and administrators, and experts in computer science and applied mathematics. Meetings were held for each of the DOE's six Office of Science (SC) program offices, as follows:

- The High-Energy Physics (HEP) review was held in June 2015.
- The Basic Energy Sciences (BES) review was held November 3-5, 2015.
- The Fusion Energy Sciences (FES) review was held in January 2016.
- The Biological and Environmental Science (BER) review was held in March 2016.
- The Nuclear Physics (NP) review was held in June 2016.
- The Advanced Scientific Computing Research (ASCR) was held in September 2016.

The overarching goal was to determine the requirements for an exascale ecosystem that includes computation, data analysis, software, workflows, high-performance computing (HPC) services, and other programmatic or technological elements that may be needed to support forefront scientific research.

Each Exascale Requirements Review resulted in a report prepared by DOE for wide distribution to subject matter experts and stakeholders at DOE's ASCR facilities, including the Argonne and Oak Ridge Leadership Computing Facility centers (ALCF and OLCF, respectively), the National Energy Research Scientific Computing Center (NERSC), and the Energy Sciences Network (ESnet).

1.1.1 Previous DOE Requirements-Gathering Efforts: "Lead with the Science"

DOE has experienced definite value in implementing its previous requirements-gathering efforts. As noted by Helland (2015), such review meetings have served to:

- Establish requirements, capabilities, and services.
- Enable scientists, programs offices, and the facilities to have the same conversation.
- Provide a solid, fact-based foundation for service and capability investments.
- Address DOE mission goals by ensuring that DOE science is supported effectively.

1.1.2 National Strategic Computing Initiative (NSCI)

The National Strategic Computing Initiative (NSCI) was established by Executive Order on July 30, 2015. Helland (2015) identified the NSCI's following four guiding principles:

1. The United States must deploy and apply new HPC technologies broadly for economic competitiveness and scientific discovery.
2. The United States must foster public-private collaboration, relying on the respective strengths of government, industry, and academia to maximize the benefits of HPC.

3. The United States must adopt a whole-of-government approach that draws upon the strengths of and seeks cooperation among all executive departments and agencies with significant expertise or equities in HPC while also collaborating with industry and academia.
4. The United States must develop a comprehensive technical and scientific approach to transition HPC research on hardware, system software, development tools, and applications efficiently into development and, ultimately, operations.

NSCI objectives echo plans already under way in DOE's current exascale computing initiatives. In fact, DOE is among the NSCI's three lead agencies (along with the U.S. Department of Defense and the National Science Foundation), which recognizes these agencies' historical roles in pushing the frontiers of HPC and in helping to keep the United States at the forefront of this strategically important field (Helland 2015).

1.2 BES Exascale Requirements Review

DOE SC convened an Exascale Requirements Review for the BES program, which took place November 3–5, 2015, in Rockville, Maryland, and brought together leading BES researchers and program managers, scientific and HPC experts from the ASCR facilities and scientific computing research areas, and DOE BES and ASCR staff (see Appendix A for the list of participants). These participants:

- Identified forefront scientific challenges and opportunities in basic energy sciences that could benefit from exascale computing over the next decade.
- Established the specifics of how and why new HPC capability will address issues at various BES frontiers.
- Promoted the exchange of ideas among application scientists in the basic energy sciences, computer scientists, and applied mathematicians to maximize the potential for use of exascale computing to advance discovery in the basic energy sciences (see Appendix B for the meeting agenda).

Outlines and input from white papers and case studies (Appendices C and D, respectively) authored by the participants and submitted to the BES Organizing Committee chairs in advance of the meeting guided the discussions in general sessions and topical breakouts. Committee members and review participants collaborated at the meeting to identify the grand challenges, priority research directions, and computing requirements for their fields of research — communicating these requirements to the DOE SC offices and ASCR facilities. This report therefore reflects extensive and varied forms of input from many voices in the BES community regarding HPC requirements for BES's world-class initiatives.

The review afforded a rare opportunity for the nearly 100 participants to interact and learn about each other's areas of expertise, challenges faced, and the exciting opportunities to be made possible by the exascale computing environment.

Exascale Requirements Reports Will Meet Multiple Needs

DOE managers will use the Exascale Requirements Review reports to guide investments and budgeting, complete their strategic planning, and respond to inquiries, including specifically in their efforts to:

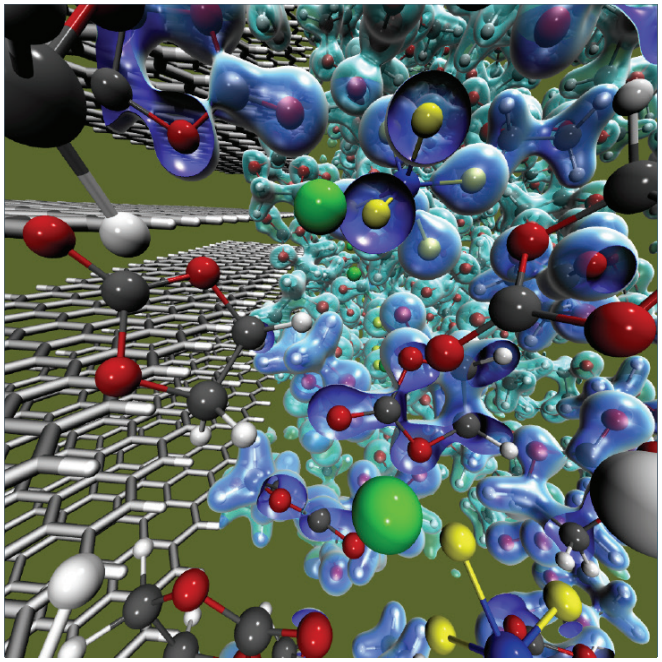
- Articulate the case for future needs to DOE and SC management, the Office of Management and Budget, and the U.S. Congress.
- Identify emerging hardware and software needs for SC, including for research.
- Develop a strategic roadmap for the facilities based on scientific needs.

BES program managers may also use the reports to inform their work. Although balancing such varied end uses can be a challenge, the reports are intended as an information tool that can be used by many stakeholders.

1.3 Report Organization

In the balance of this Exascale Requirements Review, Section 2 provides an overview of the BES vision and grand challenges facing the field of basic energy sciences. Section 3 addresses eight areas of scientific challenge and opportunity, along with the priority and cross-cutting research directions and computing needs and requirements associated with each. Section 4 outlines a path forward for successful collaboration to occur among DOE's ASCR facilities (i.e., the LCF centers and NERSC). References and the acronyms/abbreviations used in the report are listed in Sections 5 and 6, respectively, followed by the appendices mentioned previously.

A VIEW OF THE SCIENCE



Snapshot of QMD Simulation of Anode-Electrolyte Interface (graphite, EC, $LiPF_6$). Figure provided by J. Pask, Lawrence Livermore National Laboratory.

2 BASIC ENERGY SCIENCES: VISION AND GRAND CHALLENGES

2.1 BES Vision

The research disciplines that the BES program supports — condensed matter and materials physics, chemistry, geosciences, and aspects of physical biosciences — are those that lead to discovery of new materials and design of new chemical processes. These disciplines touch virtually every aspect of energy resources, production, conversion, transmission, storage, efficiency, and waste mitigation. BES enables this science by fielding a suite of cutting-edge User Facilities that provide the very best tools to thousands of scientific teams annually. BES research provides a knowledge base to help understand, predict, and ultimately control the natural world and serves as an agent of change in achieving the vision of a secure and sustainable energy future.

The energy systems of the future — whether they tap sunlight, store electricity, or make fuel from splitting water or reducing carbon dioxide — will revolve around materials and chemical changes that convert energy from one form to another. Such materials will need to be more functional than today's energy materials. To control chemical reactions or to convert a solar photon to an electron requires coordination of multiple steps, each carried out by customized materials with designed nanoscale structures. Such advanced materials and chemical processes are not found in nature; they must be designed and fabricated to exacting standards using principles revealed by basic science.

BES has a long history of working with the scientific community to develop the vision and future directions of BES research. For example, in 2005, the subcommittee on theory and computation of the BES Advisory Committee (BESAC) led a community effort that culminated in the report, *Opportunities for Discovery: Theory and Computation in Basic Energy Sciences*, that observed that the time was ripe for investments in theory and computation due to the confluence of scientific success using theory and computation, scientific frontiers requiring innovative theory, development of new experimental capabilities that challenged theory and computation, and increased computational and algorithmic capabilities. In addition, nine emerging scientific themes were identified with opportunities to expand investments in theory and computation. Several of these themes are incorporated in the BES grand challenges. The report also concluded that "[c]onceptual theory and computation are not separate enterprises" and that there must be a balanced approach to theory and computation. The report recommended that "BES should provide support for the development and maintenance of shared scientific software" — a recommendation that is still as valid today as it was in 2005.

2.2 BES Grand Challenges

In 2007, DOE SC charged a panel of ranking experts with determining a set of key areas where science researchers were poised in the coming years to make profound impacts on our ability to control matter and energy. The panel's conclusions were presented in a landmark report titled *Directing Matter and Energy: Five Challenges for Science and the Imagination* (DOE-BESAC 2007a). The report identified five grand challenges that, if realized, would lead to the control of matter and energy at the molecular, atomic, and quantum levels that could spark revolutionary changes in technologies to help us meet some of humanity's most pressing needs, including the need for renewable, clean, and affordable energy. These five grand challenges are as follows:

1. How do we control material processes at the level of electrons?
2. How do we design and perfect atom- and energy-efficient synthesis of revolutionary new forms of matter with tailored properties?

3. How do remarkable properties of matter emerge from complex correlations of the atomic or electronic constituents and how can we control these properties?
4. How can we master energy and information on the nanoscale to create new technologies with capabilities rivaling those of living things?
5. How do we characterize and control matter away — especially very far away — from equilibrium?

Since that report was issued in 2007, a number of key advances have been made in these five areas. These advances include, for example, the proliferation of a variety of low-dimensional, chalcogenide-based materials that have expanded the potential for the control of new transport devices, generalizing and extending "beyond graphene," which dominated much of the materials landscape in 2007. Also heralded are new hybrid organic perovskite-based materials that have offered new routes toward more efficient solar cells and harvesters. The development of new chemical antennae has enabled control of the molecular absorption of light and conversion to energy. The advent of new metal organic frameworks, zeolites, porphyrins, and phthalocyanines has enabled single-site catalysis and control of properties such as gas management, transport, separation, and photosynthesis. The roles of topology, correlation, and coherence have also risen since then as among the key concepts dictating new phases of matter possessing robust transport properties, such as the family of topological insulators displaying new paradigms for materials control.

Today, we have a thousand times more computational power than was available in 2007… we are on the verge of standing up 100-petaflop machines (a petaflop machine completes 1 quadrillion or 10^{15} floating-point calculations per second).

In addition to these science advancements, a number of key landscape changes have occurred that have helped to hone the pursuit of these grand challenges with more finely expressed goals. These key changes have largely come in the form of new theoretical models, extremely rapid progress in our computing ability, and the rapid unveiling of new BES facilities that have opened new avenues of research that were not available at the time of the 2007 *Directing Matter and Energy* report (DOE-BESAC 2007a). Some of these advancements have been highlighted in the 2010 report, *Computational Materials Science and Chemistry: Accelerating Discovery and Innovation through Simulation-Based Engineering and Science* (DOE-SC 2010).

For example, today we have a thousand times more computational power than was available in 2007. At that time, 100-teraflop supercomputers — meaning machines capable of performing 100 trillion floating-point calculations per second — were the gold standard; however, today we are on the verge of standing up 100-petaflop machines (a petaflop machine completes 1 quadrillion or 10^{15} floating-point calculations per second). Furthermore, the Office of Science envisions reaching the exascale level of computing within the next decade (with machines capable of performing a quintillion or 10^{18} floating-point calculations per second).

This enormous growth in theoretical capability and computational power, coupled with major advances in other fields, such as the development of coherent light sources and increased imaging resolution, is accelerating the pace of materials and chemical sciences research. In 2007, the Linac Coherent Light Source (LCLS) and the National Synchrotron Light Source II (NSLS-II) were visions on paper; today, they are operational national user facilities (at Stanford University and Brookhaven National Laboratory, respectively) bringing new, unique capabilities to materials and chemistry research. For example, since commencing user operations in 2009, LCLS has afforded direct observation of matter and energy at the natural time and length scale for a variety of phenomena relating to catalysts and catalytic processes, ultrafast switching of phases such as magnetism and structure, and the manipulation of strongly correlated phases of matter using pulsed fields. As NSLS-II comes on-line, we are witnessing a change in the paradigm of "science of observation" to the "science of control" (Figure 2-1) at unprecedented length and time scales. The upgrade of the Advanced Photon Source (APS) at Argonne National Laboratory and potential

upgrades of other BES facilities in the next 5–10 years will support more detailed and expanded experiments, producing an exponential growth of data and the need to model interactions with computational explanatory and predictive capabilities. To manage and extract useful scientific information from those data will increase the requirement for engagement with ASCR facilities and resources.

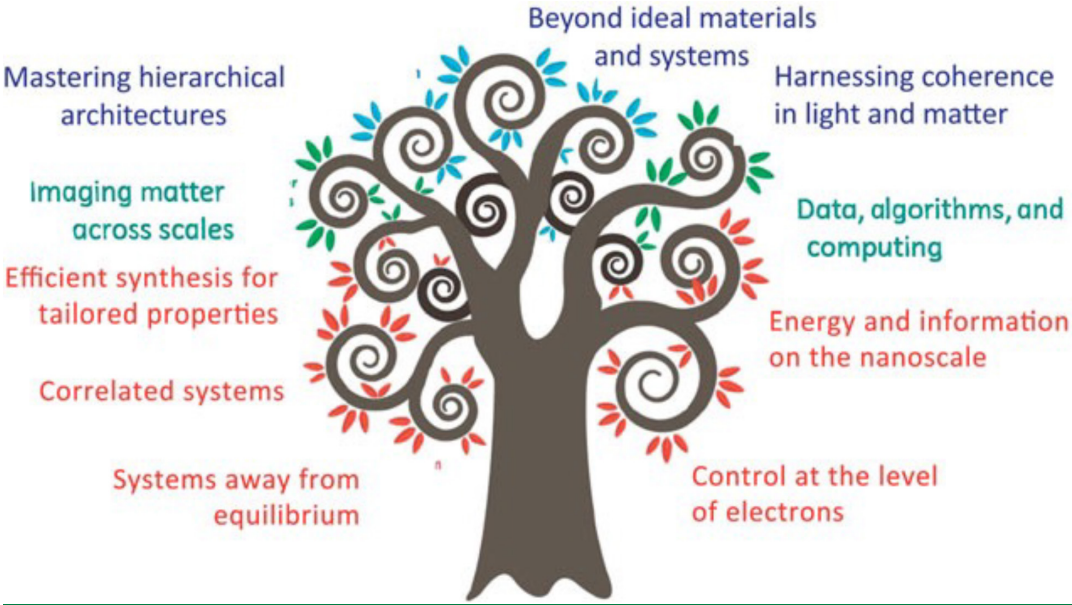


Figure 2-1. Opportunities at the Frontiers of Discovery Science for Matter and Energy (Source: Sarrao and Crabtree 2015).

In view of these advancements, the Office of Science charged BESAC with assessing the progress that has been made on all five grand challenge fronts and identifying what new knowledge opportunities exist to advance energy science. A new report, entitled "Challenges at the Frontiers of Matter and Energy: Transformative Opportunities for Discovery Science" (DOE-BESAC 2015), has recently been issued that expressed five new transformative opportunities in science that have sprung from the original five grand challenges in the environment of new opportunities in computing and light source facility developments. These five new opportunities consist of three new themes:

1. Mastering Hierarchical Architectures and Beyond-Equilibrium Matter
2. Beyond Ideal Materials and Systems: Understanding the Critical Roles of Heterogeneity, Interfaces, and Disorder
3. Harnessing Coherence in Light and Matter

In addition, there are two cross-cutting opportunities:

4. Achieving Revolutionary Advances in Models, Mathematics, Algorithms, Data, and Computing
5. Exploiting Transformative Advances in Imaging Capabilities across Multiple Scales

Common themes among these five new opportunities include the ability to realize targeted functionality in materials by controlling methods of synthesis and assembly, the evolution of mesoscale structures in both the time and space domains that determine heterogeneities in real materials, as well as the ability to exploit the "quantumness" of materials to construct macroscopically coherent states that control the outcome of certain chemical reactions or the instantaneous state of a material.

Implicit in these challenges is the need for a strong connection between advanced theoretical/algorithmic developments, new computing landscapes, and experimental toolsets to realize these transformational opportunities, as evidenced in the two cross-cutting opportunities.

2.3 Mapping the BES Grand Challenges to the Computing Ecosystem and Exascale

The convergence of theoretical, mathematical, computational, and experimental capabilities has put us on the brink of greatly accelerating our ability to predict, prior to synthesis, how to control new materials and chemical processes, to characterize existing processes, and to understand the complexities of matter across a range of scales. A goal of these efforts is to make substantial headway in the "inverse design and synthesis" of matter whereby a class of matter can be selected from a wide variety of uncharted materials and reactive chemistries to have a desired set of properties and products, moving past the science of "serendipity." Similarly, enhanced mathematical and computational capabilities significantly enhance our ability to extract physical and chemical insights from vastly larger data streams gathered during multimodal and multidimensional experiments using advanced characterization facilities.

The BES Exascale Requirements Review focused on eight areas in which advancement in the grand challenges can be achieved via key sustained efforts in computation, simulation, and advanced tools to characterize matter. These eight areas are outlined below.

2.3.1 Novel Quantum Materials and Chemicals

Our nation's future depends on having reliable, affordable, and environmentally acceptable means of producing, transmitting, and storing energy to serve an ever-developing demand. Fortunately, technology is evolving rapidly, with chemical compounds that were once laboratory curiosities now figuring prominently in new technologies ranging from wind turbines and solar energy collectors to power lines and electric cars, and in devices ranging from new bolometric sensors to nascent quantum computers. In addition, our nuclear and industrial wastes often contain heavy elements that must be recycled for further use in the energy cycle or for safe storage; this recycling requires control over the many-electrons interactions in heavy elements. Technologies of the future will rely upon specific combinations of elements, compounds, materials, and phases whose behaviors are emergent: synergistically enhanced and not predictable from studying the components in isolation. Emergent phenomena are observed in a broad array of materials and are derived from novel, and sometime unpredictable, interactions.

Compounds and materials observed "in the wild" display intriguingly enhanced and useful properties, such as the dissipationless surface currents protected by topology in certain insulators with strong spin-orbit coupling, of the highly efficient light harvesting capabilities of certain classes of lead- or tin-halide perovskites. The need is to tame and capture these behaviors: to understand the causes, learn how to optimize properties, define synthesis and growth paradigms for them, and most importantly, to predict new classes of materials and chemicals with even better properties. Quantum chemistry, the control over how atoms combine into molecules and solids, is the key discipline. Theories that span the scales from atomic electronic states and reactions of quantum chemistry to the macroscopic emergent properties of technological relevance are needed. These problems are fundamentally computational in nature to capture a wide range of phenomena at their natural time and length scales. This need will absolutely require the development of new predictive algorithms.

2.3.2 Catalysis, Photosynthesis and Light Harvesting, and Combustion

Efforts to understand and control chemical transformations lie within the core of the BES mission and are essential for advances in the energy economy. Theoretical methods and computational tools provide a framework for interpreting experimental measurements, for guiding the discovery

of catalysts with unprecedented activity and selectivity, and for designing energy-efficient and environmentally friendly new engines for transportation and power generation. Current computational limitations, however, severely restrict the size and complexity of systems that can be studied with sufficient fidelity. Emerging computing ecosystems, along with advances in theoretical methods and algorithms, will enable the study of realistic heterogeneous environments, at the long time and length scales relevant for applications. The ability to execute computations for large ensembles of systems and conditions will significantly enhance the reliability of the predictions. This capability will also allow us to model catalytic, photosynthetic, and combustion phenomena prior to their experimental realization, thus guiding future experiments.

2.3.3 Materials and Chemical Discovery

The foremost scientific challenges to realizing the vision of materials and chemicals by design are threefold: (1) computational discovery of novel materials and chemicals with target properties (including hierarchical structures with multiple functionalities), (2) prediction of pathways to the synthesis of these materials and chemicals (with consideration of sustainability and green chemistry principles), and (3) prediction of their stability and degradation pathways. Surmounting these challenges will involve a synergistic interplay of advances in predictive modeling capabilities, hardware resources, and experimental techniques — allowing for the characterization of spatial and temporal fluctuations and of short-lived intermediates along synthesis and degradation pathways.

With existing resources, over the next 5 years, we will see the prediction of size and shape distributions of semiconductor nanocrystals for thermoelectrics and of nucleation pathways and subsequent growth for specific systems involving environments of limited complexity, for example, the modification of nodes and ligands in metal-organic frameworks to tune capture and/or catalytic properties. An exascale computing environment will enable the shift to an adaptive and self-consistent multiscale-modeling paradigm.

2.3.4 Soft Matter

Soft matter provides unique and critical materials behavior in a wide range of industrial products. Polymers, surfactants, electrolytes, and microheterogeneous fluids have long been key components in a multitude of applications, including energy storage (e.g., batteries and capacitors) and energy production (e.g., photosystems), chemical separations, enhanced oil recovery, food packaging, chip manufacturing, and health care products. Soft materials composed of molecular and/or modular building blocks can provide the hierarchical complexity and tunability for making paradigm-shifting materials that can accomplish multiple tasks. The complexity of soft materials — liquids, gels, or amorphous solids (as opposed to typically crystalline hard matter materials) — presents scientific as well as computational challenges that will make exascale computing a pivotal resource in achieving the goal of designing functional matter.

Biological systems are multifunctional and highly responsive due to complex, hierarchical structures and the associated dynamics inherent in these structures. Computational researchers are just beginning to be able to simulate simpler versions of hierarchical structures, whether biological in origin or synthetic versions inspired by biology. Hierarchical soft matter is characterized by inherent structures covering different length scales that evolve over a broad range of timescales in response to external constraints and stimuli. This responsiveness arises from the ability of these structures to transfer energy between different forms; undergo controlled energetic and structural transformation; and store, alter, and transmit information. Information transmission, in particular, typically involves non-equilibrium and chaotic processes for which fundamental theoretical principles are still lacking. For the first time in the history of science and engineering, many interesting phenomena in soft matter are expected to become accessible with exascale resources, enabling the modeling of many fundamental phenomena that will have a significant broader impact on critical national needs.

2.3.5 Advances in Algorithms for Quantum Systems

Quantum mechanics (QM) touches virtually all of computational chemistry, biology, physics, and materials, either directly or indirectly. In addition, most modern classical force fields are derived entirely or partially from quantum mechanics. The bottleneck in applying quantum methods is the high scaling of these methods with system size, especially methods that account for electron correlation. To facilitate the use of these high-level QM methods, it is essential to reduce the scaling of these methods with system size *and* to develop multilevel parallel algorithms for them.

Implicit in many of the BES mission phenomena is the need to develop truly multiscale methods that can span multiple time and length scales in a seamless and self-consistent manner. Because of their high cost, direct application to complex materials and chemical problems will remain out of reach in the foreseeable future. True predictive power for these complex problems will require development of robust hierarchical theories and algorithms to treat electron correlations across all relevant length scales. Furthermore, developing these new algorithms will require the collaboration of application developers with applied mathematicians and computer scientists and engineers.

2.3.6 Computing and Data Challenges at BES Facilities

The BES user facilities operate more than 240 different instrument types that enable scientific discoveries by their user communities. The heterogeneity of instrumentation contributes to the richness of the capabilities of the BES user facilities and to the breadth of the user base. It also results in the need for many different data acquisition and analysis methodologies. In the past, and to some extent still today, detectors and end stations generated datasets that were relatively small and readily transferred to users' portable storage devices and to be taken to their home institutions for analysis. The availability of high-performance networking facilitated this transfer to a user's home institution. Once the user took possession of the data, the operational responsibility of the BES user facilities ended.

This paradigm has shifted with the introduction of rapid and high-performance detectors at some beamlines. These detectors routinely produce upwards of 10^2 to 10^3 frames/second; some will be capable of performing at 10^4 to 10^5 frames per second within the next 5 years — and detector performance will only accelerate. As a result, users can no longer readily transfer their large datasets to their home institutions. Furthermore, the heterogeneous nature of instruments at BES user facilities means that no one solution will be sufficient to address all of the data challenges. The growing complexity of the analysis process (mixing fast data analysis and numerical modeling) at other instruments will require capabilities beyond the petascale-level capabilities that the ASCR facilities offer. Addressing these challenges by providing for the analysis, management, and storage of user data signals a fundamental change in the operation and responsibility of BES user facilities.

In addition, the computational capabilities of the future will provide a platform for real-time modeling and simulation so that experiments can be augmented and understood as they are in progress. This coupling, along with the need to manage data and the ability to steer and make decisions during the experiment to optimize the scientific outcomes, will require significant "on-demand" exascale-types of computational resources to deliver the necessary feedback and insights in real time as the experimental process unfolds.

2.3.7 Mathematics and Computer Science Transforming BES Science

Bridging the gap between BES scientific goals and ASCR computing capabilities will fundamentally depend on the collective abilities of our science domains and facilities to deliver transformative breakthroughs in mathematics and computer science — an objective that will require investing in and capitalizing on state-of-the-art evolving mathematics and computer science. Indeed, this linkage cannot be overemphasized: the mathematics component provides the language

and blueprints for transforming models into equations, approximations, and algorithms that in turn set the stage for taking advantage of ASCR's computing portfolio; the computer science joins this enterprise by providing the theory, tools, and methods to efficiently execute these blueprints on the most advanced computing architectures.

2.3.8 Next-Generation Workforce

A common theme across the different BES topical areas is that the need for workforce development poses a serious bottleneck to reaching the exascale level of computing — and thus is an essential part of the equation of success in reaching exascale. The increasing complexity and dependencies of the exascale era will require enhanced collaborations, including cross-cutting, multidisciplinary teams of domain scientists, applied mathematicians, and computer scientists, who together can cope with the broad range of computing challenges. Providing the education and training we need to prepare these future computational and domain scientists and computational software developers for exascale computing presents a major challenge.

In fact, our community's ability to "field" a sufficiently skilled workforce is considered the greatest risk factor in realizing exascale computational science. Consequently, we anticipate that significant investments must be made in training a new generation of computational scientists — individuals who are well grounded in their science and engineering disciplines, but also knowledgeable about relevant computer science and applied mathematics issues — to realize the promise of exascale computing.

2.4 Report Roadmap

The preceding discussions represent the proverbial "view from 30,000 feet" of the state of BES science and the grand challenges we are tackling as a community. From such a vantage point, it is exciting to contemplate the convergences mentioned earlier — of theoretical, mathematical, computational, and experimental capabilities — and the discovery science these developments may lead to as we pursue new materials and chemical processes and seek to understand matter's complexities across a range of scales.

The next stage is to bring these "bird's eye" views to ground level and concretize them through greater context and detail. Thus, in Section 3, we expand on the seven areas, highlighting particular challenges that BES domain scientists are trying to resolve; the computational roadblocks that are inhibiting further breakthroughs in validation and prediction; and specific resource, theoretical/intellectual, and people needs that can help us bypass these roadblocks.

Section 4 concentrates specifically on requirements, recommendations, and requests that appear in Section 3. These are grouped into four categories of opportunities for collaboration among domain scientists, academics, computer scientists, and the ASCR facilities:

- Methods development
- Computational environment
- Data
- Communication/community involvement

These items constitute the requests or "ask" that BES is making at this time so DOE can closely and actively support a path forward for BES science initiatives and thereby promote an evolving computing ecosystem leading to exascale.

.

3 RESEARCH DIRECTIONS AND COMPUTING NEEDS/REQUIREMENTS

3.1 Quantum Materials and Chemistry

3.1.1 Scientific Challenge and Opportunities

Our nation's future depends on having reliable, affordable, and environmentally acceptable means of producing, transmitting, and storing energy to serve an ever-developing demand. Fortunately, technology is evolving rapidly, with materials and chemical compounds that were once laboratory curiosities now figuring prominently in new technologies ranging from wind turbines and solar energy collectors to power lines and electric cars and devices ranging from bolometric sensors to quantum computers. In addition, our nuclear and industrial wastes often contain heavy elements that must be recycled for further use in the energy cycle or for safe storage. The key issue is that technologies of the future will rely upon specific combinations of elements, molecular motifs, materials, and phases whose behaviors are emergent: synergistically enhanced and not predictable from studying the components in isolation. These behaviors are controlled by the quantum mechanical behavior of electrons so that materials exhibiting the new and potentially important properties are often termed "quantum materials." Critical to the design of quantum materials is obtaining exquisite control over local chemistry and the ability to link this capability to the emergent behavior of technological importance. This requires the development of new predictive theories, efficient and adaptive software, and exploitation of the full capabilities of new and future computing architectures at BES/ASCR facilities, coupled synergistically with advanced experimental facilities at BES light and neutron sources and nano facilities.

Emergent phenomena are observed in a broad array of materials and are derived from novel, and sometimes unexpected, interactions. Superconductivity, the ability of a material to carry electrical current without electrical resistance at low temperatures, is a key phenomenon. Because superconducting wires do not heat up, they provide a potential pathway toward powering the energy-dense cities of the future — if the temperature at which superconductivity occurs and the total amount of current that can be carried are increased beyond present limits. A related class of materials is topological insulators, which carry no current in bulk but carry dissipation-free surface currents protected by topology and can serve as the basis of a new, low-power electronics. Because these are essentially emergent quantum mechanical phenomenon, these systems provide platforms for quantum and other forms of low-power computing as well as new kinds of quantum sensors. Other ("Mott") materials can be switched from metallic to insulating states by small changes of voltage or temperature; some Mott materials have already found application as bolometers and many other prospective applications to beyond Moore's law electronics. A third class of quantum materials has remarkably enhanced thermoelectric properties, and with modest enhancement can make the capture and use of waste heat an economically viable prospect. The situation is similar for magnetism, wherein magnetic materials play crucial roles in a diverse range of

White papers by the following authors informed the writing of this section and can be found in Appendix C, Section C.1, starting on page C-3:

- D. Ceperley and P. Kent
- A. Clark
- J. Deslippe, C. Lena, and J. Chelikowsky
- J. Freericks
- E. Gull
- J.J. Rehr et al.
- P. Sushko, T. Dunning, K, Kowalski, and N. Govind
- P. Yang

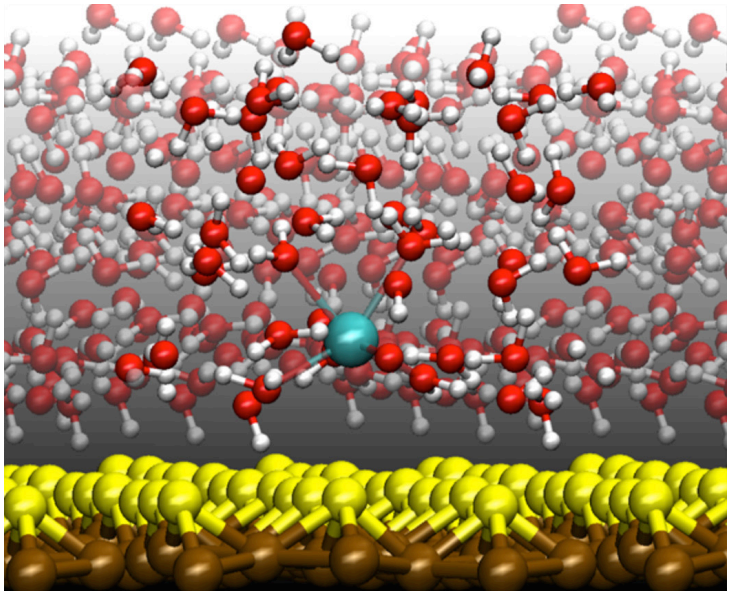
Similarly, these relevant case studies can be found in Appendix D, Section D.1, starting on page D-3 by:

- H.-P. Cheng
- A.E. Clark, C. Isborn, and T. Markland
- J. Freericks, H. Krishnamurthy, and T. Devereaux
- M. Gordon
- T. Maier and M. Summers
- A.J. Millis

applications from computer memory to antilock brakes. Unfortunately, known magnetic systems involve critical elements whose supplies lie outside of U.S. territories; identification of magnetic materials derived from earth-abundant (or at least U.S.-available) elements is a key need. Many more examples can be discussed, including light harvesting, catalysis, new multiferroic sensors and actuators, and components of electro-optical devices. But the key point is that discovery of new families of quantum materials and optimizations of properties within known material families are urgent national needs. Methods now under development in the closely linked fields of theoretical chemistry, materials science, and condensed matter physics, in combination with new data analytics paradigms, may revolutionize the discovery of new quantum materials. These new methods require exascale-level computation.

Another key priority in BES is understanding the chemical reactions and material properties associated with the nuclear energy industry. The complexities of predictive chemical and materials theory are exacerbated by the heavy elements that are essential to advanced nuclear energy systems. Progress in understanding heavy-element chemistry and materials science (including relativistic effects, high correlation, and extreme chemical complexity) will allow us (for example) to design and engineer extractants for the separation of heavy elements and unwanted species from the waste mixes. In addition, these same separation issues are important in the separation of the lanthanide materials that many of the magnetic materials depend on. However, many of these separation processes depend on nonequilibrium properties with emergent behavior that only manifests itself when the full chemical and materials environment is included. In addition to separations, new insights into materials for storing nuclear waste and their degradation processes are needed. Understanding complex materials and chemical properties such as phase diagrams is one important component toward reaching this goal.

A VIEW OF THE SCIENCE



Uranyl in Water on Mackinawite Surface. Figure courtesy of Wibe DeJong, Lawrence Berkeley National Laboratory.

The structural characterization and measurement of the physical properties of these proposed advanced materials must evolve at a rate commensurate with materials development. We also need critical understanding of the behavior of materials under *in operando* conditions in which their functionality is to be exploited and (it is hoped) controlled. The powerful probes of material properties provided by next-generation spectroscopies now coming on-line at modern synchrotron radiation facilities will provide essential experimental validation of the computational methods being developed, while the ability of the new methods to enable accurate and rapid modeling experiments over a range of their fundamental length and energy scales will provide crucial feedback to the development and execution of spectroscopic experiments at BES facilities.

Theory to meet these challenges is coming on-line. The fields of computational chemistry and materials physics are undergoing rapid development, with the standard workhorses of density functional and coupled cluster theories being superseded by more modern methods that will vastly expand our capabilities. Deploying these methods requires a vastly expanded computational ecosystem, with exascale facilities playing a central role.

3.1.2 Priority Research Directions

Within the next 10 years, we anticipate the development of quantitative and predictive theories of phase diagrams, ground and excited state physical behavior of materials dominated by strong electron-electron interactions, and multiscale correlations. This is, of course, a long-standing goal of chemistry, materials theory, and condensed matter physics; however, conceptual developments over the last decade suggest that breakthroughs are imminent. Translating the conceptual progress into working algorithms requires vast increases in computational power with concomitant development of the computational ecosystem: libraries, programmer and user training, algorithm development, and implementation. The goals are ambitious but the rewards in terms of new chemicals and materials with emergent properties to meet new national needs are promising. Here we outline a few striking examples of what might be done and what would be required, but we emphasize that the list is illustrative, not exhaustive.

3.1.2.1 Emergent Properties of Quantum Materials: Superconductivity, Magnetism, Topological Protection, and Beyond

Optimizing the properties of known quantum materials and discovering new systems with new properties are key research directions. For example, the copper oxide and iron arsenide families of high-transition-temperature superconductors retain their ability to carry current without dissipation to much higher temperatures than conventional superconductors, which are already in use for particle accelerators and medical imaging machines. These compounds were discovered serendipitously, and improvement and understanding have been based on empirical optimization of existing materials and qualitative analysis of phenomenological theories. Recent developments such as large-scale quantum Monte Carlo (QMC) simulations on the Oak Ridge Leadership Computing Facility's (OLCF's) Titan petascale supercomputer (Starr et al. 2014) suggest that a comprehensive and validated solution of a variety of model subsystems having intertwined degrees of freedom (spin, charge, lattice, and orbital) and emergent behavior can be realized. This effort will require exascale computing to obtain and cross-compare results from many other numerical methods, including multiconfigurational (or multideterminantal) QMC and complete active space self-consistent field (CASSCF) methods, (cluster) dynamical mean field, tensor network, and diagrammatic perturbation methods, as well as their multiscale extensions.

Very similar issues arise for magnetism. Although existing methods can predict geometry-spin relations for many known magnets, they sometimes incorrectly predict physical structures, are not always accurate for exchange constants, have difficulty with transition temperatures and excitation phenomena, provide qualitatively incorrect values for magnetic moments associated with "open

shell" magnetic centers, and are not adequate to the materials design problem of designing new magnets that avoid rare earth elements. Nanomagnets are candidate materials for applications in spintronics and new memory architectures. These materials typically have transition metal or rare earth cores surrounded by organic ligands, with the interplay of cores and ligand electronic interactions controlling the properties. Important challenges include design of ligands that tune the high-spin/low-spin transition to a desired temperature range and development of low-energy information technology systems requiring an understanding of anisotropies and resonant tunneling of magnetism. The aforementioned advancements (in combination with relativistic electronic structure methods and spin-orbit extensions of the many-body methods) will solve these problems on the 10-year timescale.

These are but two examples chosen from a very broad range of national needs. Very similar issues arise in the design of thermoelectrics to optimally use heat generated by other applications or Mott metal-insulator transition materials that offer new potentialities for electronics and for bolometric sensors. Light harvesting and catalytic entities require, in addition to the many-body electronic structure sketched above, new methods to describe the nonequilibrium dynamics of excited electrons, and their coupling to the lattice. Extending the remarkable properties of topological insulators to the correlated domain (with the promise of larger amplitude and more robust effects) will similarly be amenable to systematic theoretical study in the exascale era.

3.1.2.2 Heavy-Element Science

Advanced nuclear energy systems represent an essential part of the U.S. energy portfolio, and perhaps more importantly, global adoption of this technology necessitates that the United States play a strategic role in materials development for nuclear safeguards and in chemical strategies for separations protocols that are essential to environmental remediation and nuclear nonproliferation. BES plays the significant role of providing the basic chemical and materials properties and insights to determine solution, interfacial, and solid-state bonding and reactivity, as well as the fundamental science underpinning the extraction and separation of actinides (DOE SC 2014). Computational approaches for modeling involve a broad range of actinide-bearing materials in extremely challenging radiation environments and physical and chemical phases. Because of the emergent behaviors inherent in these environments, all parts of the system must be modeled together to provide accurate insights and predictions for heavy element materials and chemistry. Two critical science foci in this arena are efficient, environmentally friendly heavy-element separations and the prediction of properties for heavy-element chemicals and materials.

Advanced nuclear energy systems represent an essential part of the U.S. energy portfolio, and perhaps more importantly, global adoption of this technology necessitates that the United States play a strategic role in materials development for nuclear safeguards and in chemical strategies for separations protocols that are essential to environmental remediation and nuclear nonproliferation.

Many structural phases can influence critical physical properties, such as the pressure-volume curve and thermal conductivity. Performing the efficient, effective, and environmentally friendly separation of actinides, lanthanides, and unwanted species in complex environments is extremely challenging. Current state of the art in the computational design of extractants relies on the use of *ab initio* and molecular mechanics (MM) methods to examine the geometric strain of the extractant upon binding to the ore and the binding energy in the gas phase to help to predict which extractants will perform best. Occasionally, short time *ab initio* molecular dynamics (AIMD) calculations that include pseudopotentials are performed to gain a basic understanding of the local dynamics of one part of the process. However, it is clear that making accurate predictions requires knowing the solution properties (such as pH and anion concentration), dynamics across the solution-solid interface, and full relativistic effects. Simulations must feature all aspects of the separations process — for example, solute, solvent, counter ions, reactions, interfaces between different solvents, and uptake of the metal from the aqueous to the organic phase. This challenge is currently outside the reach of the state-of-the-art methods and

computational resources and will require exascale software and hardware, as well as exascale-aware scientific programmers.

The ability to predict — both reliably and readily — the properties of both chemicals and materials containing heavy elements and/or radionuclides will pave the way for many advances in the field of heavy-element science. These developments will, in turn, require theories that incorporate scalar relativistic and spin-orbit components and the correlated electron behavior arising from partially filled *d* and ***f*** shells. The integration of new *ab initio* electronic structure results with available thermodynamic databases will enable the prediction of phase equilibria and oxidation states in the next 10 years. As part of this advance, an important and achievable goal in the solid state is to understand the plutonium (Pu) solid-state phase diagram and its extension to compounds and solid solutions involving multiple actinides, and a fundamental understanding of thermodynamic quantities in uranium dioxide (UO_2) and plutonium dioxide (PuO_2), with extensions to multicomponent fuels. Complementing these efforts related to solid state would be the development of new theories that describe the physical properties and predict the reactive behavior of highly complex liquid solutions that result when spent fuel is removed from a reactor and must be separated into its appropriate constituents for recycling and nonproliferation storage. This latter effort will require the effective development of multiscale models that incorporate the relativistic effects as well as the complex environment and dynamics involved in these complicated mixtures to capture the emergent behavior of the systems.

3.1.2.3 Advanced Spectroscopies

Validation of concepts and techniques by comparison to experiment is essential. The new spectroscopies — including high-resolution resonant inelastic X-ray scattering (RIXS), X-ray Raman, and pump-probe techniques for nonequilibrium studies — are under development at national synchrotron radiation facilities (e.g., NSLS-II, LCLS-II, APS, Stanford Synchrotron Radiation Lightsource [SSRL]), along with the striking improvements in neutron scattering (at the Spallation Neutron Source [SNS] at Oak Ridge National Laboratory) and X-ray photoemission spectroscopy (XPS), are immensely powerful probes of chemical and material properties, providing detailed views with unprecedented resolution of atoms and complex materials in motion at meV energy resolution, nanoscale distances, and fs timescales. An example is the promise of X-ray free electron laser (XFEL) machines, such as the LCLS and LCLS-II, which enable revolutionary new possibilities for probing matter across multiple length and time scales, and will be applicable to the study of materials in action — such as for generating molecular movies of charge transfer, stimulating emergent collective excitations, and examining steps in the evolution of catalysts in chemical reactions to name a few. These experiments can be analyzed over about 10–20 ps using approximate models limited to about 100 atoms. Methods such as time-dependent density functional theory (TDDFT), many-body Green's function (dynamical mean field theory [DMFT] and beyond), and many-body perturbation theory as currently implemented in the so-called GW method and BSE (Bethe-Salpeter equation) can vastly improve the understanding of experiments — but exascale resources will be needed to exploit their potential.

With advances in instrumentation and new experimental techniques, parallel theoretical developments are critical for providing a rational basis for experimentation so that resources can be directed to the most promising regions of an exponentially increasing experimental phase space. Without such a synergy between theory and experiment, it will not be possible to exploit the full potential of powerful, high-throughput, high-resolution modern instrumentation, or to develop new, efficient paradigms for unlocking the workings of novel materials and phenomena.

3.1.3 Cross-Cutting Research Directions

All of the scientific issues of concern to BES require a solution of the quantum mechanical equations describing the behavior of many electrons in atoms, molecules, and solids. The historically separate domains of quantum chemistry and theoretical condensed matter physics are converging in topics of theoretical approach and interest, as condensed matter physics approaches the nanoscale and chemists study ever-larger molecular complexes). The key obstacle to solving the quantum many-body problem is the exponential growth of computational expense and complexity with number of electrons. The difficulties are multiplied by the need to handle materials and chemicals that may be intrinsically complex (many and/or heavy atoms in the unit cell), exhibit complex morphologies (defects, interfaces, nanostructures), or are in dynamic or fluctuating (e.g., liquid) environments. The wide range of length and time scales that must be treated means that multiscale methods are essential. While multiscale methods can in principle approach the spatial and temporal regimes needed to solve the nation's energy problems, possibly the only guarantee to success is to attain the quantum-mechanical accuracy that is currently only available with the highest-quality, most expensive methods. However, to accomplish this level of accuracy would require execution that is millions of times faster than is currently possible. In concert with the road to exascale, there must be significant algorithmic efforts and investments to determine whether this very ambitious goal can be achieved.

The standard methods of solving these problems — for example, density functional theory (DFT), coupled cluster theory at the coupled-cluster single double with approximate triple or CCSD(T) level — are not adequate for the issues the field now faces. The cross-cutting research that is urgently needed is the development, implementation, validation, and application of methods that extend and go beyond DFT/CCSD(T) methods.

Validation both in theoretical terms and by comparison to experiment is essential, thus part of the cross-cutting research is the development and use of methods for modeling the new spectroscopies. The work requires a close interplay between theory and computation. Intensive HPC activity is required as part of the development and validation of the theoretical ideas, and theoretical developments will influence the algorithms and the development of libraries. The diversity of problems and of approaches means that a heterogeneous HPC environment is required.

The cross-cutting research required to achieve these goals includes the following:

- Beyond DFT electronic structure methodologies. Some of these (e.g., DFT+DMFT and multiscale corrections, all-electron QMC, density-matrix renormalization group [DMRG]) have been implemented, and the needed work consists of improving, applying, and validating the methods. Some (e.g., GW+DMFT, Green's function methods for quantum chemistry) are now in the process of being implemented and will require substantial development and testing. New emerging theories will also be required. A key area where more research is needed is in obtaining force fields from the many-body approaches to enable realistic molecular dynamics (MD).
- Wave function techniques (such as tensor network/(infinite) projected pair-entangled states [iPEPs]/matrix product state [MPS], full configuration-interaction [FCI]-QMC, all-electron QMC for correlated situations, multireference methods), which are now being developed on model systems; further development is required to bring them to the level of modeling realistic materials and chemicals.
- Scalable extensions (such as DFT-tight binding [DFTTB]) of DFT to the fully relativistic and the spatially structured situation.

- Methods for electronic dynamics, such as in TDDFT, density matrix method, nonequilibrium DMFT, GW, and Bethe-Salpeter for response functions. In all of these cases, the methods have been introduced and the issues are scaling to obtain the needed resolution and applicability to complex systems. However, new theoretical developments are needed in this area.
- Methods for quantum molecular dynamics (QMD, such as the Born-Oppenheimer molecular dynamics [BOMD], Car-Parrinello molecular dynamics [CPMD], and optimization-free QMD methods), which are finding growing applications in chemically complex systems to capture conformational and electronic states in a dynamic manner and the effects of these changes on chemical reactivity and reaction rates. Reliance on these tools signals the need for computational methods with electronic degrees of freedom (QM methods) reaching long time simulations and large size systems. A key focus for reaching these large scales needs to be on the scalability of the algorithms to take advantage of massively parallelized supercomputers.
- Theories that incorporate scalar relativistic and spin-orbit components and the correlated electron behavior arising from partially filled *d* and *f* shells.
- New methods required to explain solution and interfacial-dominated systems that require coupling of strong and weak, anharmonic interactions (such as van der Waals interactions and hydrogen or ligand-metals bonds) as collective phenomena.
- New phase sampling methods to simulate more realistic systems and/or rare events in reasonable time frames.
- Rigorous treatment of dissipation with quantum and classical methods.

Over the next decade, we will see the complete characterization of the energetics, phase diagram, and dynamics of the standard model systems of quantum chemistry and condensed matter physics, including the Hubbard and Anderson models of correlated materials and the complete sequence of transition metal dimers. Reasonable methods for approximating the solution to more realistic and complicated situations will be developed, implemented, and validated; and in particular, we will have solutions of quasi-realistic model systems on a level that will permit researchers to perform detailed analysis of the new generation of spectroscopies emerging at the major DOE facilities. Substantial progress will also be made over the next decade on the fully *ab initio* computation (with controlled and quantified errors) of properties of moderately complex solids (e.g., La_2CuO_4, the "parent compound" of copper-oxide high-Tc superconductors), as well as the development of multiscale and approximate methods that will connect atomic-scale *ab initio* methods to the properties of a wide range of energy and environmental materials and chemicals.

There are also cross-cutting research needs in the areas of algorithms and applied math. These include:

- Parallel-in-time approaches: new methods to accelerate the sampling of rare configurations to generate ensembles with enough statistics for long-timescale sampling of rare events.
- Adaptive algorithms: updating "on the fly," for example, in doing rapid data analysis to guide experiments in real time, and in generalizing the adaptive grid and multiscale ideas that are powerful in classical partial differential equations to the sophisticated quantum calculations needed going forward.
- Improved tools for data analysis, visualization, and machine learning from multiple data sources (such as simulation and thermochemical databases).
- Tools for real-time analysis of large-scale simulations to obtain, for example, configurations and coordination numbers for metal centers, as well as pattern recognition in real time to control dynamics in order to steer systems to desired free energy minima.
- Development of communication-free computational algorithms for low-energy computing.

3.1.4 Computing Needs and Requirements

Success will depend critically on the ability of BES scientists to fully leverage the more heterogeneous architectures — both in types of hardware and in resource size and distribution — to come. The scientific topics of concern to BES have a wide array of computational requirements. Computational demands may be "bursty," requiring heavy access for durations of relatively short (few months') time periods, interspersed with periods of low access while new approaches are developed. Nearly instantaneous access with very short turnaround times is required for software development. Proposal processes as well as queuing and access procedures at HPC facilities should incorporate mechanisms to meet these needs in addition to the more common large-scale and often batch-mode provision of resources. Different computational campaigns will require different combinations of the available hardware resources with varying demands on floating point operations, memory size and bandwidth, and interconnect bandwidth. Some problems may require very long runtimes on a relatively small number of processing elements and must be a key component in the long-term strategy for the computing ecosystem.

Software development practices will need to rely on general purpose and domain–specific, accelerator-enabled scalable libraries that hide as many of the hardware specifics as possible. This will also be the best strategy to achieve architectural portability, and perhaps performance portability of scientific applications. Development and use of portable domain-specific libraries should be encouraged, as well as efficient and scalable general purpose middle and lower-level software tools, including message passing interface (MPI) libraries; directive-based threading models such as OpenMP and OpenACC; multiple accelerator-enabled mathematical libraries; and efficient, high-level I/O libraries. Distributed equidistant and non-equidistant fast-Fourier transforms are used in multiple methods. A large variety of quantum many-body methods rely, at least partially, on linear algebra. Distributed diagonalization of large dense matrices can become a bottleneck. Better support for scalable block-sparse linear algebra is of demand, as well. Many post-SCF (self-consistent field) quantum many-body methods, including coupled-cluster theory, configuration interaction, DMRG, and tensor network state methods, rely heavily on distributed tensor algebra.

Addressing these issues is part of a larger objective of pursuing portable performance, both across different machines of the same era and as architectures evolve in time. The efforts required in software development are rapidly increasing, and it is essential to ensure that as much software as possible is still useful 5–20 years into the future. Portable performance will also help bring exascale computing to applications and research codes that have shorter lives and research communities, but which are nevertheless at the forefront in the exploration of new ideas.

Certain adaptive algorithms would benefit from the availability of scalable, asynchronous dynamic runtime systems that may be task based. The implementation of distributed linear algebra is sensitive to the interconnect bandwidth, which currently grows more slowly than the Flop/s. An increased interconnect bandwidth would be highly beneficial. The newer programming models are still in their infancy, and standard interfaces are lacking. Another issue is the performance of the MPI library, which is a common communication layer for many scientific codes. In particular, the efficiency of one-sided communications and nonblocking collectives, including neighborhood collectives, still needs improvement.

Current efforts include the development of reduced-scaling algorithms for many-body methods that have controlled accuracy, based on sparse methods, low-rank representations, fast algorithms, changing representations, nonlinear approximation schemes, and matrix reconstruction, etc. These will need to be cast in high-performance algorithms that must be specified in a manner that is abstracted from architecture-specific implementations.

With the ability to perform computational screening of large collections of chemical systems or representations of complex systems, uncertainty quantification is required to transition these tools from basic research to engineering applications. Currently, these are not systematically employed, except in a few fields such as combustion.

BES Path to Exascale in Quantum Materials

- Today: Obtain solution of ground states and phase diagrams of theoretically relevant model systems.
- 2020: Develop and validate quantum Monte Carlo and tensor network methods of treating the full complexity of real materials and solutions.
- 2025: Enable quantitative verifiable prediction of materials properties, including superconducting and magnetic transition temperatures, critical currents, and topologically protected dissipationless edge currents.

BES Path to Exascale in Heavy-Elements Science

- Today: Perform fully relativistic calculations coupled to highly correlated wavefunction approaches for small molecules; also use short time AIMD with pseudopotentials.
- 2020: Incorporate use of highly scalable extensions of relativistic methods, TDDFT, DMFT, and density matrix methods; obtain force fields for large realistic simulations of complex solutions and solids.
- 2025: Develop QM methods that can reach long time simulations and large-scale systems.
- Perform long-timescale simulations that effectively capture the complex pH environments of extraction chemistry.

BES Path to Exascale in Nonequilibrium and/or Advanced Spectroscopies

- Today: Analyze experiments limited to approximately 100 atoms over 10–20 ps.
- 2020: Probe matter across length and time scales using XFEL machines (LCLS, LCLS-II).
- 2025: Exploit DFT, many-body Green's function (DMFT), GW-BSE, and other methods to realize vast improvements in understanding experiments.

3.2 Catalysis, Photosynthesis and Light Harvesting, and Combustion

3.2.1 Scientific Challenges and Opportunities

3.2.1.1 Catalysis

Catalysis — the essential technology for accelerating and directing chemical transformation — is the key to realizing environmentally friendly, economical processes for the conversion of fossil energy feedstocks and the production of fertilizers and synthetic polymers. Catalysis also is the key to developing new technologies for converting alternative feedstocks, such as biomass, carbon dioxide, and water, into fuels. To meet the demands for fuels, a deep understanding of the chemistry of complex fossil-energy feedstocks will be required together with the understanding of how to design catalysts for processing these feedstocks. To realize the full potential of catalysis for energy applications, scientists must develop a profound understanding of catalytic transformations so that they can design and build effective catalysts with atom-by-atom precision and convert reactants to products with molecular precision. Moreover, they must build and exploit novel tools to make real-time, spatially resolved measurements of operating catalysts. Ultimately, scientists must use these tools to achieve a fundamental understanding of catalytic processes occurring in multiscale, multiphase environments, summarized as follows:

> **...Major challenges in heterogeneous catalysis are to more clearly define the nature of the active sites, to engineer at the molecular level catalysts with designed properties in three dimensions, and to create new catalysts for new transformations.....Another major catalytic challenge is to create unique surfaces ('supports') and catalytic environments with three-dimensional aspects. Creating hard-matter surroundings of active catalytic centers that enhance reactivity or selectivity offers great opportunities (DOE-BESAC 2007b).**

In order to make progress on addressing these challenges, the long-term goal is to obtain an understanding of the fundamental principles of catalytic processes enabling the design of catalysts with unprecedented activity (i.e., with high reaction rates), selectivity (i.e., producing desired products), and stability (i.e., promoting many transformations without deactivating). HPC will play a central role in providing the insight needed to design catalysts and complex transformations that involve processes at multiple length and time scales.

Although significant advances have been made in reliable modeling of the behavior of catalysts and catalytic processes under idealized conditions, we currently do not know how to design coupled physical-chemical-materials systems that comprise real-world catalysts and achieve and maintain desired properties under the operating conditions in which they are expected to function. Achieving the ability to design such systems requires development of a predictive understanding of interfacial processes exhibiting complex, collective behavior in which catalyst transformations

White papers by the following authors informed the writing of this section and can be found in Appendix C, Section C.2, starting on page C-25:

- T. Baruah
- J. Chen
- W.A. de Jong
- D.A. Dixon
- L. Gagliardi
- B. Garrett and R. Rousseau
- M. Gordon
- S.J. Klippenstein
- X. Li
- A. Selloni
- L.V. Slipchenko
- E.F. Valeev

Similarly, these relevant case studies can be found in Appendix D, Section D.2, starting on page D-28 by:

- W.A. de Jong, J. Brabec, and C. Yang
- L. Gagliardi
- S.J. Klippenstein
- X. Li
- R. Rousseau and V.A. Glezakou
- L.V. Slipchenko

are coupled to mass transport and cascades of chemical reactions. Discovery and exploitation of fundamental principles of these complex, collective phenomena are central to achieving this predictive understanding.

3.2.1.2 Photosynthesis and Light Harvesting

The process of light harvesting is another important example where a predictive understanding of energy transformations in real-world conditions is lacking and therefore open to researchers pursuing transformative progress via simulation and theory. All routes for utilizing solar energy exploit the functional steps of capture, conversion, and storage. The sun's energy arrives on Earth as radiation distributed across the color spectrum from infrared to ultraviolet. The energy of this radiation must be captured as excited electron-hole pairs in a semiconductor, a dye, or a chromophore or as heat in a thermal storage medium. Excited electrons and holes can be tapped off for immediate conversion to electrical power, or transferred to biological or chemical molecules for conversion to fuel. Natural photosynthesis produces fuel in the form of sugars and other carbohydrates derived from the reduction of carbon dioxide (CO_2) in the atmosphere and used to power the growth of plants. The plants themselves become available as biomass for combustion as primary fuels or for conversion in reactors to secondary fuels like liquid ethanol or gaseous carbon monoxide, methane, and hydrogen. We are now learning to mimic the natural photosynthetic process in the laboratory using artificial molecular assemblies, where the excited electrons and holes can drive chemical reactions to produce fuels that link to our existing energy networks.

Although many routes use solar energy to produce electricity, fuel, and heat, none of these are currently competitive with fossil fuels for a combination of cost, reliability, and performance. While solar energy has enormous promise as a clean, abundant, and economical energy source, it presents formidable basic research challenges in designing materials and in understanding the electronic and molecular basis of capture, conversion, and storage before its promise can be realized (DOE-BES 2005). Several past Office of Science workshops have identified the need to develop a fundamental understanding of excited-state processes in photosynthetic and light harvesting systems. It has been clearly recommended that

> **... new theoretical and computational methods are critically needed to account for the complexities of excited state energetics applied across multiple spatial length scales and the ranges of time scales encompassing solar energy capture, conversion and storage (Olivucci and Sinicropi 2005; Dabestani et al. 1998).**

In particular, accurate electronic structure calculations for excited states are necessary for the description of photo-physical processes that may involve charge transfer, exciton transfer, and charge transport in complex environments that are relevant for photosynthesis, light harvesting, and photocatalysis processes. As in the case for catalysis, the complexity in such dynamical processes arises from the coupling between multiple electronic states of various multiplicities and their evolution, vibronic interactions, and also coupling to the environment. Understanding the complicated, light-driven molecular processes requires fast and accurate algorithms implemented efficiently in a leading-edge computational environment to model the above phenomena and account for energy losses due to the relatively fast spontaneous processes. Researchers must also account for explicit coupling of the electron and nuclear degrees of freedom to better understand the time evolution of light-harvesting processes. To propagate an atomistic level of understanding across time and length scales, accurate electronic structure methods should be coupled with multiphysics/multiscale methodologies, which will enable researchers to model photochemical processes in realistic environments.

3.2.1.3 Combustion

Combustion will continue to be a dominant mode of energy conversion for transportation, power generation, and industrial thermal processes for the next half-century. Considerations of energy and environmental security and sustainability, as well as economic competitiveness, demand accelerated development of advanced combustion technologies that combine high efficiency, low emissions, and the ability to reliably operate on an increasingly diverse range of fuels, including bio-derived and synthesis fuels, as well as an evolving feed of fossil fuels. Development of these technologies are significantly hampered by the lack of robust, predictive computational design tools for advanced combustion systems, particularly in new mixed-mode combustion regimes with evolving fuels where stringent efficiency and emissions legislation are driving future technologies.

In internal combustion engines used in transportation, innovations have been hampered by the lack of scientific understanding of how variations in fuel composition affect engine performance, and hence by the inability to predict impacts when conventional petroleum-based fuels are modified or replaced. Even for conventional fuels, existing models are often unsatisfactory, particularly near the limits of stable operation of advanced engines. Further improvements to extract efficiency come at the expense of reliable engine operation, where combustion can occur at conditions far from equilibrium and at extreme pressure. At the ragged edge, strong sensitivities to subtle differences in chemical fuel properties are amplified by the stochastic nature of turbulence, which may lead to undesirable effects such as misfire or knock. These challenges are particularly daunting because energy conversion efficiencies and exhaust emissions are governed by coupled chemical and transport processes at multiple length scales ranging from electron excitation to molecular rearrangements to nanoscale particulate formation to turbulent fuel/air mixing. Fortunately, recent advances in quantum chemistry, chemical kinetics, reactive flow simulation, high-performance computing, and experimental diagnostics suggest that first-principles-based predictive tools for optimum integration of energy conversion/control methodologies and new fuel compositions are feasible. Additional theory is required that will also treat (1) multiscale interfacial phenomena in multiphase reacting flows and (2) chemical and thermal nonequilibrium in turbulent reacting flows. Adaptive algorithms are required to treat both disparities in physical scales, as well as stiffness in chemical and transport models.

Similarly, the design and development of clean and efficient fuel-flexible, gas-fired turbines for electricity generation is an important application area where predictive high-fidelity combustion simulation could have a significant impact. As a result of stringent emissions regulations, industry is embracing novel combustion concepts such as lean-premixed combustion where the fuel is mixed with an excess of oxidizer prior to entering the combustion zone. While clearly attractive from an emissions and efficiency perspective, lean premixed combustion poses serious design challenges such as avoiding flashback and thermo-acoustic instability. At a fundamental level, premixed flame propagation into partially oxidized, autoigniting mixtures or into vitiated products of combustion is poorly understood.

3.2.2 Priority Research Directions

3.2.2.1 Catalysis

With current capabilities, we are able to describe catalytic processes in environments of limited complexity (e.g., small to medium-size localized phenomena and phenomena in relatively short timescales). It will be possible to elucidate how an active site works, what the nature of the reaction kinetics are, and what transformations of reactive sites occur during these processes.

On the other hand, we currently do not have the theoretical and computational tools that will allow us to design real-world heterogeneous catalytic systems, which are more complex than a collection of active sites. The science challenges that need to be addressed drive the need for theoretical, algorithmic, and computational advances in the priority research directions outlined below.

1. **Coupling accurate electronic structure, statistical mechanics, and kinetics for progressively larger and more complex systems, including the description of nonequilibrium systems.** The science challenge driving this priority research direction is the need to control individual reactions through design of active sites with defined structure and multiple chemical functionalities. The initial elements of a framework for coupling the theories currently exist, but advances are needed in the theories and algorithms to allow implementation on high-performance computers.
2. **Developing theoretical frameworks and computational tools to describe transport processes over time and length scales larger than those of processes at active sites.** The science challenges driving this priority research direction are the needs to control the delivery of electrons, ions, atoms, and molecules to active sites and the coupling of these transport process to reactions at active sites. Transport of the different species as well as coupling to reactive processes occur over multiple scales requiring development of self-consistent multiscale approaches, which do not currently exist. The theoretical frameworks, theories, and algorithms must be developed before the design of high-performance computational approaches can be addressed.
3. **Developing theoretical frameworks for describing dynamical transformations of complex interfaces in heterogeneous catalysis.** The science challenge driving this priority research direction is the need to control the mutual influence of catalysts and environments (e.g., solvent arrangement around catalysts and catalyst transformation). Catalysts evolve under operating conditions, and this evolution can occur over various timescales from nanosecond to hours. These processes are influenced by collective behavior over multiple length scales. The theoretical frameworks, theories and algorithms must be developed before the design of high-performance computational approaches can be addressed.

In order to make advances in these priority research directions, the development and coupling of the theoretical frameworks will require integration of applied mathematics with computational and theoretical chemistry. Computational tools that can integrate theoretical frameworks for all of these priority research directions will require computational capabilities well beyond those that will be available in the near future.

Development of new methods is a crucial element to address these challenges: investments are needed in theory to develop accurate methods that scale less steeply with the number of atoms in the molecule (e.g., the most accurate, widely used method for single reference systems, the CCSD(T) method, scales as N^7, where N represents the size of the system, which prohibits its use for very large molecules). We need to go beyond DFT and TDDFT and develop multireference quantum chemical methods for medium to large systems. It is also important to model the coherent dynamics of excited states that may intervene in certain catalytic processes, as well as

the spontaneous processes and theoretical frameworks that work at different scales and bring all of those frameworks together. Finally, it is important to sample rare events (e.g., developing metadynamics methods). Exascale computers offer the possibility to study much larger molecules and assemblies (Figure 3-1), but only if investments are simultaneously made in theory and software.

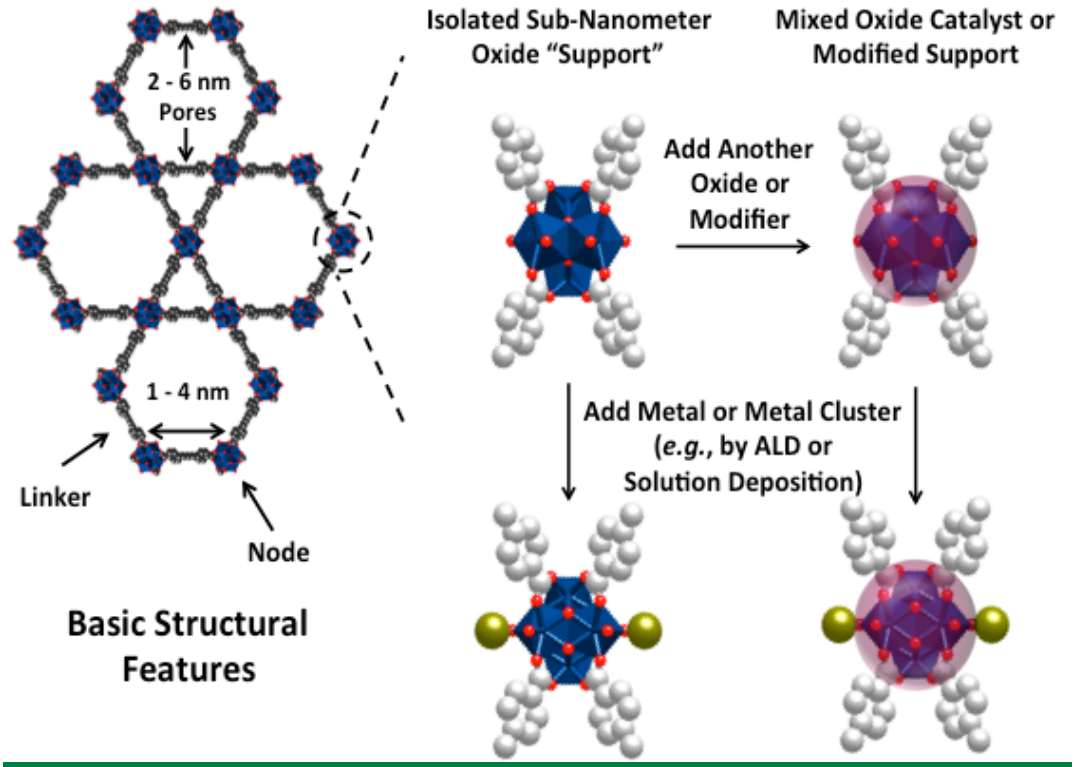


Figure 3-1. Metal and Metal-Oxide Catalyst of Atomic Precision Anchored at a Support for Natural Gas Conversion Catalysis. In order to model full catalytic processes of this type, exascale resources will be needed (Source: ICDC 2016).

3.2.2.2 Photosynthesis and Light Harvesting

The availability of accurate methodologies for excited state simulations at exascale will play a key role in understanding, predicting, and ultimately controlling matter and energy at the electronic, atomic, and molecular levels. The contemporary challenges in photochemistry do not fit into one methodological category. By their nature, these problems are extraordinarily complex and require simulation models that span a broad range of different length and time scales. They have many of the same issues as catalysis but also have their own unique challenges. The accurate electronic structure methods capable of capturing excited-state correlation effects will play a critical role in future computational ecosystems. However, given the inherent numerical complexity of existing electronic structure methods, algorithmic improvements that take advantage of the sparsity of the matrices appearing in quantum mechanical equations, as well as recent advances in applied math and programming models, should lead to successful deployment of current methods into future architectures. Furthermore, for these algorithm-related performance improvements to take place, significant investment in human resources will be required. New theoretical models should be developed and implemented to provide unbiased description of time evolution and dynamics on multistate potential energy surfaces. They will allow us to make advances in the following priority research directions:

1. **Sampling potential energy surfaces inherent to more complex systems** by coupling between different representations of quantum mechanics and using well-parameterized effective Hamiltonians.
2. **Transitioning to larger systems by sampling of longer timescales due to an increase in relaxation times**, which can be reached using on-the-fly rectification of low-order theories with more accurate approaches.
3. **Developing a methodology for simulating spectroscopies, such as core-level excitations, to build a natural synergy with existing and future experimental facilities.** The theoretical models should be validated by stringent integration with experimental workflows.

3.2.2.3 Combustion

Fundamental research in combustion science is essential to understand and predict the behavior of a diverse range of fuels in aero-thermochemical environments representative of the emerging low-temperature, premixed combustion engines. Many of the controlling "turbulence-chemistry" interactions are governed by mixed modes of combustion — including flame propagation into a pre-igniting mixture; lifted diesel flame stabilization by low-temperature–ignition, cool-flame chemistry; and soot formation and burnout, flashback, emissions formation — and often in the presence of condensed phases and thermal radiation.

Global chemical models, which are central to low-temperature combustion, typically consist of thermochemical and transport properties for hundreds of species together with rate coefficients for the thousands of reactions that connect these species within the combustion environment. The fidelity of the full simulations naturally depends on the accuracy of the parameters that make up the chemical model. Priority research directions, then, include the following:

1. **Developing automated procedures** for efficiently and accurately predicting thermochemical kinetics parameters.
2. **Re-exploring the foundations of the chemical models** because a number of the core assumptions appear to provide strong limits on the predictive accuracy of the modeling.
3. **Developing theoretical foundations for multiscale methods** to accurately account for thermal and chemical nonequilibrium in turbulent reacting flows and to treat interfacial phenomena (e.g., sprays, soot formation, catalytic-gas phase interactions, shock-turbulence interactions).
4. **Developing and implementing high-order adaptive mesh refinement and adaptive chemical and transport models** to computationally address the wide disparity of spatial and temporal scales exacerbated at high pressure with exascale computing resources.
5. **Developing uncertainty quantification.** Assessing the quality of turbulent combustion models is fraught with difficulties associated with mesh dependence, the large number of uncertain chemical and transport parameters, and the scarcity and uncertainty of experimental data. This situation strongly motivates a focus on uncertainty quantification in combustion models in order to enable rational assessment of the predictive power of models, the robust selection of predictive chemical/turbulent-combustion models, and sensitivities to key chemical parameters (Figure 3-2).

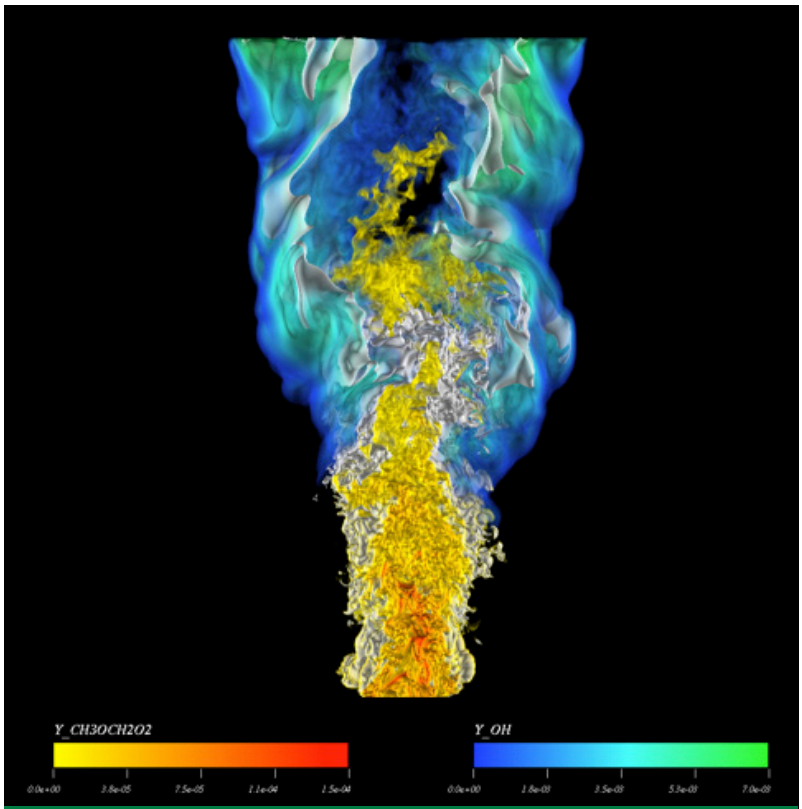


Figure 3-2. Direct Numerical Simulation of a Turbulent Di-methyl Ether Lifted Jet Flame with Multistage Ignition in the Negative Temperature Coefficient Regime at 5 atm. The cool-flame marker is a mass fraction of $CH_3OCH_2O_2$ (yellow-red), and the flame marker is a mass fraction of OH (blue). Extending this to high pressure (40 atm) with n-dodecane and including cool-flame ignition and dynamics of lifted flames and soot generation will require exascale resources, chemical model development and reduction strategies guided by uncertainty quantification, and numerical algorithmic development (Source: Minamoto and Chen 2016).

3.2.3 Cross-Cutting Research Directions

Current practice in any given sub-field of computational science tends to be pyramidal with a significant portion of research accomplished through the use of relatively fast, albeit lower-fidelity computational models, while a smaller portion of researchers use more computationally intense, higher-fidelity methodologies. Leading-edge computing systems, more capable and robust basic-energy computational modeling software, and better interfaces to these computational tools will dramatically enhance the complexity of the problems that can be addressed by individual researchers. These advancements will enable a large percentage of those scientists who are currently using a lower-fidelity method to migrate to high-fidelity methods. While clearly in the best interests of the basic energy sciences mission, this point underscores the importance of maintaining a computing and software ecosystem that encourages a combination of bold new advances in methods through invention of lower-scaling algorithms; development of more efficient software implementation for higher-fidelity methods; and extension of the fidelity of already-fast methodologies.

There is a cross-cutting need for the development of new algorithms and software for studying many chemical phenomena, including the energetics and dynamics of excited states. In particular, there is a need to improve the implementation and scaling of multiconfiguration electronic structure

methods (Szalay et al. 2012; Mok et al. 1996), so that molecular systems whose electronic structure cannot be described to a good approximation by a single orbital configuration are computationally feasible for much larger systems than is possible today. Examples of inherently multiconfigurational systems include many transition metal atoms as well as molecules and solids containing these atoms, molecules with partially broken bonds (e.g., diradicals), most excited states of molecules, and many transition states for chemical reactions (Mills et al. 2011; Christou et al. 2000). Among excited states, the doubly excited states, Rydberg states, and charge transfer states are particularly challenging with current methodologies.

To achieve predictive capabilities with a level of robustness and speed that allows the design of chemical systems for specific basic-energy processes, it will be necessary to re-engineer many existing and concept-stage molecular simulation tools. Of particular interest is the development of turnkey, massively parallel software that are capable of readily adapting to the computer ecosystems of the future. It is highly desirable to have quantum-mechanical methods with predictive accuracy that can account for the electronic, spin, and vibrational degrees of freedom for any arbitrary system. Its accuracy should be comparable to that achieved with coupled-cluster methods for single reference systems, but applicable to open-shell and multireference systems, and it must be nearly as fast as current DFT codes and scale better as a function of system size and across the available processors. It is desirable to be able to couple such tools to time-dependent integration schemes that allow for simulation on the order of milliseconds and spawn all pathways when spontaneous transitions disrupt the classical trajectories that are otherwise being followed. Major investments in algorithm and software development will be necessary to achieve these goals.

In the continuum regime where transport couples with chemistry, there is also a cross-cutting need for the development of asynchronous numerical algorithms that treat computationally stiff chemistry and multiscale interfacial phenomena (e.g., soot particulate formation, sprays, catalytic surface reactions) coupled with transport. These algorithms and their implementation need to be aware of the high level of asynchrony in future exascale environments if they are to be effective at integrating large coupled systems of stiff nonlinear partial differential equations coupled with Lagrangian stochastic differential equations for particle transport. Programming paradigms that treat the nonuniform loads and communication patterns associated with adaptive mesh refinement and adaptive chemistry models are needed. Incorporating uncertainty quantification will also enable identification of deficiencies in key elementary rates. Regarding end-to-end requirements, the dramatic increase in the volume of simulation data will necessitate a large movement toward *in-situ* data processing and visualization.

3.2.4 Computing Needs and Requirements

A common theme across the different areas is that workforce development is a serious bottleneck. The changes envisioned in computing technologies over the coming decade will require a new generation of computational scientists who are well grounded in their science and engineering disciplines but also knowledgeable about the major computational issues being addressed in computer science and applied mathematics. Providing the appropriate education and training of these computational scientists and computational software developers presents a big challenge.

3.2.4.1 Catalysis

Investments are needed in software because none of the current molecular electronic structure packages, especially those using more accurate methods than DFT, will be able to take advantage of the capabilities of exascale computing without undergoing major re-engineering. If these latter investments are not made, we will not be able to make predictions for systems that are any larger than the ones we can consider currently.

3.2.4.2 Photosynthesis and Light Harvesting

The needed investments in software will allow us to make predictions on electron and charge transport phenomena on systems much larger than those currently feasible. To model both the inhomogeneity and interfaces realistically would require systems that are approximately ten times larger. The N^3 scaling in the number of atoms of current algorithms, and hence the factor of more than a thousand-fold increase in computational resource needs makes this an exascale computing challenge. However, both algorithmic and implementation advancement are needed in order to overcome such difficulty and enable efficient exascale computing of electronic structure theory. Such a method would also require time-dependent integration schemes, as well as inclusion of all the pathways spawned by spontaneous transitions.

3.2.4.3 Combustion

Exascale direct numerical simulation (DNS) and large-eddy simulation (LES) capabilities are required to achieve the high Reynolds number, high pressure, and complex chemical kinetics required to address underlying "chemistry-turbulence" interactions in gas turbine and internal combustion engines. Larger computational domains and longer simulations are needed to provide statistical convergence, and a larger dynamic range of scales is required to resolve high Reynolds number flame/turbulence phenomena at high pressure. In addition to gas-phase kinetics, multiphysics coupling with reactive sprays, thermal radiation, soot particulate formation and oxidation, and catalytic interactions also requires enormous computing resources. Finally, adjoint sensitivity and uncertainty quantification (UQ) in reactive flow simulations are needed to analyze sensitivities of heat release rate and emissions to key chemical rate and transport parameters. This information is required to guide further research into predictions and measurements of key rates limiting chemical model parameters. UQ is also required in analyzing sensitivities of combustion models to LES boundary conditions and numerical algorithmic parameters.

3.2.4.4 Other Needs and Requirements

3.2.4.4.1 Method Development Needs

The scientific challenges outlined in Section 3.2.2 demand not only the massive increase in computational capability but also significant advancement of existing methodology for describing electronic structure and dynamics capable of accessing realistic length and time scales. These advances must aim for a wide spectrum of capability, from near-exact treatment of electronic structure for 20–50 atoms coupled to post-DFT treatment of hundreds of atoms to efficient hybrid/exact-exchange DFT methodologies for thousands of atoms. Some of the critical areas of need for methodological development are as follows:

- Robust electronic-structure methods that can treat strong electron correlations featured in most thermally and photo-driven chemistries; this area includes a need for traditional multireference methods (perturbation theory, configuration interaction, coupled-cluster), Quantum Monte-Carlo, tensor network methods (e.g., density matrix renormalization group), and emerging ideas (e.g., multireference-DFT) (Li Manni et al. 2014; Garza et al. 2015). These methods must be full featured, that is, they must:
  - Be able to treat ground states as well as valence and core excited states on even footing;
  - Feature a full suite of capability, including analytic forces, hessians, and nonadiabatic couplings, and electromagnetic response properties;
  - Render robust 2- and 4-component treatments of effects of special relativity;
  - Treat relativity using variational methods and with the analytical gradients available; and
  - Provide efficient elimination of all sources of numerical discretization error, for example, by explicit correlation.

- Seamless multiscale methodologies, for example, for integration of high-end with lower-fidelity methods via quantum embedding and QM/MM.
- Efficient classical and semiclassical *ab initio* dynamics on one or more potential energy surfaces.
- Seamless treatment of open and periodic (in one-, two-, and three-dimensional) boundary conditions.
- Methods for improving the sampling of complex high-dimensional energy landscapes and for extending the timescales of MD simulations.
- Enabling of elements of the ground- and excited-state dynamics driven by high-accuracy methods coupled to realistic environments.
- Tensor libraries that take advantage of exascale architectures that embrace new programming models.

3.2.4.4.2 Data

- Portal with data made public: Being able to do "meta" experiments (with other groups' data as well). Develop new algorithms that can go beyond the original questions that the researchers who collected the data were asking.
- Machine learning: Computers find the interesting things in a research group's data and are able to judge the data in terms of quality and information content (unsupervised and supervised learning).

3.2.4.4.3 Hardware and Software

- Development of new strategies for memory management and low-communication highly parallelizable algorithms to achieve load balancing and effectively use the fast multicore platforms.
- Increased data rates/quantities (data management, storage, access, transfer) (federated identity [ID]/logins, data format).
- Combining of multiple different experiments/data collection types/synchrotrons, synthesizing and comparing and combining the data (spectral, spatial, etc.).

BES Path to Exascale in Catalysis

- **Today: Model electronic structures and reaction kinetics on catalytic surfaces, including homogeneous as well as nano- and meso-structured materials.**
- **2020: Perform computational screening of thousands of candidate materials based on databases of accurate elementary reaction rates to guide laboratory-scale system calibration. Utilize multiscale, multiphysics methods to describe catalyst structures and reactions accurately over the necessarily long timescales.**
- **2025: Enable end-to-end, system-level descriptions of multifunctional catalysis. Uncertainty quantification and data integration approaches will enable inverse problems for catalytic materials design.**
- **Enable integration of accurate, multiscale simulations into industrial, process-level descriptions of energy production and manufacturing.**

BES Path to Exascale in Photosynthesis and Light Harvesting

- Today: Perform *ab initio* modeling of electronic phenomena of systems with hundreds of atoms.
- 2020: (petascale) Perform *ab initio* modeling of electron transport and electron-phonon coupling of complex organic polymer systems of the order of tens of thousands of atoms. These simulations model individual components, ignoring interfaces between components that affect the performance of the devices.
- 2025: (exascale) Develop a model that accurately describes the electronic charge transport in the polymers and across electrical contacts, as well as the thermal transport across the warm and cold fluids through the ceramic layers and the organic polymers at appropriate levels of detail.

BES Path to Exascale in Combustion

- Today: Perform 3D simulation of turbulent combustion at ambient and moderate pressure with small oxygenated hydrocarbons (e.g., ethylene and di-methyl ether) at low to moderate Reynolds number.
- Today: Perform 3D simulation of high-pressure, low-temperature, turbulent-lifted diesel jet flames with n-dodecane or reactivity-controlled compression ignition (RCCI) with fuel blending of alternative C1-C2 fuels and natural gas. Continue to explore the limits of high-pressure, turbulent combustion with increasing Reynolds number.
- 2020: Perform 3D simulation of high-pressure, low-temperature turbulent-lifted diesel jet flames with hydrocarbon and biodiesel and homogeneous charge compression ignition (HCCI)/RCCI with fuel blending at engine-relevant conditions including turbulent Reynolds number. Use DNS and experimental data to validate and improve LES sub-grid combustion models for multistage ignition and mixed combustion regimes with strong turbulence-chemistry interactions and with UQ to identify uncertainties in chemical and combustion models.
- 2025: Fully incorporate multiphysics, multiscale combustion science into validated, predictive simulation capability that can be used by industry in the design cycle, allowing industry to reduce the development time for efficient engines.

3.3 Materials and Chemical Discovery

3.3.1 Scientific Challenges and Opportunities

The foremost scientific challenges that need to be surmounted to realize the vision of materials and chemicals by design are threefold: (1) the computational discovery of novel materials and chemicals with target properties (including hierarchical structures with multiple functionalities), (2) the prediction of pathways to the synthesis of these materials and chemicals (with consideration of sustainability and green chemistry principles), and (3) the prediction of their stability and degradation pathways. Surmounting these challenges will involve a synergistic interplay of advances in predictive modeling capabilities, hardware resources, and experimental techniques allowing for the characterization of spatial and temporal fluctuations and short-lived intermediates along synthesis and degradation pathways.

Along the ladder of increasing complexity, the computational discovery of materials and chemicals initially started with prediction of the energetic stability of crystalline matter at zero Kelvin and the heats of formation of chemical compounds in the gas phase. The current state of the art allows for the prediction of the thermodynamic and mechanical stability of more complex bulk phases (but without extended structural disorder and microheterogeneity) and of reaction pathways in homogeneous fluids and at ordered fluid-solid interfaces. Beyond the prediction of synthesis pathways for relatively simple chemical compounds, the computational synthesis of stable and metastable materials involves either highly simplified models or focuses on specific steps along the pathway to larger-scale structures. With extant resources, over the next five years, we will see the prediction of size and shape distributions of semiconductor nanocrystals for thermoelectrics and of nucleation pathways and subsequent growth for specific systems involving environments of limited complexity, for example, the modification of nodes and ligands in metal-organic frameworks to tune capture, separation, and/or catalytic properties. An exascale computing environment will enable the shift to an adaptive and self-consistent multiscale-modeling paradigm (involving integration of algorithms and models ranging from electronic structure with correlated wave function methods, to MD and Monte Carlo with enhanced sampling approaches for Kohn-Sham density functionals and molecular mechanics force fields, to dissipative dynamics of nanoscale objects). Thus, exascale resources will enable scientists to take pivotal steps toward the reliable prediction of the very complex synthesis pathways involved, for example, in the hydrothermal synthesis of hierarchical zeolites and carbonaceous materials (e.g., 3DOM carbons; Stein Research Group 2016) where the pathways are influenced by not only the composition, temperature, and pressure of the reacting system but also by the pH; local concentration gradients and fluctuations; structure-directing agents at multiple length scales; and mechanical agitation, external fields, and strong heating/cooling that keeps the evolving system far from equilibrium. A similar progression of complexity is involved in predictive modeling of the stability and degradation pathways of complex materials and chemicals under operating conditions that may include extreme environments.

White papers by the following authors informed the writing of this section and can be found in Appendix C, Section C.3, starting on page C-58:

- H.-P. Cheng
- M. Fernandez-Serra
- C.J. Mundy and G. Schenter
- J. Pask
- J.I. Siepmann
- B.G. Sumpter

Similarly, these relevant case studies can be found in Appendix D, Section D.3, starting on page D-49 by:

- G. Galli and F. Gygi
- C.J. Mundy and G.K. Schenter
- J.I. Siepmann
- M. Stevens

In order to obtain complete control of modern synthesis, a detailed understanding of how complex free energy landscapes, including long-lived metastable states, influence the desired outcomes (e.g., limiting blocked pores in sorbents or blocked junctions in electronic materials, favoring 2D over 3D crystal growths) is needed. Reaching this capability will require obtaining a theoretical understanding of the principles that couple scales and lead to control of the synthetic process, comparison with existing

experimental information, access to experimentalists to perform specific experiments to verify theoretical capabilities, and on-the-fly coupling of theoretical and experimental *in situ* characterization of chemical assemblies and materials. To date, there are no scientific principles or frameworks that incorporate this level of detail; for example, current frameworks for assembly are based on mean-field concepts, such as DLVO (Derjaguin-Landau-Verwey-Overbeek) theory, that will not hold as the nanometer scale is approached where distributions/fluctuations at solid-fluid interfaces are essential for the coupling. A recent review has highlighted the importance of competing pathways for nucleation and assembly in nanomaterials (Smeets et al. 2015).

Thus, exascale resources will enable scientists to take pivotal steps toward the reliable prediction of the very complex synthesis pathways involved, for example, in the hydrothermal synthesis of hierarchical zeolites and carbonaceous materials.

An integral part of the efforts toward designing materials and chemicals with target properties is gaining the capability to characterize heterogeneous, disordered, and/or defective materials experimentally with emergent phenomena that relies on computational approaches for on-the-fly analysis of real- and reciprocal-space images and spectroscopic signals. First, there is the need for enhanced reliability of the computational techniques in such a way that they can accurately (and rapidly) address the above complex functionalities; provide the precision necessary for discriminating between closely competing behaviors; and are capable of achieving the length scales necessary to bridge across features such as domain walls, grain boundaries, and gradients in composition. Second, there is a need to take full advantage of all of the information contained in experimental data to provide input into computational methods to predict and understand new materials. This capability includes integrating data efficiently from different characterization techniques to provide a more complete perspective on materials structure and function. For example, scanning probe microscopy reveals position-dependent functionality data, while transmission electron microscopy provides position-dependent electronic structure information. Techniques like inelastic neutron scattering allow for direct measurements of space and time-dependent response functions, which in principle can be compared directly with theoretical calculations. However, many of these leading-edge experimental techniques come with extreme requirements for full analysis and utilization of instrumental data. For example, improvements in detector technology and data capture rates will result in huge datasets across spectroscopy and imaging, providing enormous amounts of potentially insightful experimental data for detailed analysis. Third, access to predictive models of material/chemical behaviors and fluctuations will enable new experimental characterization approaches and provide validation to connect predicted synthesis/degradation pathways with measurements. Toward these goals, experimental characterization needs to progress from focusing on average properties to a focus on distributions, fluctuations, and transients. This topic is essential for connecting scales where thermodynamic and structural fluctuations will increase with decreasing length scale.

The current state of the art encompasses *a posteriori* analysis of complex characterization, simulation of time-resolved spectroscopy of model/simple systems with well-defined states (ground or simple excitations), solution-phase simulations of dilute systems or idealized interfaces without generality, or collisions of nanoparticle with liquid surfaces (Figure 3-3). There are computational limitations related to algorithmic requirements and scaling of almost all computational methods required for characterization. By 2020, extant petascale computing capabilities will open opportunities to direct simulation of characterization (imaging or spectroscopy) data as additional and continual output of large-scale simulations. Furthermore, analysis of measured characterization will define the timescale to connect measurement with interpretation as a target for optimization. Computing requirements for characterization will be beyond additive given the need to simulate excited states (various spectroscopies) in addition to ground state properties (imaging). The exascale computing capabilities will enable combining information gained on free energy landscapes relevant to material/chemical processes/pathways with predicted distinguishing characteristics of metastable or kinetically trapped states and, hence, will permit coupled validation of synthesis efforts, both theoretical and experimental. This level of functionality will provide

confidence limits on the predicted outcomes of proposed synthesis/materials realization pathways and lead to continually simulated experimental output for validation and also design of tandem experiments to produce more robust theories and methods.

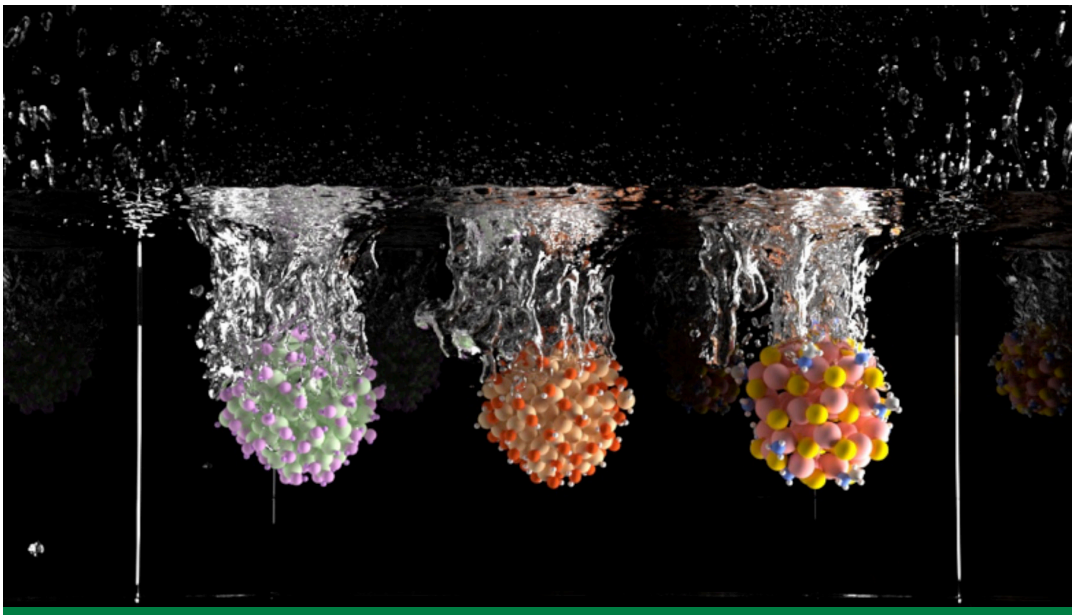


Figure 3-3. Nanoparticles Entering a Solution Phase as Part of a Complex Synthesis Pathway. High-fidelity modeling of the synthetic pathway of nanostructured solids is still of out of reach for atomistic and first-principles simulations (Source: Figure courtesy of N. Brawand, University of Chicago [The Galli Group undated]).

3.3.2 Priority Research Directions

The predictive modeling capabilities for materials and chemical discovery and for synthesis and degradation pathways are most urgently needed in the following priority research directions:

1. **Solar energy harvesting** (light matter interactions involving electronic band alignment and solid-liquid interfaces and excited state dynamics).
2. **Energy storage involving electronic, chemical, and thermal means** (relating specific interfacial details with gross measurements of electronic and thermal conductivity and capacitance).
3. **Complex chemical assemblies and materials with quantum phenomena, multiferroics, and emergent states/properties.**
4. **Hierarchical materials with multiple functionalities,** such as porous materials with mesoporous regions for facile transport and microporous regions for sorption/catalytic selectivity.
5. **Robust structural and molecular models of complex, heterogeneous materials.** Real materials used in applications are, almost always, heterogeneous. Despite this condition, most computational materials discovery today is performed on or for idealized, single-phase materials and compounds. Describing, controlling, and predicting the nature of heterogeneous systems are crucial to addressing key energy and information challenges, such as energy conversion and storage processes and next-generation electronic materials. These complex systems offer new functionalities (e.g., solid-liquid interfaces with catalysts) and properties that will challenge existing computational techniques. For example, predicting transport and dynamical properties in heterogeneous systems will require advances in developing predictive calculations of transport and quantum dynamical models and software.

We must extend current computational resources and methods to address these more realistic models of materials. The computational prediction of heterogeneous materials and their properties will affect nearly all important areas of technological materials and chemicals. Energy harvesting and storage applications have significant urgency as a result of climate change and energy-independence issues that are well documented. New electronic materials are urgently required owing to the end of Moore's Law and the replacement of CMOS (complementary metal oxide semiconductor) devices, as well as to maintain economic competitiveness.

Examples of the ubiquitous role of heterogeneous systems are illustrated in Figure 3-4 in the areas of (1) energy storage via high-energy density battery systems, and (2) energy harvesting via nanostructured thermoelectric materials.

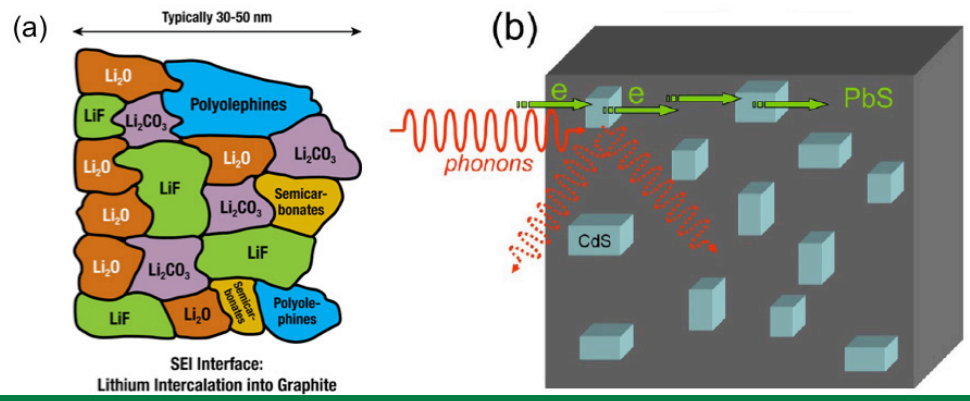


Figure 3-4. The Ubiquitous Role of Heterogeneous Systems: (a) The solid-electrolyte interface (SEI) is a critical component in electrochemical energy storage. Because of the high reactivity between the electrolyte and the electrodes at the SEI interface, Li-ion batteries show limited calendar and cycle life, much lower than the time required for enabling this technology in vehicles (Source: Argonne 2012). (b) Thermoelectric materials convert heat into useful electricity, and so have enormous potential for waste heat recovery. Heterogeneous, nanostructured thermoelectric materials are some of the highest efficiency thermoelectric materials today. These materials effectively scatter heat-carrying phonons, while allowing electrical carriers to transport easily through the material. Computational prediction and design of these heterogeneous systems will enable the next generation of high-efficiency thermoelectrics (Source: Zhao et al. 2012).

6. **Additional overlapping computational research needs that are crucial to both chemical and materials discovery,** as well as to a variety of other research directions, including these:

   - The calculation of charge/heat/mass transport through complex, heterogeneous materials. We need both new modeling capabilities, as well as significantly increased computational resources to compute these challenging properties.
   - More accurate "computational spectroscopy" to provide validation and connect with characterization and facilities.
   - Electronic structure methods beyond standard DFT to include additional correlation effects. These methods are a prerequisite to achieving the accuracy required for general prediction across wide classes of materials types.
   - "Structure problem." Atomistic computational methods rely on knowledge of the crystal structure of the system of interest, which, in many cases, poses a significant challenge:
     - **Missing structures.** Materials discovery today is largely based on known, experimentally derived crystal structures, assembled in databases. However, these databases are largely incomplete: for example, the Inorganic Crystal Structure Database (ICSD) contains ~150,000 entries of known crystal structures; however,

the Powder Diffraction File contains ~300,000 entries for material synthesized and diffraction measured. This discrepancy implies that in a very large number of cases where materials have been synthesized, we do not know the structure — perhaps even in the majority of cases! Crystal structure prediction methods exist today and could help solve these unknown structures; however, these methods are extremely expensive computationally, and today's computational resources cannot tackle these ~10^5 cases of "missing structures."

- **Structures of interfaces.** Heterogeneous systems are necessarily composed of interfaces: solid/solid, solid/liquid, or solid/gas. The atomic-scale structure of these interfaces is almost always lacking or completely unknown. We must develop new methods to predict the structures of these interfaces.
- **Combinatorial explosion.** As we consider higher-order multicomponent materials in our discovery efforts, the number of possible stoichiometries and structures explodes. For instance, considering ~100 elements in the periodic table, there are ~5,000 binary combinations (e.g., 5,000 binary alloy systems); however, if we increase system size to four-component systems, now ~4 million quaternary systems become possible. This profusion of possibilities, while it increases the chances of finding exciting and novel materials and functionalities, also increases the space over which we must perform these searches. The historical rates of materials discovery are illustrated in Figure 3-5.

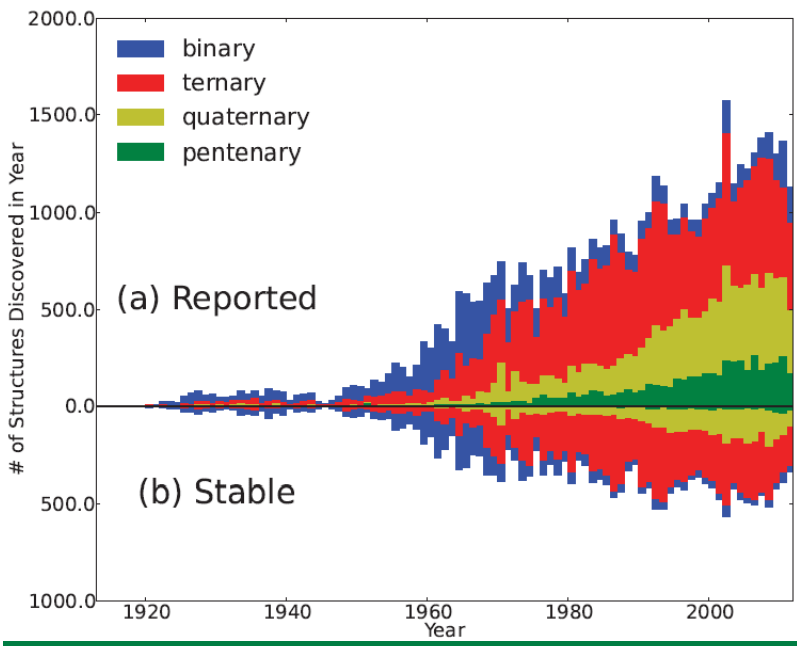


Figure 3-5. Historical Rates of Materials Discoveries: (a) the total number of compounds discovered within the ICSD by year, and (b) the number of stable (T = 0 K) compounds discovered in the ICSD, where the stability is assessed by the Open Quantum Materials Database (OQMD) energies (Source: Kirklin et al. 2015).

- The prediction of synthesis/degradation pathways will benefit from both descriptor discovery via machine learning techniques (extraction of hidden correlations) and theories underpinning coupling of different properties and multiple length and timescales. Understanding mechanisms, pathways, and intermediates are key to achieving control, with hierarchical materials comprising an opportunity for development. In addition, theory and computation will need to compare with existing experimental information and will also require very specific new experiments to verify theoretical capabilities.
- Investment in training and availability of excellent chemical and material scientists with extensive knowledge of physical models, mathematics, and high-performance computational science working in tandem with mathematicians, computer scientists, and computational engineers.

3.3.3 Cross-Cutting Research Directions

Advances in characterization are pushing into regimes that are more realistic or coupled to applied energy-relevant phenomena, employing *in situ*, *operando*, and time-resolved techniques. Theory, modeling, and simulation currently struggle to keep up with these advancements in terms of the complexity of systems under study and multiplicity of external controls/pumps.

3.3.3.1 Validation and Uncertainty Quantification in Predictive Simulations

A systematic validation of models and approximations is a fundamental aspect of the development of reliable computational discovery tools (Figure 3-6). An exascale computational environment will greatly accelerate the validation and UQ process and thus lead to computational methods of unprecedented fidelity and robustness.

Simulations involve a wide variety of approximations at all levels of description of matter. At the microscopic level, quantum mechanical approximations include various approximate models of electron-electron interactions, such as DFT, DMFT, and multiple forms of many-body perturbation theory (MBPT), as well as QMC methods. Atomistic simulation methods rely on accurate and transferable force fields and efficient sampling algorithms. The description of transport properties (charge, spin, or heat) requires elaborate models that involve both atomic trajectories and electronic properties. Furthermore, numerical approximations, such as finite basis set in electronic structure computations or potential truncation in force field-based simulations, also affect the reliability of predictions. In addition, length and time scales investigated by computations are often orders of magnitude smaller than those of the corresponding experimental systems.

A systematic validation of such a broad range of approximations requires considerable computational resources and involves a hierarchy of computations performed with models of increasing accuracy.

Current resources allow for the validation of relatively cost-efficient DFT approaches for the description of ground-state properties of simple periodic solids. MBPT approximations require much larger resources and have only been validated in a few systems of very limited size. Comparisons of DFT-based approaches with wave function methods used in quantum chemistry are only feasible for molecules and clusters. The description of liquids and solid/liquid interfaces — essential to understanding many energy conversion processes — requires MD or Monte-Carlo simulations, which further increase the computational cost of the validation process. The validation of hybrid-DFT methods for the description of liquids is currently limited to simple systems and requires the largest resources currently available, namely, via INCITE [1] awards on leadership-class supercomputers.

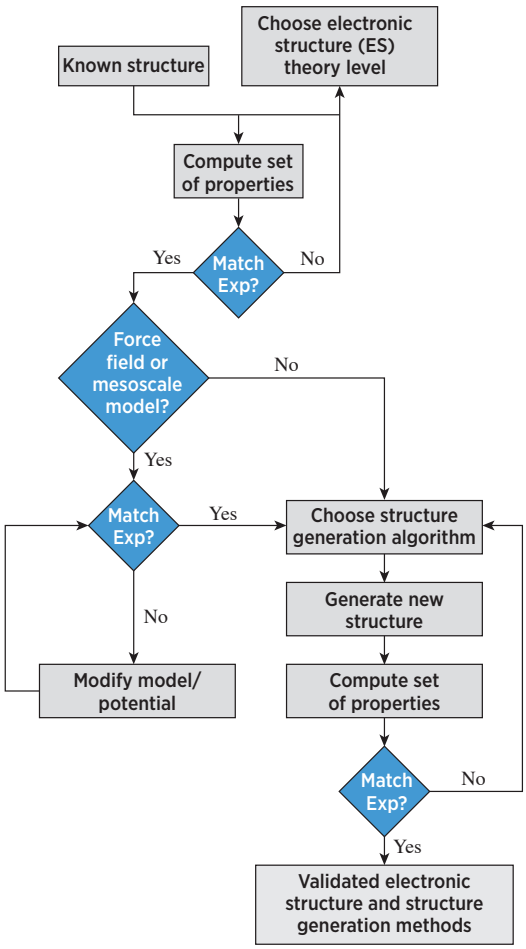


Figure 3-6. Decision Tree Diagram of Structure Determination and Computing (adapted from diagram available at miccom-center.org/validation.html).

Validation of force fields also requires long simulations given that they involve the testing of transferability to various problems. UQ resulting from model parameters will require large numbers of simulations to cover high-dimensional parameter spaces.

[1] INCITE stands for Innovative and Novel Computational Impact on Theory and Experiment awards and are administered by the ASCR User Facilities.

Currently, a number of important validation procedures are simply inaccessible with existing computational resources. They include the validation of models of transport properties, as well as the validation of linear- or reduced-scaling electronic structure methods in condensed-phase problems. One limiting factor in the validation of linear-scaling algorithms is the need for reference results obtained with more costly algorithms.

An exascale computational environment will open the door to the validation of MBPT approaches in liquids and disordered solids, as well as nonadiabatic approximations for the description of chemical reaction pathways in condensed phases.

A successful implementation of the validation process will require the ability to run scalable AIMD or Monte-Carlo simulations on large computing platforms. Workflow tools and online comparison tools will be essential when making the resulting reference data available to the research community.

3.3.4 Computing Needs and Requirements

Computing needs enabling advances in the area of predictive modeling and detailed characterization of synthesis/degradation pathways are inherently heterogeneous. In some cases, memory requirements are limiting multireference wave function calculations for highly correlated electronic systems. In other cases, it is complex workflows for self-consistent multiscale modeling with adaptive resolution where seamless and low-latency connections of rather different computer programs with different scalabilities are required. However, in many cases, ensemble computing already successfully exploits petascale capabilities and will be transferable to exascale resources. For such calculations, the speed of individual compute nodes is often most essential.

High-throughput computational databases are responsible for many successful examples of materials discovery today (e.g., OQMD, Materials Project, aflowlib). We have the capability to scan computationally through thousands of materials to find those with desired properties. However, the current status of these databases is limited in terms of the properties that are calculated (mainly $T = 0$ K energetic properties) and level of accuracy of the methods (mainly standard DFT).

Three distinct, significant directions in high-throughput materials databases are enabled by computational power that is 100–1,000 times greater than present capabilities:

1. More accurate methods (than $T = 0$ K DFT)
2. More complex properties
3. Combinatorial expansion of material space

These future computational resources will allow significant extensions of these databases, which will greatly enhance our ability to predict new materials with tailored properties:

1. The use of methods beyond DFT, such as hybrids, GW methods/many-body, or QMC typically require ~100–1,000 times more computational resources than standard DFT, and hence fit precisely into the scale of future computational architectures. These methods also often have more parallelism than standard DFT (e.g., parallel over bands vs. band pairs, walkers, etc.).
2. Inclusion of a much richer variety of properties in these databases (at either the DFT level or beyond), such as defects, alloys, phonons, disorder, free energies, interactions, and spectroscopy (e.g., photoemission, X-ray, neutrons, Raman, absorption, electron energy loss spectroscopy [EELS]). These properties are more expensive than a single DFT energetic calculation, often again in the 100–1,000 times range, and hence are well suited to future computational expansions.

3. Combinatorial expansion, as detailed above, encompasses the structure prediction problem, missing structures, predicting structures of interfaces, etc., and will each require approximate sampling of ensembles of structures, often in the range of hundreds or thousands of structures.

The issue of "execution management" of these ensemble jobs will also need to be addressed. The current high-throughput databases have developed their own management systems for standard $T = 0$ K DFT (e.g., qmpy for the OQMD, Fireworks for Materials Project, and aflow for aflowlib). However, the inclusion of new methods and properties will require development of new execution management strategies.

BES Path to Exascale in Materials and Chemical Discovery

- **Today: Perform computational materials discovery on or for idealized, single-phase (i.e., homogeneous) materials and compounds.**
- **2020: Predict transport and dynamical properties in heterogeneous systems.**
- **2025: Model synthesis, stability, and degradation of heterogeneous or hierarchical materials and complex chemical assemblies.**
- **Perform "computational spectroscopy" to validate simulation results and connect with characterization and facilities.**
- **Close the gap on crystal materials in the Inorganic Crystal Structure Database with unknown structures.**

3.4 Soft Matter

3.4.1 Scientific Challenges and Opportunities

Soft matter provides unique and critical materials behavior in a wide range of industrial products. Polymers, surfactants, electrolytes, and microheterogeneous fluids have long been key components in a multitude of applications, including energy storage (e.g., batteries and capacitors) and energy production (e.g., photosystems), chemical separations, enhanced oil recovery, food packaging, chip manufacturing, and health care products. However, as the complexity of applications increases, so does the challenge to gain greater control over static and dynamic properties through a fundamental understanding of material types and properties and conditions of, for example, polyelectrolytes, complex structures in solution, active matter, soft/hard composite materials, transport in soft matter, processing of soft materials, and response to mechanical deformation. Further, many developments in supramolecular material sciences focus on mimicking the self-organizing nature, dynamics, and complex interactions within and with the environment of living systems. These and other grand challenges described in the early BESAC report (DOE-BESAC 2007a) and more recently in the report *Challenges at the Frontiers of Matter and Energy: Transformative Opportunities for Discovery Science* (DOE-BESAC 2015) are intimately tied to soft matter. In particular, *creating materials with capabilities rivaling those of living systems* is at the heart of soft matter research. Soft materials composed of molecular and/or modular building blocks can provide the hierarchical complexity and tunability for making paradigm-shifting materials that can accomplish multiple tasks. The complexity of soft materials presents scientific as well as computational challenges that make exascale computing a pivotal resource in achieving the goal of designing functional matter, which requires not only orders of magnitude greater scalability in both dimensional and time scales, but also seamless integration with exabyte big data analytics and mining so as to extract maximal scientific knowledge.

Biological systems are multifunctional and highly responsive due to complex, hierarchical structures and the associated dynamics inherent in these structures. Computational researchers are just beginning to be able to simulate simpler versions of hierarchical structures, whether biological in origin or synthetic versions inspired by biology. Hierarchical soft matter is characterized by inherent structures covering different length scales that evolve over a broad range of timescales in response to external constraints (e.g., chemical potentials, pressure, temperature, electric or magnetic fields) and stimuli (e.g., shear or changes in any of the external constraints). This responsiveness arises from their ability to transfer energy between different forms. The ability to undergo controlled energetic and structural transformation also underlies the ability to store, alter, and transmit information. For example, DNA is the most well-known macromolecular information storage system. While in equilibrium, information can be understood in terms of entropy; however, information transmission typically involves nonequilibrium and chaotic processes for which fundamental theoretical principles are still lacking. For solar energy conversion, systems such as photosystem II serve as key exemplars for energy production and will involve complex excited state crossings in addition to the intricate structural changes that accompany energy transmission (Figure 3-7).

A white paper by the following authors informed the writing of this section and can be found in Appendix C, Section C.4, starting on page C-75:

- P. Vashistha (with R.K. Kalia and A. Nakamo)
- T. Panagiotopoulos

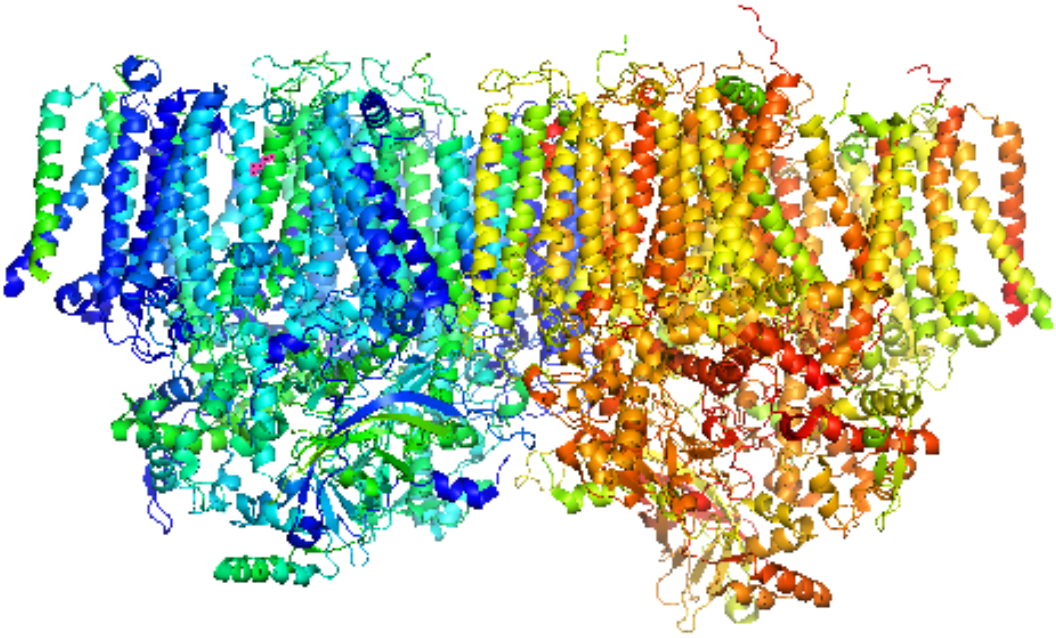


Figure 3-7. Photosystem II (Source: Ferreira et al. 2004).

While "hard matter" commonly involves crystalline materials that can be simulated using unit cells or super cells with linear dimensions of less than 100 nm or, more often, less than 10 nm, soft materials typically are liquids, gels, or amorphous solids with internal structures spanning from nanometers to micrometers or even beyond. Moreover, soft matter is often heterogeneous at multiple length scales (e.g., preferential solvation on the molecular length scale, self-assembly into vesicles or fibrils reaching micrometer dimensions, to multiphase systems with macroscopic dimensions). Correspondingly, soft matter typically possesses an extraordinarily broad spectrum of timescales for equilibrium processes and an even broader spectrum when considering nonequilibrium assembly, responsive behavior, and degradation. It is a steep challenge to simulate large enough systems that adequately represent the hierarchy of structures and dynamics in soft materials.

The ability of soft matter to undergo controlled energetic and structural transformation also underlies its ability to store, alter, and transmit information… While in equilibrium, information can be understood in terms of entropy; however, information transmission typically involves nonequilibrium and chaotic processes for which fundamental theoretical principles are still lacking.

For the first time in the history of science and engineering, many interesting phenomena in soft matter are expected to become accessible with exascale resources. The advent of exascale computing will enable modeling of many fundamental phenomena in soft matter that will dramatically advance our understanding and have a significant broader impact on critical national needs. For example, it will be possible to systematically investigate the connection between the molecular building blocks (e.g., repeat units in a polymer), their sequence (e.g., block copolymers, metal-organic frameworks [MOFs] or DNA and proteins), and composition (e.g., dispersity within the same type of polymer and blending of different polymers) of complex soft matter systems and their thermophysical properties and functional attributes. This knowledge will lead to the development of novel soft matter systems that offer significantly improved performance compared to today's technological solutions or emergent functional properties that are synergistically enhanced and not predictable from studying the components in isolation.

3.4.2 Priority Research Directions

In this section, specific classes of soft matter systems and important computational advances are discussed, although it must be emphasized that the list is illustrative, not exhaustive. Advances in theory and algorithms for alternative separations, multicomponent fluid mixtures, composite polymeric materials, and rational design of polymer dielectrics for energy storage will directly address many of the grand challenges in soft matter discussed earlier.

3.4.2.1 Alternative Separations

At present, separations in the chemical and petrochemical industry are responsible for about 10% of the energy consumption in the United States. Most of the large-scale separations utilize energy-intensive distillation as the separation process. Thus, there is a great need for alternative separation processes that require significantly less energy. These alternative separation technologies may involve extraction with designer solvents, pervaporation using membranes, or numerous forms of chromatography that all employ soft matter as the mass-separating agent. Within the framework of the Materials Genome Initiative and as described in the DOE BES report, *Computational Materials Science and Chemistry: Accelerating Discovery and Innovation through Simulation-Based Engineering and Science* (DOE-SC 2010), the vision is to use predictive modeling to discover these alternative mass-separating agents.

The computational challenge in developing a screening methodology for soft matter is the complexity of the steps required to obtain thermodynamic and transport properties. The relationship between chemical composition and state and resultant properties is much less transparent than in the case of hard materials. Property predictions intrinsically require conformational sampling through molecular dynamics (MD) or Monte Carlo (MC), or variants such as nonequilibrium MD (NEMD). Currently, prediction of soft matter properties is most often made through a series of manual steps like those shown in Figure 3-8 for MD/MC simulations. In contrast to the computational screening of hard materials, where the location of all atoms is fixed *a priori* in a lattice and interactions are dominated by strong electrostatic forces, the collective properties of soft materials are frequently driven by a combination of relatively weak interactions between molecules and intramolecular interactions resulting from covalent bond networks.

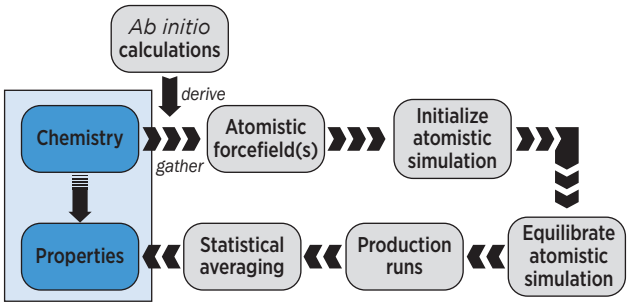


Figure 3-8. Prediction of Many Soft Materials Properties from Chemical Structure/State (shown as downward arrow on the left) Consists of Multiple Steps. Depending upon availability, a force field may need to be derived which involves *ab initio* calculations (Source: Figure courtesy of Clare McCabe, Vanderbilt University).

Algorithmically, the challenge is to automate all the steps shown in Figure 3-8, and moreover to develop the data tools needed to archive and retrieve force fields, *ab initio* calculations, and molecular simulations in order to minimize repetition of calculations; all the methods need to be scriptable to be used in a material genome environment. Once developed, the computational infrastructure will require exascale resources to screen a given class of soft materials, such as ionic liquids, synthetic membranes, MOFs, and deep eutectic solvents, against property targets, such as the ability to capture/extract CO_2, valuable bio-renewable compounds, mercury, or radioactive elements from complex mixtures. It is easy to envisage screening scenarios for alternative separations in which millions of simulations are managed simultaneously, thus saturating an exascale computer over long periods of time (measured in days of wall-clock time).

3.4.2.2 Multicomponent Fluid Mixtures

Multicomponent fluid mixtures are found in many key technologically important areas such as reservoir fluids (gas and oil), lubricants, personal and health care products, and food. Key characteristics of these fluids are that they contain tens, hundreds, or thousands of different interacting colloidal particles, polymers, molecular and polymeric electrolytes, surfactants, solvents, and other functional additives. Small changes in composition, temperature, pressure, or external stimuli (e.g., shear) can cause large changes in the system's phase behavior and thermophysical, structural (often thermodynamically metastable states), and transport properties. In reservoir fluids, for example, knowledge of phase, wetting, and flow behavior is of utmost importance to ensure high recovery and safe operation. Complex lubricant formulations are blended to achieve desirable performance metrics, with certain compounds preferentially adsorbing at specific surfaces and other compounds governing viscosity-temperature and viscosity-pressure effects. Here, a grand challenge is predicting lubricant performance/degradation under realistic operating conditions that involve complex nonequilibrium flows and chemical reactions.

For example, while *n*-dodecane and 6-methylundecane possess very similar molecular properties and form nearly ideal liquid solutions with extremely small excess properties, sufficiently long polymers formed by the corresponding repeat units become immiscible in the melt phase and exhibit rather different dynamic properties. Similarly, surfactants formed by chemically linking a polar headgroup to one of the methyl groups in *n*-dodecane and/or to the methyl branch of 6-methylundecane yield dramatically different surfactant aggregates and phase behavior.

To make the modeling of these complex mixtures more tractable, the compositional complexity is often reduced by the introduction of pseudo-components representing an entire group of chemical compounds. Even with this reduction in complexity, computational constraints still make use of equation-of-state–based approaches as the primary way of modeling phase behavior and transport properties of reservoir fluids. Again, exascale computing will enable researchers to tackle this complexity by allowing for increases in system size for MD and MC simulations using molecular mechanics force fields, and will also allow for scenarios where thousands of simulations are managed concurrently to probe variations in composition, temperature, pressure, shear, and other external stimuli. Predicting and understanding how their complex interactions govern performance and stability presents an exascale challenge.

3.4.2.3 Composite Polymeric Materials

Exascale computing will also enable large, heterogeneous, polymeric systems to be explored in greater detail. In particular, hydrogels — lightly crosslinked polymers in water — are important technologically in many medical applications (drug delivery, wound healing, contact lenses), environmental remediation (metal absorption, membrane separation), electrical storage, and conducting fabrics. These applications arise from the heterogeneous structure of the hydrogel, which can undergo large-scale responses to small perturbations in pH, salt concentration, or temperature. The length between crosslinks is quite long, and many such lengths are required in a simulation cell. As a consequence, billion-atom simulations with long run times are needed in many cases; that is, exascale computing is required. Many composite materials that have an intrinsic large separation distance require large simulation cells. For example, simulating carbon nanotubes in a polymer melt will require system sizes that are much longer than the carbon nanotubes themselves and the average separation between them. This simulation can only be executed using exascale resources.

The distinguishing phenomena for polymer melts or surfactant solutions from small-molecule liquids involve slow chain entanglement and aggregate size dynamics, as well as important collective and cooperative interactions. In order to be able to design functional polymeric

materials, it is critical to simulate entanglement dynamics — the dynamics that connect variations in structures — and their effect on polymer mobility and viscoelasticity. Similarly, the design of surfactant formulations requires the ability to simulate the formation and destruction of micellar aggregates. To treat polymer entanglements and aggregate size dynamics, atomistic simulations must be able to reach beyond the 1-µs timescale. Although much progress has been made in understanding entanglements using coarse-grained models, only a few atomistic simulations have been performed of entanglement dynamics because of the long timescale. Presently, 100-ns simulations are quite feasible using GPUs. With the further development of better parallelized accelerator architectures and codes, performing 1-µs simulations will enable the simulation of entanglement dynamics at the atomistic level. This advance will be transformative, because we will be able to connect the molecular architecture of a polymer with the fundamental dynamic process — entanglements — that governs the physical properties. Such simulations will enable the tuning and design of polymeric materials, which to date has not been possible. In addition, the processing of polymeric materials is controlled by entanglement dynamics. A robust understanding of entanglements would enable the development of better processing, which can save energy as well as produce better materials.

3.4.2.4 Rational Design of Polymer Dielectrics for Energy Storage

The demand for high-energy-density capacitors has increased in recent years, courtesy of the ongoing electrification of land (Nalwa 1999, Ennis et al. 2007) and sea (Ennis et al. 2009) transportation, as well as other military and civilian systems (Bluhm 2006). Whereas ceramics could conceivably be used as dielectrics in capacitive energy-storage applications, polymers provide a clear advantage as they display "graceful failure" at high electric fields. Because the energy stored in a capacitor is proportional to the dielectric constant and the square of the electric field, dielectric polymers of interest should display a high dielectric constant and high electrical-breakdown field. For example, biaxially oriented polypropylene (BOPP), with a high breakdown field of about 700 MV/m and a dielectric constant of about 2.2, is the current state-of-the-art polymer dielectric in high-energy-density (metalized) film capacitors. Attempts to improve upon BOPP, based on poly(vinylidene fluoride) (PVDF) and its copolymers, polymer nanocomposites, multilayers, and so on, have suffered from one weakness or another. Huan et al. (2016) have recently reviewed the history of capacitor materials, including recent advances and persistent challenges underlying new materials development.

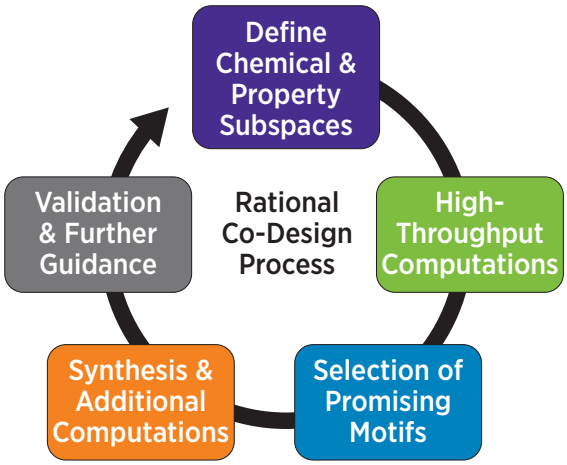


Figure 3-9. The Primary Steps Involved in a Rational Collaborative Design (or Co-design) Approach.

Clearly, strategies are needed to identify new promising polymer dielectrics. Given the vastness of the polymer chemical (and configurational) space, it is safe to assume that significant untapped opportunities exist, and that several new polymer dielectrics are waiting to be discovered. One way to approach this problem is to synergistically combine computational guidance with experiments in a "co-design" approach (Mannodi-Kanakkithodi et al. 2016). For the rational computation-guided co-design approach to work, the problem has to be amenable to rapid high-throughput computations. Moreover, it should be possible to specify the chemical subspace of interest clearly, and to state the (initial) screening criteria in terms of calculable properties. This technique can lead to a shortlist of potentially useful candidate materials. Only at this point are any benchtop experiments carried out, and attempts are made to produce the few selected materials, which can provide iterative feedback to the initial chemical subspace search step. A possible workflow that captures these notions is portrayed in Figure 3-9. This workflow has been utilized recently to design several new organic polymer

dielectrics within known generic polymer subclasses and to uncover the untapped potential inherent in entirely new and unanticipated chemical subspaces offered by organometallic polymers (Mannodi-Kanakkithodi et al. 2016).

Nevertheless, several challenges remain. How can researchers go about significantly expanding the chemical space (beyond the miniscule subspace explored thus far) without encountering the pervasive combinatorial explosion? More importantly, how can researchers account for the role played by the physical, or morphological, degrees of freedom? Data-driven approaches may be brought to bear to address the former aspect, and new methodological developments (e.g., efficient and high-fidelity force fields) will be required to address the latter. Finally, several other enormously important factors — such as the dielectric loss at application-relevant (ca. kHz) frequencies, morphological evolution and the progressive creation/dynamics of defects *in the presence of a persistent large electric field*, and dielectric breakdown — are largely "unsolved" problems. The hope for continued progress and ingenuity in this field will be determined largely by new methodology developments to address these unsolved problems based on synergies between advanced characterization and materials modeling efforts. Here, quantum molecular dynamics and Monte Carlo simulations will play an important role; however, widely used DFT methods do not describe the dielectric properties of polymers adequately. Thus, the above approaches need to be accompanied by developments in improved DFT methods such as computationally efficient treatment of exact exchange or self-interaction correction or in much more efficient wavefunction methods. Exascale computing will be required to enable the quantum molecular dynamics or Monte Carlo simulations to be performed in a timescale of use to experiment.

3.4.3 Cross-Cutting Research Directions

Particle-based simulation methods (MD and MC) are not specific to soft matter, and there are many aspects that apply to chemical systems and materials in general. For example, development of enhanced sampling algorithms and transferable force fields will have cross-cutting impact. Validation of force fields needs to involve experimental comparisons of thermodynamic, structural, and transport properties. However, not all soft matter phenomena can be captured using force fields and classical partition functions, and cross-cutting activities need to advance capabilities for molecular simulations by solving the electronic structure problem "on the fly" and treating nuclei as quantum particles. For example, interfaces often involve quantum and dynamical effects (e.g., van der Waals interactions and proton hopping) that require a better physical description than classical methods can provide.

3.4.3.1 Developing Polarizable and Reactive Potentials

Polarizable potentials are key to the development of quantitative models for aqueous systems and for solvent-free ions containing polymers, which are important for energy storage devices and many consumer products. However, including polarizable interactions adds expense to the simulation, which is why few treatments have occurred. In addition, while the first generation of polarizable force fields has been developed for biomolecular systems, a much smaller level of effort has been expended on materials systems. The physics requires improvements in the first-generation force fields for material systems leading to the development of more expensive but more accurate versions. However, accelerators have great potential to speed up the computation of more complicated potentials.

Presently, most atomistic soft matter simulations have a fixed topology; however, changing the topology by breaking and/or forming bonds is fundamental to many processes. One of the hallmarks of soft materials is their responsive behavior, and often this responsiveness manifests because of reactions within the system responding to external stimuli. In addition to treating nonequilibrium processes, treating reactions is critical. Thus, development and implementation

of reactive potentials, such as the ReaxFF potential in open source MD and MC codes, are major priorities. Related to polarizability and reactive potentials is treatment of pH in aqueous systems. Present implementations of pH are typically distinct from either of these two and need development for exascale. Therefore, polarizable reactive force fields are essential to future progress.

3.4.3.2 Developing Methods for Long-range Electrostatics for Exascale Architectures

Typically, the computationally most demanding part of MD and MC simulations is the treatment of long-range Coulombic electrostatic interactions between charged or partially charged atoms. The currently popular Particle-Mesh-Ewald (PME) method (Darden et al. 1993; Essmann et al. 1995) for long-range electrostatics relies on Fast Fourier Transforms (FFTs), which require expensive global communications. Extending PME from the petascale to the exascale will be challenging because of bottlenecks in FFT-based Poisson solvers (Pronk et al. 2013; Abraham et al. 2015; Brown et al. 2011, 2012). Alternatives to PME need to be explored. Possible methods include multigrid (Sagui and Darden 2001), multipole (Kurzak and Pettitt 2006), and multilevel summation (Hardy et al. 2015) methods. Fast electrostatic computations based on tree data structures have also been used widely for evaluating Coulomb matrices in quantum mechanical calculations (Challacombe, Schwegler, and Almlof 1996). Exascale will enable some simulations to treat systems larger than the Debye length, where long-range electrostatics will become screened. This capability suggests the possibility of developing a mathematically distinct method from those mentioned previously.

3.4.3.3 Developing a Framework for Multiscale Science

Many problems in soft matter require atomistic detail in one region but can become progressively more coarse-grained as the separation increases from this region. For example, interfaces are best treated using full atomistic detail within the interfacial region (potentially with both quantum and classical methods); and then at distances progressively farther from the interface, only coarse-grained detail is required. Other supramolecular systems will require varying levels of detail in different regions of space and possibly even varying in time. General multiscale frameworks and methods that couple two levels of treatment need to be developed with exascale architectures in mind.

In addition, new developments in coarse-grained models are essential to being able to treat the long time and large length scales inherent in soft materials. Recent work has developed more advanced coarse-grained models derived from the underlying atomistic model. These finer-grained models preserve important chemical details of the system and enable more predictive simulations with the coarse-grained simulations. However, the most optimal manner to map a group of atoms from the atomistic model to a particle in the coarse-grained model is not known. Furthermore, the conditions for transferability of the coarse-grained force field to other thermodynamic states needs to be better determined. With respect to replicating the underlying dynamics of the atomic system, there are known weaknesses in present methods as the local energy barriers are altered by the coarse-graining process and result in different dynamics. Developing coarse-graining methods that maintain the underlying atomistic dynamics while maintaining efficient functional evaluations is critical.

3.4.4 Computing Needs and Requirements

While the current multipetaflop/s computers have enabled 10^8–10^9-atom reactive molecular dynamics (RMD) (Shekhar et al. 2013; Nomura et al. 2016), and 10^4-atom QMD simulations (Shimamura et al. 2014), their length (10^{-7} m) and time (10^{-10} s) scales need to be extended by several decades on exascale computers. This need will require computational approaches that will continue to scale and achieve portable performance on current and future computer architectures, that is, simultaneously achieving nearly perfect speedup on billions of cores and a high percentage of peak floating-point performance within each core. A promising approach is to use globally scalable and locally fast solvers based on the global-local separation concept (Romero et al. 2015). Another requirement is the reproducibility of results for billion-way parallelism. This need is a challenge since floating-point arithmetic is "nonassociative" due to rounding error, and accordingly sums become a random walk across the space of possible rounding errors (Figure 3-10) (Small et al. 2016). Most importantly, exascale simulations need to be seamlessly integrated with exabyte big data analytics and mining so as to extract maximal scientific knowledge. This need will require the merger of hardware and software stacks between exaflop/s high-performance computing and exabyte big data (Reed and Dongarra 2015), as was promoted by the recent executive order by President Obama (2015) on the National Strategic Computing Initiative. At the application level, this need will require the development of *in situ* data analytics approaches (Romero et al. 2015) based on scientific "discovery informatics" (Gil et al. 2014).

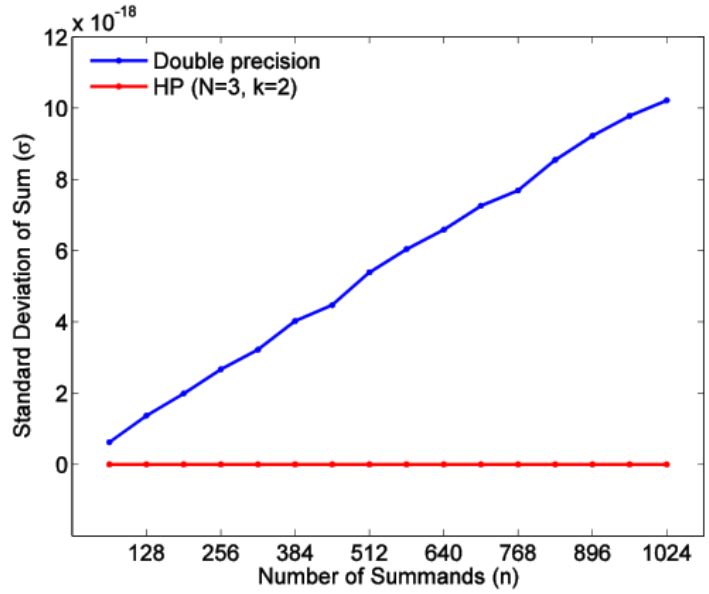


Figure 3-10. Observed Standard Deviation for the Sum of Sets of *n* Random Numbers. A standard double-precision arithmetic result (blue) is compared with that of an order-invariant sum method named HP (Source: Small et al. 2016).

One of the generic computing needs of soft matter simulations is reaching long timescales, and in terms of computing, this means long runs. Week-long runs are typical for many soft matter simulations. The very short queue times presently used on the DOE computer centers is a major deterrent to conducting soft matter research. The ability to continue jobs automatically through the submission scripts alleviates only a small bit of the difficulty of performing full simulations. Many shorts runs increase the likelihood of damaged output files and require more sophisticated analysis codes to deal with the increased number of files to be processed. This difficulty results in writing analysis codes to deal with the output situation instead of writing code to advance the science or the exascale computing. Details matter and accounting for them is a progress-limiting factor.

One of the changes with the advent of exascale computing is that the amount of generated data will be so large that post-simulation analysis of, for example, MD and MC configuration files will *require* parallel computations. In the past, these computations have typically been performed on serial computers, or researchers used simple parallelization that will no longer be practical. In some cases, where parallel efficiency is high, these computations can become part of the simulation and no longer a post-simulation-only task. Realizing this capability does mean that routines to perform these calculations must be part of the standard packages. However, not all computations belong on exascale computers, and there will be a need to parallelize many analysis codes so that they can run on smaller-scale computers.

Within the next 10 years, we anticipate the development of hybrid multiscale predictive simulations of soft matter that provide quantitative and detailed information on length scales ranging from 10^{-10} to 10^{-4} m (length of chemical bonds to size of systems containing 10^{+12} atoms) and timescales ranging from 10^{-14} to 10^{+1} s (bond vibrations to self-assembly of complex aggregates). This capability represents, of course, a long-standing goal of chemistry, materials theory, and condensed matter physics; however, conceptual developments over the last decade suggest that breakthroughs are imminent. Translating the conceptual progress into working simulations requires vast increases in computational power with concomitant development of the computational ecosystem: libraries, algorithms, workflows, and programmer and user training. While the goals are ambitious, the rewards in terms of developing soft matter substances with functional attributes to meet new societal needs are promising.

BES Path to Exascale in Soft Matter

- **Today: MD and MC simulations of simplified heterogeneous systems accessing sizes of up to 10^{-8} m and lengths of time of up to 10^{-7} s.**
- **2020: Use of polarization, reactive force fields for MD and MC simulations of thermochemical, structural, and short-timescale dynamical properties in heterogeneous systems.**
- **2025: Hybrid multiscale predictive simulations that provide quantitative detailed information on length scales of 10^{-10} to 10^{-4} m and timescales ranging from 10^{-14} to 10^{+1} s for complex heterogeneous systems.**

3.5 Advances in Algorithms for Quantum Systems

3.5.1 Challenges and Opportunities

Quantum mechanics (QM) touches virtually all of computational chemistry, biology, physics, and materials, either directly or indirectly and including science areas that are in other sections of this report. Popular methods such as Hartree-Fock (HF), DFT, second-order perturbation theory (MP2), coupled-cluster theory (CC), multireference methods and QMC are all examples of widely used quantum approaches. In addition, most modern classical force fields are derived entirely or partially from quantum mechanics. The bottleneck in applying the quantum methods noted above to important problems in areas such as catalysis, photochemistry/photobiology, and mesoscale simulations is the high computational cost of scaling the quantum methods with system size, especially the methods that account for electron correlation like MP2, CC, QMC, and multireference approaches. Consequently, in order to facilitate the use of these high-level QM methods for (for example) MD simulations, nonadiabatic dynamics to describe photophysical phenomena, the study of heterogeneous catalysis, the investigation of quantum materials, or the investigation of processes at the mesoscale, it is essential to reduce the scaling of these methods with system size and to develop multilevel parallel algorithms for them.

True predictive power for these complex problems will require development of robust hierarchical theories and algorithms to treat electron correlations across all relevant length scales.

Implicit in many of the BES mission phenomena is the need to develop truly multiscale methods that can span multiple time and length scales in a seamless and self-consistent manner. For example, substantive improvements in the treatment of electron correlation in model systems has recently been achieved through the development of new algorithms (greatly improved solvers for dynamical mean-field theory [DMFT] methods and their extensions), aided by the adoption of agile, open-source software practices. However, due to their high cost, direct application to complex materials and chemical problems will remain out of reach in the foreseeable future. True predictive power for these complex problems will require development of robust hierarchical theories and algorithms to treat electron correlations across all relevant length scales.

Furthermore, developing new algorithms as discussed in the previous paragraphs will require the collaboration of application developers with applied mathematicians and computer scientists and engineers, as exemplified by the SciDAC (Scientific Discovery through Advanced Computing) program. As discussed in Section 3.8, Next-Generation Workforce, investments must be made in human resources for emerging scientists to integrate expertise in a chosen field such as chemistry or physics with applied mathematics and computer science to advance the field.

3.5.2 Priority Research Directions

Methodological developments are required for all these algorithms. As discussed in context in the other sections of this report, the substantial challenges that remain in the underlying methods used for quantum simulations will not be solved by purely computational advances. Examples of research directions include the following:

- The treatment of strong electron correlation for ground and (especially) excited states for complex or extended systems is underdeveloped: thus, in extended systems, chemical accuracy has only recently been achieved for the simplest of materials, while the optical equivalent for excited states has not been reached. For molecular systems, the treatment of highly multiconfigurational electronic structures remains a significant challenge.

White papers by the following authors informed the writing of this section and can be found in Appendix C, Section C.5, starting on page C-80:

- E. Carter
- K. Kowalski
- M. Jarrell

Similarly, these relevant case studies can be found in Appendix D, Section D.5, starting on page D-65 by:

- K. Kowalski
- A.M.N. Niklasson

- Related problems exist for model (non-*ab initio*) Hamiltonians and electronic phase diagrams.
- New quantum methods are needed that will have significantly reduced computational costs that are more in line with the cost of density functional theory-type methods and realize the accuracy of multi-configurational methods.
- Theories are needed that incorporate relativistic components and the correlated electron behavior arising from partially filled *d* and *f* shells.
- There needs to be a systematic improvement of errors associated with numerical representations, for example, explicit correlation.
- Methods should be developed that include electronic dynamics such as in TDDFT, DMRG, DMFT, GW, QMC, and emerging technologies.
- The community also needs methods that model new spectroscopies, such as those that include core-level excitations and time resolution, in the time frame of the experiment.

Overall, the problems listed above limit the attainable accuracy of quantum methods regardless of computational cost. Except for niche areas, the accuracies achievable today limit the possibilities for chemical and materials design and for speeding up the development process.

Except for niche areas, the accuracies achievable today limit the possibilities for chemical and materials design and for speeding up the development process.

Multiscale calculations, in both time and length, are essential for many areas of chemistry, materials, physics, and biology. Examples are heterogeneous catalysis in solvents, photosynthesis, and the growth of films on surfaces. The most common approach to the theory and computation of multiscale phenomena proceeds in isolated steps. First, a short timescale/small-length-scale calculation is performed with a high level of electronic structure theory, such as coupled cluster theory or multireference perturbation theory. The output of this first calculation (e.g., energies, gradients, vibrational frequencies) is then passed as input parameters to a typically lower level of theory (e.g., a classical force field or a lower level of electronic structure theory) that is capable of treating larger length scales and/or longer timescales. In particular, classical MD currently consumes many computational resources, and this consumption will continue to grow with the availability of exascale computers. However, reactive force fields and those that are parameterized for nonequilibrium properties will be more important. While such sequences of calculations may be useful, they are not true multiscale solutions given that the multiple time and length problems are not fully integrated. In particular, robust, low-scaling methods for solving the different scales in a self-consistent manner have yet to be developed. Such methods are essential for the accurate study of the dynamics of multiscale problems. Examples of the research directions required on this front include:

- Systematic, self-consistent frameworks to couple accurate electronic structure, classical mechanics, and continuum presentations with statistical mechanics methods and kinetics for large, complex systems.
- Force fields from many-body approaches to enable realistic MD.
- Coupling of strong and weak, anharmonic interactions as collective phenomena.
- Advanced sampling methods to sample important and/or rare-event parts of the phase space for large systems.
- Parallel-in-time methods.

Methods such as those described in the preceding paragraph will be very demanding of computational resources. It is not obvious *a priori* that they will be able to take advantage of the petascale systems that already exist, not to mention the planned exascale computers. To take advantage of such computer systems, it is necessary to develop highly parallel, low-scaling

algorithms for the employed multiscale methods. In particular, since computer configurations have ever-increasing numbers of computational elements on each node, the newly designed algorithms must be able to take advantage of multilevel parallelism with multiple layers of memory and communication hierarchies.

3.5.3 Cross-Cutting Directions

Effective Use of New Architectures

A key requirement of the exascale era is to develop portable, performant, scalable, and sustainable implementations of scientific algorithms. These algorithms challenge not only the system software and languages, but also each of the application domains. In most applications, the algorithms and their implementations are strongly mixed, which limits the ability for transformational approaches. Techniques for implementing transformational changes are either not well developed or not broadly communicated, and restricted to narrow application domains. Methods to achieve flexibility in execution and data layouts are required. For example, in the space of density functional theory-based calculations, many extant codes use a single set of data layouts for all calculations, when additional flexibility would enable both greater performance for individual calculations and unlock greater performance portability.

Today, only a tiny percentage, if any, of the major applications implementing state-of-the-art algorithms are ready for exascale architectures. The complexity and resource requirements of effectively exploiting even current petascale architectures is already causing research groups to choose between implementing new algorithms and functionality and supporting multiple architectures. Improved standards-based languages, tools, appropriate training, and programming support are required to facilitate implementation and support on all architectures with high performance. Projects such as Kokkos (GitHub 2016; SNL 2015), which has demonstrated portable high performance for classical MD, are promising but must become standardized and supported in order to attract users and avoid technical dead ends.

Broadly stated, in 10 years, applications should not be written as they are now. The increasing complexity of computational nodes requires adoption of runtime systems that can hide the details from the typical applications programmer. If these runtime systems are not developed or adopted, substantially greater programmer effort will be required, and simulation software that is more fragile overall will result, or will even fail to be deployed. Gains from techniques such as autotuning and adaptive algorithms are expected to be much greater in the exascale era. An important start in this direction is the development and adoption of domain-specific toolkits such as those developed by the many-body chemistry and condensed matter theory community, for example, tensor contraction libraries for chemistry and solvers for model Hamiltonians in condensed matter theory (e.g., TRIQS, ALPS).

3.5.4 Computing Needs and Requirements

QM-based algorithms and applications have a long and successful history of effective utilization of new generations of supercomputers. The first teraflop and first petaflop-scale calculations were DOE–supported, QM-based applications. These won the Institute of Electrical and Electronics Engineers (IEEE) Gordon Bell prizes for fastest scientific application in 1998 (Locally-Self-consistent Multiple-Scattering [LSMS]) and 2008 (DCA++), respectively. Both applications were not “one offs” crafted for peak performance but later were used productively for research. To extend this path to the exascale era and to address the discontinuous changes in the computational architectures will require significant developments in the computing and software ecosystems, as well as commitments to address the education, development, and availability of a skilled workforce.

One of the greatest challenges to application performance in the exascale era is the adoption of increasingly deep and complex memory hierarchies. Effective data placement and use of appropriate memory hierarchy-aware algorithms will enable order-of-magnitude improvements in effectiveness.

Addressing the pace of hardware change is proving challenging to applications teams, even in the 10-petaflop era. To make the best use of their limited resources, many application teams are developing codes for specific architectures. In most cases, there is a substantial investment in effort required to develop or port a code to both CPU and GPU architectures, for example; and with limited resources, this investment must be traded off against the implementation of new theories and algorithms that would immediately lead to new science. This approach inevitably postpones the re-architecting of applications for *all* exascale architectures. These barriers must be reduced by taking action at all levels of the software stack: by development and education of practical performance portable languages; by porting of standard library functionality to new architectures; and by identifying and extracting common domain-specific functionality to new libraries or computational frameworks. Importantly, mechanisms to communicate these developments to the application teams should be developed. To ensure broad relevance, effectiveness, developer acceptance, and ready identification of deliverables, we recommend that these efforts should be supported by competitively awarded, combined theory development and software development centers that, already at the proposal stage, identify and work with multiple applications within one or more related domains. Utility for multiple applications are essential to avoid customized solutions that are not general solutions for the community. Assessment criteria should include the extent to which the aforementioned challenges will be addressed and improved software instantiated in production domain science codes.

Even in the exascale era, research teams will still have production computing workloads that will require running jobs at all scales. Many important scientific problems are addressed using codes and computational methods that will not initially scale well to the full size of an exascale system, but are nonetheless crucial to the success of BES research programs and user facilities. In addition, runs of moderate size are needed for code and algorithm development, debugging, parameter studies, code validation, and scientific validation of results. These workflows need to run at all ASCR facilities. Making effective use of all the available computational resources requires common middleware — compilers, schedulers, interfaces, libraries, and access mechanisms — across the facilities. Currently, the centers have different access mechanisms and scheduling software, adding to developer cost and requiring additional implementation in workflow automation tools.

Broad adoption of multiscale approaches also increases demands on facilities. At the facility level, techniques for coscheduling different calculations are required: exascale calculations are envisaged to depend on many prior tera- and petascale calculations. These must be run either locally or remotely, with appropriate data transfers scheduled between facilities.

Access to the increasing range of computational scales may benefit from more flexible allocation mechanisms. Currently, the ASCR Leadership Computing Challenge (ALCC) and INCITE award programs have annual periods, while “director’s discretionary” allocations at each center are perceived to be targeted primarily toward ALCC and INCITE readiness, and not to bursts of scientific activity. Adoption of a “rapid scientific response” mode will enable pressing scientific problems to be addressed and new techniques to be applied on more suitable timescales.

Despite the availability and potential advantages of advanced software runtime systems, domain-specific languages (Fowler 2010), and execution models (e.g., the POSIX Threads library) most quantum simulations are still programmed traditionally using explicit parallelism via MPI, OpenMP, or a combination of both. These newer runtime systems are expected to reduce demand on the application programmer and are better suited to achieving performance portability than legacy approaches such as MPI. However, currently available runtime systems are largely perceived as not being production-level tools. They are also not directly deployed or supported by the computational

facilities, and there can be a lag before new architectures are supported, limiting their uptake. Identification and support of a very few runtime systems and standardized interfaces as part of the computing ecosystem will help applications reach higher performance in a more portable manner and enable developers to adopt a more sustainable development strategy.

One of the greatest challenges to application performance in the exascale era is the adoption of increasingly deep and complex memory hierarchies. Effective data placement and use of appropriate memory hierarchy-aware algorithms will enable order-of-magnitude improvements in effectiveness. This sensitivity is already evident in today's GPU architectures. While some current algorithms have a clear mapping to complex hierarchies, many do not. In addition to developing appropriate language features and libraries that can exploit these features, development of effective and practical tools to characterize the performance of real-world applications is essential. Sufficiently performance-portable and expressive programming models are urgently required; today these are at best developmental research projects.

BES Path to Exascale in Algorithms for Quantum Systems

- Today: Implement sequences of calculations in isolated steps to incorporate theory and computation of multiscale phenomena, but which are not true multiscale solutions.
- Perform QM calculations that rarely scale to petascale and have high computational and system resource costs.
- 2020: Develop highly parallel, low-scaling QM algorithms.
- Couple the QM with multiscale methods in a systematic framework with a view to taking advantage of future exascale systems.
- Establish combined theory development and software development centers that, already at the proposal stage, identify and work with multiple applications within one or more related domains.
- 2025: Support the emergence of a whole new way of writing applications, such that runtime systems hide details from the typical applications programmer, given the increasing complexity of computational nodes.

3.6 Computing and Data Challenges @ BES Facilities

3.6.1 Challenges and Opportunities

The BES User Facilities operate more than 240 different instrument types that enable the scientific discoveries of their user communities. BES User Facilities provide photons, neutrons, or electrons and the means to manipulate and detect those particles after interaction with user samples. The heterogeneity of instrumentation contributes to the richness of the capabilities of the BES User Facilities and to the breadth of the user base. It also results in the need for many different data acquisition and analysis methodologies. In the past, and to some extent still today, detectors and end stations generated datasets that were relatively small and readily transferred to users' portable storage devices and to be taken to their home institutions for analysis. The availability of high-performance networking facilitated this transfer to a user's home institution. Once the user took possession of the data, the operational responsibility of the BES User Facilities to the user ended.

In recent years, the introduction of rapid and high-performance detectors at some beamlines has led to changes in users' needs. These detectors routinely produce upwards of 10^2 to 10^3 frames/second, with some to become capable of performing at 10^4 to 10^5 frames per second within the next 5 years — and detector performance will only continue to accelerate. As a result, user datasets have started growing such that users can no longer readily transfer data to their home institutions. Multiple 10- to 100-GB datasets per shift are being produced by some tomography and scattering instruments, and tens of terabytes from the newest electron microscopes. In a three-year "pilot" period ending December 2015 at one BES User Facility, the data from two beamlines have produced more than 244,000 datasets amounting to 1.7 PB and have launched >3.5 million jobs at NERSC. In another example, a single scattering station has occupied OLCF's entire GPU resource (on Titan). We expect only more of these kinds of demands going forward, and the facilities and their users are either not prepared or are only minimally prepared to address these data-intense environments. Going forward, users cannot be expected to manage and analyze such data volumes and computations on their own. Experiments have become so complex that users require expert data science and mathematics help to develop suitable analysis methods.

Accelerating detector performance has led to other challenges, such as the need to perform:

- Streaming analysis to enable experimental steering and decision making during the experiment to optimize the scientific outcome.
- Multimodal analysis of concomitant experiments.
- Numerical modeling and simulation.
- Background knowledge from curated archives to drive new scientific discoveries.

White papers by the following authors informed the writing of this section and can be found in Appendix C, Section C.6, starting on page C-90:

- M. Chan, P. Darancet, S. Gray, I. McNulty, S. Sankanarayanan, and M. Sternberg
- A. Hexemer and D. Parkinson
- A. Perazzo
- M. Stevens
- B. Toby

Similarly, these relevant case studies can be found in Appendix D, Section D.6, starting on page D-73 by:

- D.Y. Parkinson, A. Hexemer, and C.E. Tull
- J. Thayer and A. Perazzo

Each of these approaches will require significant "on-demand" exascale types of computational resources to deliver the necessary feedback and insights in real time as the experimental process unfolds.

The heterogeneous nature of instruments at BES User facilities means that not one solution will be sufficient to address all of the data challenges. Some instruments may lag in the requirement for data-intense management and analysis. In those cases, "beamside" computation may suffice, particularly if

affordable and easily managed petascale "boxes" become available, together with higher-speed networking architectures. Such systems must, however, be "user friendly." The growing complexity of the analysis process (mixing fast data analysis and numerical modeling) at other instruments will require capabilities beyond the petascale-level capabilities that the ASCR facilities offer. Addressing these challenges in support of users to facilitate and provide for the analysis, management and storage of the data signals a fundamental change in the operation and responsibility of BES User Facilities toward users.

In addition, the computational capabilities of the future will provide a platform for real-time modeling and simulation so that experiments can be augmented and understood as they are in progress. This coupling, along with the need to manage data and the ability to steer and make decisions during the experiment to optimize the scientific outcomes, will require significant "on-demand" exascale types of computational resources to deliver the necessary feedback and insights in real time as the experimental process unfolds.

BES User Facilities provide photons, neutrons, or electrons and the means to manipulate and detect those particles after interaction with user samples. The heterogeneity of instrumentation contributes to the richness of the capabilities of the BES User Facilities and to the breadth of the user base. It also results in the need for many different data acquisition and analysis methodologies.

As users cannot be expected to "go this course alone," BES User Facilities must partner with the resources of ASCR facilities and research to make the process more efficient. Without such partnerships, in fact, the operational budgets of BES User Facilities will soar, as additional staff will need to be hired who are facile with high-performance computing, data management, and analysis. The commissioning, maintenance, and operation of high-performance computers and storage systems are not trivial and thus are not a simple addition to existing BES User Facilities staff members' responsibilities. In an attempt to achieve these partnerships, BES User Facilities have begun to pilot, where appropriate, the interaction with ASCR User Facilities, applied math resources, and computer scientists to establish initial working relationships that can produce tangible improvements to the experience of the user base of BES User Facilities. Partnering with ASCR Facilities will also allow BES User Facilities to explore ways to collaborate more closely and efficiently on the development and sharing of analytical tools because each BES Facility would not have to develop its own suite of tools. Cutting-edge computing, data management, and computation are necessities to achieve world-class science. Additional resources from both BES and ASCR will be required, in a partnership, to prepare the BES User Facilities to support the world-class science endeavors of its users.

3.6.2 Priority Directions in User Support

During the Exascale Requirements Review, three topics dominated the discussions: streaming analysis and steering of experiments, multimodal analysis of results from different instruments, and long-term data curation.

3.6.2.1 Streaming Analysis

In time-resolved (e.g., *in situ* or *in operando*) or time-sensitive experiments, scientists need to make critical decisions based on the:

- Quality assessment of the data being recorded,
- State of the sample and/or process,
- Statistics of exposure, and
- Overall scientific goal of the experiment.

Against a background of growing data volumes and more complex experimental setups, these tasks can no longer be performed manually in a time-effective manner. Streaming analysis that enables steering by the scientists and/or that provides immediate feedback to the data acquisition (DAQ) systems and/or beamline and instrument control systems can enable scientific experiments and investigations that are impossible with current, manual approaches. Examples of such experiments include those where:

- Samples are quickly destroyed by the X-ray or electron probe, or
- Transient events must be captured, such as during
  - Crack propagation,
  - Crystallization, or
  - Fluid flow.

In addition, now-novel approaches like running a full simulation of an experiment, running a "digital twin" to the experiment, or performing multimodal data collection will become routine with exascale computing and new sources. Such parallel investigations will predict interesting features to investigate or guide control of the experiment. This capability can be useful when trying to measure rare and hard-to-capture time points or regions of interest.

3.6.2.1.1 Driving Use Cases

For advanced spectroscopies, validation of concepts and techniques by comparison to experiment is essential. The new spectroscopies — including high resolution RIXS, X-ray Raman, and pump-probe techniques for nonequilibrium studies — are under development at national synchrotron radiation facilities and, along with striking improvements in neutron scattering and X-ray photoemission spectroscopy, are immensely powerful probes of chemical and material properties, providing unprecedentedly detailed views of atoms and complex materials in motion at meV energy resolution, nanoscale distances, and fs timescales. An example of these vastly improved capabilities are XFEL machines like the LCLS and LCLS-II that generate molecular movies of catalytic processes. These experiments can be analyzed over about 10–20 ps using approximate models limited to about 100 atoms. Methods exist that can vastly improve scientists' understanding of experiments, including time-dependent density functional theory, many body Green's function (DMFT and beyond), and the GW and BSE — but exascale resources will be needed to exploit their potential.

Another example is ptychography, in which a series of X-ray scattering patterns are collected while a sample is scanned across the X-ray spot (Figure 3-11). Ptychographic reconstruction, which is computationally intensive, must be carried out to yield a final real-space image; however, without real-time reconstruction, researchers are "flying blind" in their experiments, having little concrete information about the exact state of their sample.

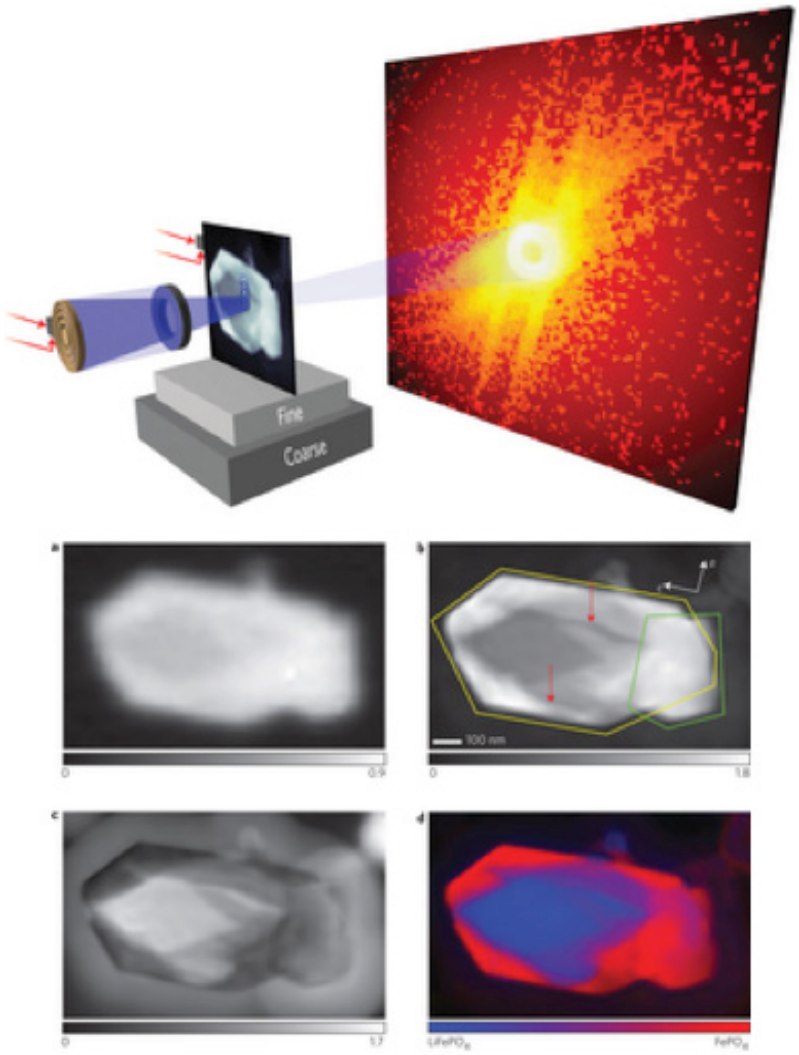


Figure 3-11. Chemical Composition Mapping with Nanometer Resolution by Soft X-ray Microscopy. a, b, Optical density maps from conventional scanning transmission X-ray microscopy (STXM) (a) and ptychographic (b) imaging at an energy that presents maximum contrast between the two chemical components present. The resolution of the STXM is not adequate to visualize cracks in the particle, indicated by red arrows. The crack is only on the surface and reveals an unreacted domain below. The pixel size is 4.0 nm and scale bar is 100 nm. (c) Phase contrast from the ptychographic reconstruction showing maximum relative phase shift between chemical components. (d) Colorized composition map calculated from the full complex refractive index. Only material containing the two chemical components is apparent (Source: Shapiro et al. 2014).

3.6.2.1.2 State of the Art

There are two distinct types of streaming analysis workflows: (1) the static analytical pipeline, created once and then run many times without human input; and (2) the adaptive, hypothesis-driven steering of an experimental process by scientists informed by streaming data analysis and interpretation support. For the former, many social media and tech-oriented companies (e.g., LinkedIn, Facebook, Google, Paypal, IBM, GridGain, etc.) have developed a wide range of infrastructure tools and mathematical libraries that enable users to create such streaming workflows. This type of pipeline will work well in situations where there are agreed-upon community standards and methods of experimental analysis (e.g., in X-ray crystallography). More commonly

needed in science, however, are adaptive processes that can be tailored to the requirements of a specific experiment to enable scientists to prove or disprove their hypotheses effectively and to maximize the science outcomes from their experiments. Examples of techniques for which the adaptive workflow is needed include soft X-ray small-angle X-ray scattering (SAXS)/wide-angle X-ray scattering (WAXS)/grazing-incidence small-angle X-ray scattering (GISAXS); micro-tomography; and hard X-ray nano-crystallography: use of these techniques involves sufficiently high data rates; compute-intensive analysis; and diversity of sample, experimental technique, and novel experiments. Nevertheless, the very heterogeneity of these techniques makes standardizing real-time feedback difficult, and therefore "one-size-fits-all" solutions are next to impossible to develop. Initial efforts are under way to provide some customized real-time feedback at light source beamlines. Moving forward, it will be critical to develop approaches that enable scientists to steer the analytical and data taking process adaptively so as to optimize their scientific environments.

These analytical pipelines incorporate many different components, from the initial mathematical algorithms, to the data and information visualization, to the event-processing framework, all of which involve coordinating workflow and feedback processes to the instrument. All of these have to work together effectively to create a high-performance streaming analysis environment that delivers its results in a timely, reliable, and accurate manner. Many of the underpinning research and challenges, in particular in extreme-scale data environments, have been discussed in a number of recent workshops (i.e., the DOE Accelerated Scientific Knowledge Discovery, DOE ASCR Storage Systems and I/O Workshop, DOE ASCR The Future of Scientific Workflows, DOE ASCR and NSF STREAMS15).

3.6.2.1.3 Computing Needs and Requirements

Requirements in 5 years:

- Streaming analysis on high-data-volume beamlines allowing scientists to see high-quality results on timescales sufficient to influence experimental decisions at the beamline or end station.
- A structured engagement allowing computer science and mathematics experts (e.g., a dedicated SciDAC Institute) to work with BES User Facilities to create fully functional and performant analysis pipelines including algorithms, adaptive work flows, streaming data and information visualization, provenance, effective programming pattern, performance assessment, and optimization.
- Automated workflow scheduling systems that enable the effective placement of workflow tasks to minimize and meet demanding response time requirements of running experiments. This capability would include advice on optimal placement and use of compute and networking resources at specific experiments (e.g., the use of petascale systems near the beamline), depending on their analysis pipeline characteristics (e.g., placement of additional computing resources at the instrument to handle the initial data reduction or compression).
- Real-time visualization of results that are viewable at the experimental end station and available/accessible to collaborators at remote locations.

Requirements in 10 years:

- Streaming analysis feeding automated feedback systems that steer DAQ and instrument control systems to facilitate high-quality, time-resolved experiments.
- Streaming analysis and decision support systems that enable scientists to steer their analysis and data taking adaptively and thus optimize scientific outcomes.
- Digital twins integrated with streaming analysis and potentially with DAQ and control systems to steer experiments to temporal, spatial, or parameter regions of maximum scientific interest.

3.6.2.2 Multimodal Analysis

Understanding complex material problems often requires the use of multiple probes in different locations. Today, many researchers already use multiple beamlines at multiple user facilities around the world (Figure 3-12); however, the integrative analysis of such results remains difficult and is dominated by either ad hoc or nonscalable solutions. True multimodal analysis will require that user teams analyze identical samples at different BES User Facilities. Moreover, information from each facility may be needed at the other facility to steer the experiments in real time. The emerging examples of this type of analysis are digital twins, where full simulations of an experiment are running in step with the experiment. Such parallel investigations will predict interesting features to investigate or guide control of the experiment. This capability can be useful when trying to measure rare and hard-to-capture time points or regions of interest. Not only will there be real-time demands on the ASCR User Facility to support analysis at both BES User Facilities; but on-demand, high-performance networking will be required to enable this inter-facility operation. In addition, new analytical approaches need to be developed that enable the easy integration, correlation, and comparative analysis of these disparate results, potentially across different user facilities if the datasets are large. It is expected that simultaneous simulations will play a major role in this type of experiment, and their results need to be included in the analysis in the same way as experimental results. In the future, the meeting participants could even foresee the need to execute such experiments contemporaneously at different instruments and facilities.

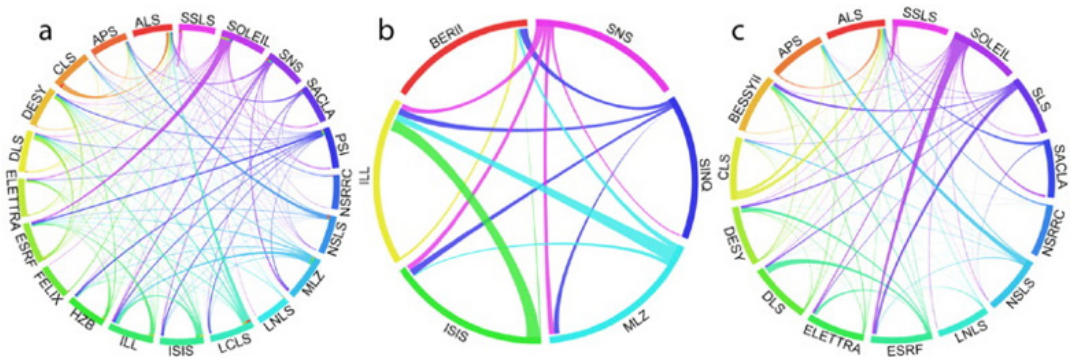


Figure 3-12. Common Publications (DOI): (a) across all facilities; (b) for neutron sources; (c) for synchrotrons (Source: Research Data Alliance Photon and Neutron Science Interest Group 2015).

3.6.2.2.1 Driving Use Cases

In heterogeneous catalysis, metallic nanoparticles are supported on substrates, and the interaction between the metal and support can be critical in their overall function (Cargnello et al. 2013). During catalysis, an incoming gas stream is converted from one molecular state to another, mediated by the interaction of the molecule with the catalyst and support. The goal of heterogeneous catalysis is to enhance the rate of reaction and the selectivity of the reaction or to create catalysts that function as well as existing catalysts but with cheaper materials. In order to understand how catalysts function, it is necessary to measure multiple aspects of the system, using combinations of photons, electrons, and neutrons. These include physical structure (through diffraction, imaging, and spectroscopy), electronic structure (spectroscopy), and gas-characteristics (spectroscopy and gas chromatography), all when the catalyst is in a reaction environment (i.e., at elevated temperature and pressure). At present, aspects of these measurements are made at multiple user facilities, but rarely are they intimately coordinated owing to a lack of integrated analysis modalities. It would be transformative to the field if users, on the same catalyst, could (1) perform all of these measurements in operando conditions (Li et al. 2015) in a directly coordinated fashion; (2) analyze and correlate the data in near real time from multiple imaging

modalities; (3) match this with computational and theoretical insights from exascale computation, again in near real time; and (4) use this information to improve the reactivity or selectivity of the catalyst through autonomous experimentation (Matsumaru 2004).

3.6.2.2.2 State of the Art

Traditionally, datasets have been small enough so that the relevant data from the current and prior experiments can be easily stored on a user's hard drive. In most cases, the analysis of multiple datasets is actually not carried out in an integrated way but is sequential, with the user determining whether a given model is consistent with all experimental observations rather than truly co-analyzing all of the results because of a lack of algorithms and tools.

3.6.2.2.3 Computing Needs and Requirements

Requirements in 5 years:

- Ability to run digital twin experiments routinely, including with comparative analytical capabilities during the experiment. This capability will include multisource streaming analysis in distributed computing environments.
- Access to experimental, computational, and network resources co-scheduled across different ASCR and BES facilities.
- Structured support from computer science and mathematics experts (e.g., a dedicated SciDAC-like Institute) for BES facilities to create fully functional and performant multisource analysis pipelines, including algorithms, adaptive workflows, streaming data and information visualization, provenance, effective programing pattern, performance assessment, and optimization.
- Automated workflow scheduling systems that enable the effective placement of workflow tasks to minimize and meet demanding response time requirements of running experiments. This capability would include advice on optimal placement and use of compute and networking resources at specific experiments, depending on their multisource analysis pipeline characteristics (e.g., optimized distribution of workloads depending on data source locations and response time requirements).
- Real-time, streaming visualization of multisource results that are viewable at the experimental end station and available/accessible to collaborators at remote locations.

Requirement in 10 years:

- The necessary tools and infrastructure to elevate multimodal data analysis from "one-off" studies to routine and rigorous projects that pull full significance from data. This capability will require a large-scale framework development effort that allows researchers in different experimental domains to develop algorithms and codes that can be combined to fit a single, comprehensive model to all data collected on sets of related samples. Such work must anticipate current and future platforms to allow for the computational demands.

3.6.2.3 Data Curation

In 1991, the International Union for Crystallography (IUCr) developed a then-state-of-the-art data format for exchange of data and results (Hall et al. 1991). Over the next decade, a data dictionary was developed that carefully defined fields for specification of raw data, metadata, and limited amounts of provenance first for small-molecule, single-crystal diffraction (Brown and McMahon 2002) and later for areas such as macromolecular and powder diffraction crystallography (Hall and McMahon [eds.] 2005), and work continues in areas such as magnetism and superspace structure definitions. During the same period, the IUCr encouraged software developers to support this format and made the use of this format a prerequisite for publication in IUCr journals. The

result has been that all crystallographic results, comprising more than 500,000 known structures and five different databases, are available in a single widely implemented format, enabling projects such as the Materials Genome Project. A "CIF" has become a slang synonym for a crystal structure result. Raw data can be imported into analysis programs directly and automatically from journals for alternate hypothesis testing. This level of data sharing and mining is needed to make full use of the data produced by the nation's user facilities; however, it should be noted that enabling this functionality is more than a simple file format issue: this level of curation and reuse will not be possible without a similar large-scale effort.

3.6.2.3.1 Driving Use Cases

In 2013, the Office of Science and Technology Policy (OSTP) released a directive on *Increasing Access to the Results of Federally Funded Scientific Research* (OSTP 2013). Its implementation has been linked to providing sufficient supporting information with scholarly publications to enable their validation. To enable compliance, scientists need selective curation — the ability to easily select suitable raw data, analyzed data, metadata, provenance, and tools from their scientific work that led to the insights in the paper in question. As the crystallography case shows, the results of a scientific experiment can be the basis of future research, either as a direct starting point, or as a basis to identify new research directions. To facilitate this reuse, the data have to be sufficiently standardized, documented, discoverable, and immediately usable. The more complex the data collection and analysis process, the more help scientists will need to leverage the results for reuse easily. Furthermore, with large data volumes, fast and targeted discovery methods are vital. New analytical methods, such as the streaming analysis and decision making required to steer complex experiments to more optimized science outcomes, will increasingly rely on fast access to context-relevant data to support scientists in decisions about whether observations are new or how they are different from prior results. Given the time-critical nature of the decisions that the scientists need to make, fast discovery and extraction of relevant information and delivery are essential. Science relies on scientific reproducibility to validate scientific discoveries, so in a world with exponential growth in scholarly publications, data tools are needed not only to identify research that has produced similar insights but also to facilitate comparison and contrast of those results.

It would be transformative to the field if users, on the same catalyst, could (1) perform all of these measurements in operando conditions (Li et al. 2015) in a directly coordinated fashion; (2) analyze and correlate the data in near real time from multiple imaging modalities; (3) match this with computational and theoretical insights from exascale computation, again in near real time; and (4) use this information to improve the reactivity or selectivity of the catalyst through autonomous experimentation (Matsumaru 2004).

3.6.2.3.2 State of the Art

In BES, data curation at present is regarded as a task for the individual principal investigator (PI), and often consists of preserving lab notebooks and associated hard drives with the associated raw data. Sustaining this approach is becoming more and more challenging as data sizes increase; this system also does not allow the reuse of data by the community. In the general field of data curation, the UK Digital Curation Center (http://www.dcc.ac.uk/) is the premier source for information and training on data curation, providing guidelines on assessing existing data assets, planning the setup of a data curation facility, and assessing the effectiveness of an existing facility. The Open Archival Information System (OAIS) is an International Organization for Standardization (ISO) standard that describes the organization of people, processes, and systems required to run a long-term data preservation service (OAIS 2012 update), (http://public.ccsds.org/publications/archive/650x0m2.pdf). However, while standards and guidelines have been in existence for some time, no reference implementations or standardized tools exist today for digital preservation and curation. Research, development, and deployment of data curation services have largely focused on the preservation of the data and less on the active support of its reuse.

3.6.2.3.3 Computing Needs and Requirements

Requirements at present:

Guidelines and methods for sustainable data curation and reuse — it is unclear today what data, metadata, provenance, and other information are required to support not only data curation but the identified use cases for active reuse of the curated data in support of scientific progress. The community needs support in developing suitable guidelines based on firm evidence.

Requirements in 5 years:

The community wants to be able to find curated data easily, with good metadata, provenance, data quality, and uncertainty assessment. To realize this requirement, the community needs:

- Access to methods that easily collect sufficient metadata and provenance throughout the experimental and analysis process.
- Access to raw data, analysis results, metadata, and provenance information described in a common, searchable format, thereby enabling full scientific analysis (including data mining, data reuse, and comprehensive studies).
- Access to sustainable data curation resources.
- Access to effective data discovery and access from anywhere (this includes the need for a fast, high-performance science data network, i.e., ESnet).
- Structured support from computer science experts (e.g., a dedicated SciDAC-like Institute) for BES facilities to create customizable, effective data curation solutions.

Requirements in 10 years:

The community wants to be able to find what is of high importance to their immediate scientific tasks with ease, including tools for accessing/analyzing the data. To realize this requirement, the community needs:

- Capabilities to discover and access context-relevant data and information in time-critical scenarios (e.g., in support of immediate use in real-time analysis and experimental steering applications).
- Methods to support scientific reproducibility and validation across different scientific methods.
- Push notification for new datasets of interest that have become available, including the ability to link up with the creator.

3.6.2.4 Accelerator Simulation

BES accelerator facilities, including both light sources and neutron sources, are among the most productive and important large-scale scientific facilities in the nation. They are also expensive, costing on the order of $1 billion to construct and with a single hour of operation costing the equivalent of >$20k. Reducing facility cost, reducing downtime, and providing beamline scientists with the computational resources and infrastructure to do their science are all essential to gaining the maximum benefit from these extraordinary facilities. Looking to the future, developing the next generation of BES accelerator facilities is essential to maintaining U.S. leadership in science and technology, strengthening U.S. economic competitiveness, and meeting national needs.

Advanced computing systems that will emerge in the next 5–10 years will greatly affect the operation and optimization of current BES accelerator facilities and their upgrades. Exascale simulation will also play a key role in developing next-generation facilities, such as fifth-generation light sources and facilities based on advanced accelerator concepts. These concepts have the potential to greatly reduce the size and cost of future facilities. The ability to build and deploy compact accelerators in a space measured in meters, instead of kilometers, would have profound consequences for science, technology, and medicine.

3.6.2.4.1 Driving Use Cases

Use Case 1: Storage-Ring Light Sources

The optimization of performance of storage-ring light sources presents very demanding computational requirements, as illustrated by the Advanced Photon Source Upgrade (APS-U) design effort. Although collective effects modeling is already fairly sophisticated, in the future we want to track (at least) 10K particles/bunch, 324 bunches, with element-by-element simulation, multiple impedance elements (at the actual location of the device), multiple radio frequency (RF) cavities with different modes, RF feedback systems, and beam feedback systems. Such a simulation will require about 10M core hours *for one run*. To obtain an answer in 1 minute in order to provide real-time guidance for machine operation requires about 10^9 cores, assuming perfect efficiency. In addition, although lattice optimization makes use of state-of-the-art multiobjective methods, using resources that are approximately 1,000 times larger as well as more sophisticated algorithms would provide more robust and predictable results. These methods would be critical in developing the next generation of storage-ring light sources.

Reducing facility cost, reducing downtime, and providing beamline scientists with the computational resources and infrastructure to do their science are all essential to gaining the maximum benefit from these extraordinary facilities. Looking to the future, developing the next generation of BES accelerator facilities is essential to maintaining U.S. leadership in science and technology, strengthening U.S. economic competitiveness, and meeting national needs.

Use Case 2: Linac-based X-Ray Free Electron Lasers

Light sources based on XFELs have revolutionized ultra-fast science. The LCLS at the Stanford Linear Accelerator Center (SLAC) is the first hard XFEL in the world. Its success has motivated LCLS-II, which is expected to be operational in the early 2020s. The design of future XFELs (beyond LCLS-II) will require improved modeling codes. As with the storage-ring light sources, XFELs will benefit greatly from improved parallel optimization tools. In addition, the accelerator modeling codes will need to be enhanced to model physical phenomena that cannot be modeled at present but that are likely to be important in future facilities. This need is driven, in part, by the exploration of novel "seeding" systems that prepare the beam prior to its entering the free electron laser (FEL) undulator that produces the X-ray radiation. Novel concepts for seeding systems impart very fine structure to the electron beam and furthermore couple the longitudinal and transverse degrees of freedom. As a result, certain approximate models, like the 1D model of coherent synchrotron radiation (CSR), need to be verified and possibly augmented with 3D CSR models. In addition, new simulation tools are needed to model the radiation production in proposed facilities where currently used approximations (e.g., "wiggle averaging") might not be applicable.

Use Case 3: Future Facilities Based on Advanced Accelerator Concepts

Accelerator science is now at a crossroads. With major new facilities costing hundreds of millions to billions of dollars, the development of future accelerator facilities demands that new technologies be developed to reduce their size and cost. Advanced accelerator concepts, namely those based on lasers, plasmas, and dielectrics, support extremely high gradients and could dramatically reduce the size and cost of future accelerators. The current simulations (e.g., for laser-plasma accelerators) require tens of millions of core-hours for 3D runs and involve weeks of execution time. While most attention has been focused on electron beams that are relevant to future light sources, advanced concepts in ion acceleration also have the potential to greatly reduce the size and cost of ion linear accelerators. In addition to novel approaches for high-gradient acceleration, there are also promising concepts in high-brightness beam production, manipulation, and transport. These present potentially large cost-saving opportunities, for example, in schemes to reach short radiation wavelength with a lower energy beam, in beamlines with large energy acceptance, and in high-efficiency energy recovery systems.

3.6.2.4.2 Gaps

Although sophisticated simulation tools are available, significant gaps and deficiencies are apparent when approaching the tasks described herein. There is a need for global parallel optimization using high-fidelity models, which will help produce robust optimized accelerator designs for present and future projects and with less uncertainty, risk, and guesswork. Real-time, high-fidelity modeling of both storage rings and FELs is needed to anticipate and resolve complex operational issues. Current week-long runtimes for many simulations inhibit using modeling for exploration and insightful design; to be truly useful, runtimes instead need to be on the order of minutes. In addition to reducing runtimes, new and computationally demanding models must be implemented for new regimes where existing models might be inadequate (e.g., 3D CSR).

BES Path to Exascale in Streaming Analysis @ BES Facilities

- Today: Support two types of streaming analysis workflows: the static analytical pipeline and the adaptive, hypothesis-driven steering of an experimental process by scientists informed by streaming data analysis and interpretation support.
- 2020: Deploy streaming analysis on high-data-volume beamlines; place mathematics and computer science personnel at BES facilities to focus on analysis pipelines; automate workflow scheduling systems; and enable real-time visualization of results.
- 2025: Deploy streaming analysis to steer DAQ and instrument control systems; decision support systems to optimize scientific outcomes; and digital twins integrated with streaming analysis.

BES Path to Exascale in Multimodal Analysis @ BES Facilities

- Today: Implement ad hoc or nonscalable solutions for integrative analysis when using probes at different facilities.
- 2020: Run “digital twin” experiments routinely; enable projects co-scheduled across different ASCR and BES facilities; place mathematics and computer science personnel at BES facilities to focus on analysis pipelines; automate workflow scheduling systems; and enable real-time visualization of results.
- 2025: Elevate multimodal data analysis from “one-off” studies to routine and rigorous projects that pull full significance from data.

BES Path to Exascale in Data Curation @ BES Facilities

- Today: Preserve data (with ever-increasing datasets) as opposed to reusing data, where the PI is chiefly responsible for its storage and transport.
- 2020: Support/provide/enable access to data curation resources and solutions.
- 2025: Support abilities and tools that enable finding what is of high importance to immediate scientific tasks with ease, including ability to link up with creators of datasets.

BES Path to Exascale in Accelerator Simulation @ BES Facilities

- Today: Incorporate simulation in design of light source upgrades to reduce facility cost and downtime.
- 2020: Acquire (in the future) ability to track (at least) 10k particles/bunch, 324 bunches, with element-by-element simulation, multiple impedance elements (at the actual location of the device), multiple RF cavities with different modes, RF feedback systems, and beam feedback systems.
- 2025: Address need for global parallel optimization using high-fidelity models to help produce robust optimized accelerator designs for present and future projects.

3.7 Mathematics and Computer Science Transforming BES Science

3.7.1 Challenges and Opportunities

Bridging the gap between BES scientific goals and ASCR computing capabilities will fundamentally rest on transformative physical models, mathematics, and computer science. Mathematics and computer science are how scientific theory, experiment, and computers talk with one another. Obtaining near-term scientific goals and realizing future visions will mean investing in and capitalizing on state-of-the-art evolving mathematics and computer science.

Indeed, this linkage cannot be overemphasized — for it is mathematics that provides the language and blueprint to transform models into equations, approximations, and algorithms that set the stage to take advantage of ASCR's computing portfolio; and it is computer science that provides the theory, tools, and methods to efficiently execute these blueprints on the most advanced computing architectures.

Figure 3-13 shows how this transformative role is embedded throughout advanced scientific discovery.

Figure 3-13. Mathematics and Computer Science Embedded in Advanced Scientific Discovery (Source: Diachin and Sethian 2015).

3.7.2 Goals

To this end, we have identified four key 10-year goals that will help position mathematics and computer science to accelerate science throughout Basic Energy Sciences.

Goal 1: Simulation. *Develop the mathematics that enables order-of-magnitude improvements in speed and accuracy in predictive materials and chemistry modeling.*

Goal 2: Experiment. *Deliver the mathematical algorithms and unified software environments that allow fast, multimodal analysis of experimental data across different imaging modalities and DOE facilities.*

Goal 3: Software. *Build the tools that will make efficient programming of tomorrow's machines as straightforward as programming today's laptops.*

Goal 4: Linkage. *Combine these three advances together to significantly advance our understanding in scientific domains.*

The case study most relevant to this section can be found in Appendix D, Section D.7, starting on page D-86 and is by:

- J. Donatelli, A. Hexemer, D. Kumar, R. Pandolfi, D. Parkinson, V. Venkatakrishnan, P.H. Zwart, and J.A. Sethian

3.7.2.1 Simulation

Over the past decade, mathematical models and computational algorithms in material science and chemistry have become increasingly sophisticated, producing powerful new capabilities. New algorithms that significantly increase accuracy, tackle new physical regimes, or reduce computational cost are continually being discovered and implemented. Although these advances are significant, the exascale computing capabilities that will be delivered in the next decade provide opportunities to constantly push the envelope of what is possible with modeling and simulation. Close collaboration between domain scientists and mathematicians will be required to capitalize on this opportunity with the development of new models and formulations, fundamentally new algorithms to solve both new and existing models, and continued improvements to existing algorithms and implementations to reduce solution times and increase robustness and accuracy.

Goal 1: *Develop the mathematics that enables order-of-magnitude improvements in speed and accuracy in predictive materials and chemistry modeling.*

The examples of new algorithms that improve time to solution, accuracy or robustness and are enabled by close collaboration between domain scientists and applied mathematicians are numerous. The range of collaboration covers model formulation/analysis and algorithm development and implementation. However, much more remains to be accomplished, and examples of the new work possible in each of these areas highlighted in the meeting include:

- *Model formulation and analysis.* For example, new multireference formulations for strongly correlated systems and new impurity solvers for materials that are far from equilibrium.
- *Algorithm development and implementation.* For example, parallel-in-time algorithms; improved sampling methods for rare events; dimensional reduction including reduced sample spaces, better approximations, and discontinuous basis sets; adaptivity in algorithms, physics, basis sets, and meshes; and optimization of accelerator design problems.

New algorithms that significantly increase accuracy, tackle new physical regimes, or reduce computational cost are continually being discovered and implemented. Although these advances are significant, the exascale computing capabilities that will be delivered in the next decade provide opportunities to constantly push the envelope of what is possible with modeling and simulation.

A recurring theme at the meeting was the orders-of-magnitude improvements that are made possible through the use of multiscale methods. These techniques significantly increase the spatial and temporal scales of simulation that can be computed and are now prevalent in many different areas in chemistry and materials modeling ranging from chemistry and material design and soft matter studies to interfacial phenomenon and combustion. However, there is no single multiscale methodology that works for all BES applications, and a number of different techniques have been developed including QM/MM methods, embedding methods, and combined coarse- and fine-grained approaches. Even with this wealth of approaches, there are still significant open research questions that can best be solved when applied mathematicians work closely with domain scientists, together addressing for example, such questions as:

- "How does a given research team accomplish the rigorous transfer of information across scales and different algorithmic methodologies?"
- "How does a given research team know when to apply which approximation scheme?"
- "What is the error associated with each different approximation technique and across scales?"

Goals identified at the meeting include developing:

- Techniques to use both algorithm and mesh adaptivity in a broader range of multiscale/multiphysics simulations; this effort will include research into the development of indicators to understand when and how to invoke the right levels of approximation.

- A deep understanding of how to transfer information across different scales. In this case, information is broadly defined to include, for example, physical parameters that may need to be conserved in continuum simulations, the errors introduced when applying different types of approximation in different regimes, and particle interaction information to avoid duplication of interactions in multiscale particle simulations.
- In the longer term, new multiscale formulations to increase the range of problems that can be solved. Examples of new problems that could benefit from such techniques include, but are not limited to, modeling material and chemical microstructures, catalytic energy conversion and storage devices, and the structure and folding mechanisms of peptides.

Another common theme that emerged at the meeting was the use of UQ techniques in BES applications. UQ methods have long been used in the DOE engineering and nuclear stockpile communities to understand error in individual and coupled components, thereby increasing our confidence in predictive simulations. They are increasingly used in the scientific community, and several different groups at the meeting, ranging from the combustion and electronic structure communities to the layered material genomics community, identified UQ as an important component of their simulation workflow. To date, applications in BES are primarily making use of nonintrusive, or sampling-based, UQ methodologies in which parameters are varied and the simulation is run repeatedly to determine sensitivities and to quantify the margins of error. While these methods have been used successfully in many different contexts for several years, open research questions remain for their use in material science and chemistry applications.

- In the near term, we must push to expand the range of applications to which UQ is applied. However, the challenges associated with effectively employing UQ depend heavily on system complexity, for example, inflow/boundary conditions, chemical model parameters, transport properties, subgrid scale models, etc. Every new application reveals unique considerations, and collaboration between domain scientists and experts in the technology are required to determine how to make best use of existing methodologies in new material science and chemistry applications.
- Over the longer term, we must continue to advance our understanding of how to develop new UQ methodologies that are applicable to chemical and materials modeling. This topic has been the focus of "hot topic workshops" hosted by the Institute for Mathematics and its Applications (IMA) and the National Institute of Standards & Technology (NIST) in recent years. Challenges include combining models with very different scales and descriptions, such as particle-based and continuum modes; capturing intrinsic variability at atomic and microstructural levels; dealing with uncertainties with multiple origins that need to be quantified and propagated; and leveraging and effectively using scarce experimental data. Ideas for advanced techniques that were specifically mentioned at the meeting include adaptive UQ methods that employ low-rank, sparse representations; advanced statistical techniques for code calibration, prediction, and UQ; advanced sampling methodologies; and error analysis and propagation.

3.7.2.2 Experiment

Ever more powerful experimental facilities are creating vast amounts of data — far more than have ever existed before. Experimentation is an integral part of scientific investigation, and BES facilities such as synchrotron radiation light source facilities, neutron scattering facilities, and nanoscience centers generate vast amounts of data. For example, beam science is undergoing a rapid change as facilities probe matter at higher and higher physical resolutions and rapid timescales. These experiments generate massive amounts of data; in the future, they may generate multiple terabytes of data per sample run. The data are often statistical in nature and replete with noise, poor contrast, and signal dropout. Advances in science using these facilities require fundamental advances in the mathematics associated with data science.

Goal 2: *Deliver the mathematical algorithms and unified software environments that allow fast, multimodal analysis of experimental data across different imaging modalities and DOE facilities.*

Fundamental statistical, mathematical, algorithmic, and computational methods are needed to extract information from murky data, interpret experimental results, and provide on-demand analysis as information is being generated. To make sense of this information, new algorithms that fuse different branches of mathematics are at work. For example, algorithms may combine dimensional reduction, graph techniques, and computational harmonic analysis to perform robust reconstructions from scattering data, and merge partial differential equation methods with machine learning to analyze experimental image data.

Fundamental statistical, mathematical, algorithmic, and computational methods are needed to extract information from murky data, interpret experimental results, and provide on-demand analysis as information is being generated.

At the same time, the landscape of experimental facilities is rapidly changing. In some situations, quick and rough results are desirable while an experiment is under way. In other situations, considerable computation time can be dedicated so as to provide the most accurate reconstruction and analysis possible.

The desire for immediate results from algorithms embedded close to detectors spawns different mathematical questions from those involved in post-processing aided by high-speed networks and extreme-scale computing:

- One end of the spectrum aims at "-demand" computational tools for analysis, data reduction, and feature extraction next to facilities, using embedded advanced algorithms and special-purpose hardware. Here, questions that arise include: What is the minimum/fastest computational model/algorithm that gives (at least some) useful information? Can users quickly determine whether data are useful, are not useful, or are in between? By taking advantage of powerful increases in core hybrid CPUs and general-purpose computing on graphics processing units (GPGPUs), can users quickly perform an analysis in order to steer ongoing experiments to more optimal configurations or output?
- The other end of the spectrum aims at post-processing using reconstruction, inter-comparison, simulation, and visualization using high-performance and extreme-scale computing. Here, different questions arise, including: What is the maximum amount of information that can be measured, processed, organized, and displayed to help understand and shed light on further experiments? Can data be transformed to initialize computational models, with output framed to complement the experiment?

In the near term, identified goals include:

- Building mathematical tools and software environments for real-time streaming analysis of experimental data generated at BES facilities. This desired capability can range from algorithms running on hardware directly embedded with detectors to codes analyzing data efficiently that are shipped using ESnet from facilities to ASCR's most advanced computing platforms.
- Developing "triage" algorithms that can quickly determine which data are useful, are not useful, and are in between in order to steer the experiment as it is performed and to reduce the amount of data that is sent across networks.
- Devising multimodal algorithms that fuse information from multiple imaging modalities.
- Constructing efficient and robust data and dimension reduction methods, including the use of methods to detect key features in data automatically, perform pattern matching to search across datasets, and employ past information to analyze new results.

In the long term, scientists will best make use of an integrated environment that seamlessly fuses experiments across multiple facilities and compute environments. As an example, in this view, users can sit at one facility and conduct experiments across a range of beamlines, automatically processing samples and fusing results to vastly increase our ability to understand chemicals and materials with great complexity. Keys to this vision include:

- Advanced mathematical algorithms that can extract information from time-resolved, high-resolution data from multiple imaging sources simultaneously.
- Coordinated scheduling of local and remote resources across the full spectrum of the available compute ecosystem.
- Unified software and data curation environments that can allow data to be analyzed across the DOE landscape.

3.7.2.3 Software Development and Optimization for Extreme-Scale Computers

The requirements that were strongly articulated at the meeting by BES scientists were largely focused on improving the ease-of-use of future computers: researchers do not want to spend years of time rewriting their software to utilize the machines, thereby hindering the progress of their scientific endeavors.

The extreme-scale systems that will be delivered in the eight- to ten-year time frame will offer unprecedented compute power and challenging parallelism because the machines will contain millions of heterogeneous cores with deep memory hierarchies, but with relatively slow interconnects between nodes. Issues that will significantly complicate software development include the need to explicitly manage data movement, depth and types of memory, and burst buffers and parallel file systems, as well as working with billions of threads, optimizing for power usage, and understanding the behavior of algorithms in the presence of hard and soft faults. For these reasons, refactoring application codes to utilize these architectures effectively will be a massive undertaking. The requirements that were strongly articulated at the meeting by BES scientists were largely focused on improving the ease-of-use of future computers: researchers do not want to spend years of time rewriting their software to utilize the machines, thereby hindering the progress of their scientific endeavors. Thus, our 10-year goal for ASCR computer science support is:

Goal 3: *Build the tools that will make efficient programming of tomorrow's machines as straightforward as programming today's laptops.*

Achieving this goal requires significant development and changes to the entire software stack, from operating system software and run time systems to cross-cutting applied math libraries and application software. Ideally, programming models and languages will emerge that make obtaining efficient performance on the highest-end computers as easy to accomplish as on today's laptops.

This goal is complicated by the fact that computer architectures are becoming increasingly diverse, where code that is optimized to perform well on one computer may not perform well on a different architecture. Tools that allow application software developers to manage this complexity are in high demand. However, the desired realization of this goal has many different forms, ranging from a python-like or domain-specific languages (DSLs) that hide all of the details of the architecture from the application code developer to supplements to the MPI that exposes those details and makes them accessible to the application scientist. Goals in the area of programming models are to provide a range of options, including the following:

- In the short term, it will be necessary to enable the continuity of development of existing codes while ensuring the forward-looking move toward the longer-term goals. This effort will require continued support and development of existing programming models (such as MPI + X where X is OpenMP, OpenACC, or CUDA) and providing tools (such as profilers and debuggers) that will enable developers to analyze their software to prepare for the move to new architectures.

- Development must proceed on community-based standards such as MPI and OpenMP that provide low-level programming models and libraries and that allow effective use of the features of computer architectures. Examples here would include standard methods to deal with new storage and memory hierarchies, which may encompass diverse devices like stacked memory, dynamic random access memory (DRAM), video random access memory (VRAM), and solid state drive (SSD). These developments are already under way, and DOE scientists are playing a key role in their definition. Education and outreach to the application scientists as the standards evolve will ensure that the latest functionalities are as widely used as possible.
- Constructs and tools are needed that work within existing languages, such as C++, to provide higher-level abstractions that will hide the details associated with obtaining performance portability across different computer architectures by decoupling the specification of the science application from how it is mapped onto target platforms. Examples of this capability that are under development today include the Kokkos, Resource-Adaptive Java Agent (RAJA), and Legion tools at Sandia National Laboratories (SNL), Lawrence Livermore National Laboratory (LLNL), and Stanford University and Los Alamos National Laboratory (LANL), respectively. These tools are currently used in a wide variety of math libraries and application codes and are showing significant promise in providing performance portability across different many-core architectures. Incorporation into application codes can further extend the impact of these tools.

Ideally, programming models and languages will emerge that make obtaining efficient performance on the highest-end computers as easy to accomplish as on today's laptops.

- Runtime-centric frameworks are needed that allow for specification of data-dependence (such as graph methods) and asynchronous task-based programming models to significantly improve performance, facilitate portability, increase the level of resiliency, and ultimately improve scientific productivity. Expressing node-level tasks that can be executed asynchronously can lessen the impact of operating system noise, variations within the performance of individual cores, and overall nonuniformity.
- Development must also proceed on workflow tools that enable a suite of interoperable capabilities such as embedded DSLs that emit high-performance tasks for a myriad array of chemical properties, chemical mechanisms, high-order adaptive mesh refinement (AMR) partial differential equation (PDE) stencils, interpolants and operands for a range of turbulence and combustion physics, analytics, UQ algorithms, topological segmentation and tracking algorithms, multivariate statistics, and visualization. Policies and mechanisms need to be developed to manage the execution of end-to-end workflows under strict power and performance constraints while still maximizing the throughput of scientific research.
- Longer term, we can develop DSLs and the associated compilers that provide high-level programmability for particular application areas. Examples of these languages currently exist in materials science for the generation of custom materials (e.g., Matriarch developed at the Massachusetts Institute of Technology [MIT]) but are in limited use. In quantum chemistry, tensor contraction-based DSLs are starting to be used extensively in the community. Research on the use of DSLs to manage performance portability has been explored in relevant applied math and algorithm areas (e.g., structured AMR), but these implementations are still research prototypes and are not yet widely available for use.

Related to this are tools that are being developed by the computer science community to help application scientists improve their productivity when programming for exascale computers. This effort encompasses a broad array of tools, such as those that allow software developers to understand deepening memory hierarchies, data motion and cache use, the trade-offs in performance and power usage, debugging and profiling at scales that contain millions of processes, and optimization of the performance of the software. It is critical that such tools be available on the extreme-scale computing architectures and that they are robust and easy to use for adoption by the

application community. This pursuit can be enabled by a co-design type of process where these tools are designed with application specialists and applied mathematicians to develop effective strategies for the broad scientific community. For example, domain-specific scheduling policies may be appropriate for some software while autotuning-driven approaches might be very effective for others. In addition, there must be an effective mechanism for taking the research-level software to hardened software with long-term maintenance.

It is critical that such tools be available on the extreme-scale computing architectures and that they are robust and easy to use for adoption by the application community. This pursuit can be enabled by a co-design type of process where these tools are designed with application specialists and applied mathematicians to develop effective strategies for the broad scientific community.

Even with effective computer science tools to help maximize the use of extreme-scale computers, the architectures are changing in such a way that many of the algorithms currently used by the material science and chemistry communities may not scale or perform well. Thus, existing algorithms may need to be recast, or in some cases radically redesigned, to leverage key architectural features. In particular, as data motion becomes increasingly expensive compared to floating-point operations, algorithms that have high arithmetic intensity (many floating-point operations for each byte moved from memory) become increasingly attractive. In many cases, these algorithms have been known to exist for many years; however, they were not considered attractive on previous computer architectures. Thus, the community is reconsidering "expensive" but highly accurate methods such as Green's function methods and AIMD methods and finding that they will be increasingly cost effective on the machines of the future.

Similarly, many algorithms in chemistry and material science scale steeply with the number of particles (e.g., N^3 to N^6), and even significant increases in hardware capability result in minimal increases in the number of particles that can be simulated. Methods that scale linearly or quadratically exist for many application areas; however, until a certain threshold number of particles is reached, these methods have overheads that make them more expensive than their more steeply scaling counterparts. The machines of today and tomorrow are putting us at that cross-over threshold point for many applications. Thus, we are witnessing a revival of linear scaling algorithms with successes for certain classes of applications, such as first-principle electronic structure calculations. Additional research is needed to expand the regimes and problem spaces for which these methods are suitable. Collaborative development of prototypical comprehensive simulation code(s) representing the range of application motifs in BES would be helpful to further flesh out the key requirements for next-generation programming environments and runtimes so these can be adapted to changing architecture designs and new algorithmic approaches.

Finally, it will be important to leverage as much parallelism as possible in the problem formulation. For example, in the case of the GW-BSE approach, it is possible to leverage band parallelism, parallelism over the frequency space, and plane wave basis set parallelism. Parallel-in-time methods also provide parallelism in time as well as in the energy/gradient evaluations.

3.7.3 Cross-Cutting Issues

3.7.3.1 Use of Applied Math and Computer Science Software Libraries and Tools

BES scientists make considerable use of a wide range of applied math and computer science software libraries and tools. Examples include, but are not limited to, linear algebra kernels in the Basic Linear Algebra Subprograms (BLAS) libraries; diagonalization and linear system solution in libraries such as Linear Algebra Package (LAPACK); Scalable LAPACK (ScaLAPACK); portable, extensible toolkit for scientific computation (PETSc); Quantum and Magma; and FFT routines in libraries such as Fastest Fourier Transform in the West (FFTW) and Hierarchical Data Format (HDF5), etc. Such tools provide key functionality and shield application scientists from the details of needing to implement such functionality themselves.

As we move toward extreme-scale computing, these tools will continue to be invaluable to application scientists if they meet the following requirements. In the near term, the tools must be:

- Efficient and portable across different exascale computing environments; in particular, this software must perform well in the context of the application data layouts and in concert with each other.
- Easy to use. The software libraries should have significant documentation and must be robust and "productized." In some cases, graphical user interfaces (GUIs) are the most effective and easiest mechanisms for accessing these libraries' underlying functionality. In other cases, a "grey box" approach that has parameters set for naïve users but allows experts to dive in as needed would be the best path for software use.
- Community codes with a clear path for development and long-term availability. Users do not want to invest several months to a year in adopting new tools, languages, or libraries, only to have development cease at the end of a three- or five-year project life cycle. Long-term support for such software is needed.

Over the long term, new library developments are needed in areas such as scalable tensor libraries and linear scaling algorithms, as well as in facilitation of arithmetic manipulations of multicomponent electronic wave functions.

BES Path to Exascale in Mathematics and Computer Science

- **Today: Continue to consolidate advances of past two decades while exploring utilization of multiscale solutions.**
- **2020: Put structures in place for close collaboration between domain scientists and mathematicians to capitalize on the opportunities presented by new algorithms and multiscale-related capabilities.**
- **2025: Regarding *simulation*, develop the mathematics that enables order-of-magnitude improvements in speed and accuracy in predictive materials and chemistry modeling.**
- **Regarding *experiment*, deliver the mathematical algorithms and unified software environments that allow fast, multimodal analysis of experimental data across different imaging modalities and DOE facilities.**
- **Regarding *software*, build the tools that will make efficient programming of tomorrow's machines as straightforward as programming today's laptops.**
- **Regarding *linkages*, combine advances in the three areas above with those in other science domains to advance our understanding significantly in science domains.**

3.8 Next-Generation Workforce

3.8.1 Meeting the Exascale Challenge Is Grounded in Workforce Development

A common theme across the different BES topical areas is that the need for workforce development poses a serious bottleneck to reaching the exascale level of computing — and thus is an essential part of the equation of success in reaching exascale computing. The changes envisioned in computing technologies over the coming decade will require a new generation of computational scientists who are well grounded in their science and engineering disciplines but also knowledgeable about the major computational issues being addressed in computer science and applied mathematics. Likewise, the field also needs computer science and mathematics experts who are knowledgeable about scientific computing and the HPC needs of domain scientists.

In fact, having available and sustaining a sufficiently skilled workforce is considered the greatest risk factor in realizing exascale computational science. Consequently, we anticipate that significant investments must be made in training a new generation of computational scientists.

While BES researchers have an historic, strong, and recognized synergy between theory, math, algorithms, implementation, and computing, the increasing complexity and dependencies of the exascale era will require these collaborations to be further enhanced. The complexities and multiple layers of hierarchy of next-generation programming environments will require cross-cutting, multidisciplinary teams of domain scientists, applied mathematicians, and computer scientists, who together can cope with the broad range of challenges, from developing the physical and mathematical models, to expressing the scientific workflow, to developing the numerical algorithms, to decomposing the algorithm to the optimal level of task granularity, to expressing fine-grained parallelism in DSLs, and to ensuring that all of the layers of the programming model and runtime have the right abstractions to enable flexibility and performance. Providing the kinds of education and training deliverables we need to prepare these future computational and domain scientists and computational software developers for exascale computing presents a major challenge. In fact, having available and sustaining a sufficiently skilled workforce is considered the greatest risk factor in realizing exascale computational science. Consequently, we anticipate that significant investments must be made in training a new generation of computational scientists — individuals who are well grounded in their science and engineering disciplines, but also knowledgeable about relevant computer science and applied mathematics issues — to realize the promise of exascale computing.

3.8.2 Priority Directions to Promote the Next-Generation Workforce

As stated, staffing and producing the next generation of domain scientists with the requisite training in computer science will be increasingly critical in the near future in pursuit of exascale computing.

Several efforts can be implemented at the DOE/programmatic level. For instance, partnerships between DOE's BES and ASCR that fund computational scientists in the facilities and help develop clear career paths will encourage students and scientists to pursue those career paths. Furthermore, to increase domain scientists' interactions with graduate and undergraduate students, DOE should consider:

- Increasing the number of Computational Science Graduate Fellowship (CSGF) graduates. In fact, expansion of the CSGF program is regarded as one very useful way of pursuing workforce development.
- Creating joint programs between DOE, the National Science Foundation, and universities.
- Having the ASCR facilities, NERSC, and universities share graduate students to foster training and knowledge transfer.

Additional ideas for fostering workforce development, particularly with regard to integrating mathematics and computer science knowledge, include:

- Embedding math/computer science expertise in domain-specific science groups.
- Establishing a center of mathematicians and computer science experts connected to multiple domain science teams.
- Assigning "roaming" architectural experts to visit science groups for 3- to 4-month terms.
- Establishing resident math and computer science experts in the facilities.
- Pursuing funding mechanisms or programs to fill the need for cross-trained graduate students.

HPC training programs similar to the Argonne Leadership Computing Facility's ATPESC (Argonne Training Program for Extreme-Scale Computing) should be expanded, as well as hack-a-thons and coding contests in universities to attract and encourage students. Only by taking actions on all of these fronts can DOE and the universities hope to grow the next generation of domain specialists with the skills needed for achieving success in the exascale computing era.

3.8.3 Cross-Cutting Opportunities to Promote the Next-Generation Workforce

Opportunities exist across the BES and ASCR domains to promote workforce development — including adapting existing collaboration and/or training mechanisms toward this end.

3.8.3.1 Collaboration Opportunities

For example, the SciDAC program exemplifies the collaboration that is possible between application developers, applied mathematicians, and computer scientists and engineers in the development of new and better algorithms. These interdisciplinary code development teams are supported (and should continue to be supported) over long time periods to guarantee better code design, increase the lifespan of codes, and give more flexibility to future developments. Expanding this program and developing similar programs to address the training of future scientists (e.g., graduate students and postdocs) would be highly desirable, because the ability of emerging scientists to understand the importance of integrating expertise in a chosen field such as chemistry or physics (for example) with applied mathematics and computer science will result in needed advances in development of the next-generation workforce.

Although collaboration mechanisms currently exist at scales ranging from the individual PI to national labs and institutions with national reach, collaboration between fields should be particularly encouraged. At the individual PI level, the CSGF program is regarded as highly successful. It receives many more strong applicants than can currently be funded. At the research group level, the SciDAC program is also successful; however, SciDAC's reach has been limited by the number of opportunities and scope of the program. At the national level, additional mechanisms to bring key stakeholders together should be encouraged. One possible mechanism, particularly in the domain of new algorithms, is the CECAM (for the Centre Européen de Calcul Atomique et Moléculaire) workshop model. CECAM is responsible for sponsoring more than 100 workshops, schools, and tutorials each year that typically focus on specific applied topics, with attendance limited to less than 100. NERSC and the ASCR facilities are in a particularly attractive position to lead these efforts with their continuity in expertise in the hardware, runtime environments, mathematics, computer science, and domain science (e.g., catalysts and liaisons). Adapting this format may provide the required mechanism for moving computational research ideas and tools developed in computer science over to the applications domains.

Another mechanism to facilitate information exchange and adoption of new strategies for scientific software design and implementation within the community is to create a shared communication mechanism where domain scientists, mathematicians, and computer scientists could describe their algorithmic challenges and mathematical processes for calculations at scale in a virtual forum. Not only would this facilitate training of the next generation workforce, it would provide continuing education and development opportunities for the current workforce.

3.8.3.2 At the Science and Computing Interface: The Application Engineer Position

The increasing complexity of theory, applications implementation, maintenance, tools, languages, and computational facilities is creating a strong need for new stable career positions at the interface between science domains and computing. Increasingly, our challenges require a position that rarely exists today: "application engineers" who are neither purely programmers or software engineers, nor purely domain scientists. The assessment criteria for these positions are distinct from purely scientific work and should involve consideration of the user base of each code and its strategic mission importance, as well as code citations. For these positions to be successful, institutional support and funding must continue. The consequence of the absence of these positions is clearly illustrated by the "maintenance challenge" faced by research groups where new techniques are supported by research grants, but there are few options for ongoing maintenance and porting efforts after the period of the initial grant. This gap in ongoing support limits the sustainability and efficiency of each application. The increasing complexity of the exascale era is expected to exacerbate this problem. A greater availability of "glue" grants for ongoing support of applications teams may address this challenge.

Extending this path to the exascale era and addressing the discontinuous changes in the computational architectures will require not only significant developments in the computing and software ecosystems, but also commitments to address the education, development, and availability of a skilled workforce.

BES Path to Exascale in Next-Generation Workforce

- **Today: Examine existing student/workforce development/training programs (e.g., SciDAC, CSFG, CECAM) for best practices to adapt or use to promote building exascale-relevant skillsets.**
- **2020: Stand up program(s) to train (and embed) exascale-oriented graduate students and scientists at ASCR facilities.**
- **2025: Realize employment of computer scientists in an Application Engineer type of role or position at the ASCR facilities, academic institutions, and other HPC centers.**

4 PATH FORWARD

For researchers to move forward in addressing the scientific challenges documented in this review, an evolving computing ecosystem must support them. This computing ecosystem includes computational and data systems; scientific applications and software; and the infrastructure for data transfer, sharing, access and analysis — each of which must undergo further investment and development on the path to exascale. Realization of advances in real-time computing, flexible queuing structures, complex workflows, and other policies coupled to the operation of the computing ecosystem must also occur. The coupling of an exascale ecosystem, along with a convergence of theoretical, mathematical, computational, and experimental capabilities, will bring many opportunities for new scientific breakthroughs at an unprecedented scale.

Collaboration between BES and ASCR scientists and facilities staff will help ensure development and deployment of an effective, realistic computing ecosystem that enables revolutionary discoveries in areas described in this report. The computing ecosystem requirements resulting from this review will form the basis to direct future investments of time and resources. These requirements fall into broad categories: methods development; computational environment; data and workflow; and communication and community involvement.

4.1 Methods Development

The advancing complexity of computer hardware requires BES researchers to have more scalable, performant algorithms and applications that are capable of efficient execution on future computing architectures fielded by ASCR facilities. Meeting participants discussed those computing ecosystem aspects that will accelerate or impede their progress in the next 5–10 years. Participants named application codes as well as models and algorithms as key factors, requiring significant methods development activity. A representative list of the methods development topics discussed by the review participants is as follows (see Section 3 for a more detailed overview of the methods development topics presented by the review participants):

- Strong-electron correlation methods for ground and excited states.
- Theoretical foundation for multiscale coupling of different representations of QM and well-parameterized, validated effective Hamiltonians.
- Inclusion of electronic dynamics such as in TDDFT, DMRG, DMFT, GW, QMC, and emerging methods.
- Full incorporation of relativistic effects.
- Prediction of structures at interfaces.
- Parallel-in-time approaches and other new methods to accelerate the sampling of rare configurations and more realistic systems to generate ensembles with enough statistics for long-timescale sampling of rare events.
- Uncertainty quantification.
- Mathematics that enables order-of-magnitude improvements in speed and accuracy in predictive materials and chemistry modeling.

Other methods development requirements suggest that innovations are needed in order to use the computing resources more effectively and to develop pipelines to and from the resources:

- New strategies for memory management and low-communication/highly parallelizable algorithms to achieve load balancing and effectively use the fast multicore platforms.

- Adaptive algorithms that update "on the fly," for example, in performing rapid data analysis to guide experiments in real time and changing grids and multiscale methods.
- Fully functional and performant analysis pipelines including algorithms, adaptive work flows, streaming data and information visualization, provenance, effective programming patterns, performance assessment, and optimization.

One of the methods development requirements points to the need to provide a structured engagement allowing computer science and mathematics experts (e.g., a dedicated SciDAC Institute) to work with BES user facilities and researchers.

A close dialogue between BES and ASCR researcher and facilities staff will streamline and promote research and development through the exchange of information about computing ecosystem roadmaps and application requirements and the availability of systems for simulation and testing.

4.2 Computational Environment

Requirements for the access, scheduling, and software ecosystem identify an evolving use-model. The "traditional" HPC model, defined as a large simulation generating data that is then post-processed, is no longer the only primary use-model for many BES projects. Emerging demands, such as for complex workflows and near-real-time computing, are changing the landscape.

New requirements for the computing ecosystem include the following.

- Real-time and near-real-time computing (i.e., nearly instantaneous, with short turn-around times), to provide support for streaming analysis on timescales sufficient to influence experimental decisions, steer DAQ systems, and visualize experiment progress in real-time.
- Proposal/award processes to support the wider array of requirements, including flexible allocations mechanisms that allow for "on-demand" allocations for bursty computational needs and software development.
- A user-friendly development environment, with uniform environments among DOE HPC centers supporting portable, high performance across systems with improved and new runtime systems that mask HPC complexity from application programmers, and training aimed at all levels of HPC developers, including nontraditional HPC users.
- High-performing languages, libraries, and DSLs; and profiling and debugging tools for complex workflows.

4.3 Data

The scale of data generated from both BES simulations and experiments has created an opportunity and a challenge. Even with increased network speeds, datasets from experimental facilities as well as from simulations can no longer be feasibly transferred to a user's home institution for further analysis. As a result, ASCR and BES facilities must create more data-centric environments with highly effective data analytics tools for their users. Development of such environments and tools will require expertise from domain scientists, data scientists, and applied mathematicians. Continued collaboration will be required to assess proper deployment of the environments as computing resources evolve.

Requirements related to data generation, storage, transport, curation, and exploration include the following:

- Improved tools for data analysis and visualization including machine learning and deep learning techniques for pattern recognition and data quality assessment.

- Tools for real-time analysis of large-scale simulations to obtain, for example, configurations and coordination numbers for metal centers, as well as pattern recognition in real-time to control dynamics in order to steer systems.
- Improved methods and guidelines for data tracking and provenance with searchable and consistent data and metadata from experiment to simulation, analysis, and curation. Push notification for new datasets of interest that have become available, including the ability to link up with the creator(s).
- The ability to carry out "meta" experiments (with other groups' data as well), including the option to develop new algorithms that can go beyond the original questions that the researchers who collected the data were asking.
- The necessary mathematics, tools, and infrastructure to elevate multimodal data analysis from "one-off" studies to routine and rigorous projects that pull full significance from data. This capability will require a large-scale framework development effort that allows researchers in different experimental domains to develop algorithms and codes that can be combined to fit comprehensive models to all data collected on sets of related samples.
- Improved data access, transfer, storage, and management capabilities that include federated logins across BES experimental and ASCR facilities and high-performance networking from experiment to HPC system.

4.4 Communication and Community Involvement

To foster development of the requisite exascale-level skills and to disseminate this learning widely throughout the community, DOE (with the ASCR facilities) must seek to create or make use of existing initiatives that promote the following:

- Workforce development (education and training).
- Collection and sharing feedback from involvement with standards committees.
- Development of better training materials including best practices, examples, etc.

These activities are ongoing today in multiple institutions; however, efforts to connect them to the larger science community have been attempted on an "ad hoc" basis to date. ASCR facilities can explore new or improved communication channels and activities. In addition, experience has shown some of the best impact from strong collaborations. The previously identified structured collaborative efforts could focus more attention on this important mechanism for community involvement.

4.5 Conclusion

Requirements that are key to an evolving computing ecosystem have been identified in areas of methods development, computational environment, data, and communication and community involvement. These areas are a collaborative research opportunity across much of the computational ecosystem. Structured collaborative efforts between BES and ASCR are a path to address the growing complexity of the science and the computational resources.

5 REFERENCES

Abraham, M.J., T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl., 2015, "GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers," *SoftwareX*, Volumes 1–2, pp. 19–25.

Argonne (Argonne National Laboratory), 2012, *Center for Electrical Energy Storage: Tailored Interfaces Energy Frontier Research Center*, Chemical Sciences and Engineering Division, June, http://www.cse.anl.gov/%5C/cees/index.html, accessed February 7, 2016.

Bluhm, H., 2006, *Pulsed Power Systems*: Principles and Applications, Berlin/Heidelberg, Germany: Springer Verlag.

Brown, I.D., and B. McMahon, 2002, "CIF: The Computer Language of Crystallography," *Acta Crystallographica*, Section B-Structural Science, 58: p. 317.

Brown, W.M., P. Wang, S.J. Plimpton, and A.N. Tharrington, 2011, "Implementing Molecular Dynamics on Hybrid High Performance Computers – Short Range Forces," *Computer Physics Communications*, Volume 182, pp. 898–911.

Brown, W.M., A. Kohlmeyer, S.J. Plimpton, and A.N. Tharrington, 2012, "Implementing Molecular Dynamics on Hybrid High Performance Computers – Particle-Particle Particle-MeshComputer Physics Communications," Volume 183, pp. 449–459.

Cargnello, M., V.V. Doan-Nguyen, T.R. Gordon, R.E. Diaz, E.A. Stach, R.J. Gorte, P. Fornasiero, and C.B. Murray, 2013, "Control of Metal Nanocrystal Size Reveals Metal-Support Interface Role for Ceria Catalysts," *Science*, Volume 341, pp. 771–773.

Challacombe, M., E. Schwegler, and J. Almlöf, 1996, "Fast Assembly of the Coloumb Matrix: A Quantum Chemical Tree Code," *The Journal of Chemical Physics*, Volume 104, 4685. DOI: 10.1063/1.471163.

Christou, G., D. Gatteschi, D.N. Hendrickson, and R. Sessoli, 2000, "Single-Molecule Magnets," *MRS Bull.,* Volume 25, pp. 66-71.

Dabestani, R., R.J. Reszka, and M.E. Sigman, 1998, "Surface Catalyzed Electron Transfer from Polycyclic Aromatic Hydrocarbons (PAH) to Methyl Viologen Dication: Evidence for Ground-state Charge Transfer complex Formation on Silica Gel," *J. Photochem. Photobiol.* A, Volume 117(290), pp. 223−233.

Darden, T., D. York, and L. Pedersen, 1993, "Particle Mesh Ewald: An N•log(N) Method for Ewald Sums in Large Systems," *The Journal of Chemical Physics* Volume 98, Issue 12, 10089.

Diachin, L., and J. Sethian, 2015, figure captioned "Mathematics and Computer Science Embedded in Advanced Scientific Discovery," Lawrence Livermore National Laboratory (L. Diachin).

DOE-BES (U.S. Department of Energy-Office of Basic Energy Sciences), 2005, *Report of the Basic Energy Sciences Workshop on Solar Energy Utilization*, Office of Science, April 18–21, http://science.energy.gov/~/media/bes/pdf/reports/files/seu_rpt.pdf, accessed February 6, 2016.

DOE-BESAC (Basic Energy Sciences Advisory Committee), 2007a, *Directing Matter and Energy: Five Challenges for Science and the Imagination*, prepared by the BESAC Subcommittee on Grand Challenges for Basic Energy Sciences, December 20, http://science.energy.gov/~/media/bes/pdf/reports/files/gc_rpt.pdf, accessed April 7, 2016.

DOE-BESAC, 2007b, *Basic Research Needs: Catalysis for Energy*, Office of Science, December 20, http://science.energy.gov/~/media/bes/pdf/reports/files/cat_rpt.pdf, accessed February 6, 2016.

DOE-BESAC, 2012, *From Quanta to the Continuum: Opportunities for Mesoscale Science*.

DOE-BESAC, 2015, *Challenges at the Frontiers of Matter and Energy: Transformative Opportunities for Discovery Science*, November, http://science.energy.gov/~/media/bes/besac/pdf/Reports/CFME_rpt_print.pdf, accessed April 7, 2016.

DOE-SC (Office of Science), 2010, *Computational Materials Science and Chemistry: Accelerating Discovery and Innovation through Simulation-Based Engineering and Science*, http://science.energy.gov/~/media/bes/pdf/reports/files/cmsc_rpt.pdf, accessed June 12, 2016.

DOE-SC, 2014, "Heavy Element Chemistry," http://science.energy.gov/~/media/bes/pdf/brochures/bes-cras/2014-feb/cra_23_heavy_element_chemistry.pdf, accessed March 1, 2016.

Ennis, J.B., F.W. MacDougall, X.H. Yang, R.A. Cooper, K. Seal, C. Naruo, B. Spinks, P. Kroessler, and J. Bates, 2007, "Recent Advances in High Voltage, High Energy Capacitor Technology," in *16th IEEE International Pulsed Power Conference*, IEEE, Piscataway, New Jersey, United States, pp. 282–285.

Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, 1995, "A Smooth Particle Mesh Ewald Method," *The Journal of Chemical Physics*, Volume 103, 8577.

Ferreira, K.N., T.M. Iverson, K. Maghlaoui, J. Barber, and S. Iwata, 2004, "Architecture of the Photosynthetic Oxygen-Evolving Center," *Science*, Volume 303, pp. 1831–1838.

Fowler, M., 2010, *Domain Specific Languages*, Addison-Wesley, http://martinfowler.com/books/dsl.html, accessed February 2, 2016.

The Galli Group, undated, figure captioned "Nanoparticles Entering a Solution Phase as Part of a Complex Synthesis Pathway," N. Brawand, University of Chicago, http://galligroup.uchicago.edu/News/news.php, accessed May 11, 2016.

Garza, A.J., I.W. Bulik, T.M. Henderson, and G.E. Scuseria, 2015, "Synergy between Pair Coupled Cluster Doubles and Pair Density Functional Theory," *J. Chem. Phys.*, Volume 142(4).

Gil, Y., M. Greaves, J. Hendler, and H. Hirsh, 2014, "Amplify Scientific Discovery with Artificial Intelligence," *Science*, Volume 346, Issue 6206, pp. 171–172.

GitHub, 2016, "Kokkos," https://github.com/kokkos/kokkos, accessed March 1, 2016.

Hall, S.R., et al., 1991, "The Crystallographic Information File (CIF) — A New Standard Archive File for Crystallography," *Acta Crystallographica*, Section A, 47: p. 655.

Hall, S.R., and B. McMahon (eds.), 2005, "International Tables for Crystallography," Volume G: *Definition and Exchange of Crystallographic Information*, Springer: Dordrecht.

Hardy, D.J., Z. Wu, J.C. Phillips, J.E. Stone, R.D. Skeel, and K. Schulten, 2015, "Multilevel Summation Method for Electrostatic Force Evaluation," *Journal of Chemical Theory and Computation*, Volume 11, Issue 2, pp. 766–779.

Helland, B., 2015, *BES Requirements Review*, presentation at the Basic Energy Sciences Exascale Requirements Review Meeting, U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, Rockville, Maryland, November 3.

Huan, T.D., S. Boggs, G. Teyssedre, C. Laurent, M. Cakmak, S. Kumar, and R. Ramprasad, 2016, "Advanced Polymeric Dielectrics for High Energy Density Applications," *Progress in Materials Science*, Volume 83, pp. 236–239. DOI: 10.1016/j.pmatsci.2016.05.001.

ICDC (Inorganometallic Catalysts Design Center), 2016, http://www.chem.umn.edu/icdc, accessed February 6, 2016.

Kirklin, S., J.E Saal, B. Meredig, A. Thompson, J.W Doak, M. Aykol, S. Rühl, and C. Wolverton, 2015, "The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies," *npj Computational Materials*, Volume 1, article number 15010. DOI:10.1038/npjcompumats.2015.10.

Kurzak, J., and B.M. Pettitt, 2006, "Fast Multipole Methods for Particle Dynamics," *Molecular Simulation*, Volume 32, Issues 10–11, p. 775–790.

Li, Y., D. Zakharov, S. Zhao, R. Tappero, U. Jung, A. Elsen, Ph. Baumann, R.G. Nuzzo, E.A. Stach, and A.I. Frenkel, 2015, "Complex Structural Dynamics of Nanocatalysts Revealed in Operando Conditions by Correlated Imaging and Spectroscopy Probes," *Nat. Comm.* **6**, 7583.

Li Manni, G., R.K. Carlson, S. Luo, D. Ma, J. Olsen, D.G. Truhlar, and L. Gagliardi, 2014, "Multi-Configuration Pair-Density Functional Theory," *J. Chem. Theory Comp.,* Volume 10(9), pp. 3669–3690.

MacDougall, F., J. Ennis, X.H. Yang, R.A. Cooper, J.E. Gilbert, J.F. Bates, C. Naruo, M. Schneider, N. Keller, S. Joshi, T.R. Jow, J. Ho, C.J. Scozzie, and S.P.S. Yen, 2009, "High Energy Density Capacitors for Pulsed Power Applications," in *IEEE Pulsed Power Conference*, IEEE, Piscataway, New Jersey, United States, p. 774.

Mannodi-Kanakkithodi, A., G.M. Treich, T.D. Huan, R. Ma, M. Tefferi, Y. Cao, G.A. Sotzing, and R. Ramprasad, 2016, "Rational Co-Design of Polymer Dielectrics for Energy Storage," *Advanced Materials*. Volume 28, Issue 30, pp. 6277–6291. DOI: 10.1002/adma.201600377.

Matsumaru, N., 2004, "Self-Adaptive Scouting—Autonomous Experimentation for Systems Biology," in *Applications of Evolutionary Computing*, Berlin: Springer-Verlag.

Mills, D.P., F. Moro, J. McMaster, J. van Slageren, W. Lewis, A.J. Blake, and S.T. Liddle, 2011, *Nat. Chem.,* Volume 3, 454.

Minamoto, Y., and J. Chen, 2016, "DNS of a Turbulent Lifted DME Jet Flame," *Combustion and Flame* **169**, 38–50, July, http://www.sciencedirect.com/science/article/pii/S0010218016300542, accessed May 11, 2016.

Mok, D.K.W., R. Neumann, and N.C. Handy, 1996, "Dynamical and Nondynamical Correlation," *J. Phys. Chem.,* Volume 100(15), pp. 6225–6230.

Nalwa, H.S. (editor), 1999, *Handbook of Low and High Dielectric Constant Materials and Their Applications*, San Diego, California: Academic Press.

Nomura, K., R.K. Kalia, Y. Li, A. Nakano, P. Rajak, C. Sheng, K. Shimamura, F. Shimojo, and P. Vashishta, 2016, "Nanocarbon Synthesis by High-Temperature Oxidation of Nanoparticles," *Scientific Reports*, Volume 6, 24109.

Obama, B.H., 2015, *U.S. Presidential Executive Order — Creating a National Strategic Computing Initiative*, The White House, Office of the Press Secretary, Washington, D.C., July 29.

Olivucci, M., and A. Sinicropi, 2005, "Computational Photochemistry," *Theor. Comput. Chem.,* Volume 16, pp. 1−33.

OSTP (Office of Science and Technology Policy), 2013, *Increasing Access to the Results of Federally Funded Scientific Research*, Executive Office of the President, February 22, https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf, accessed February 7, 2016.

Pronk, S., S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, 2013, "GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit," *Bioinformatics*, Volume 29, Issue 7, pp. 845–854.

Reed, D.A., and J. Dongarra, 2015, "Exascale Computing and Big Data," *Communications of the ACM*, Volume 58, No. 7, pp. 56–68.

The Research Data Alliance Photon and Neutron Science Interest Group, 2015, *Synchrotron Radiation News*, Volume 28(2), pp. 43–47.

Romero, N.A., A. Nakano, K.M. Riley, F. Shimojo, R.K. Kalia, P. Vashishta, and P.C. Messina, 2015, "Quantum Molecular Dynamics in the Post-Petaflops Era," *Computer*, IEEE, Volume 48, Issue 11, pp. 33–41.

Sagui, C., and T. Darden, 2001, "Multigrid Methods for Classical Molecular Dynamics Simulations of Biomolecules," *The Journal of Chemical Physics*, Volume 114, 6578.

Sarrao, J.L., and G.W. Crabtree, 2015, *MRS Bulletin*, Volume 40(11), pp. 919–922, http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=10030177&fileId=S0883769415002651, accessed April 7, 2016.

Shapiro, D.A., Y.-S. Yu, T. Tyliszczak, J. Cabana, R. Celestre, W. Chao, K. Kaznatcheev, A.L. David Kilcoyne, F. Maia, S. Marchesini, Y. Shirley Meng, T. Warwick, L. Lisheng Yang, H.A. Padmore, 2014, "Chemical Composition Mapping with Nanometre Resolution by Soft X-ray Microscopy," *Nature Photonics*, Volume 8, pp. 765–769, doi:10.1038/nphoton.2014.207.

Shekhar, A., K. Nomura, R.K. Kalia, A. Nakano, and P. Vashishta, 2013, "Nanobubble Collapse on a Silica Surface in Water: Billion-Atom Reactive Molecular Dynamics Simulations," *Physical Review Letters*, Volume 111, 184503.

Shimamura, K., F. Shimojo, R.K. Kalia, A. Nakano, K. Nomura, and P. Vashishta, 2014, "Hydrogen-on-Demand Using Metallic Alloy Nanoparticles in Water," *Nano Letters*, Volume 14, Issue 7, pp. 4090–4096.

Small, P.E., R.K. Kalia, A. Nakano, and P. Vashishta, 2016, "Order-Invariant Real Number Summation: Circumventing Accuracy Loss for Multimillion Summands on Multiple Parallel Architectures," *Proceedings of the International Parallel and Distributed Processing Symposium*, IEEE, pp. 152–160.

Smeets, P.J.M., K.R. Cho, R.G.E. Kempen, N.A.J.M. Sommerdijk, and J.J. De Yoreo, 2015, "Calcium Carbonate Nucleation Driven by Ion Binding in a Biomimetic Matrix Revealed by *In Situ* Electron Microscopy," *Nature Materials*, Volume 14, pp. 394–399, January 26.

SNL (Sandia National Laboratories), 2015, *Sandia Mission Computing: 2015 HPC Annual Report*, p. 36, http://www.sandia.gov/news/publications/computing_reports/_assets/documents/HPC_AnnualReport2015_FNL.pdf, accessed March 1, 2016.

Starr, P., T. Maier, and T.C. Schulthess, 2014, "Two-particle Correlations in a Dynamic Cluster Approximation with Continuous Momentum Dependence: Superconductivity in the 2D Hubbard Model," February 18, http://arxiv.org/pdf/1402.4329.pdf, accessed February 7, 2016.

Stein Research Group, 2016, *Stein Publications Page*, Andreas Stein, Department of Chemistry, University of Minnesota, http://www.chem.umn.edu/groups/stein/publications.html, accessed May 11, 2016.

Szalay, P.G., T. Muller, G. Gidofalvi, H. Lischka, and R. Shepard, 2012, "Multiconfiguration Self-Consistent Field and Multireference Configuration Interaction Methods and Applications," *Chem. Rev.*, Volume 112(1), pp. 108–181.

Wang, Y.-G., D. Mei, V.-A. Glezakou, J. Li, and R. Rousseau, 2015, "Dynamic Formation of Single-Atom Catalytic Active Sites on Ceria-Supported Gold Nanoparticles," *Nature Comm.*, Volume 6, 6511.

Zhao, L.-D., J. He, S. Hao, C.-I. Wu, T.P. Hogan, C. Wolverton, V.P. Dravid, and M.G. Kanatzidis, 2012, "Raising the Thermoelectric Performance of p-Type PbS with Endotaxial Nanostructuring and Valence-Band Offset Engineering Using CdS and ZnS," *Journal of the American Chemical Society*, Volume 134 (39), pp. 16327–16336.

6 ACRONYMS AND ABBREVIATIONS

1D, 2D, 3D	one-, two-, three-dimensional
AIMD	*ab initio* molecular dynamics
ALCC	ASCR Leadership Computing Challenge
AMR	adaptive mesh refinement
APS	Advanced Photon Source (Argonne National Laboratory)
ASCR	Advanced Scientific Computing Research (DOE Program Office)
BES	Basic Energy Sciences (DOE Program Office)
BESAC	Basic Energy Sciences Advisory Committee
BOPP	biaxially oriented polypropylene
BSE	Bethe-Salpeter equation
CC	coupled-cluster theory
CCSD(T)	coupled-cluster single double (triple)
CECAM	Centre Européen de Calcul Atomique et Moléculaire
CPU	central processing unit
CSGF	Computational Science Graduate Fellowship
CSR	coherent synchrotron radiation
DAQ	data acquisition
DFT	density functional theory
DMFT	dynamical mean-field theory
DMRG	density-matrix renormalization group
DNS	direct numerical simulation
DOE	U.S. Department of Energy
DSL	domain-specific language
FEL	free electron laser
FFT	fast Fourier transform
GPU	graphics processing unit
HPC	high-performance computing

ICSD	Inorganic Crystal Structure Database
ID	identity
IMA	Institute for Mathematics and its Applications
IUCr	International Union of Crystallography
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
LAPACK	Linear Algebra Package
LCF	leadership computing facility
LCLS	Linac Coherent Light Source (Stanford University)
LES	Large Eddy simulation
MBPT	many-body perturbation theory
MD	molecular dynamics
MM	molecular mechanics
MOF	metal-organic framework
MP2	second-order perturbation theory
MPI	message passing interface
NERSC	National Energy Research Scientific Computing Center (Lawrence Berkeley National Laboratory)
NSCI	National Strategic Computing Initiative
NSLS	National Synchrotron Light Source (Brookhaven National Laboratory)
OLCF	Oak Ridge Leadership Computing Facility (Oak Ridge National Laboratory)
OQMD	Open Quantum Materials Database
PI	principal investigator
PME	Particle-Mesh-Ewald
QM	quantum mechanics
QMC	quantum Monte Carlo
QMD	quantum molecular dynamics
RCCI	rate-controlled compression ignition
RF	radio frequency
RIXS	resonant inelastic X-ray scattering

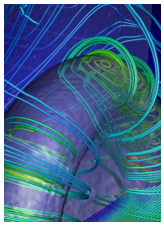
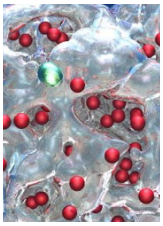
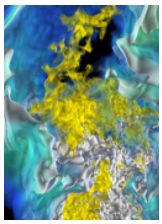
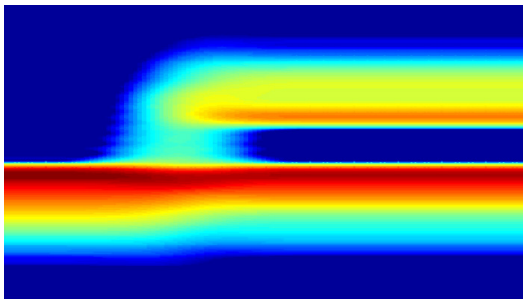
SC	Office of Science (DOE)
SciDAC	Scientific Discovery through Advanced Computing
STXM	scanning transmission X-ray microscopy
TDDFT	time-dependent DFT
UQ	uncertainty quantification
XFEL	X-ray free electron laser

Units of Measure

atm	atmosphere(s)
fs	femtosecond(s)
GB	gigabyte(s)
k	thousand(s)
K	Kelvin
m	meter(s)
M	million
meV	mill-electron volt(s)
MV	megavolt(s)
nm	nanometer(s)
ns	nanosecond(s)
PB	petabyte(s)
ps	picosecond(s)
s	second(s)

BES

BASIC ENERGY SCIENCES

APPENDICES: MEETING MATERIALS

An Office of Science review sponsored jointly by
Advanced Scientific Computing Research and Basic Energy Sciences

APPENDIX A: BASIC ENERGY SCIENCES ORGANIZING COMMITTEE AND MEETING PARTICIPANTS

A.1 BES ORGANIZING COMMITTEE

Michael Banda, Lawrence Berkeley National Laboratory (MJBanda@lbl.gov)

Thomas Devereaux,[b] Stanford Linear Accelerator Center (tjd@slac.stanford.edu)

Theresa Windus, Iowa State University/Ames Laboratory (twindus@iastate.edu)

Julia C. White, Oak Ridge Leadership Computing Facility (whitejc@ornl.gov)

A.2 BES SPONSORS AND REPRESENTATIVES

James Davenport, DOE/Office of Basic Energy Sciences (James.Davenport@science.doe.gov)

Matthias Graf, DOE/Office of Basic Energy Sciences (Matthias.Graf@science.doe.gov)

Carolyn Lauzon, DOE/Office of Advanced Scientific Computing Research (carolyn.lauzon@science.doe.gov)

Eliane Lessner, DOE/Office of Basic Energy Sciences (Eliane.Lessner@doe.gov)

Mark Pederson, DOE/Office of Basic Energy Sciences (Mark.Pederson@science.doe.gov)

A.3 BES MEETING PARTICIPANTS

Katie Antypas	National Energy Research Scientific Computing Center
Michael Banda	Lawrence Berkeley National Laboratory
Tunna Baruah[a]	University of Texas, El Paso
Anouar Benali	Argonne Leadership Computing Facility
Steven Binkley	U.S. Department of Energy
Michael Borland	Argonne National Laboratory
David Ceperley[a]	University of Illinois
Christine Chalk	U.S. Department of Energy
Jackie Chen[a]	Sandia National Laboratories
Hai-Ping Cheng[a,b]	University of Florida
Aurora Clark[a,b]	Washington State
Richard Coffey	Argonne Leadership Computing Facility
Joaquin Correa	National Energy Research Scientific Computing Center
Vince Dattoria	U.S. Department of Energy
Jim Davenport	U.S. Department of Energy
Juan de Pablo	University of Chicago

[a]Authored or co-authored a White Paper (Appendix C).
[b]Authored or co-authored a Case Study (Appendix D).

Wibe DeJong[a,b]	Lawrence Berkeley National Laboratory
Jack Deslippe[a]	National Energy Research Scientific Computing Center
Thomas Devereaux[b]	Stanford Linear Accelerator Center
Lori Diachin	Lawrence Livermore National Laboratory
David Dixon[a]	The University of Alabama
Thomas Dunning[a]	Pacific Northwest National Laboratory
Markus Eisenbach	Oak Ridge Leadership Computing Facility
Marivi Fernandez-Serra[a]	Stony Brook University
James Freericks[a,b]	Georgetown University
Laura Gagliardi[a,b]	University of Minnesota
Giulia Galli[b]	University of Chicago/Argonne National Laboratory
Bruce Garrett[a]	Pacific Northwest National Laboratory
Timothy Germann	Los Alamos National Laboratory
Dave Goodwin	U.S. Department of Energy
Mark Gordon[a,b]	Iowa State University/Ames Laboratory
Matthias Graf	U.S. Department of Energy
Stephen Gray[a]	Argonne National Laboratory/ Center for Nanoscale Materials
Emanuel Gull[a]	University of Michigan
Francois Gygi[b]	University of California, Davis
Robert Harrison	Brookhaven National Laboratory
Bill Harrod	U.S. Department of Energy
Robin Hayes	U.S. Department of Energy
Barb Helland	U.S. Department of Energy
Alexander Hexemer[a,b]	Lawrence Berkeley National Laboratory
Mark Jarrell[a]	Louisiana State University
Paul Kent[a]	Oak Ridge National Laboratory
Kerstin Kleese van Dam	Brookhaven National Laboratory
Stephen Klippenstein[a,b]	Argonne National Laboratory
Gabriel Kotliar	Rutgers University
Karol Kowalski[a,b]	Pacific Northwest National Laboratory
Jeff Krause	U.S. Department of Energy
Carolyn Lauzon	U.S. Department of Energy
Steven Lee	U.S. Department of Energy

[a]Authored or co-authored a White Paper (Appendix C).
[b]Authored or co-authored a Case Study (Appendix D).

Xiaosong Li[a,b]	University of Washington
Dmitry Liakh	Oak Ridge Leadership Computing Facility
Thomas Maier[b]	Oak Ridge National Laboratory
Andrea Manning	Argonne National Laboratory
Paul Messina	Argonne Leadership Computing Facility
Andrew Millis[b]	Columbia University
Chris Mundy[a,b]	Pacific Northwest National Laboratory
Habib Najm	Sandia National Laboratories
Lucy Nowell	U.S. Department of Energy
Thanos Panagiotopoulos[a]	Princeton University
Dula Parkinson[b]	Lawrence Berkeley National Laboratory
John Pask[a]	Lawrence Livermore National Laboratory
Amedeo Perazzo[a,b]	Stanford Linear Accelerator Center
Robinson Pino	U.S. Department of Energy
David Prendergast	Lawrence Berkeley National Laboratory
Thomas Proffen	Oak Ridge National Laboratory
John Rehr[a]	University of Washington
Katherine Riley	Argonne Leadership Computing Facility
Nichols Romero	Argonne Leadership Computing Facility
Lauren Rotman	ESnet
John Russell	U.S. Department of Energy
Robert Ryne	Lawrence Berkeley National Laboratory
Sonia Sachs	U.S. Department of Energy
John Sarrao	Los Alamos National Laboratory
Richard Scalettar	University of California, Davis
Thomas Schulthess	ETH Zurich
Andy Schwartz	U.S. Department of Energy
William Scullin	Argonne Leadership Computing Facility
Annabella Selloni[a]	Princeton University
Jamie Sethian[b]	University of California, Berkeley
Ilja Siepmann[a,b]	University of Minnesota
David Skinner	National Energy Research Scientific Computing Center
Lyuda Slipchenko[a,b]	Purdue University
Eric Stach	Brookhaven National Laboratory
Dale Stansberry	Oak Ridge Leadership Computing Facility

[a]Authored or co-authored a White Paper (Appendix C).
[b]Authored or co-authored a Case Study (Appendix D).

Mark Stevens[a,b]	Sandia National Laboratories
T.P. Straatsma	Oak Ridge Leadership Computing Facility
Rangan Sukumar	Oak Ridge Leadership Computing Facility
Bobby Sumpter[a]	Oak Ridge National Laboratory
Ceren Susut	U.S. Department of Energy
Brian Toby[a]	Argonne National Laboratory
Craig Tull[b]	Lawrence Berkeley National Laboratory
Ed Valeev[a]	Virginia Tech
Priya Vashishta[a]	University of Southern California
Jack Wells	Oak Ridge Leadership Computing Facility
Julia White	Oak Ridge Leadership Computing Facility
Stuart Wilkins	Brookhaven National Laboratory
Theresa Windus	Iowa State University/Ames Laboratory
Chris Wolverton	Northwestern University
Ping Yang[a]	Los Alamos National Laboratory

[a]Authored or co-authored a White Paper (Appendix C).
[b]Authored or co-authored a Case Study (Appendix D).

APPENDIX B: BASIC ENERGY SCIENCES MEETING AGENDA

MONDAY, NOVEMBER 2

7:00 Evening meeting in the Lincoln Room for chairs and breakout leads to discuss the next day's activities.

TUESDAY, NOVEMBER 3

7:30 Registration, Refreshments

8:30 Welcome and Introductions
Michael Banda, Tom Devereaux, and Theresa Windus

8:40 View from Basic Energy Sciences
Harriet Kung, Associate Director, BES

9:20 Genesis of this Meeting
Barb Helland, ASCR

9:50 Transformative Opportunities to Enable the Transformative Opportunities
John Sarrao, Los Alamos National Laboratory

10:20 Break

10:40 BES Facilities: Data Production, Analysis, and Storage
Craig Tull, Lawrence Berkeley National Laboratory

11:10 Requirements Gathering for ASCR Facilities
Katherine Riley, Argonne National Laboratory

11:50 Charge to Working Groups
Michael Banda, Tom Devereaux, and Theresa Windus

12:10 Working Lunch

1:00 Breakout Sessions: Science Drivers

Quantum Materials, Core Challenges in Heavy-Element Chemistry, Exotic States, and Emergence
Aurora Clark, Washington State University, and Andy Millis, Columbia University

Catalysis, Photosynthesis, Light Harvesting, and Combustion
Laura Gagliardi, University of Minnesota, and Thanos Panagiotopoulos, Princeton University

Materials and Chemical Discovery
Ilja Siepmann, University of Minnesota, and Chris Wolverton, Northwestern University

Computing & Data Challenges @ BES Facilities
Thomas Proffen, Oak Ridge National Laboratory

2:30 Break

2:50 Breakout Sessions: Science Drivers (cont.)

4:00 Break

4:15 Preliminary Reports: Breakout Leads, 15 minutes each

5:30 Questions and Answers
Harriet Kung, Associate Director, BES, and Steve Binkley, Associate Director, ASCR

6:15 Dinner on your own

WEDNESDAY, NOVEMBER 4

8:00 Refreshments

8:30 Computational materials science: a multi-disciplinary enterprise
Thomas Schulthess, Director, Swiss National Supercomputing Centre, ETH Zurich

9:10 Summary by Chairs, Outline of Report
Michael Banda, Tom Devereaux, and Theresa Windus

9:30 Break

10:00 Breakouts

Continuation of 1:00 PM Tuesday Sessions; Outline Sections, Start Writing

Quantum Materials, Core Challenges in Heavy-Element Chemistry,

Exotic States, and Emergence
Aurora Clark, Washington State University, and Andy Millis, Columbia University

Catalysis, Photosynthesis, Light Harvesting, Combustion
Laura Gagliardi, University of Minnesota, and Thanos Panagiotopoulos, Princeton University

Materials and Chemical Discovery
Ilja Siepmann, University of Minnesota, and Chris Wolverton, Northwestern University

Computing & Data Challenges @ BES Facilities
Thomas Proffen, Oak Ridge National Laboratory

12:00 Working Lunch

1:00 Breakout Sessions

Computing Landscape

Next Generation Programming
Jackie Chen, Sandia National Laboratories, and Anouar Benali, Argonne National Laboratory

Advances in Quantum Algorithms
Mark Gordon, Iowa State University, and Paul Kent, Oak Ridge National Laboratory

Math & Computer Science
Jamie Sethian, University of California, Berkeley, and Lori Diachin, Lawrence Livermore National Laboratory

Data Management, Analytics, Visualization & Preservation
Craig Tull, Lawrence Berkeley National Laboratory, and Kerstin Kleese van Dam, Brookhaven National Laboratory

Soft Matter, Biochemistry, Bioinspired Materials: Included here are Multiscale Methods & Algorithms and Classical Molecular Dynamics
Priya Vashishta, University of Southern California, and Mark Stevens, Sandia National Laboratories

2:30 Break

4:00 Break

4:20 Reports on Wednesday Breakouts, Breakout Leads, 15 minutes each

5:20	Summary and Thanks from Chairs
5:45	End for Most Participants

THURSDAY, NOVEMBER 5

8:30	Co-chairs, Leads, Writers meet to continue working on report
12:00	End of Meeting

APPENDIX C: BASIC ENERGY SCIENCES WHITE PAPERS

The following white papers were submitted by the authors listed below in advance of the Exascale Requirements Review to guide both the agenda and meeting discussions.

C.1 White Papers Addressing Novel Quantum Materials and Chemicals

C-3 David Ceperley (University of Illinois) and Paul Kent (Oak Ridge National Laboratory)

C-5 Aurora Clark (Washington State University)

C-7 Jack Deslippe (NERSC); Charles Lena and James Chelikowsky (University of Texas)

C-9 Jim Freericks (Georgetown University)

C-12 Emanuel Gull (University of Michigan)

C-15 J.J. Rehr, et al. (University of Washington)

C-19 Peter Sushko, Thom Dunning, Karol Kowalski, and Niri Govind (Pacific Northwest National Laboratory)

C-22 Ping Yang (Los Alamos National Laboratory)

C.2 White Papers Addressing Catalysis, Photosynthesis and Light Harvesting, and Combustion

C-25 Tunna Baruah (University of Texas, El Paso)

C-29 Jackie Chen (Sandia National Laboratories)

C-32 W.A. de Jong (Lawrence Berkeley National Laboratory)

C-35 David A. Dixon (The University of Alabama)

C-38 Laura Gagliardi (University of Minnesota)

C-41 Bruce Garrett and Roger Rousseau (Pacific Northwest National Laboratory)

C-44 Mark Gordon (Ames Laboratory)

C-46 Stephen J. Klippenstein (Argonne National Laboratory)

C-49 Xiaosong Li (University of Washington)

C-52 Annabella Selloni (Princeton University)

C-54 Lyudmila V. Slipchenko (Purdue University)

C-56 Edward F. Valeev (Virginia Tech)

C.3 White Papers Addressing Materials and Chemical Discovery

C-58 Hai-Ping Cheng (University of Florida)

C-61 Marivi Fernandez Serra (Stony Brook University)

C-63 Christopher J. Mundy and Greg Schenter (Pacific Northwest National Laboratory)

C-65 John Pask (Lawrence Livermore National Laboratory)

C-68 J. Ilja Siepmann (University of Minnesota)

C-71 Bobby G. Sumpter (Oak Ridge National Laboratory)

C.4 White Paper Addressing Soft Matter

C-75 Priya Vashishta (University of Southern California, in collaboration with Rajiv K. Kalia and Aiichiro Nakano)

C-78 Thanos Panagiotopoulos (Princeton University)

C.5 White Papers Addressing Advances in Algorithms for Quantum Systems

C-80 Emily Carter (Princeton University)

C-83 Mark Jarrell (Louisiana State University)

C-87 Karol Kowalski (Pacific Northwest National Laboratory)

C.6 White Papers Addressing Computing and Data Challenges @ BES Facilities

C-90 Maria Chan, Pierre Darancet, Stephen Gray, Ian McNulty, Subramanian Sankaranarayanan, and Michael Sternberg (Argonne National Laboratory)

C-93 Alexander Hexemer and Dula Parkinson (Lawrence Berkeley National Laboratory)

C-96 Amedeo Perazzo (Stanford Linear Accelerator Center)

C-99 Mark Stevens (Sandia National Laboratories)

C-103 Brian Toby (Argonne National Laboratory)

Mathematics and Computer Science Transforming BES Science

Although many of the white papers discuss issues related to mathematics and computer science, no one paper is about this topic only.

Next-Generation Workforce

Although many of the white papers mention the need for a next-generation workforce and workforce development, no one paper is about this topic only.

C.1 White Papers Addressing Novel Quantum Materials and Chemicals

BES White Paper – David Ceperley (University of Illinois)
and Paul Kent (Oak Ridge National Laboratory)

1. **Please specify the current science drivers for your field of research.**

Properties of materials are at the basis of much of science, technology, and applications as recognized in the National Genome Initiative.[1] Today, accurate methods have been developed to calculate properties of material *ab initio ,* that is, based on the fundamental laws of physics that describing atoms, molecules, and solids, without the necessity of using empirical information. Among the algorithms are those based on mean field, density functional theory, quantum Monte Carlo, and perturbation theory. Because of advances in algorithms and computer hardware, impressive progress has been made. More accurate methods are increasingly able to be applied, particularly to materials such as transition metal oxides. These energy-relevant materials have complex electronic phase diagrams that require the use of more advanced methods for sufficiently accurate predictions.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

We can confidently expect that significant progress will be made. Precise quantum simulations will be made of small systems (containing less than a thousand electrons) for materials composed of most of the atoms in the periodic table. The methods will be tuned to achieve quantitatively accurate energetic calculations for selected systems. However, calculations of larger systems and for properties needed to confront experiment will be lacking in many cases. Significant improvements over extant systems are required, in general. In the specific case of quantum Monte Carlo, extant computing ecosystems will limit the method's application to simpler systems, to lighter elements, and to simpler properties. For example, energy gradients (forces) will not be routinely applied to complex or heavy element systems, and inclusion of spin-orbit effects will be limited.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

It is unlikely that "Materials Design" will be widely feasible with the extant computing ecosystems, especially for challenging materials such as those containing transition metal elements. Current methods to determine stability of structures, the first question to be asked in materials design, are based on using either empirical potentials or on density functional calculations. However, for most materials, the accuracy of these methods,[2–5] unless they are tuned by experiment, is insufficient to determine whether a given structure is stable. Experimental scanning for materials design is expensive. Computational methods that would be accurate enough, such as quantum Monte Carlo or those based on perturbation theory such as the GW method, take several orders of magnitude more computer time; such calculations are just feasible now for looking at individual structures, but are too costly to be used to scan over a multitude of combinatorial possibilities needed for materials discovery and design. Many experimental and technologically relevant properties remain difficult to access by simulation, particularly those related to the lifetime of materials. These properties require both larger computing systems and improved methods to treat the complexities of the many possible processes.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Large increase in computing resources.	We have not reached threshold where scanning thousands of materials with high accuracy is feasible or where, in general, predictions can be made with sufficient confidence.
2. Development of algorithms for MGI	Individual methods such as quantum Monte Carlo need to be extended. Connections between the methods need to be made.
3. Development of QMC software	Software to take advantage of the hardware and improved algorithms need to be developed. Today application of the software is largely supervised by experts. More automated and robust approaches are required.
Impede	**Why?**
1. Lack of resources	Without a coordinated development of computers, algorithms and software, materials design will not happen.
2. Career development	Positions need to be available to reward software development activities.
3. Computer environment	Different methods have different computing requirements. We need a heterogeneous environment, stretching from the desktop to the large scale facilities, able to accommodate research needs.

5. **Characterize the data ecosystem aspects**

During the last few years, several projects have begun to develop large databases of materials properties computed using empirically tuned density functional theory calculations. It will be important to develop mechanisms to attach or link high accuracy results to these databases and enable their discovery.

6. **References**

1. https://www.whitehouse.gov/mgi. [Materials Genome Initiative description and motivation.]
2. G. Hautier et al., "Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability" Phys. Rev. **B 85**, 155208 (2012).
3. L. Wang et al., "Oxidation energies of transition metal oxides within the GGA+U Framework," Phys. Rev. **B 73**, 195107 (2006).
4. A. Jain et al., "Formation enthalpies by mixing GGA and GGA+U Calculations," Phys. Rev. **B 84**, 045115 (2011). [In systems with both localized and delocalized states, both GGA and GGA+U fail to give an accurate description.]
5. A. Narayan, A. Bhutani, D. P. Shoemaker, and L. K. Wagner, "Towards finding new ternary selenides and sulphides" (2015). [Shows how unsuccessful DFT prediction is for these materials.]

BES White Paper – Aurora Clark (Washington State University)

1. Please specify the current science drivers for your field of research.

Several science drivers illustrate where theory and simulation are under-utilized but could play a transformative role. This pertains to DOE interests that encompass complex, multi-component chemical systems – defined by non-ideality; multiple equilibria; and potentially driven by external perturbations, excited state reactivity, or the presence of interfaces.

Interfacial phenomena — Interfaces exhibit properties associated with the nanoscale in one dimension, and the macroscale in others. This leads to unique physicochemical characteristics and reactivity essential to the success of a device or process. Consider nanoparticles used in solar energy: here large surface area-to-volume ratios, surface defects, and strong chemical coupling to solvent necessitate computational methods beyond those used for the bulk.

Charge transfer in complex environments – The ability to accurately model charge transfer phenomena in molecules/materials would lead to great strides in optimizing systems for capturing and storing solar energy, and for using light to catalyze chemical reactions. This requires a fundamental understanding of the interplay between photo-excitation and the resulting proton and electron transfer processes that occur over a range of timescales in complex and fluctuating environments, such as in solution, at surfaces, and in disordered polymers.

Multicomponent solutions – Multicomponent solutions often exhibit emergent behavior. This is typified in DOE environmental management and waste processing. Here, high ionic strength, extreme pH, complex crystallization and dissolution, micro-environments and phase separation may be combined. Major investments in theory are required for techniques that provide information about the composition, diffusion and flow of solutes, gases, precipitation and dissolution of solids, and nucleation events in the complex environments of waste forms, waste tanks, and the subsurface.

Data-enabled prediction — State-of-the-art experiments provide vast multi-dimensional data; however, new data fusion, analytics, and computational capabilities are required to allow for knowledge inference. Field-scale measurements and computational predictions can represent combined chemical, biological, and physical parameters from a vast and temporally evolving data ecosystem.

2. Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.

Interfacial phenomena – Reactive phenomena including redox reactivity, optical excitations, and solvent effects will need accurate quantum mechanical (QM) treatments (embedding methods and others) that can capture the often localized nature of these processes while employing realistic large periodic boundary conditions. We must move beyond cluster models as representing interfaces.

Charge transfer in complex environments – It should be possible to compute electronic excitations while treating a large portion of the environment with QM methods, therefore we can directly study the effect of the environment on the excitation process. With ab initio molecular dynamics, we can treat both electrons and nuclei with QM methods to better understand how the quantum nature of the solution affects the charge transfer process.

Multicomponent solutions – Force field development is a long-standing challenge, particularly those transferable to complex solutions. Interdisciplinary methods that leverage information science are needed to elucidate correlating relationships between local and extended structure and dynamics of complex solutions, enabling predictive theories for energy and mass transfer, aggregation/phase phenomena that hinder the rational design of new catalysts, materials, and optimal conditions.

Data-enabled prediction — Analytics approaches will need to be applied to fused data-sets so as to reveal hereto unknown correlations of behavior and transformation from the molecular to geologic

scale, as well as the feedback and cross-relation of seemingly unrelated phenomena and properties from molecular to field scale measurements and simulations.

3. Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

Without a major paradigm shift in computational methods or infrastructure, it will be difficult to fully address some of the major challenges described above, including:

- Moving beyond QM cluster models for representing reactions and localized phenomena at interfaces.
- Long-time scales for ab initio molecular dynamics that take into account electron and proton transfer on the time scale of solvent rearrangement (nanoseconds to microseconds).

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Application codes	Enable facile transfer between systems of interest (changing solute, solvent, surface, etc.)
2. Models and algorithms	Moving past cluster models for interfaces and allowing simultaneous treatment of local and periodic systems
3. Hardware resources	Large memory and parallel processing capabilities (such as GPUs) will enable more realistic QM calculations, and longer accessible timescales

Impede	Why?
1. Application codes	Codes cannot handle complex, multicomponent systems; for complicated solutions they must be modified or interfaced with other codes.
2.Data workflow	Management and analysis of data can become burdensome in multiscale simulations.
3. Workforce development	Training students and post-doctoral scholars is time-intensive. Computational training resources and workshops may help.

5. (Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.

Simulation data only, created at high velocity (MD) with large volume (trajectories) but little variety. Current plan to use on the fly analysis (using intermolecular network theory) to aid in rare event sampling and equilibration identification. Analysis can often create wider variety of data set.

1. References

DOE BES Basic Research Needs Report, “Basic Research Needs for Solar Energy Utilization” (2005).
DOE BES Basic Research Needs Report, “Basic Research Needs for Advanced Nuclear Energy Systems (2006).
DOE BES Basic Research Needs Report, “From Quanta to the Continuum: Opportunities for Mesoscale Science.”

BES White Paper – Large-Scale Calculations for Ground State and Excited State Properties with the PARSEC and BerkeleyGW Codes:

Jack Deslippe (NERSC); Charles Lena and James Chelikowsky (University of Texas)

Motivation

A number of materials science grand challenges that address the DOE mission require quantitative predictive simulations of increasing complexity. For example, the design of photovoltaics, thermo-electrics, batteries, efficient lighting, and nanoscale electronics are often associated with material properties of nanostructures at interfaces, or at junctions, or with defects. Such systems often require one to consider large-scale calculations. Moreover, for a simulation to be predictive, one must have a method that accurately describes complex many-body physics such as an accurate description of the dielectric screening.

Simulations based solely within common approximations to density functional theory (DFT) can accurately account for ground-state properties, although they are known to both quantitatively and qualitatively fail for predictions of important excited-state quantities of interest. Higher-level methods, like the GW-Bethe-Salpeter Equation (GW-BSE) approach, are essential to predict and understand excited-state device properties once the structural properties are known.

To this end, electronic-structure simulations are trending toward larger-scale calculations using more sophisticated physics. These simulations may include GW-BSE where the starting wave functions may be obtained from hybrid functionals. The confluence of these many-body approaches with new high-performance computational approaches is opportune as such approaches are well-positioned to take advantage of pre-exascale and exascale HPC hardware to a much larger degree than traditional approaches.

Whereas electronic-structure calculations in the past decade or so have been dominated by European developed codes, domestic DOE-supported applications like PARSEC and BerkeleyGW are leaders in large-scale ground-state and excited-state properties calculations. We therefore have the ability to closely couple energy-related applications with the DOE exascale strategy.

Here, we describe the requirements for using PARSEC and BerkeleyGW for 10,000-atom, excited-state simulations in the 2020–2023 time frame.

PARSEC

PARSEC is a real-space pseudopotential DFT code specifically designed for handling both confined and extended large systems with a variety of boundary conditions, e.g., confined systems such as nanocrystals, nanowires and bulk crystal, as well as generating the large numbers of electron orbitals, which are a key ingredient for GW calculations. PARSEC is currently capable of simulating 10,000+ atom systems.

PARSEC's use of a real-space basis eliminates the all-to-all communication typically present in plane-wave DFT codes when computing the Hamiltonian operating on the wave function as part of an iterative solution to the Schrödinger equation:

$$H\psi = \left[\frac{-\nabla^2}{2} + V_{ion} + V_H + V_{xc} \right] \psi = E\psi$$

Plane-wave-based codes typically need to perform many parallel FFTs from reciprocal space to real-space and back in order to apply the local potential to each orbital. In PARSEC, these potentials can be applied to each eigenstate via a dot-product that involves no communication, and the laplacian operator can be applied using a finite difference approach that involves only nearest-neighbor communication.

PARSEC additionally implements a spectrum slicing approach that allows different energy regions to be computed independently and in parallel, i.e., it can be used as a "parallel eigensolver." In this way, one can efficiently generate thousands of occupied and empty orbitals in parallel by breaking the computation into energy (or band) groups.

BerkeleyGW

The GW-BSE approach takes advantage of a large number of layers of parallelism for crystalline environments. For example, there are typically multiple levels of "band parallelism" (corresponding to occupied and unoccupied electron orbitals) as well as parallelism over frequency (corresponding to energy) and multiple levels of plane-wave or basis set parallelism. Contrast this with a typical DFT computation for crystals where parallelism is traditionally done over plane-waves and k-points (which are vanishingly small in large complex systems) and band parallelism is still considered a novelty.

The method has been shown to be able to effectively utilize internode parallelism, on-node parallelism in terms of threads and on core parallelism in terms of vectorization. BerkeleyGW has been shown to scale to 100,000+ CPU core calculations.

The traditional GW-BSE approach depends on a large number of node parallel FFTs, large GEMM type operations as well as matrix reduction operations that can be well blocked into various levels of the memory hierarchy. The FFTs are treated locally rather than the vast distributed FFTs in plane-wave DFT applications.

A Combined Approach to 10,000-atom Exascale GW Calculations

On pre-exascale and exascale HPC machines, we plan to use a combined approach where we couple PARSEC calculations on 10,000+ atom systems (including orbital generation) with BerkeleyGW excited-state property calculations.

We expect to run at concurrencies in excess of 1 million cpu cores utilizing both MPI and OpenMP (both loop level and task level) parallelism or an emerging task-based parallelism model. We expect to be able to take heavy advantage of vector processing units on exascale hardware.

Unlike traditional electronic structure calculations in applications that are crippled by all-to-all communication and a lack of multi-layered parallelism, we believe our approach is primed for exascale computing.

BES White Paper – Nonequilibrium Dynamical Mean-Field Theory:
Jim Freericks (Georgetown University)

1. **Please specify the current science drivers for your field of research.**

One of the main scientific themes of the Office of Science is to understand the behavior of materials driven far from equilibrium and possibly how to control this behavior. Experimental results from DOE facilities, like LCLS, are probing materials at ultrashort times and are driving them out of equilibrium with increasingly powerful driving fields. Yet a full understanding of the excitation process, a classification of possible transient or steady states, and a model for how the strongly correlated quantum materials relax after pumping remains a challenge to the theoretical/computational community. We are only starting to understand both how to simulate this behavior numerically and how to interpret it.

This work has wide-ranging potential. It may be possible to change the character of materials via nonequilibrium pumping. One can create nonequilibrium steady states (or long-lived transients) with different properties than the equilibrium phases (so-called hidden phases) which can be employed in devices that can have impacts on energy. One can study how energy flows in materials driven with ultralarge fields, and gain an understanding of the complex quantum-mechanical processes responsible for relaxation of excitations back to thermal equilibrium.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

The biggest need in the field is the development of robust, accurate, and efficient algorithms that allow strongly correlated nonequilibrium systems to be numerically solved. For nonequilibrium dynamical mean-field theory, the key is to find an impurity problem solver. Exact methods are known for a few models, but they suffer from not being able to be extended to long enough times to be effective in studying the evolution from excitation to transient behavior to relaxed or steady state. To date, nearly all results are either for noninteracting systems or are based on strong or weak-coupling perturbation theory. Many questions that can only be posed today, can be solved if such an algorithm can be developed. I believe at least on such algorithm is likely to be developed in the next 5–10 years.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Systems that are in nonequilibrium and inhomogeneous are likely to require resources that go beyond those available in the next decade. The ability to solve such problems would allow researchers to be able to properly calculate switching behavior in devices, amongst other things. Even with good algorithms to solve impurity problems, coupling them together to solve inhomogeneous problems requires far more computer power than what would be available in the near term, unless new algorithms are developed that can address and greatly improve upon the required efficiency.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Impurity algorithms	We cannot solve problems without them.
2. More efficient algorithms	Codes tend to be more compute-time limited than memory limited, although memory access latency is often the limiting factor in the efficiency of a code when scaled to large size.
3. Real-time visualization collaboration tools	These can make it easier for groups to collaborate on visualizing and interpreting data in real time at disparate locations.

Impede	Why?
1. Requiring all codes to scale to the full extent of the machine	Some algorithms have inherently scalar aspects to them which can impede scaling beyond some number of processors. They will nevertheless solve important problems and should not be neglected in the race to build and utilize the biggest and best machines.
2. Balancing clock cycle speed versus memory access speed through the bus	Most programs are memory limited in the sense that memory access is slower than the clock cycle speed which impeded computational performance.
3. Challenges with developing new efficient algorithms	Some hard problems may require significant team efforts to develop efficient algorithms to solve them, and support for these efforts might not be available.

5. (Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.

Our data needs tend to be modest in the high-performance computing arena. We do not produce nor do we need to analyze terabytes of data. Usually data analysis can be performed off-line after computation has been concluded. Some real-time collaboration tools might be helpful for analyzing data in the post-processing stage.

6. References

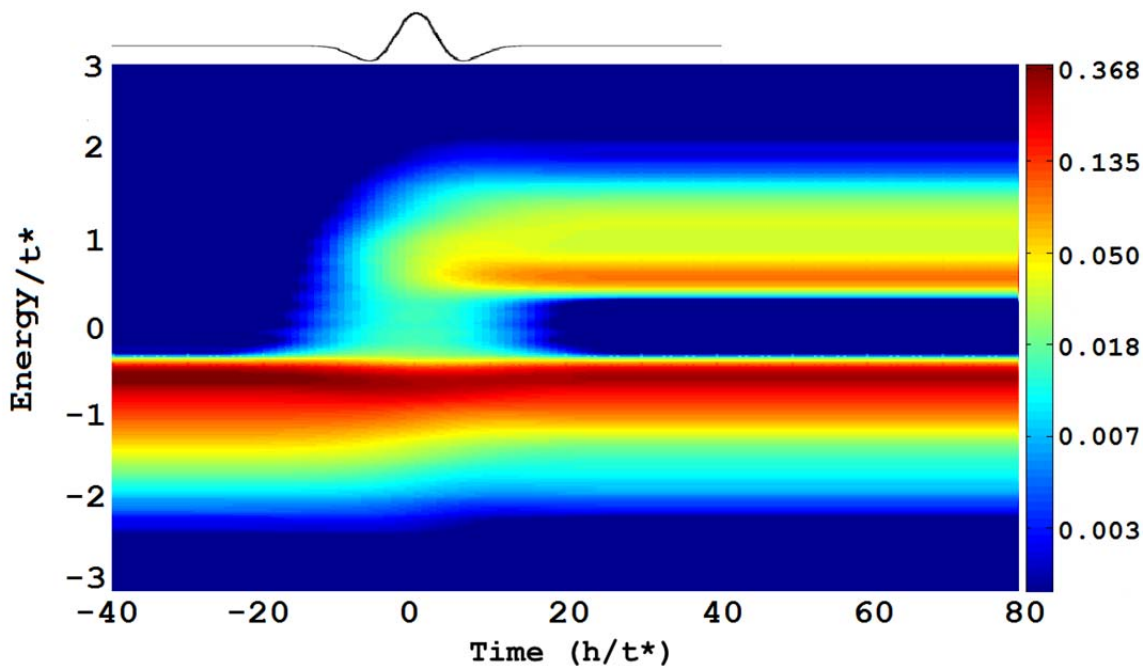
J. K. Freericks, V. M. Turkowski, and V. Zlatic', *Nonequilibrium dynamical mean-field theory*, Phys. Rev. Lett. **97**, 266408 (2006). 10.1103/PhysRevLett.97.266408.

J. K. Freericks, *Quenching Bloch oscillations in a strongly correlated material: The dynamical mean-field theory approach*, Phys. Rev. B **77**, 075109 (2008) 10.1103/PhysRevB.77.075109.

J. K. Freericks, H. R. Krishnamurthy, and Th. Pruschke, *Theoretical description of time-resolved photoemission spectroscopy: application to pump-probe experiments,* Phys. Rev. Lett. **102**, 136401 (2009). 10.1103/PhysRevLett.102.136401.

B. Moritz, A. F. Kemper, M. Sentef, T. P. Devereaux, and J. K. Freericks, *Electron-Mediated Relaxation Following Ultrafast Pumping of Strongly Correlated Materials: Model Evidence of a Correlation-Tuned Crossover between Thermal and Nonthermal States*, Phys. Rev. Lett. **111**, 077401 (2013). 10.1103/PhysRevLett.111.077401.
M. A. Sentef, A. F. Kemper, B. Moritz, J. K. Freericks, Z.-X. Shen, and T. P. Devereaux, *Examining Electron-Boson Coupling Using Time-Resolved Spectroscopy,* Phys. Rev. X **3**, 041033 (2013). 10.1103/PhysRevX.3.041033.
Wen Shen, Yizhi Ge, A. Y. Liu, H. R. Krishnamurthy, T. P. Devereaux, and J. K. Freericks, *Nonequilibrium "melting" of a charge density wave insulator via an ultrafast laser pulse,* Phys. Rev. Lett. **112**, 176404 (2014). 10.1103/PhysRevLett.112.176404.
M. A. Sentef, M. Claassen, A. F. Kemper, B. Moritz, T. Oka, J. K. Freericks, and T. P. Devereaux, *Theory of Floquet band formation and local pseudospin textures in pump-probe photoemission of graphene,* Nature Commun. **6,** 7047 (2015).

7. (Optional) Image



Wen Shen, Yizhi Ge, A. Y. Liu, H. R. Krishnamurthy, T. P. Devereaux, and J. K. Freericks, *Nonequilibrium "melting" of a charge density wave insulator via an ultrafast laser pulse,* Phys. Rev. Lett. **112**, 176404 (2014). 10.1103/PhysRevLett.112.176404.

BES White Paper – Emanuel Gull (University of Michigan)

1. **Please specify the current science drivers for your field of research.**

Lattice / Low-energy effective models: We are now at the point where static properties of lattice (Hubbard, tJ, bose Hubbard, etc.) models are essentially understood and can be computed numerically. The open questions are in dynamical aspects (excitations, spectral functions, etc.) and their relation to experimental probes of the compounds for which the low-energy model has been constructed. Also still open are details of the interplay between phases where many correlated phases exist.

Realistic materials simulations: Numerical methods for uncorrelated systems are well established and simulations are routinely done using, for example, density functional or GW techniques. Methods for correlated electron systems that take into account the full electronic structure have been devised (the LDA+DMFT method is one of them) but contain adjustable parameters that limit their predictive power (functionals, double counting correction, selection of orbitals, level of self-consistency, etc.).

Non-equilibrium dynamics: Theoretically, the calculation of the response of systems to conditions far outside of equilibrium remains a difficult problem. Even for the most simple quantum impurity system, initial transients, steady-state behavior, and and long-time relaxation dynamics of correlated electron systems could only be tackled in the last few years. For extended systems, the situation is worse: even the "simple" non-equilibrium dynamical mean field theory can only be solved with approximate numerical solvers. It is therefore not clear if discrepancies with experiment are due to the construction of model systems, the approximation of these model systems with a numerically tractable, or the approximate solution of the time-dependence of the model systems. Exciting new developments in non-equilibrium algorithms have only now started to appear.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Lattice / Low-energy effective models: The biggest open question in this field is the relation of low-energy effective models with experiment. I expect that in addition to the structural and energetic calculations that are done at present (which are expected to become more precise and applicable to more general/complex systems), it will become possible to compute experimentally relevant quantities to much higher accuracy, both on the single-particle level (ARPES, STM) and on the two-particle level (neutron, NMR, RIXS, Raman).

Realistic materials simulations: The main qualitative difference to present-day simulations will be the elimination of adjustable parameters and the assessment of uncertainties. Where a present–day, state-of-the-art simulation of, e.g., spectral function of a correlated material starts from a density functional calculation (choice of functional), followed by a "downfolding" procedure to an effective model (choice of orbital, choice of model, choice of interaction parameters) and an approximate solution of that model (DMFT? solver? Finite cluster?), realistic simulations will become more systematic, will eliminate free parameters, and will become reproducible (code/package/implementation independent).

Non-equilibrium dynamics: Non-equilibrium dynamics is in many ways similar to equilibrium 15 years ago: progress is hampered by an exponential scaling in all system parameters, so that large systems, low temperatures, or long times are exponentially difficult to obtain. Based on our progress in equilibrium, I expect that the next 5–10 years will bring larger systems and longer times but that the numerical cost will remain exponential, so that we will be severely limited in the phase space we can access. Whether the accessible phase space is large enough to be predictive and useful for experiment is not clear to me.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Challenge: Black-box materials simulation toolkit for correlated materials with predictive power: For many weakly correlated systems, DFT is a workhorse that provides good enough predictions of a large number of properties and systems that experimentalists care about. Codes are straightforward and

simple to run, results are reproducible. A similar toolbox for strongly correlated systems is currently out of reach. While algorithmic research is ongoing and promising, the testing, improving, packaging, and engineering of these methods will take more than 5–10 years.

Challenge: Long-time dynamics of strongly correlated materials:

Solving the time dependence in addition to the strong correlation nature of non-equilibrium systems is difficult. The multiple exponential barriers present in the system may be partially overcome but at this time, a realistic simulation of the time-dependent behavior of extended systems is not in sight.

Challenge: Large-scale searches of strongly correlated materials:

Large-scale searches of parameter space using standard electronic structure methods are well under way. For properties that depend on the strong correlation aspect of a material, where we have no black-box methods at present, a large-scale "blind" search of materials would require unbiased simulation without hand-tailored user input. This is presently not possible, and I see no way to enable it without first constructing a black-box materials science code that does not depend on user input.

What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Algorithms	Historically, algorithmic progress has by far beaten Moore's law. One recent example: with the development of continuous-time DMFT codes, a typical "single-site" impurity problem had its time to solution decreased by a factor of 27'000.
2. Reusable components and libraries	More and more good/domain-specific libraries are becoming available. Libraries encapsulate "difficult" but "reusable" physics. Long-term investment into libraries/reusable components is not compatible with "instant gratification" of publish or perish.
3. Open source codes	Open source libraries and application codes (ALPS, TRIQS, etc.) allow us to preserve knowledge over different generations of students and postdocs, and allow us to make codes and results verifiable and reproducible.

Impede	Why?
1. Accelerators / New programming language / new programming paradigms	Most of our new capability comes from new algorithms and new theory, not from running the same algorithms on faster/larger computers. Time that has to be spent for, say, adapting codes to graphics cards or accelerators is time that is not spent on algorithmic development.
2. Workforce development	Basic computer science skills are often missing for our students/postdocs. Programming is taught (if taught) by non-professionals. "Best practices" are usually 20–30 years behind industry standards (nobody knows about unit tests!).
3. Hardware resources	Scaling up a calculation from a trial code to a large-scale simulation takes 2–3 months – far shorter than a typical application period, with demand quickly ramping down to zero if a new/better/different idea appears. It is therefore vital to have access to large resources in a minimal amount of time.

Algorithmic speedup illustrated by problem size in a typical quantum impurity problem: from 1986 (black curve) to 2005 (blue curve) and 2006 (red curve), the numerical problem size was reduced by a factor of 30, corresponding to a time to solution decreased by 27000 or 25 years of Moore's law. From Rev. Mod. Phys 83, 349 (2011).

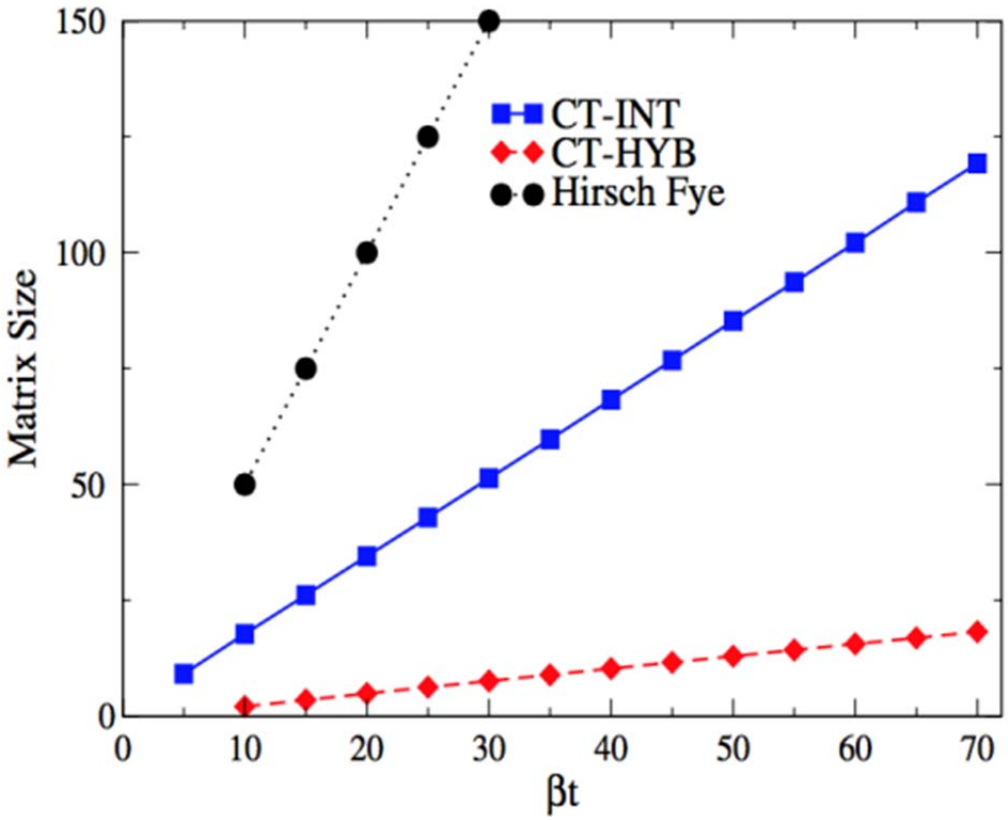


FIG. 18 (color online). Upper panel: Bethe lattice, single-site DMFT, scaling of matrix size with temperature at $U/t = 4$ for the Hirsch-Fye, CT-INT, and CT-HYB algorithms. For the Hirsch-Fye algorithm, the resolution $N = \beta U$ has been chosen as a compromise between reasonable accuracy and acceptable speed, while the average matrix size is plotted for the continuous-time solvers. Lower panel: Scaling of matrix size with U/t for fixed

BES White Paper – J.J. Rehr, et al. (University of Washington)

Please specify the current science drivers in your field of research

Excited states of materials: An understanding of excited states of materials [a] is generally key to understanding and predicting many of their physical and chemical properties. Such understanding and predictive capability are among the DOE challenges to design and control advanced materials. These excited-state properties include dielectric behavior, response to electromagnetic fields, real-time response, and thermodynamic behavior. At present, computational demands have led to simplified models and qualitative results, often with little resemblance to experiment. Exascale capabilities and compatible algorithms and software are essential to enable more realistic models and conditions. In particular, exascale computing would permit an exploration of several new dimensions, including real-time, real-space, and finite temperature regimes that are currently intractable computationally. Such advances would also enable a database of excited-state properties such as spectra and dielectric response as an extension of the Materials Project.

Describe the science challenges expected to be solved in the 2020–2025 using extent computing ecosystems.

Beyond ground-state DFT: A serious limitation of current materials simulations is the reliance on ground-state, zero-temperature, density functional theory of electronic structure [b]. Thus, a key objective of exascale computing for materials simulation would be to make advanced functionals, such as hybrid, van der Waals, and finite temperature, routine. Exascale capabilities would also enable simulations of advanced materials using Green's function methods. Examples include the GW approximation and the cumulant expansion [a,c] (Fig. 1). While parallel algorithms have been proposed, optimization for exascale computing is needed. Realistic methods and models based on large-scale systems are often essential for quantitative prediction. A typical problem involves impurities and vacancies: small simulation cells (and high concentrations) reduce computational cost but can yield "unphysical" results. Another problem is the lack of nuclear motion and thermal vibrations. Zero-temperature simulations ignore thermal effects that are crucial in many properties. A common technique for including thermal effects is via *ab initio* molecular dynamics (AIMD) [b]. However, such simulations are extraordinarily demanding computationally. Real-time and real-space algorithms are desirable, but must be optimized and parallelized for exascale computing.

Next-generation spectroscopic techniques: Spectroscopies are immensely powerful probes of material properties [d] that are used extensively at modern synchrotron radiation facilities. However, their interpretation can be extremely computationally demanding. One reason is that spectroscopic quantities, such as X-ray absorption or photoemission, involve many-body response over a range of energy and time scales, thus requiring state-of-the-art excited state theories [a,d]. Second, while theories of excited-state electronic and nuclear dynamics are currently available, many are now computationally prohibitive. Thus, advanced approaches such as the GW-BSE [e] for calculating optical and X-ray spectra in complex systems demand exascale capabilities. Third, simulations of complex systems such as liquids, interfaces, and correlated materials require sophisticated models with many hundreds of atoms and/or electrons. Next-generation spectroscopic methods (e.g., at the LCLS) pose additional challenges. New algorithms such as real-time TDDFT with improved functionals may help. Exascale computing will likely address many of these problems by 2025.

Describe the science challenges that probably cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

Chemical reactions and catalysis: Thermal excited states are the driving force behind chemical reactions and catalysis [f]. One challenge in theoretical catalysis is that no single piece of software can take advantage of exascale facilities for this purpose. The theoretical and computational tools of systems such as liquids, solvated molecules, or interfaces are often computationally intractable because of their complexity. Even short simulations (over tens of picoseconds) require on the order of 100K CPU-hours for relatively simple systems. In addition, the coupling between vibrational and electronic structure requires many such calculations. Fortunately, such calculations parallelize naturally, and can take advantage of exascale computing facilities. This problem also emphasizes the need for advanced workflow control that exploits a range of computing capabilities. Development should focus on those codes that scale up well to many thousands of processors. Such scalability should be favored over other efficiencies. Thus, codes that do one step well (e.g., AIMD) should be favored over those that do other steps slowly. Although developments to enhance scalability vs. model size are often useful, systems like nanocatalysts require inherently small models, where most of the computational effort goes into the dynamic structure [g] (Fig. 2). Thus, efforts are needed to develop parallelized simulations in the time domain. By allowing AIMD simulations in the ns regime (which require 10–100M CPU-hours), this development will enable simulations of reactivity in real time. Consequently, exascale computing will greatly help, but may not fully solve this problem by 2025.

What *top three* computing ecosystem aspects will accelerate or impede progress in the next 5–10 years? Why?

Accelerate	Why?
1. Advanced theory	Limitation of DFT and approximate theories
2. Efficient algorithms and parallelization	Enable realization of exascale capability
3. Workflow tools	Facilitate multi-faceted HPC calculations

Impede	Why?
1. Lack of fundamental theory in some materials	Correlated systems are not yet fully understood.
2. Lack of massively parallel codes and algorithms	Cannot take advantage of exascale otherwise.
3. Computational complexity of some problems	Complex, non-parallel algorithms are needed.

Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data. Big data handling and data analysis are both involved in excited state and spectroscopy calculations and experiments. Quantitative analysis of experimental spectra is essential to extract reliable information. With few exceptions (e.g., EXAFS), a quantitative analysis is ill-conditioned and computationally difficult. However, parallel fitting and analysis algorithms scale well, and exascale capabilities may help solve these problems. Moreover, this effort would mitigate the Big Data problem by permitting a real-time data analysis to help separate signal from noise.

References

a. Onida, G., Reining, L. & Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **74**, 601 (2002).
b. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).

c. Kas, J. J., Rehr, J. J., and Reining, L., Cumulant expansion of the retarded one electron Green's function. *Phys. Rev. B* **90**, 085112 (2014).
d. Rehr, J. J. *et al.* Ab initio theory and calculations of X-ray spectra. *C. R. Phys.* **10**, 548 (2009).
e. Vinson, J., Rehr, J. J., Kas, J. J. & Shirley, E. L., Bethe-Salpeter equation calculations of core excitation spectra. *Phys. Rev. B* **83**, 115106 (2011).
f. Frenkel, A. I. *et al.* Critical review: Effects of complex interactions on structure and dynamics of supported metal catalysts. *J. Vac. Sci. Technol., A* **32**, 020801 (2014).
g. Rehr, J. J. & Vila, F. D. Dynamic structural disorder in supported nanoscale catalysts. *J. Chem. Phys.* **140**, 134701 (2014).

Images:

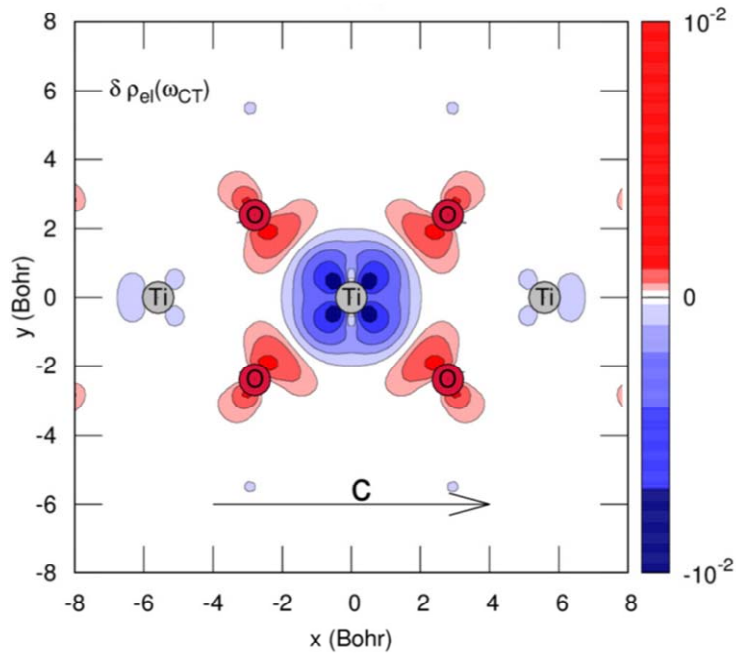


Figure 1: Excited state electron density $\delta\rho(r,\omega_{CT})$ in the Ti-O plane at the charge-transfer (CT) excitation energy $\omega_{CT} \approx 14$ eV (see the arrow in the middle plot) based on the real-time real-space cumulant method. The in-plane structure with the c axis along x is superimposed in the density plot. From Kas *et al.* Phys. Rev. B 91, 121112(R) (2015).

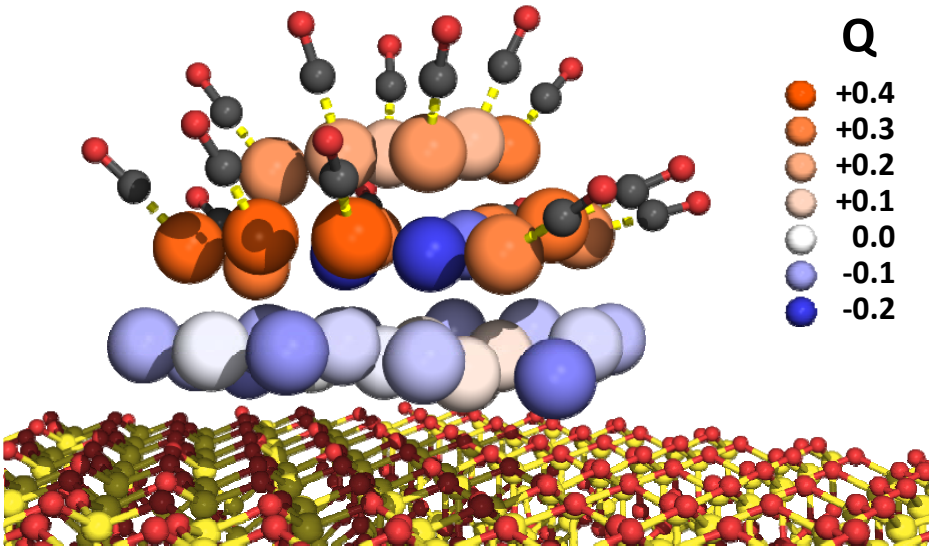


Figure 2: Charge transfer in nanocatalysts upon CO adsorption in a Pt_{37} cluster on SiO_2 support. The orange atoms are positively charged while the blue ones are negative. The black-red dimers characterize the CO molecules. From Elsen *et al., J. Phys. Chem.* (in press, 2015).

BES White Paper – Exascale Computing Requirements for Predictive Materials Synthesis, Functionality, and Performance:

Peter Sushko, Thom Dunning, Karol Kowalski, and Niri Govind (Pacific Northwest National Laboratory)

1. Current science drivers

The forthcoming 2015 BESAC report on Transformative Opportunities for Discovery Science[1] identifies *mastering hierarchical architectures*, *beyond-equilibrium matter*, and *understanding the critical roles of heterogeneity, interfaces and disorder* as areas poised to transform the way we understand and control the properties of matter. These scientific challenges are exemplified by processes occurring in *heterogeneous systems*, such as liquid/solid interfaces. For example, synthesis of branched and hierarchical nano- and meso-scale materials; functioning of materials in chemical and electrochemical energy capture, conversion, and storage devices; and materials degradation under operating conditions take place at structurally and chemically complex interfaces and occur under nonequilibrium conditions. Yet, understanding and controlling processes in heterogeneous systems is a cornerstone of the rational design and synthesis of robust materials for energy capture, conversion, and storage.

A principal challenge in understanding collective effects in these systems stems from their complexity. While advanced *in situ* characterization techniques are now approaching the resolution needed to observe interfacial phenomena at the length- and time-scales in which they occur, understanding the underlying mechanisms of these phenomena will require close collaboration between computational modelers and experimentalists. Computational modeling of heterogeneous systems is, of course, a "grand challenge"—the models must be able to account for differences in the spatial variations of the electron density, localization of vibrational modes, and differences in electron and ion mobilities in the two phases and at their interface, as well as accurately predict the thermodynamics and kinetics of electro-, photo-, and thermo-stimulated atomic-scale processes. A further complication is that the processes in the energy capture, conversion, and/or storage cycle can alter the structure and properties of the material—in the real world, the system never returns to its initial state upon completing the cycle. The mechanisms by which physical and chemical processes accumulate and induce qualitative changes in the materials are system-specific and largely not understood.

In order to achieve control over these phenomena, one has to develop a microscopic to mesoscopic[2] understanding of (i) the underlying mechanisms of *materials synthesis* (e.g., growth and assembly in heterogeneous environments), (ii) the fundamental principles of *materials functionality* (e.g., energy storage and catalytic reactions, capture, and separation), and (iii) the mechanisms of *materials damage and degradation* under *in operando* conditions,[3] extreme environments, as well as the effects of ageing and weathering in functional and structural materials.

2. Science challenges that can be solved by 2020–2025 using extant computing ecosystems.

Existing methods and HPC capabilities provide the means to predict materials properties given their structures and to predict material structures given their chemical compositions, although additional computational resources will be needed to fully address the heterogeneous materials systems and processes discussed above. The accuracy of *ab initio* methods, size of the systems amenable to modeling, and time-scale of dynamics simulations remain bottlenecks in the use of the existing computing systems. Nonetheless, progress to-date and advances envisaged for the next 10 years will enable computational modeling to: (i) *explain* the observed materials properties and phenomena and *interpret* experimental data by providing insight into the electronic structure, mechanisms of mass and charge transfer, interfacial disorder and reactivity, *etc.,* and (ii) *predict* new structures, properties, and

mechanisms of elementary processes, providing insights into new materials synthesis and their functionality.

3. Science challenges that cannot be solved by 2020–2025 using extant computing ecosystems.

The computational design of materials for the broad range of demanding energy applications articulated in #1 above necessitates an understanding of *the interplay among energy transfer and accumulation, materials chemistry, the environment in which the materials function, the changes induced in the material while operating, and the effects of external stimuli*. Addressing these issues requires an ability to connect molecular-scale details of the interactions involved to the mesoscale behavior of the material along with an understanding of the interplay between the various processes involved, which is predicated on the ability to provide sufficient accuracy in the models and access the relevant energy-, length- and time-scales simultaneously. This will require computing resources well beyond those currently available. In addition, as the time, space, and energy resolutions of experimental techniques advance, storing and analyzing the raw data may become a bottleneck.

4. *Top three* computing ecosystem aspects to accelerate/impede progress in the next 5–10 years

Accelerate	Why?
1. Highly scalable implementations of accurate *ab initio* methodologies	Approaches routinely used in *ab initio* calculations do not provide a sufficiently accurate description of the electronic structure (both ground and excited states) of the materials of interest, cannot take full advantage of existing HPC infrastructure, and/or have a limited range of applicability.
2. Versatile multi-scale methodologies	New paradigms in modeling extended and disordered heterogeneous systems are necessary to bridge the gap between elementary processes in idealized systems and behavior of real-world materials.
3. Data storage and processing	The vast amount of data accumulated from both computational and experimental studies needs to be analyzed in a timely manner in order to validate the computational models and interpret the experimental data.

Impede	Why?
1. Develop workforce	Progress is impeded by the lack of early-career scientists trained in, simultaneously, applied mathematics, numerical computation and data analysis, and high-performance computing. In addition, more early-career scientists are needed who are knowledgeable about the implementation and use of advanced computational models, e.g., many-body and multiscale methods, to tackle the most challenging problems in materials science.

2. Codes and code infrastructure	Materials science is plagued with a large number of modeling codes that cannot take full advantage of existing computing technologies, let alone the technologies of tomorrow. Focused efforts are needed to develop the next generation of modeling codes that can take full advantage of future computing systems. In addition, there are a large number of "home-made" codes that never get wide use because of the effort/cost associated with advertising their capabilities and/or parallelization. A creative solution to creating and maintaining a code infrastructure for materials science that addresses both of these problems should be sought.

5. Characteristics of the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.

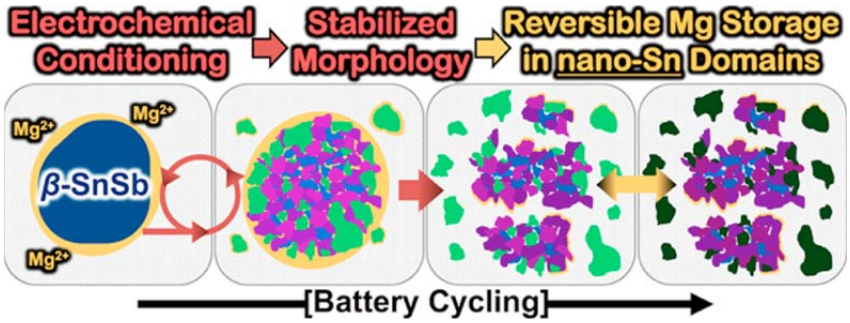
The data ecosystem is currently not a major issue in this area, although it is expected to become more important in time with the growth in both variety and volume of data in the future.

6. References

1. BESAC, July 7–8, 2015; science.energy.gov/~/media/bes/besac/pdf/201507/Sarrao_BESAC_July.pdf.
2. From Quanta to the Continuum: Opportunities for Mesoscale Science, BESAC Report, 2012.
3. Parent, Cheng, Sushko, Shao, Liu, Wang, Browning, *Nano Letters* **15**, 1177 (2015).

7. (Images)

Electrochemical nanostructuring of SnSb electrodes: Mg intercalation, followed by Sn substitution and ejection, results in the formation of ~30 nm compositionally segregated particles: electrochemically active Mg-Sn and passive Mg-Sb alloys.[3]



BES White Paper – Nuclear Energy: Ping Yang (Los Alamos National Laboratory)

1. **Please specify the current science drivers for your field of research.**

Nuclear energy is a critical component of the national energy mix due to its low greenhouse gas emission and reliability. With the rapid development of high-performance computing, we can now predict the structures and properties of small- to-medium-size physical systems up to a few hundreds atoms, depending on the level of theory, using first-principle methods. The results provide insightful information to interpret experimental characterization data and predict their chemical and physical behaviors. For medium to large systems that contain more than a thousand atoms, we can carry out fully atomistic simulations using molecular dynamics with interatomic potentials. Predictive simulation and modeling was identified as a critical priority research directions for advanced nuclear energy, as pointed by the BESAC reports of Advanced Nuclear Energy[1] and Computational Materials Science and Chemistry.[2]

However, a big gap remains between the time and length scale that can be simulated today and those necessary for truly predictive science. Quantum computations are required for electronically complex systems, such as heavy element systems and strongly correlated systems. For many important scientific areas, such as heavy element separation, waste management, and environmental contamination, there is a clear need to simulate larger systems with thousands of atoms for longer timescales (hundreds of picoseconds) while retaining first-principle accuracy. For example, scientists now compute relative energies for various configurations of metal-ligand complexes in separation systems. However, in order to design a separation agent, one would need to include the separation ligands, counter ions and solvent molecules explicitly, moving beyond the implicit models that are currently used. Furthermore, a predictive simulation capability for actinide bearing materials at the nanoscale is highly demanded to design better nuclear materials and interfaces with radiation, temperature, and corrosion resistance and to understand the fate of nuclear waste in the environment;[3] but the required sizes and times often remain prohibitive. New approaches are therefore required in order to understand and design complex nuclear energy-related systems.

The above areas lie at the core missions of the U.S. Department of Energy. Given the difficulty and cost of extensive experimental investigation (especially when dealing with radioactive materials), predictive computing and modeling is of central importance to the understanding of the underlying fundamental chemistry and physics.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using existent computing ecosystems.**

In the next five to ten years, with the current developing speed of supercomputer, we will be able to investigate larger and more complex systems of hundreds to thousands atoms in order to understand key phenomena occurring in actinide-bearing materials. With advances in accelerated methods development, we will be able to tackle extended time scales for small to medium systems, allowing direct comparison with experimental measurements. Larger and more complex nanomaterials of thousands and more atoms will become amenable to direct simulation, albeit at limited accuracy afforded by the use of first-principle based tight-binding methods and empirical potentials.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using existent computing ecosystems.**

We expect that severe limitations on size and time scales will remain. For example, it is likely that we will not be able to directly simulate ligand exchange reactions of metal complexes in solution with full

first-principle accuracy. Furthermore, current simulation algorithms and implementations for the type of high-level simulation required to accurately describe strongly correlated systems do not scale past hundreds of processors in parallel supercomputers, further limiting the range of size and timescales that can be studied.[4] Therefore, the computational design on an efficient nuclear fuel cycle will still be a challenge.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Availability of massive computing resources	The exponential increase in the available computing power will allow the systematic exploration of wide families of compounds, enabling the design of systems with tailored properties.
2. Models and algorithms	Current algorithms used in quantum chemistry do not parallelize well beyond hundreds of cores, so large MPI machines are of little benefit. Radically redesigned methods and algorithms are required in order to enable the massive parallelism expected from future architectures. Such development, e.g. linear scaling, is presently underway, but sustained effort is required.[5]
3. Hardware resources	Memory fat nodes will offer greater opportunities for performance improvement of quantum chemical codes than massive MPI parallelism.

Impede	Why?
1. Changing architecture landscape	Currently, porting of legacy quantum mechanical simulation codes to new architectures occurs on timescales that are on the order longer than the lifetime of a particular architecture. Better programming models and libraries are needed to decrease the cost of porting codes to new architectures.
2. Workforce development	Most graduate students now use simulation codes as a black boxes. Few have the training in advanced algorithms and parallelization that is required to improve codes and ports them to new architectures. This limits the pace of innovation. More adapted academic programs are needed.
3. Middleware	Most quantum chemical calculations carried out today are not very scalable. Efficiently harnessing massive computing resources requires the development of standardized middleware that can generate, manage, analyze, and archive, huge numbers of simulations. The development of such middleware is still in its infancy.

In addition, the following two areas should be supported and improved in order to take full advantage of the new developments of high performance computing:

[1] Libraries and compilers for new architectures: In view of upcoming architectures much different from traditional ones, it would be useful to have a group supported by DOE tasked with facilitating the porting of complicated codes to the new machines. This could be done with the development and maintenance of ported libraries of general use, such as LAPACK, ATLAS, etc., optimized for using the new hardware. The benefit is that porting codes would be much simplified and possible. At the moment, porting a code for modeling quantum chemistry or quantum materials can take more than the lifetime of the hardware making it unfeasible and the developers decide to ignore the advantages of the new advanced architecture.

[2] Bridge between algorithm and methods developers: Currently, the funding is roughly divided among groups developing algorithms and groups developing new methods. The former usually work with experimental codes that are not very applicable for actual research beyond the algorithm. The later group usually has codes that are slow but can compute properties and do simulations that are state of the art. It would be beneficiary if these two communities could be bridged by people bringing state-of-the-art algorithms to the new computational methods.

5. References

1. *Basic Research Needs for Advanced Nuclear Energy Systems*. (DOE Office of Science, Basic Energy Sciences, Washington, DC, 2006). at <http://science.energy.gov/~/media/bes/pdf/reports/files/anes_rpt.pdf>.
2. *Computational Materials Science and Chemistry for Innovation*. (DOE Office of Science, Basic Energy Sciences, Washington, DC, 2010).
3. Kalmykov, S. N. & Denecke, M. A. *Actinide Nanoparticle Research*. (Springer Berlin Heidelberg, 2010). doi:10.1007/978-3-642-11432-8.
4. *Scientific Grand Challenges*. 1–117 (2009).
5. WTEC. *International assessment of research and development in simulation-based energineering and science*. 1–426 (2008).

C.2 White Papers Addressing Catalysis, Photosynthesis and Light Harvesting, and Combustion

BES White Paper – Tunna Baruah (University of Texas, El Paso)

1. Please specify the current science drivers for your field of research.
The main science driver is the electronic-structure-based understanding of how chemical systems undergo energy conversion processes, such as light harvesting, and how the environment affects such processes. The simulation problems can be broadly classified as the accurate prediction of quantum states and dynamical processes in molecular systems. This includes the need to efficiently and accurately calculate excited states of molecules that are interacting with an electric field and to account for all spin-dependent processes, electron-vibron interactions, and the effects of a polarizable environment. Examples of problems that require such capabilities include electron dynamics, energy transfer, and coherent processes. For chemical systems that are being driven by low levels of radiation, such as sunlight, methodologies that include the dissipative dynamics also need to be developed. To achieve predictive capabilities with a level of robustness and speed that allows design of chemical systems for specific basic-energy processes, it is necessary to carefully convert many existing and concept-stage chemical simulation tools into turn-key, massively-parallel software that is capable of adapting to computer ecosystems of the future. It is reasonable to expect that computationally designed light harvesting molecules will have components similar to those found in photosystem I and photosystem II but may have substituents that are better for light harvesting but less likely to be amenable to fully supporting the life of a man-made photosystem.

Taking the cue from the structures of such systems (e.g., cyanobacteria photosystem II), it reasonable to expect that such systems will also contain spin-carrying transition metal centers or possibly, for thermal stability, rare-earth f-electron systems that are expected to more readily maintain their magnetic ground state. Current fast quantum-mechanical methods do not have the full flexibility required to self-consistently account for all magnetic excitations in these intrinsically broken symmetry systems due to size restrictions, scientific gaps, and software limitations. The ability to address the full periodic table requires non-empirical “well beyond Hubbard-U” methods for DFT and full accounting of relativistic affects for heavier atoms in a way that still allows for Hellmann-Feynman forces. It is further clear from the structure of these systems that the appendages and/or antennae often consist of helical chains which, as a single unit, are not amenable to computational inquiry with existing cluster codes or three-dimensional periodic codes. Understanding the complicated light-driven molecular processes will also require fast and accurate capabilities for calculating electron excited states; understanding how conical intersections and vibrational and spin dependencies affect stimulated absorptions; and accounting for energy losses due to the relatively fast spontaneous processes that are easy to account for with master equations but less simple to include in time-dependent DFT. Moreover, some of the most interesting features associated with chemical conversion happen in parts of phase space where the Kohn-Sham electronic structure, due to underestimated HOMO/LUMO gaps, overstates the degree of metallicity of the molecule. This feature, due to incorrect treatment of the long-range coulomb problem, significantly complicates the computational simulation at these potentially most interesting parts of configuration space and would tend to overestimate polarizability and reactivity in these regions. Often, light-harvesting systems deposited on a substrate represent 2-dimensional systems but the existing technologies that can be applied are either cluster codes or 3D periodic codes. Similarly, one-dimensional periodic codes will be extremely useful for simulations of photosystems. Such approaches can be useful beyond the light harvesting systems. For example, metal-organic frameworks have quasi one-dimensional channels that contain spin-carrying open transition-metal adsorption and reaction

centers. Also, simulations of electron transport across molecular wires may be important for learning about field-driven electron transport in artificial photosynthetic devices, and these systems could benefit from such a capability.

2. Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.

Problems, solvable in the next five years, for the aforementioned photosystems and many other BES-relevant chemical systems, on computational systems that will be available in 2025 that would help address these problems, include:

- Development of new basis sets and removal of all basis-set dependencies in existing codes.
- Calculation of excited states and excited state processes.
- Improvements of current quantum-mechanical methods and ensuring that such methods are formulated in ways that correctly treat all known behaviors of correct quantum-mechanical methods.
- Trend toward wide use of methods that treat both core and valence electrons on an equal footing and with chemically numerical precision.
- Development of codes that are specifically designed to exploit one- and two- dimensional periodicity associated with molecular wires, helical units, and catalytic surfaces and that are tuned to problems where the known decay of electronic densities into the vacuum region are part of basis-set choice.
- Development of new strategies for memory management and low-communication highly parallelizable algorithms to achieve load balancing and effectively use the fast multi-core platforms.

3. Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

Future lines of effort will need to combine many different advances. As mentioned above, it is desirable to have a quantum-mechanical method with predictive accuracy that can account for all electronic, spin, and vibrational degrees of freedom for any arbitrary system. It should have high accuracy comparable to coupled-cluster accuracy, and must be intrinsically faster than current DFT codes, but scale better as a function of system size and available processors. It is desirable to be able to couple such a tool to a time-dependent integration scheme that allows for simulations on the order of milliseconds and spawns all pathways when spontaneous transitions disrupt the classical trajectories that are otherwise being followed.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

While all of these topics are important, the most important need is to strengthen a workforce that has the scientific knowledge (g) to understand the models and algorithms (b) needed to implement and further develop portable applications codes (a).

Accelerate	Why?
1. Application codes	Application codes with the functionalities such as excited states, basis sets for accuracy and efficiency, and time-dependent processes can help in screening materials.
2. Models and algorithms	Newer models such as 1D, 2D codes for reduced dimension. Memory management algorithms can make codes more efficient/scalable in terms of time and memory.
3. Libraries/tools	Standardized format for data sharing which will allow easy I/O exchange between different codes. Basis set libraries (e.g., Gaussian basis set libraries), which can be incorporated easily into codes. Libxc is one such example, which contains many functionals.

References

1. *Massively Parallel Simulation of Light Harvesting in an Organic Molecular Triad* , T. Baruah, M. R. Pederson, W. A. Anderson, HPCMP USER Group Conference, 11–17 (2005). doi:10.1109/DODUGC.2005.39.
2. *DFT Calculations on Charge-Transfer States of a Carotenoid-Porphyrin-C_{60} Molecular Triad*, T. Baruah and M. R. Pederson, J. Chem. Theory and Comp. 5, 834 (2009).
3. *Charge transfer excitation energies by perturbative delta self consistent method*, T. Baruah, Marco Olguin, and R. R. Zope, Journal of Chemical Physics, 137 084316 (2012).
4. *Charge transfer excitation energies in cofacial porphyrin fullerene complexes*, R. R. Zope, Marco Olguin, and Tunna Baruah, Journal of Chemical Physics, 137, 084317 (2012).
5. *The effect of structural changes on charge transfer states in a light-harvesting carotenoid-diaryl-porphyrin-C60 molecular triad,* Marco Olguin, Luis Basurto, Rajendra R. Zope, and Tunna Baruah, Journal of Chemical Physics, 140, 204309 (2014).
6. *Crystalline Alloys of Organic Donors and Acceptors Based on TIPS-Pentacene,* Monica Kai Jes, Tunna Baruah, Parking Guan, Sean Parkin, Balaji Purushothaman, Rajendra Zope, and Michael Chabinyc, J. Phys. Chem C 119, 20823 (2015).
7. *Electronic structure and charge transfer excited states of a multichromophoric antenna*, Luis Basurto, Rajendra R. Zope, Tunna Baruah, arXiv:1306.2627.
8. *Electronic Transport as a Driver for Self-Interaction-Corrected Methods*, Anna Pertsova, Carlo Maria Canali, Mark R. Pederson, Ivan Rungger, Stefano Sanvito, Advances In Atomic, Molecular, and Optical Physics, 64, 29–86 (2015).
9. Modelling the Photosynthetic Water Oxidation Center: Preparation and Properties of Tetranuclear Manganese Complexes Containing [Mn402]6+17+,8+ Cores and the Crystal Structures of Mn402(02CMe)6(bipy)2 and [Mn402(02CMe)7(biPY)21(c104), John B. Vincent, Cheryl Christmas, Hsiu-Rong Chang, Qiaoying Li, Peter D. W. Boyd, John C. Huffman, David N. Hendrickson, George Christou, Journal of American Chemical Society 111, 2086 (1989).
10. Towards quantifying the role of exact exchange in predictions of transition metal complex properties, E. I. Ioannidis, and H. J. Kulik, Journal of Chemical Physics 143, 034104 (2015).

(Optional) Images

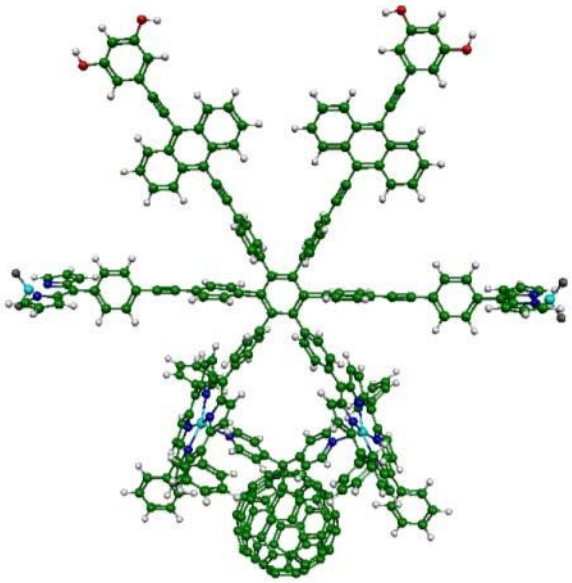


Fig. 1: Multichromophoric heptad antenna. This is representative of the systems that can be studied using traditional DFT and current technologies.

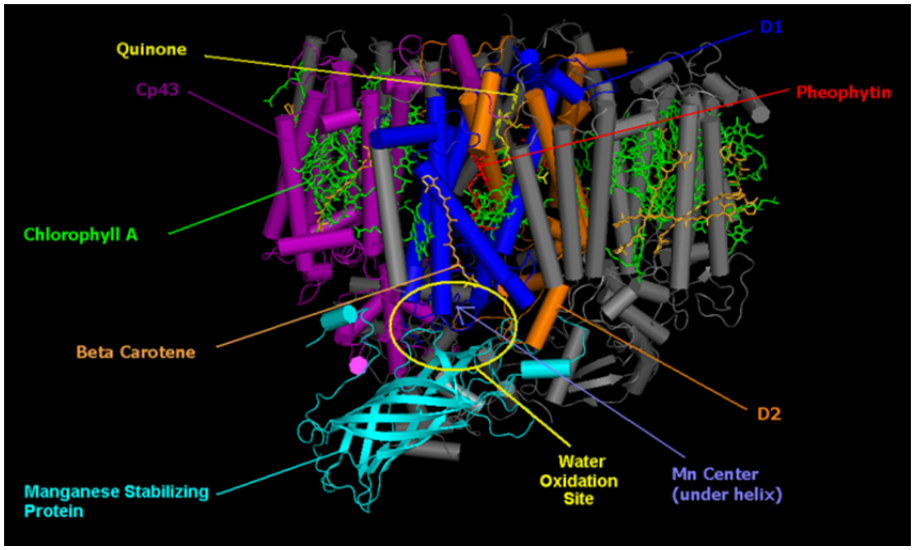


Fig. 2: cyanobacteria photosystem II monomer. Source: Wikipedia. It shows the photosystem that contains spin centers.

BES Exascale White Paper: Combustion Science– Jacqueline Chen (Sandia National Laboratories)

Please specify the current science drivers for your field of research.
The development and utilization of clean and efficient hydrocarbon and bio-derived fuels used in transportation and power generation depends upon understanding fundamental "turbulence-chemistry" interactions in combustion [1]. Combustion in regimes in which the next generation of advanced efficient and clean engines will operate intrinsically involves low-temperature chemistry coupled with transport at conditions far from equilibrium [2] and at extreme pressures.

With exascale computing, longstanding combustion science challenges related to advanced gas turbines (for high-efficiency electrical power production) and reciprocating engines (for ground transportation) can be addressed with a combination of direct numerical simulation (DNS) (for physical insight and benchmark simulations), large-eddy simulation (LES) (for more complicated boundary conditions and parametric sensitivity analysis), and theory and experimentation (benchmarks for chemistry/transport coupling). For example, the mechanisms of flame stabilization, flashback, pollutant formation in gas turbines, and the effects of fuel composition and fuel spray parameters on multistage ignition and soot formation in engines at realistic pressures and turbulence scales can finally be addressed. A key challenge for high-fidelity simulation is to understand the role of low-temperature ignition of a cool flame and its subsequent transport toward fuel-rich conditions (as either a spontaneous ignition front or as a deflagration wave) to establish the conditions for hot-flame ignition to occur. Under certain conditions, controlled by the competition between transport and low-temperature chemistry, ignition may occur earlier than predicted in a homogeneous system. Controlling ignition delay through tailoring the design of low-temperature chemistry would influence the premixedness of the reactants before the onset of combustion, which, in turn, affects soot formation processes in a diesel jet flame.

Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.
Today's leadership-class petascale capabilities applied to DNS has enabled our ability to interrogate fine-grained turbulence-chemistry interactions in canonical configurations. In particular, three-dimensional DNS, at moderate Reynolds numbers and with hydrocarbon chemistry, is providing unprecedented levels of detail to isolate and reveal fundamental causality between turbulence, mixing, and reaction [3]. This information is leading to new fundamental understanding, providing benchmark data for model development and assessment, providing ideas for new closure hypotheses, and assisting in the interpretation of statistics obtained from lower-dimensional measurements. In the 2020–2025 timeframe, the focus is computationally constrained to small-hydrocarbon (C1-C5) chemistry, small laboratory-scale flames at atmospheric pressure and at moderate turbulence levels. Within these constraints, fundamental transient processes including extinction and reignition, flame stabilization in autoigniting flows, premixed and stratified turbulent burning velocity, and boundary layer flashback can be studied using first-principles DNS. These highly intermittent turbulence-chemistry interactions determine the stability, emissions, and efficiency characteristics in advanced efficient, clean combustors.

Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.
Computations of multi-phase sprays with intense turbulence and complex hydrocarbon chemistry (larger than C5) at high pressure (greater than 10 bar) cannot be tackled with extant computing in the

2020–2025 timeframe. Understanding and modeling the dynamics of ignition and establishment of a lifted diesel jet flame at high pressure (40–100 bar) from high-fidelity combustion simulation is a challenge. At high pressure, the disparity between turbulence and flame scales increases. At the same time, low-, intermediate-, and high-temperature chemical kinetics are required to adequately describe multistage ignition and flame chemistry, often requiring large chemical models with greater than 100 species and thousands of elementary reaction steps even after mechanism reduction. The dynamic range of turbulence also increases with pressure as the dissipation scales decrease with pressure. As a result, much greater spatial resolution is required near high-gradient flames or ignition kernels, and the number of degrees of freedom increases as well because of larger mechanism sizes. There is a need for both adaptive mesh refinement and adaptive chemical and transport model reduction methods. Multiscale approaches are also required to treat liquid-gas phase interfacial phenomena and chemical and thermal nonequilibrium in continuum reacting flow.

Accelerate	Why?
1.Application codes for compressible and low-Mach reacting flows with adaptive mesh refinement and adaptive chemistry	Treat disparities in scale between turbulence and flame scales at high pressure with detailed chemical models.
2.Hardware resources at exascale including processors, interconnect, storage, memory, I/O	Need exascale machine to resolve all temporal and spatial scales of a turbulent flame at high pressure.
3. Framework/DSL for PDE solvers to hide complexities of task-based programming environments needed to deal with myriad latencies across machine	Ease of programmability, portabllity, and extensibility for applications programmer with asynchronous task-based programming environment to address with myriad latencies and resiliency.

Impede	Why?
1.Bulk synchronous software stack	Bulk synchronous approach will lead to inefficient use of inherently asynchronous exascale resources (won't deal with power management, NUMA issues, interconnect latencies, I/O latency).
2.Synchronous algorithms for PDEs and for analytics	Lack of asynchrony-tolerant PDE stencil methods causes unnecessary synchronization and lack of fault tolerance.
3. Workforce development	Lack of early career computational scientists trained in HPC to accelerate application code development; need better and wider-reaching training programs.

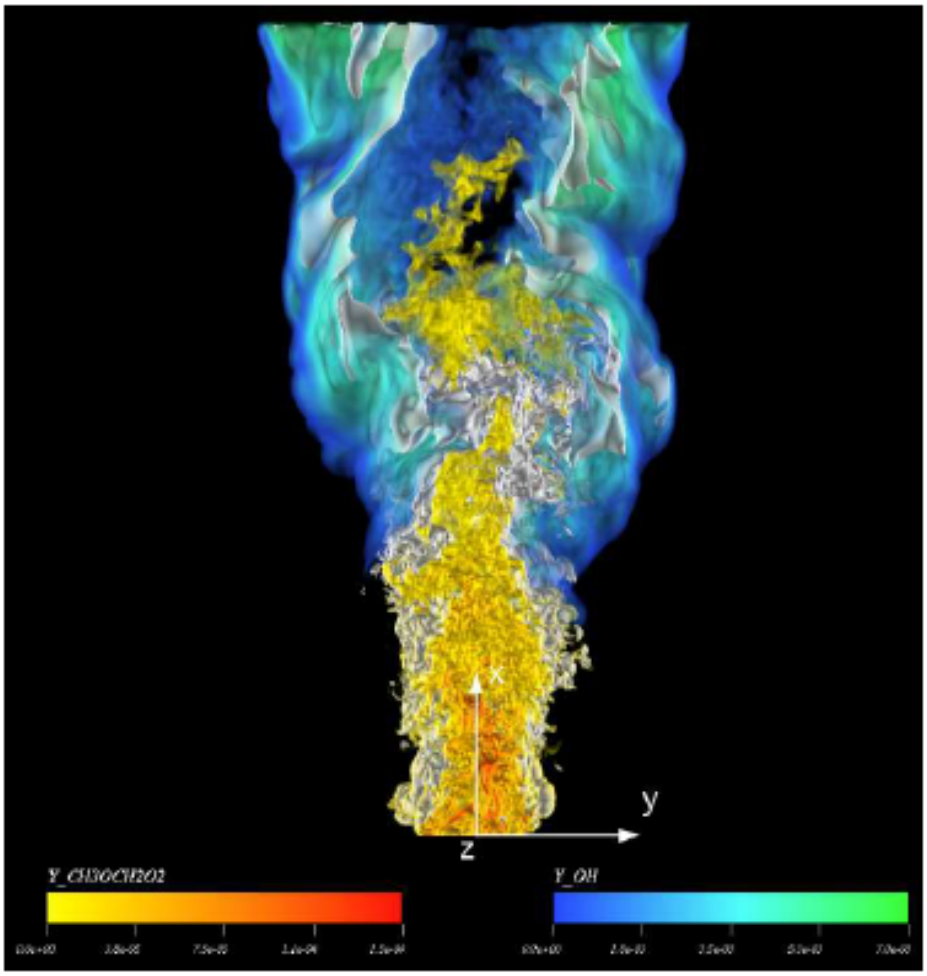
References

1. Basic Energy Sciences Workshop, "Basic energy needs for clean and efficient combustion of 21st century transportation fuels," U.S. Dept. of Energy (2007). http://www.science.doe.gov/bes/reports/files/CTF_rpt.pdf.

2. Basic Energy Sciences Advisory Committee, “Directing matter and energy: five grand challenges for science and the imagination,” U.S. Dept. of Energy (2007). http://www.science.doe.gov/bes/reports/files/GC_rpt.pdf.
3. Chen, J.H., “Petascale DNS of turbulent combustion — fundamental insights towards predictive models,” Proceedings of the Combustion Institute 33 (2011) 99–123.

Image

Instantaneous volume rendering from DNS of a turbulent di-methyl ether lifted jet flame showing a low-temperature heat release marker, YCH3OCH2O2, and a high-temperature flame marker, YOH. The lifted flame exhibits a pentabrachial flame structure due to low-temperature heat release and negative-temperature coefficient behavior, and the flame stabilization is assisted by low-temperature ignition intermediates (Courtesy of Y. Minamoto and J. Chen, submitted to *Combustion and Flame*, 2016).



BES White Paper – W.A. de Jong (Lawrence Berkeley National Laboratory)

1. Current science drivers.

Development of sustainable sources of essential energy, chemicals, and materials as an alternative for, and supplement to, fossil fuels is crucial to meet the increased demands for consumables of our growing society. The key scientific and engineering challenge in the production of sustainable products is to develop efficient, environmentally friendly and cost-effective ways at industrial scale. One is catalysis for biomass conversion. Catalysts are central to overcoming the engineering and scientific barriers to economically feasible routes for the conversion of biomass-derived and solar-mediated fuel and chemicals into usable products. An example is the conversion of cellulose into sugars and bio-oils, which through catalytic processes can be converted into biofuels or building blocks for industrial applications such as plastics. Heterogeneous catalysis has a rich history of facilitating energy efficient selective chemical transformations and contributes to 90% of chemical manufacturing processes. Other areas that rely on controlling chemical transformations with energy applications include the development of new and novel fuel cells and chemical processes to generate hydrogen and the development of chemical processes for the separation and production of nuclear fuels. Accurate kinetics is required to improve our models of combustion processes that can lead to the development of more fuel-efficient engines. Access to accurate simulations of the complex chemical transformations occurring in our atmosphere will greatly enhance modeling their role on the changing climate. In silico design utilizing high performance computing resources is critical in accelerating the development of new catalysts and chemical reaction and transformation processes. Accurate simulations of the kinetics and thermodynamics of chemical transformations enable scientists to discover new and novel ways to predict, control, and design optimal – industrially viable – catalytic activity and selectivity by rapidly scanning the large design space. To be predictive requires the capability to model chemical reaction landscapes with very high accuracy, to determine and discover dynamic reaction pathways that can exhibit complex behavior over a wide range of time scales. These large compute intensive simulations require next-generation scale computing resources of exascale and beyond.

2. Science challenges expected to be solved in the 2020–2025 timeframe.

Utilizing a petascale machine thermodynamics and kinetics of organic molecules of less than 100 atoms are feasible. Considering the O(N^6-N^7) scaling of coupled cluster methods, molecular systems of 2.5-–3x larger will be feasible, more if advances in reducing the computational complexity due to sparsity can be achieved. The first target for biomass catalysts at the exascale computing systems would be to obtain accurate thermodynamic and kinetic properties for the chemical transformation of two widely used bio-oil products free fatty acids (50 atoms) or triglyceride reacting (160 atoms) with methanol and a catalyst to for biodiesel. To model the catalyst one will have to include at least the active part of the catalytic material (which could be a zeolite or clay-like material).

3. Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

Complex transition states will need to be considered and multi-reference methods are needed that increase work by a factor of 10–100 beyond what was mentioned above. Future studies of other biomaterials and catalysts may require a further increase of the number of atoms and electrons that need to be considered in the simulations.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Hardware resources	Truly accelerating scientific discovery requires the generation of large numbers of simulation data of realistic models for analysis and deep learning in a reasonably short timeframe.
2. Programming frameworks and libraries	Programming languages and runtime schedulers that enable developers to express work concurrency and data movement in a system agnostic way will be instrumental in developing efficient scalable applications.
3. New novel algorithms in applications	Algorithms need to be developed that can exploit the drastically increasing concurrency and can handle increasingly expensive and more dynamical data movement.

Impede	Why?
1. Workforce development	Within chemistry and materials fewer and fewer students are properly trained with the essential software development, computing and algorithm development skills. Those that did receive training often leave the field for higher paying industrial opportunities.
2. Absence of integrated application development	Co-design of the applications developers, computer scientists, and applied mathematicians is critical in identifying the most suitable discretization and solver techniques, in addition to developing simulation software that will take optimal advantage of the extant platforms in the 2020–2025 timeframe to enable new scientific discoveries.
3. Hardware resources and utilization thereof	The efficiency and scalability of computational chemistry and materials applications are driven by the ability to move data into the cache of the processor. Processing ability is increasing in next-generation architectures, but the ability to store and move data is progressing much slower, which will impact application performance. Another impediment is a focus on hero runs and the development of hero applications that have limited science impact and a small user base.

5. **Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**

The increase in the problem size accessible on the exascale platforms will inherently lead to increased data files of the discretized wave functions being generated by the simulations. Most of the essential simulation results volume tends to be small, but an increasing amount of information needs to be stored in databases and coupled with experimental data for sharing, analysis and deep learning to go beyond the Materials Project style of scientific discovery. Some post processing and visualization of the wave function data will be performed.

BES White Paper – David A. Dixon (The University of Alabama)

1. Current science drivers

Electronic structure simulations are broadly used throughout the HPC community to address problems in molecular science, materials science and discovery, and biology. The most commonly used approaches are density functional theory and correlated molecular orbital theory.

A workshop sponsored by BESAC[1] reached the conclusion that "*the Grand Challenge for Catalysis Science in the 21st Century is to understand and thereby control the relationship between catalyst structure and catalytic chemistry (both activity and selectivity)*," i.e., a molecular-level understanding of the detailed steps in chemical reaction mechanisms with knowledge of the thermodynamics and kinetics of the various steps. Computational chemistry is key to the development of new catalytic processes as their complexity and diversity demand a revolution in the way catalysts are designed and used. Catalysis is at the core of many of the U.S. Department of Energy's missions.[2,3] Improved catalysts can increase energy efficiency, while increasing product selectivity and concomitantly decreasing wastes and emissions. Catalysis can help us meet the challenges of creating alternative fuels including the solar generation of and chemical storage of hydrogen, cleaning up the environment and preventing future pollution, dealing with the causes of global warming, keeping us safe from the release of toxic substances and infectious agents, and creating safe pharmaceuticals. Catalysis is governed by a delicate balance between a myriad of competing processes. For heterogeneous catalysis, these include adsorption, bond breaking/making, desorption, and surface diffusion that can occur at active catalytic centers whereas for homogeneous catalysts, solvation effects, diffusion in solution, and separations play a role in addition to bond breaking/making. The physical phenomena that underlie catalytic behavior occur at different spatial and temporal scales and there is a need to go beyond integrating the concepts of homogeneous and heterogeneous catalysis.[4] The computational design of practical catalysts for real applications requires the ability to predict, at the molecular level, the detailed behavior of large complex molecules as well as solid-state materials together with their reaction environments. Although intermediate level computations can often provide insight into how a catalyst works, especially the prediction of trends, the true computational design of practical catalysts for industrial and commercial applications will require the ability to predict accurate thermodynamic (better than ± 1 kcal/mol or less) and kinetic (rate constants to better than an order of magnitude initially) properties. Except for small to moderate size systems, this is difficult to do today for real systems that are catalytically relevant. Accuracy is important as a factor of 2 to 4 in improving catalyst efficiency can determine the economic feasibility of a process. An error of only 1.4 kcal/mol in reaction energies leads to an error in predicting an equilibrium constant of a factor of 10 and the same error in the activation energy leads to an error of a factor of 10 in a rate constant at 298 K. The requirement for such accuracy means that we must be able to predict thermodynamic and kinetic quantities including the effects of entropy in complex systems to high accuracy, currently a daunting computational task, particularly for systems that involve metal atoms.

Advanced nuclear energy systems involve the production and exposure of a broad range of actinide bearing fuel and structural materials to extremely challenging radiation environments. The ability to reliably and readily predict the properties of both chemicals and materials containing heavy elemetns and/or radionuclides using advanced computational methods will (1) speed up the design of nuclear fuel systems which typically require decades to optimize, (2) provide improve understanding of the materials of construction for next generation reactors, (3) provide new insights into materials for storing nuclear waste and the degradation processes that can occur, and (4) understand the fate of radionuclides in the environment. A critical need is the ability to reliably predict the thermodynamics of the systems and the kinetics of critical reactions and processes. Compounds containing heavy elements require a proper treatment of relativity which includes both scalar relativistic and spin-orbit components. Materials composed of atoms and molecules with open 4f and *5f* shells exhibit strongly correlated electron behavior, which thus far has prevented reliable predictions of how the physical properties of a material system changes in response to external conditions such as temperature, pressure, and impurities.

2. Science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.

- Predicting the thermodynamics of ensembles of weakly-coupled complex systems, e.g., solutions with electronic structure methods.
- Predicting dynamics and kinetics in condensed media and at interfaces with electronic structure methods (possible but not completely likely).
- Simulation of redox reactions and proton-coupled electron transfer thermodynamics with chemical accuracy.
- Simulation of the behavior of complex reactions temporally and spatially.
- Accurate and usable Quantum Monte Carlo approaches with 1st and 2nd derivative methods.
- Predicting solid state structures for all systems *ab initio.*

3. Science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

- Predictive capability (e.g., chemical accuracy for equilibrium constants and rate constants) for modeling solutions and interfacial phenomena for actinide-containing systems under extreme conditions of pressure, ionic strength, temperatures, pH, and high radiation fields for aqueous media as well as other solvents and other media such as molten salts and ionic liquids.
- Multireference methods for complex spin systems with many electrons, e.g., multiple metals, 1st row metals actinides.
- Simulation of condensed-state reaction kinetics to a factor of 2 as a function of temperature.
- Accurate simulations of phase changes.
- Simulation of homogeneous and heterogeneous reactions in real world conditions.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years?

Accelerate	Why?
1. Need new approaches to deal with multireference systems (theory)	Multiple metals or even single actinide complexes require multireference techniques and these are not currently available for > ~ 15 electrons.
2. F12 correlated methods for all electron relativistic calculations hold promise for reducing the basis set needed to reach the CBS limit.	Large systems with complex ligands will benefit from improved scaling to reach chemical accuracy.
3. Spin orbit approaches	Needed for all molecules closed or open shell with heavy atoms (Z > ~ 40) to get chemical accuracy.

Impede	Why?
1. Lack of memory and local memory bandwidth	Reliable correlated molecular orbital calculations using current algorithms require substantial memory.
2. Processor architectures do nto get too simple	Have complicated mathematical algorithms and cannot reprogram for every generation of chip architecture.
3. Fast I/O	Current algorithms for accurate calculations require large, fast, local I/O. Global I/O does not work.

Fault-tolerant architectures are an additional aspect to be considered.

5. Data ecosystem aspects

- Data is mixed: simulation and experiment but not highly integrated.
- Planned solutions for integrated data sets have not been funded.
- Data security only when dealing with proprietary corporate data.

6. References

[1] "Opportunities for Catalysis Science in the 21st Century", BESAC-sponsored workshop, May 14–16, 2002, Gaithersburg, Maryland, Report issued 2003. www.sc.doe.gov/BESAC/CatalysisReport.pdf.
[2] "Catalysis Looks to the Future," Panel on New Directions in Catalytic Science and Technology, Alexis Bell, Chair, National Academy Press (1992).
[3] http://science.energy.gov/~/media/bes/images/reports/cat-xlg.jpg.
[4] "Beyond relationships between homogeneous and heterogeneous catalysis," D. A. Dixon, A. Katz, I. Arslan, B. C. Gates, *Catal. Lett.* **2014**, *144*, 1785–1789.

BES White Paper – Laura Gagliardi (University of Minnesota)

1. **Please specify the current science drivers for your field of research.**

Where we are today? Be sure to include broad impact, DOE interest, ties between experiment/theory, etc.

Study of complex systems with quantum mechanical methods: enzymes; periodic systems; heterogeneous catalysis; go beyond DFT; perform statistical sampling with quantum mechanics. Having detailed information of the local structure of enzymes in proteins and their reactivity is important and still represents a challenge for modern quantum chemistry and also classical simulations. Catalytic processes in materials need to be understood at a more detailed level. Atomic precision of the reactive processes would be useful. However, it is difficult to obtain that information with conventional DFT calculations. Periodic calculations can be performed only with certain functionals. Sometimes more advanced methods would be useful, like multireference methods. However, they are not affordable for periodic systems. Is using cluster models, instead of the full periodic system, a good alternative? It would be ideal to be able to study the full periodic system with the full accuracy that we can afford on a cluster model, but this is not possible nowadays.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What will probably be solved in the next 5–10 years? Why is this important to the field?

Heterogeneous catalysis; solar energy challenges; photochemistry; potential energy surfaces of excited states.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What might not necessarily be solved in the next 5–10 years? Again, what is the importance?

Predicting crystal structures; protein structures directly from quantum chemical calculations.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?** *Suggested topics include the following.*
    a. Application codes (implementation, development, portability, etc.)
    b. Models and algorithms
    c. Hardware resources (at all scales) including I/O, memory, etc.
    d. Data workflow (including sharing, transmitting, archiving, etc.)
    e. Visualization and analysis resources
    f. Internal/external libraries/frameworks
    g. Workforce development
    h. Other

Accelerate	Why?
1. a	It will be possible to investigate more complex systems than it currently is.
2. b	This will allow researchers to address chemistry and material science challenges that are not affordable nowadays.
3. e	These are very useful tools for complex systems.

Impede	Why?
1. c	Quantum chemical software has not developed in the same way as state-of-the-art hardware. My calculations take similar time to what they took several years ago (at least similar order of magnitude).
2.d	

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.** For example:
   - Classify the data as simulation, experimental, both, or something else.

One of the challenges in my field is still I/O. CPU time of the calculations improves, but the elapsed is still a bottleneck.

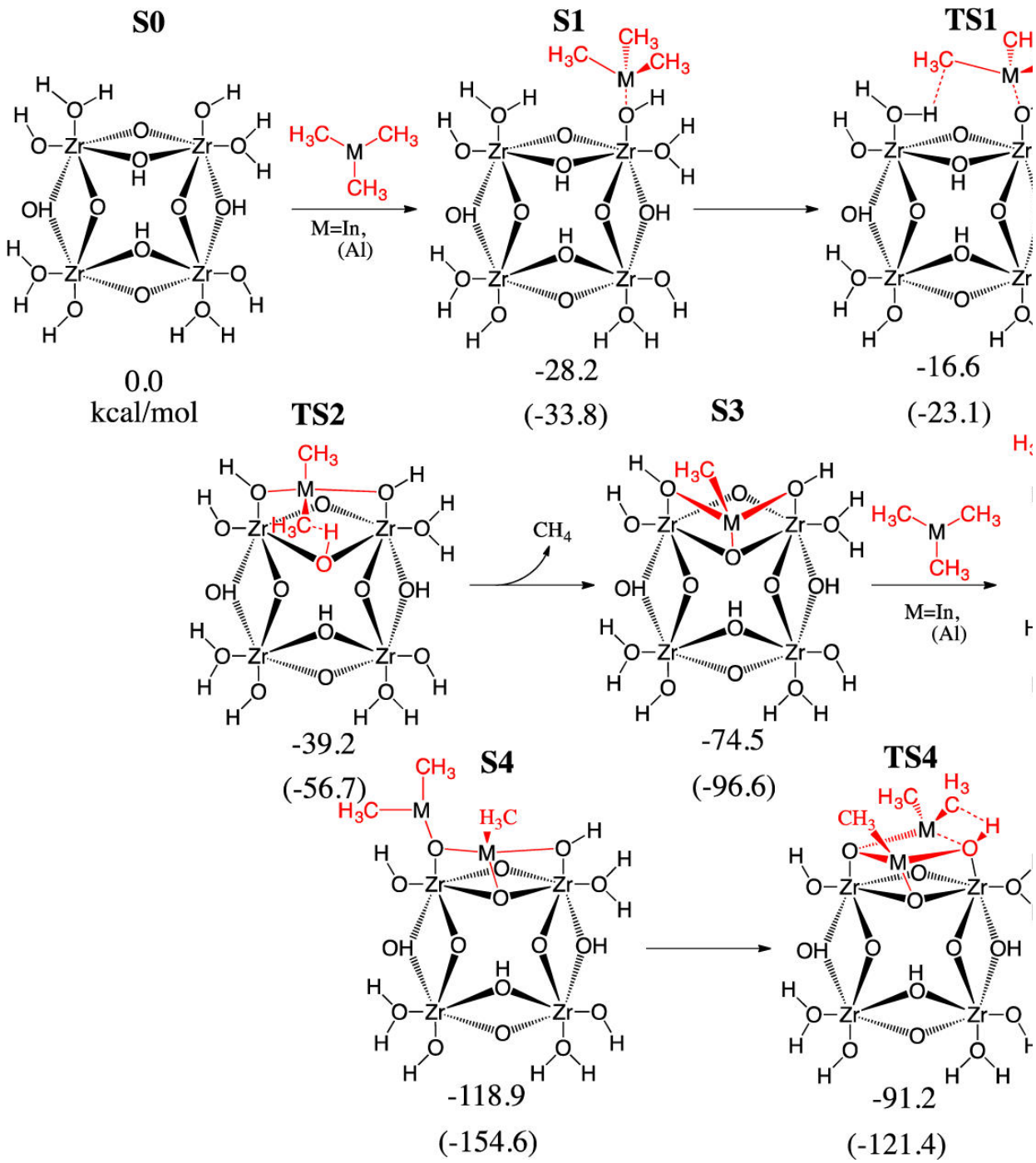
6. **References**

Image taken from:

I. S. Kim, J. Borycz, A. Platero-Plats, S. Tussupbayev, T. Wang, O. Farha, J. Hupp, L. Gagliardi, K. Chapman, C. Cramer, and A. Martinson, Targeted Single-site MOF Node Modification: Trivalent Metal Loading via Atomic Layer Deposition, Chem. Mater., 27 (13), 2015, pp 4772–4778.

7. **(Optional) Image (next page)**

Back to Article Zoom: 100% | Fit width | Fit height | 100% Close Window



BES White Paper – Exascale Computing Requirements for Catalysis at Real-World Interfaces:
Bruce Garrett and Roger Rousseau (Pacific Northwest National Laboratory)

Please specify the current science drivers for your field of research.

A common challenge for advancing many energy applications is the need to design systems in which physical, chemical, and materials processes occur under non-equilibrium conditions in heterogeneous environments. For example, most catalytic conversions of feedstock materials (e.g., petroleum, biomass) into energy carriers (e.g., hydrocarbon fuels) occur in heterogeneous environments at elevated temperatures, high pressures, and with high fluxes of reactants to and product from solid-fluid interfaces. Under these conditions, the catalytic materials transform over time, often changing significantly from their original form during operation. Although significant advances have been made in understanding the behavior of catalysts and catalytic processes under idealized conditions, we currently do not know how to design coupled physical-chemical-materials systems that (1) comprise real-world catalysts and (2) achieve and maintain desired properties under the operating conditions in which they are expected to function. Designing such systems requires development of a predictive understanding of interfacial processes exhibiting complex, collective behavior in which catalyst transformations are coupled to mass transport and cascades of chemical reactions. Discovery and exploitation of fundamental principles of these complex, collective phenomena (i.e., the goal of mesoscale chemical and materials sciences[1]) is central to achieving this predictive understanding.

Predictive understanding implies knowledge embodied in computational tools that allow predictions of properties and processes of *de nova* systems. The need to predict mesoscale behavior requires computational tools that describe phenomena across different scales. Although theoretical frameworks exist for phenomena at most scales, mathematically consistent approaches to seamlessly integrate them across frameworks are only now emerging. Moreover, at the length scale of molecular ensembles, where interfaces impose structure on surrounding media that, in turn, introduce potential energy gradients on the system, we do not understand the essential physics. Experimental studies are essential to identify mesoscale phenomena, delineate the physics that must be included in models, provide model input, and validate models at specific scales and integrate between them. Therefore, we must advance theoretical, computational, and experimental approaches that can both bridge from molecular scales to ensemble outcomes and understand interfacial dynamics, structure, and composition at length scales where the operative physical principles are poorly understood.[2]

Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.

Today's capabilities in computational chemistry provide the means to understand active sites of catalysts: small, local (1–10 nm) arrangements of atoms and molecules, the kinetics of reactions at these active sites, how the kinetics at these active sites can couple with kinetics at other active sites, and mass and heat transport mechanisms. They can describe dynamical processes of active sites during reactive processes showing how active sites transform to promote reaction; the local chemical environment couples to catalytic activity ultimately defining the nature of the *in operando* active site; and how reactions couple kinetically to influence product activity/selectivity. These approaches will provide critical insight into how chemical environments determine the structure and function of active catalytic species and allow us to control reactivity by manipulating both local and extended material structure.

Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

We currently do not have the computational tools that will allow us to design real-world catalytic systems, which are more complex than a collection of active sites at the interface of a catalyst material (1- to 10-nm length scale). Real catalytic systems require conditioning periods to prepare the active catalyst and transform under the operating conditions. Operating catalyst systems often represent a metastable state with overall performance determined by interfacial transformations. Sizes of atom/molecular systems required to capture the coupled physical, chemical, and materials processes that occur across different time and length scales are far beyond what can be treated computationally today or in the next decade if the current rate of improvement of computational tools persists.

What *top three* computing ecosystem aspects will accelerate/impede progress in the next 5–10 years? Why?

Accelerate	Why?
1. Theoretical frameworks for coupling models across scales	Current *ad hoc* approaches for multiscale coupling can lead to different models that are inconsistent with one another. We need approaches that treat the coupling in a self-consistent manner.
2. Internal/external libraries/frameworks	Scalable algorithms that are highly parallel are needed for electronic structure on progressively larger systems; efficient approaches are needed to implement 2-way coupling between scales; low-level software is needed to exploit hardware.
3. Hardware	Rapid evolution of architectures will allow for the simulation of larger, more complex models, but only if there is a parallel development in application codes, algorithms, and libraries that allows for them to be exploited to their maximum potential.

Impede	Why?
1. Workforce development	We don't have a sufficient supply of early-career scientists trained in high-performance computing to accelerate the development of the domain-specific codes needed.
2. Application codes	Need to build on 'best' chemistry/materials codes to implement new theories/algorithms and maximally exploit the capability of the new computing ecosystems.

Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.

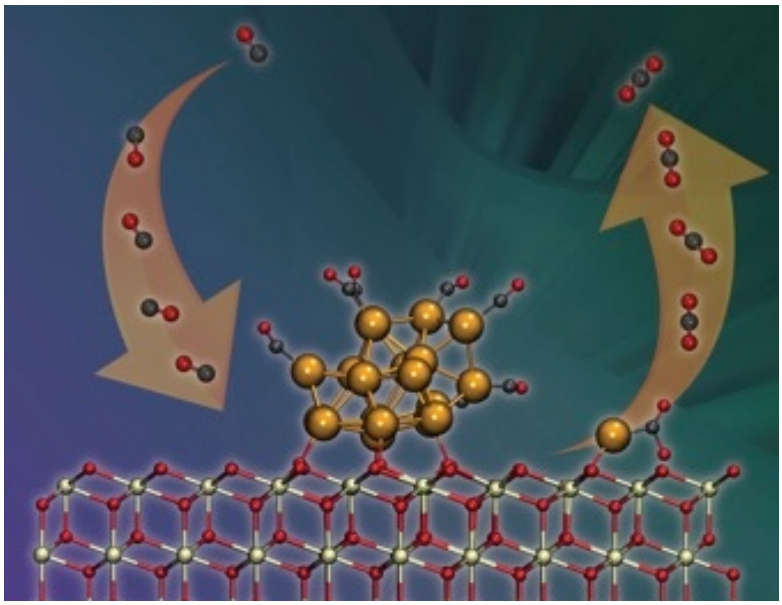
The data ecosystem is not a high priority in this area.

References

1. From Quanta to the Continuum: Opportunities for Mesoscale Science, BESAC Report, 2012.
2. Basic Research Needs: Catalysis for Energy, report from US DOE BES Workshop, 2007.
3. Wang, Mei, Glezakou, Li, Rousseau, *Nature Comm.* **6**, 6511 (2015).

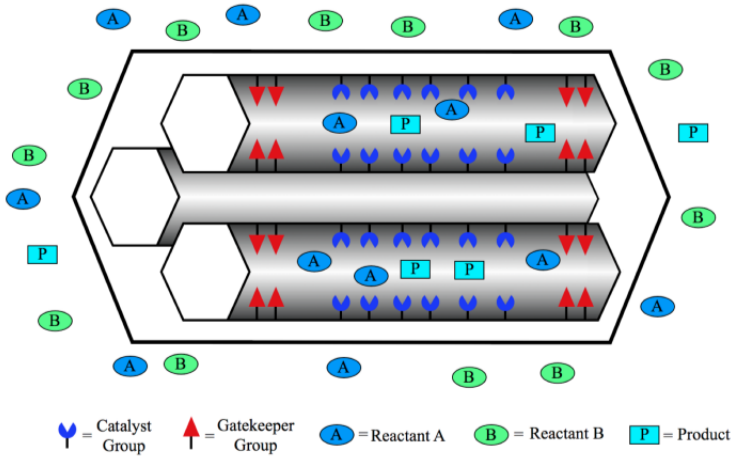
Image

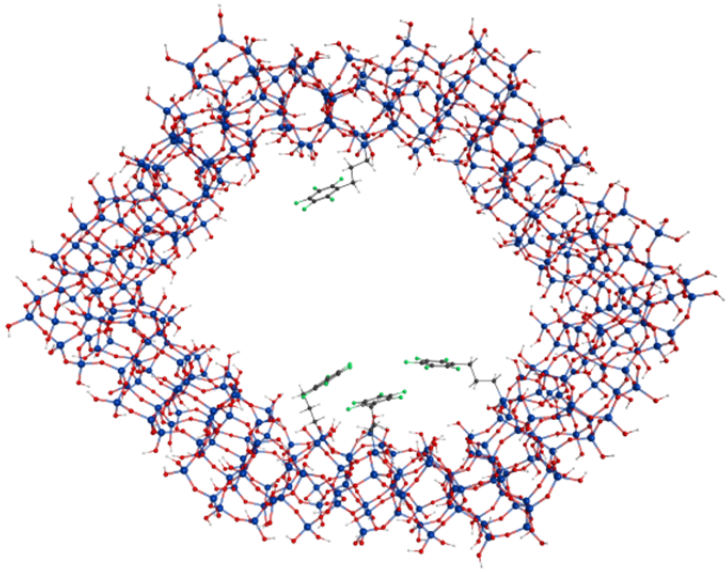
The oxidation of CO to CO_2 can be catalyzed by a gold nanoparticle on a cerium dioxide support. The challenge is determining what happens during the reaction. Combining computers with appropriate molecular representations, scientists found that carbon monoxide affects the catalyst. In attaching, the carbon monoxide changes the gold particle's structure and surface atoms. This interaction results in the detachment of a single gold ion from the gold particle carrying carbon monoxide with it.[3]



BES White Paper – Mark Gordon (Ames Laboratory)

Heterogeneous Catalysis. Mesoporous silica nanoparticles (MSN) and similar materials, many of which are synthesized and experimentally characterized in the Ames Laboratory, have been shown to be excellent, selective catalysts. A schematic of an MSN is shown in the top figure below. In order to simulate such a system with accurate quantum chemistry





methods, one must include at least a reasonable size section of the MSN itself, including the gatekeeper groups that keep undesirable species from entering the pore plus the reactants. A solvent is typically present as well and must be included in the simulation. The lower figure shows a "small" section of an MSN that contains 1,770 atoms. For a reasonable atomic basis set, there would be on the order of 35,000 basis functions or more. Even a single-point energy calculation on such a system with an accurate level of theory (say second-order perturbation theory or coupled cluster theory) would require several tens of

thousands of cores, assuming that algoritms are available that could provide sufficient multi-level parallelism. Mapping out the potential energy surface for a catalytic reaction in the presence of a solvent would increase the problem by orders of magnitude. We use our fragment molecular orbital (FMO) method for these calculations. Because the FMO method scales to tens of thousands of processors, we can greatly reduce the computational cost. Nonetheless, the size of the MSN model remains limited. In order to include solvent and to study dynamics inside an MSN, much more computer power is needed.

Ionic Liquids. Ionic liquids have a wide variety of uses, including for separations of heavy (f-block) metals. The latter use is very important since the properties of f-block metals are so similar to each other that they are extremely difficult to separate. Many of these metals are referred to as "critical" materials, since they are needed for components in many electronic devices and they are very rare. As part of an INCITE grant, we are using our FMO method to run molecular dynamcis simulations on a series of ionic liquids. However, even with the access to the ANL BGQ, and the ability of the FMO method to scale to tens of thousands of cores, we are still limited with regard to the size of a system (with regard to both the sizes of the cations and the number of moleucules included in the simulation) that can be studied.

Bottlenecks. The obvious bottleneck is the limitation on the number of cores that are availabe for the MD simulations. More cores would reduce the CPU time/time step and would therefore allow us to do larger simulations. So, exascale computing capability would be nice. We are, of course, continually improving our algorithms and methods However, the most serious bottleneck is power/energy consumption. The cost of the power to run exascale class computer systems is a killer. Unless this problem is solved, exascale computing will not become a reality. To this end, we and others have been vigorously investigating alternative architectures, such as graphical processing units (GPUs), Intel Phi, and ARM to investigate the trade-offs between performance (e.g., time to solution) and power consumption. Much more is needed in this direction.

BES White Paper – Stephen J. Klippenstein (Argonne National Laboratory)

1. Please specify the current science drivers for your field of research.

The predictive simulation of internal combustion engines (PreSICE) is a long term grand challenge driver for combustion science.[1] Fuel economy, emissions requirements, and environmental and natural resource pressures are demanding more efficient engines fueled by alternate fuels. Recent years have seen enormous progress towards the truly predictive simulation of internal combustion engines. As a result, there is now considerable interest in the use of simulations as engine design tools,[2,3] but, to be truly effective in reducing the number of expensive and time consuming prototypes that need to be built, the predictive accuracy of such simulations needs to continue to improve.

Such PreSICE simulations require the coupling of chemical models for the conversion of the fuel into combustion products with computational treatments of the fluid dynamics of reacting flows. Until recently, global simulations of the combustion process required drastic simplifications in one or the other aspect of the problem. Consequently, these two components (chemistry and fluid dynamics) of combustion modeling have evolved as independent research efforts with little communication between them. Advances in computational algorithms and hardware now allow for simulations that employ meaningfully accurate treatments of both aspects of the problem.[4,5]

On the chemical side, the models should describe not only the conversion of the fuel into oxidation products, but also the formation of various pollutants. The basic research needs for the chemical aspects were described in a 2006 DOE-BES workshop report.[6] The global chemical models typically consist of thermochemical and transport properties for hundreds of species together with rate coefficients for the thousands of reactions that connect these species within the combustion environment. The fidelity of the full simulations naturally depends on the accuracy of the parameters that make up the chemical model. This need for improved accuracy for thousands of parameters is driving the community to develop automated ways for obtaining high level theoretical predictions for the key thermochemical kinetics parameters. There is also a need to re-explore the foundations of the chemical models as a number of the core assumptions appear to provide strong limits on the predictive accuracy of the modeling.

2. Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.

Combustion chemistry is hierarchical with mechanisms for larger fuels depending on submechanisms for smaller fuels. The mechanism for H_2/O_2 provides the foundation for all combustion mechanisms. Similarly, mechanisms for the small C1 hydrocarbons CH_4, CH_2O, and CH_3OH are important components of mechanisms for almost all fuels. Including all fuels containing up to 3 or 4 C atoms (e.g., C_2H_6, C_2H_4, C_2H_2, C_3H_8, C_3H_6, C_3H_4, C_2H_5OH, CH_3CHO, etc.) provides a core mechanism that is of utility in modeling realistic fuels. Mechanisms for moderately larger species, like heptane and iso-octane, are valuable in understanding system size dependent variations.

The current computing ecosystem should allow us to develop a core mechanism for combustion that is of sufficient accuracy for the needs of the PreSICE. Such a development will require feedback between theory, modeling, and experiment incorporating sophisticated levels of uncertainty analysis. Although there is still major progress to be made, efforts along these lines are already showing great promise.[7,8]

The current ecosystem is also leading to realistic mechanisms beyond the core fuels. These mechanisms are becoming increasingly complete and formally correct, especially as automated mechanism generators[9] continue to improve their functionality. Coupling with uncertainty analysis will contribute to enhanced fidelity of these mechanisms. However, considerable empiricism will remain and large scale, on-the-fly, mechanism reduction for inclusion in CFD simulations will be required.

3. Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

Practical fuels are complex mixtures of many components, with many of the components having 10 to 20 heavy atoms. These complex mixtures are typically modeled with a surrogate fuel consisting of a handful of component fuels designed to mimic the properties of the real fuel.

Our current computing ecosystems cannot automatically produce a high accuracy mechanism for such surrogate fuels in the timeframe required to influence engine design because of multiple difficulties: (i) our current standard algorithms for predicting the kinetics of elementary reaction are too time consuming, poorly automated, and not readily applicable to molecules of the required size, (ii) similarly, current codes for automatically generating mechanisms need further developments in scale, in parallelization, and in coupling to a priori theoretical kinetics, (iii) the electronic structure methods of requisite accuracy for core mechanisms are not readily applied to systems of 10–20 heavy atoms, (iv) the number of reactions needing characterization grows combinatorially with fuel size, and (v) the mechanism must ultimately be reliably reduced to a size appropriate for CFD modeling. Current mechanism development approaches bypass these problems by employing low accuracy, empirical estimates for the vast majority of the reactions. Improved estimates will need to be implemented before the simulations can be truly predictive.

Current chemical modeling approaches were formulated in the 1970s with foundational assumptions required by the computational capabilities of that time. For example, a core assumption is that chemical species are thermalized between their bimolecular reactions. At combustion temperatures this assumption is often inadequate and at high pressures (up to 400 atm are of interest today) the assumption of isolated binary collisions is inadequate. Improved modeling approaches that address some of these failures are required before the simulations can be predictive without empirical adjustment whose fidelity for conditions outside their range of validation is uncertain. Finally combustion chemistry is not just a gas phase problem: the breakup/evaporation of liquid fuel and the transient formation and burnout of solid soot particles involve condensed media whose additional computational demands for modeling at the level of gas phase processes is poorly known at this time.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Hardware resources	The gas phase chemical kinetics community has not really even started to take advantage of today's computational resources.
2. Push to parallelization	Industry is seeing the need/value of PreSICE and so there is a push to parallelize the CFD codes. This in turn is leading to renewed interest in large-scale automated fuel/combustion chemistry.
3. Codes for potential energy surface exploration	Codes are beginning to become available for automatically exploring the potential energy surface.[10] Such codes greatly facilitate the development of fully automated fuel/combustion kinetics algorithms.

Impede	Why?
1. Application Codes	There are no chemical kinetics codes that have been designed for leading edge scale computers. Currently it requires a great deal of human effort to predict the rates for just one reaction.
2. Models and Algorithms	Algorithms are needed for (i) performing high accuracy electronic structure calculations on systems with 10–20 heavy atoms, (ii) for evaluating anharmonic effects on partition functions, (iii) for modeling the complexities of chemical dynamics and kinetics at high temperatures and pressures.
3. Workforce Development	There are very few chemical kineticists with the required training for developing and implementing novel large scale algorithms.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**
6. **References**

[1] A Workshop to Identify Research Needs and Impacts in Predictive Simulation for Internal Combustion Engines (PreSICE), Sponsored by Office of BES, Office of Science and the Vehicle Technologies Program, Office of EERE, US DOE, March 3, 2011; http://www1.eere.energy.gov/vehiclesandfuels/pdfs/presice_rpt.pdf.

[2] Reitz, R. D. Directions in Internal Combustion Engine Research. *Comb. Flame* **2013**, *160,* 1–8.

[3] Virtual Engine Research Institute and Fuels Initiative. http://verifi.anl.gov.

[4] Anand, K.; Reitz, R. D.; Kurtz E.; Willems W. Modeling Fuel and EGR Effects under Conventional and Low Temperature Combustion Conditions. *Energy Fuels* **2013**, *27*, 7827–7842.

[5] Som, S.; Longman, D. E.; Luo, Z. Y.; Plomer, M.; Lu, T. F.; Senecal, P. K.; Pomraning, E. Simulating Flame Lift-off Characteristics of Diesel and Biodiesel Fuels using Detailed Chemical-Kinetic Mechanisms and Large Eddy Simulation Turbulence Model. *J. Energy Res. Tech. – Trans ASME*, **2012**, *134*, 032204.

[6] A Workshop on Basic Research Needs for Clean and Efficient Combustion of 21st Century Transportation Fuels, sponsored by Office of Basic Energy Sciences, US DOE, Oct. 30 – Nov. 1, 2006; http://science.energy.gov/~/media/bes/pdf/reports/files/ctf_rpt.pdf.

[7] Varga, T.; Nagy, T.; Olm, C.; Zsely, I. G.; Palvolgyi, R.; Valko, E.; Vincze, G.; Cserhati, M.; Curran, H. J.; Turanyi, T. Optimization of a Hydrogen Combustion Mechanism using both Direct and Indirect Measurements. *Proc. Comb. Inst*. **2015**, *35*, 589–596.

[8] Burke, M. P.; Goldsmith, C. F.; Klippenstein, S. J.; Welz, O.; Huang, H. F.; Antonov, I. O.; Savee, J. D.; Osborn, D. L.; Zador, J.; Taatjes, C. A.; Sheps, L. Multiscale Informatics for Low-Temperature Propane Oxidation: Further Complexities in Studies of Complex Reactions. *J. Phys. Chem. A*, **2015**, *119*, 7095–7115.

[9] Green, W. H.; West, R. H.; and the RMG Team. RMG Py – Reaction Mechanism Generator (Python Version). https://github.com/ReactionMechanismGenerator/RMG-Py.

[10] Maeda, S.; Ohno, K.; Morkuma, K. Systematic Exploration of the Mechanism of Chemical Reactions; Global Reaction Route Mapping (GRRM) Strategy by the ADDF and AFIR Methods. *Phys. Chem. Chem. Phys*. **2013**, *15*, 3683–3701.

BES White Paper – Xiaosong Li (University of Washington)

1. **Please specify the current science drivers for your field of research.**

Photochemical processes in emerging non-linear and multidimensional spectroscopies, molecular collision processes in gas phase and surface-assisted chemistries, spin-selective electronic processes for next-generation data processing and storage, charge carrier dynamics that underpin energy conversion in solar cell technology, etc., all demand a quantum dynamical description of the various field-mediated interactions to be understood from a non-empirical, first principles vantage. Over the past decades, lower-scaling methods for treating the dynamics of many electron systems have seen great successes in both predicting and explaining complex chemical phenomena.[1–5] Specifically, the real-time, time-dependent mean-field methods which reduce the correlated, many-body wave function complexity to that of an effective single-particle interaction model have shown the best compromise of computational expense and accuracy.

Applications for current quantum chemical dynamics methods that are immediately relevant to the DOE BES mission include the resolution of non-equilibrium photophysical processes such as excitonic and free electron-hole pair dynamics at the heterojunctions of semiconductor devices, charge-transfer dynamics and vibrational trapping in semiconductor quantum dots and organic materials, spin-echo dynamics in magnetically-active materials, and spin-crossover dynamics in metal complexes. These processes are ubiquitous in energy conversion, energy storage, photocatalysis, information processing and storage, etc. and are foundational to numerous proposed technologies, most notably in the areas of photonics and spintronics. The information gained from the proposed research will have broad implications for each of these technologies by granting researchers across a broad range of disciplines a deeper understanding of excited-state electronic structure/function relationships. The products of this research field could furthermore result in the development of novel materials with new or enhanced photophysical properties for application in a variety of scientific contexts from fundamental research to energy conversion. Finally, this research will provide new fundamental insights into the interplay of excited electronic state dynamics and realistic environments in chemistry and physics, deepening our understanding of this important class of interactions.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

The time evolution of quantum systems is governed by the time-dependent Schrodinger (non-relativistic limit) or Dirac (fully-relativistic) equation. For many electron systems, one must invoke approximations to systematically approach the exact solution to these equations due to the complexity of the N-body problem. Even still, the effort expended to solve the approximate models increases very steeply with increasing system size. Despite many notable successes, substantial progress is still needed to expand the applicability and functionality of numerical quantum electronic dynamics methods to faithfully model systems of interest in modern chemical and materials science research. Efficient treatment of coupled electron-nuclear dynamics, dynamical electron correlations (driven by both charge and spin carrier dynamics), and realistic effective environmental degrees of freedom via quantum/classical and quantum/continuous-medium embedding approaches are research directions that will broaden the utility of these formalisms to accurately describe real-world chemistries beyond the few-atom system size limit.

While the vast majority of the electronic energy of a system is captured by the current mean-field approaches, the systematic errors in this description are very often on the same order as the energy of chemical bonds. Correlated electronic dynamics within the time dependent configuration interaction and coupled cluster formalisms will improve upon the current state-of-the-art DFT based methods for including correlation effects into electronic dynamics, albeit with elevated computational cost.

Correlated electron-nuclear dynamics within Bohmian mechanics framework or by utilizing semi-classical nuclear wave-packet approaches will enable the theoretical description of crucial chemical processes such as proton transfer that are qualitatively misrepresented by a classical mechanical description. Finally, quantum embedding schemes that exploit modern polarizable classical force fields can give quantum dynamics for chemical and materials systems electrostatically-corrected to account for a non-equilibrium chemical environment. Similarly, solving for the explicit time-dependence in the polarization of a continuous medium encapsulating a quantum mechanically described system can resolve dynamical effects of system-environment interactions that scale far more favorably than corresponding simulations where the environment is explicitly modeled as part of the system itself.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

There are many theoretical and technical challenges that will likely prevent the realization of fully relativistic electronic dynamics (especially in explicit magnetic fields) with completely coupled descriptions of electronic and positronic dynamics within the next decade, some of which are identified below. Beyond the unification of quantum chemistry with special relativity, one may hope to go beyond the classical description of the system-field interactions and incorporate quantum electrodynamics into simulations of realistic many-body systems for chemical and materials research. These developments represent some of the grandest challenges to the applied mathematics, computer science, and scientific computing fields, and will most likely not be fully addressed for many decades to come.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	**Why?**
1. Application codes	There is a lack of high-performance quantum electronic dynamics code that exploits the massively-parallel architectures of peta-to-exa-scale compute centers.
2. Models and algorithms	There is a lack of libraries that are specifically designed to facilitate arithmetic manipulations of multi-component electronic wave function.
3. Hardware resources	Many-electronic dynamics can be greatly accelerated by hardware resources with large memory sizes and bandwidths.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.**

   N/A

6. **References**

1) *F. Ding, J. J. Goings, H. Liu, D. B. Lingerfelt, X. Li "Ab Initio Two-Component Ehrenfest Dynamics," J. Chem. Phys., 2015, 143, 114105.*

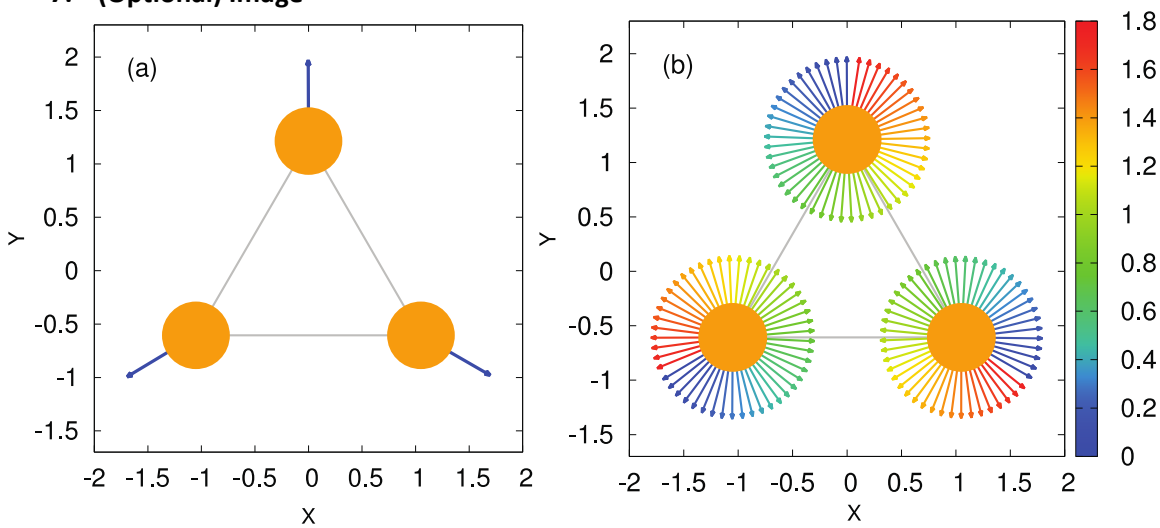
2) *B. Peng, D. B. Lingerfelt, F. Ding, C. Aikens, X. Li, "Real-Time TDDFT Studies of Exciton Decay and Transfer in Silver Nanowire Arrays," J. Phys. Chem. C, 2015, 119, 6421.*

3) *F. Ding, D. B. Lingerfelt, B. Mennucci, X. Li, "Time-Dependent Non-equilibrium Dielectric Response in QM/Continuum Approaches," J. Chem. Phys., 2015, 142, 034120.*

4) *F. Ding, J. J. Goings, M. Frisch, X. Li, "Ab Initio Non-Relativistic Spin Dynamics," J. Chem. Phys., 2014, 141, 214111.*

5) *F. Ding, E. Guidez, C. Aikens, X. Li, "Quantum Coherent Plasmon in Silver Nanowires: A Real-time TDDFT Study," J. Chem. Phys., 2014, 140, 244705.*

7. (Optional) Image



(Left) initial magnetization of Li_3 at t=0, and (Right) time evolution of the spin magnetization in a uniform magnetic field applied along z axis (the time-evolution is represented as the progression of coloration in picoseconds, and magnetization expressed in units of the Bohr magneton).

BES White Paper – Annabella Selloni (Princeton University)

1. **Please specify the current science drivers for your field of research.**

Understanding the fundamental atomic-scale mechanisms and kinetics of photocatalytic and solar energy conversion processes such as water splitting is critical for the design of new materials and architectures with very much needed improved efficiency. There are many and diverse theoretical and computational challenges in this field, ranging from the realistic description of complicated solid-liquid interfaces, including atomic geometries, electronic structures and excited state properties, to the accurate prediction of reaction pathways, free energy barriers, and reaction rates, requiring extensive sampling of the relevant potential energy surfaces. Advanced first principles methods, e.g., DFT with hybrid functionals, are essential for describing the electronic properties and excitations of the solid (often a transition metal oxide with significant correlation effects), the charge transport, the alignment of the electronic energy levels, and the chemical and electron transfer reactions at the interface. However, it is still computationally cumbersome or even prohibitive to apply these methods to the large systems that are typically of interest in this field. For instance, current studies are still unable to take into account important features such as the double layer at the semiconductor-electrolyte interface, which is likely to have a significant impact on the reactivity, or to model reactions at a given pH.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

We are currently able to carry out first principles (GGA or even hybrid DFT-based) molecular dynamics simulations on solid-liquid interface systems of several hundred atoms (depending on the materials) for times of the order 20 ps. Methods for extending the time scale of these simulations and improving the sampling of complex energy landscapes are needed in many cases. Such methods will become more easily available in the next ten years.

DFT, also with hybrid functionals, is not sufficiently accurate in several situations, e.g., for highly correlated oxides, which are of potential interest in photoelectrochemistry. More accurate quantum chemistry methods for systems of a few hundred atoms are likely to become available in the next ten years. These will be used as a benchmark for simpler approaches and for comparison with experiment.

Current methods for calculating excited state properties are still expensive and too heavy to be used routinely, thus comparison to experiment is often approximate. We expect that this difficulty will be largely overcome in the next 10 years.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Many photocatalytic materials undergo structural and compositional changes during operation that are difficult to incorporate into atomistic theoretical/computational schemes. Several other non-equilibrium and non-adiabatic effects like electron-hole recombination at trap states are also difficult to describe and it is unlikely that these difficulties will be overcome in a few years. These effects are not only scientifically interesting, but have also significant impact on the performance of the photocatalytic process. Overall, a full description of photocatalysis on realistic nanoparticles in a liquid electrolyte is a challenge that is unlikely to be solved in the next ten years.

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Models and algorithms	Essential for improving accuracy and extending the time scale of simulations; see point 2 above.
2. Hardware resources	Larger and more complex systems will be treated.
3. Application codes	Must be adapted to the hardware.

Impede	Why?
1. Workforce development	Very few long-term positions for people developing/improving codes in U.S. universities.

5. (Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.

The data ecosystem is not a high priority in this area.

BES White Paper – Lyudmila V. Slipchenko (Purdue University)

1. **Please specify the current science drivers for your field of research.**

The focus of my research program is on the development of theoretical and computational approaches targeting the electronic structure of extended systems, such as photosynthetic and fluorescent proteins, molecular solids, polymers, and bulk liquids. We use the developed techniques to investigate fundamental aspects of non-covalent interactions and the effect of the environment on electronic structure and dynamics. Below, I describe three areas that serve as science drivers for my research.

Excitation energy transfer in photosynthetic complexes

Solar radiation is the biggest energy source available for utilization by human society. Photosynthetic plants and bacteria developed supreme mechanisms of solar energy harvesting and photoprotection. Understanding of these processes holds the key to the solution of the world energy problem. Efficiencies and rates of excitation energy transfer (EET) in photosystems are controlled by the protein environment. Couplings between the electronic excited states are determined by relative orientation of chromophores, kept in place by protein matrix, and affected by protein non-uniform and time-dependent electric fields. However, details of the environment effects on the energy transfer and photoprotection in photosynthetic systems are elusive as rigorous account for protein in theoretical and computational studies of EET is challenging. Our goal is to develop universal polarizable QM/MM methods that accurately and efficiently reproduce influence of protein environment on EET, and extend this methodology to predictive modeling of EET in organics photovoltaics.

Non-adiabatic dynamics and vibronic interactions in aromatic molecules and materials

Understanding and control of dynamics of electronically excited molecules is quintessential for advances in science and technology. To explain processes such as conversion of solar to electrical energy in photosynthetic complexes and photovoltaic devices, the electronic and nuclear motions cannot be uncoupled such that the Born-Oppenheimer approximation should be abandoned. We develop theoretical models that treat vibronic interactions in a rigorous but computationally efficient way. Application of these models to increasingly more complex multi-chromophore systems (various diphenyl-alkanes, cyclo-para-phanes and cyclo-para-phenylenes) for which high-resolution experimental data are emerging (in a collaboration with Tim Zwier, Purdue) allow us to develop fundamental understanding of vibronic interactions and their spectroscopic signatures.[1]

Structure and charge-transfer dynamics in electrical energy storage materials

The current worldwide electric power generation capacity of ~20 terawatt hours is predicted to double by 2050 and triple by the end of the 21st century. Meeting this demand requires advances in the fundamental understanding and innovation of electrical energy storage (EES) technologies, including batteries and electrochemical capacitors. Key challenges include increasing energy density, capacitance, breakdown voltage, and conductivity by tailoring electrolyte and electrode structure. Addressing these challenges requires developing an improved fundamental understanding of the molecular interfacial properties that influence ion binding thermodynamics and kinetics. Our goal is to facilitate a rational design of advanced EES by providing insight into the questions such as the role of aromaticity and surface curvature in ion-binding at liquid interfaces and the influence of electrolyte (aqueous and non-aqueous) on thermodynamics and kinetics of ion binding. The work on model systems (aqueous and non-aqueous salt solutions of aromatic, mixed hydrophobic/hydrophobic and charged solutes, as well as solutes with concave and convex surfaces) is performed in collaboration with Ben-Amotz (Purdue), who is an expert in sensitive Raman measurements, and leads toward building fundamental understanding of various aspects of ion binding chemistry at interfaces.[2]

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

My expectation is that the methodology for modeling solvent effects will be established and the role of the environment on chemical structure, dynamics and optical properties will be clarified. Many fragmentation and embedding methods and polarizable force fields have emerged in recent years aiming at a more reliable description of extended systems than either classical QM/MM or implicit solvent models can provide.[3] However, the new techniques often involve complicated algorithms that are difficult to combine with implementations of advanced electronic structure methods, higher computational cost than the cost of implicit/classical explicit solvent models, and complexity in developing analytic gradients needed for dynamics simulations. Thus, there is a gap between these emerging methodologies and their practical applications to chemical structure, dynamics and photochemistry. However, my expectation is that continuous efforts toward development of software infrastructure, modular libraries, and scalable algorithms will facilitate a breakthrough in modeling extended systems within the next decade. This breakthrough will result in significant advances in understanding chemical structure and dynamics at interfaces, solvent effects on photochemistry and catalysis, etc.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

The current state of methods and algorithms for electronic excited states is far beyond the needs dictated by the progress in ultrafast spectroscopy and imaging techniques. Some of the associated challenges are low accuracy and/or high computational cost of excited state methods, as well as non-adiabtic couplings between the states that might significantly complicate modeling. Especially problematic are electronic states with double excitation character, Rydberg states, charge-transfer states (particularly for DFT functionals), and states in open-shell species. While developments of scalable tensor libraries, linear scaling algorithms and software infrastructure will certainly speed up calculations and make excited state dynamics simulations practical, the underlying deficiencies of electronic structure theories might impede the progress in this field. Thus, I doubt that in-depth understanding of excited state dynamics, e.g., in EET phenomena, will be reached in the next decade, unless revolutionizing models for describing electronic excited states emerge.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	**Why?**
1. Models and algorithms	To achieve our goals in all considered science drivers, new models and algorithms should be developed. We base our research on our own algorithms as well as the work of others.
2. Internal/external libraries/framework	Computer codes that we develop and use contain multiple components, e.g., electronic structure models, embedding models, dynamics models. Thus, libraries and frameworks for individual components simplify interface and dramatically enrich possibilities for new computational schemes.
3. Data workflow	Our modeling often includes multiple steps, e.g., preparing initial system, obtaining representative sampling, computing observables, data analysis and visualization, etc. Development of data workflows would dramatically accelerate and streamline our progress.

References

1. N. Pillsbury, N. Kidwell, B. Nebgen, L.V. Slipchenko, K. Douglass, J. Cable, D. Plusquellic, and T. Zwier, J. Chem. Phys. *141* 064316 (2014).

2. B.M. Rankin, M. D. Hands, D. S. Wilcox, L.V. Slipchenko, and D. Ben-Amotz, Faraday Disc., *160*, 255–270 (2013).

3. M.S. Gordon, S. Pruitt, D. Fedorov, L.V. Slipchenko, Chem. Rev., *112* (1), 632–672 (2011).

BES White Paper – Edward F. Valeev (Virginia Tech)

1. **Specify the current science drivers for your field of research.**

The main driver for the electronic structure theory in the context of DOE is understanding and controlling the chemistry of energy production and use. The scope of problems relevant to DOE is extremely broad, from basic chemistry of heavy-element compounds in solution and separations, to singlet fission in organic photovoltaics, to heterogeneous catalysis involved in operation of a battery/fuel cell, to understanding combustion of real fuels. Therefore, the electronic structure theory is essential to the missions of DOE and BES by providing the insight that cannot be easily obtained from experiment and by guiding new experiments.

Knowledge of electronic structure automatically provides more detailed information than can be experimentally accessed. The challenge is how to turn the electronic structure theory into an *equal* partner to the experiment in the energy context. The current mainstream tool, namely the Kohn-Sham Density Functional Theory (DFT), is often not accurate enough or even fails qualitatively. Thus, the driver for my field is the development of predictive models of electronic structure that have controlled accuracy to reveal what often are delicate energetic and dynamical factors involved in the DOE-relevant chemistries and are computationally robust to access the length/timescales necessary for building realistic models of chemical phenomena.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

This is probably best directed to application scientists, but here are some guesses:

- Quantitative understanding of structure and dynamics of water and solvated ions; this is a prerequisite for understanding how to model condensed phase chemistry in general.
- Quantitative models of natural photosynthesis.
- Quantitative simulation of surface catalysis.

These problems are hugely important, grand-challenge-type problems. Breakthroughs in these developments will results first and foremost from new methodologies and algorithms. Ab initio MD on thousands of atoms at post-DFT level will become routine within this time period.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

This is also best directed to application scientists, but here are guesses nevertheless:

- Complete simulation of a nanoelectronic device at an atomistic level.
- Catalyst design from first principles.

These problems are grand challenges.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Programming models, languages, and libraries	Modern parallel programming models (MPI, threads) are low-level tools that are not adequate for composing applications at high levels of abstraction. A new generation of programming models will allow researchers to develop software for the exascale machines at a much higher level of abstraction than the mainstream allows. Improvements in programming languages (e.g., increased role of functional-style programming in imperative languages) and development of standard libraries and frameworks will make it easier to write reusable code, and reuse other people's codes. This will also help to raise the level of abstraction without sacrificing performance.
2. Tooling	Modern tools for scientific software development are still relatively immature. I expect to see improved compilers, performance analysis, and other tools that will make it easier to develop high-performance codes.
3. Collaboration with computer science and applied math communities	Avoid reinventing the wheel and engage these communities by informing them of the needs of computational science.

Impede	Why?
1. Fault tolerance	Requires significant redesign of applications and perhaps even the algorithms, e.g., Monte-Carlo should be more fault tolerant than tightly coupled simulations.
2. Workforce	Increasing complexity of programming and theories mean greater barrier to entry; this is already a huge problem even when considering campus-level resources.
3. Hardware	More complex memory subsystem and the need to optimize for power will make programming more complicated. This again will call for redesign and co-design of software.

C.3 White Papers Addressing Materials and Chemical Discovery

BES White Paper – Hai-Ping Cheng (University of Florida)

Please specify the current science drivers for your field of research.

Interfaces and nano-junctions: The research focus of my group is on fundamental physical processes at the nano-scale. We investigate a wide array of physical problems, including interfacial properties and processes, electron transport across nano-molecular-junctions, electron relaxation dynamics and magnetic properties of nanoparticles/quantum dots, and reduction of thermal noise in optical coating materials. We construct and implement models based on first-principles for large-scale computational simulations, and we use computation and simulation for materials discovery that includes finding building blocks with intriguing physical properties as well as designing electronic and vibrational spectra as well as mechanical properties for future novel applications and high-precision measurements. An important part of our effort is to bridge between nano-scale and macroscale physics, for example, between atomistic quantum mechanical predictions and classical electromagnetism. We can make predictions for new experiments in addition to finding physical mechanisms for experimental measurements. For the next few years, we expect to be focused on understanding the fundamental physics underlying electron and spin transport through 2D junctions in the presence of single and double gates, electronic and magnetic properties of single-molecule nanomagnets, and optimal optical coating materials with minimal thermal noise.

Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.

Computer-driven discoveries: With rapid advancement in algorithms, computer hardware and software, especially the capacity of parallel computing, the extant computer ecosystem can (with sufficient allocation):

(a) *Enable* us to identify a large number of structure-property relations based on first-principles simulations. Bulk and interface structure often play vital roles in many physical and chemical processes. New functionality can potentially be discovered through materials genome initiatives as well as through a more conventional approach, i.e., by intelligent guesses. Databases can be built to shape and inform those guesses.

(b) *Enable* first-principles based engineering of electronic and magnetic structure using charge doping, impurities, and external electric fields for systems consisting 10^3–10^4 atoms (or 10^6–10^7 if limited to insulators and semiconductors). Most problems in systems made of layered 2D crystals will be solved. Additionally, we will understand many aspects of quantum dots, e.g., electron dynamics coupled with structural relaxation and phonon-electron interaction after photo-excitation. In coordination with experiment, we will discover optimal organic ligands to functionalize these quantum dots.

(c) *Enable* thorough examination of the content and implications for real materials of simplified physical models contrived prior to the computer era. Examples include Hubbard model, two-state model, classical dielectrics at nano-scale, etc.

Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

Hard Problems: (a) *Strong electron correlation* in electron transport. Currently, first-principles based theoretical treatment for electronic degrees of freedom is mainly at the level of density functional theory with generalized gradient approximations. For strong electron correlation, one can invoke several levels of greater refinement, e.g., hybrid functionals, "*+U*" methods, GW and random phase approximations, dynamic mean field theory (DMFT), etc. However, the more sophisticated the method is the more it is limited to prototype systems. Major breakthroughs are needed, not just incremental progress.

(b) *Material-specific computational design* of many-body electronic states. Despite the success of the materials genome initiative, discovery and characterization of exotic quantum states relies mostly on experiments. Theoretical and simulational predictions lag behind.

(c) *Electron dynamics beyond steady states*. More theoretical development is needed before simulation models can be constructed and tested. Some work is under way for nonequilibrium statistical mechanical simulations of high-temperature systems.

(d) *Poor scalability* (compared to quantum Monte Carlo) of electronic structure codes limits first-principles based calculations. Also, the huge human resource requirements to redevelop large electronic packages to new computer architectures are close to being a prohibitive barrier to progress. A specific problem is construction of Green functions in planewave codes. I cannot predict whether or not we can solve it within the timeline. But if successful, that would open a new avenue for tackling many-body problems.

What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why? *Suggested topics include the following.*

Accelerate	Why?
1. Computable theory and realistic model	Toy models are not enough for real-life materials.
2. Long-term, focused support for codes and algorithms	Algorithms and codes are critical to computation. It is counter productive to force science components especially experiments in one call.
3. Libraries and updates	Advancement in computer architecture.
Impede	**Why?**
1. Lack of fundamental theory for realistic systems	Theories for strong correlation are mostly built on model Hamiltonians.
2. Lack of massively parallel code and algorithm	The nature of Schrodinger equation.
3. Lack of commitment from BES	Code development should be separate components of science.

Application codes, models and algorithms, and work force development are three most important aspects to affect my research. These three are intertwined. My colleagues agree. As a computational physicist, I have three suggestions to U.S. DOE. First, we need to sustain long-term funding for algorithm and code development. The recent DOE call on software development [1] is a start. However, more funding should follow for smaller groups with demonstrated software records and more such groups. Examples are NSF SSE, SI^2, S^2I^2 programs. [2] It is important to focus on software instead of mixing with

applications beyond benchmark calculations. It is important to keep a track record on software sharing and popularity of codes (scaled to the size of the community), and use that as criteria for renewal. Second, organize efforts to rewriting major scientific software with state-of-the-art design ideas for exascale architecture. Third, organize nationwide meetings and workshops by multi-government agencies to educate researchers with modern computing philosophy and practice. The Big Data Hub [3] can be a model for a starting point. If a genuine commitment can be made at the national level, career paths can then be created in academic institutions to meet these needs.

References

[1] http://science.energy.gov/~/media/grants/pdf/foas/2015/SC_FOA_0001276.pdf.
[2] http://www.nsf.gov/pubs/2015/nsf15553/nsf15553.htm.
[3] http://www.nsf.gov/pubs/2015/nsf15562/nsf15562.htm.

BES White Paper– Marivi Fernandez-Serra (Stony Brook University)

1. **Please specify the current science drivers for your field of research.**

In computational condensed matter physics, and in particular in the area of electronic structure theory, there are two scientific directions driving the forefront research and advances in new methodologies and their applications. These are on one side research in materials for energy applications and research in quantum materials for new technological applications.

Functional materials, that is, materials whose properties can be controlled by external perturbations, allowing researchers to tailor their use at the time of synthesis, are key to both energy and new technologies applications. In order to understand how the composition and structure of the materials can determine their macroscopic properties it is necessary to have a tightly integrated theory-experimental approach.

Predictive theories are only predictive once they are confirmed by experiments. And experimental measurements are often subject to interpretations that only theory can clarify.

In the area of energy materials, the understanding of surface physics in aqueous environments or ambient conditions is critical. This involves dealing with alloy physics, impurities and liquids simultaneously. In terms of Electronic structure, ab initio molecular dynamics simulations are becoming standard. Therefore we have moved from a time when total energy methods where sufficient to a time where forces are critical and statistical methods to deal with dynamics and combined electronic and ionic time-scales are the state of the art. Calculations of non local correlations such as dispersion effects in liquid/solid interfaces, accurate calculations of band gaps and accurate ban alignments between solid and liquid semiconducting interfaces are leading the development of new simulation methodologies capable of treating the complexities and large sizes involved in the simulations.

For quantum materials, the treatment of medium to strong correlations and the treatment of localized electrons in a sea of delocalized electrons is critical. Accurate description and simulation of magnetic states and quantum phase transitions are needed for our electronic structure methods to have predictive power.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Understanding how surface effects in ambient conditions can control or affect bulk measurements is a critical challenge that currently is leading a lot of the research in oxide-based electronics. This is critical also when looking for materials for energy applications. In the next 5 years, we will be able to understand how much and why humidity (i.e. aqueous environments) affects the performance of functional materials like ferroelectric perovskites. We should be able to accurately compute band alignments and band gaps of complex structures. We shall be able to perform in-situ experiments, with dynamical probes and matching numerical simulations allowing to interpret the experimental results and modify the experiments on the fly. This will be possible as long as these experiments involve very short time scales (picosecond time scales). With the availability of extant computers we will be able to obtain “experimental accuracy data” at the level of quantum Monte Carlo simulations, that can be used to fit new density functionals to treat a specific class of materials, by looking for a functional which better fits a set of very accurate data obtained with experimental accuracy. Therefore density functional theory will become the standard method for simulation for materials and systems sizes that previously needed to be described with semiempirical methods.

Another challenge that needs to be addressed is linked to uncertainty quantification (UQ). The electronic structure community needs to set up standards for UQ and have a dialogue with experimentalists on analyzing and putting together data with the correct UQ.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

One of the most important problems yet to be solved is the treatment of long time scales and non-equilibrium behavior. Degradation of materials and predicting their life time (and potential hazards) are a challenge for current simulation techniques.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Data workflow	More data will be available, and benchmarks and tests will be done at faster paces.
2. Models and algorithms	They are always in continuous and positive progress.
3. Internal/external libraries	Developed by experts to make algorithm implementation easier.

Impede	Why?
1. Development and portability of application codes	Students need to be prepared with computational background that they currently do not have and advisors cannot keep up to date with new languages and new architectures.
2. Analysis resources	There needs to be a continuous support for small computing and not put all the money in Extent computing.

BES White Paper – Exascale Computing Requirements for Predicting Pathways of Nucleation and Assembly:

Christopher J. Mundy and Greg Schenter (Pacific Northwest National Laboratory)

Please specify the current science drivers for your field of research: One of the key challenges in chemical physics is our ability to predict and understand phenomena that occur at the solid-liquid interface. The implication of controlling processes at the liquid-solid interfaces is far reaching and is a cross-cutting science theme common to energy storage (electrode design), catalysis, geosciences, and materials synthesis. The underlying scientific principles and phenomena demand an accurate description of molecular-scale mass, charge, and electrodynamic fluctuations in the vicinity of the liquid-solid interface. It is the nature of these fluctuations that are responsible for the coupling of length-scales (*e.g.,* the *mesoscale* problem[1]) through the distinct short-range interactions and long-range collective response (both electrostatic and dispersion). The consistency of phenomena on these scales will require an intense enhancement of fundamental understanding. Science drivers are aimed at quantifying the breakdown of conventional mean-field frameworks, such as Classical Nucleation Theory, and Derjaguin, Landau, Verwey, and Overbeek (DLVO) Theory, that are an essential part of the aforementioned complex phenomena. The self-consistent incorporation of the important molecular detail into new, verifiable, frameworks is essential to future advances. The importance of these calculations is to provide a foundation for demonstrating consistency of phenomena by mapping the molecular framework (including electronic structure to represent charge density, bond breaking and polarization response) to frameworks of reduced complexity (such as continuum frameworks, generalized Langevin equations) to predict quantitative outcomes of nucleation and assembly under relevant experimental solution conditions.

Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems. Our approach to this problem is to determine the correct representation of molecular interaction and collective response that produces an accurate description of the short-range solvent response to broken symmetries of the form of interfaces and solutes[2,3]. The measurable manifestation of the solvent response will be forces between nanoassemblies, and a wide variety of structural, thermodynamic, and kinetic surface phenomena in the presence of aqueous electrolytes, all verifiable through experimental measurement. Current computational resources and algorithms (*e.g.,* CP2K) will allow for the converged calculation of solvent response to model interfaces, isolated complex ions, and the free-energetics of ion-pairing in an explicit solvent using electronic structure methods. These tools represent an essential balance between efficiency and accuracy that is required to achieve significant impact. The result of these calculations can be used to construct reduced models of the initial stages of nucleation and assembly producing a picture of how the details of molecular interaction influence the pathways and kinetics. One outcome is to produce a description of short-range many-body dispersion that is consistent and in balance with short-range many-body electrostatics.

Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems. Using a single atomistic computational framework to predict the outcomes of nucleation or assembly as a function of solution conditions is beyond the current capabilities.

What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years?

Accelerate	Why?
1. Theoretical frameworks for coupling models across scales	This is an important theoretical step of identifying when mean-field theories (*e.g.,* DLVO) are correct and where the difficult atomistic calculations need to be performed. This will allow us to focus our computer resources on the important aspect of the complex problem.
2. Acceptance of ensemble based computing as a legitimate way of utilizing exascale resources	Statistical mechanics is a naturally parallel theory. Exploiting our already efficient electronic structure code (*e.g.,* CP2K) to compute converged multidimensional free-energy surfaces of important correlated phenomena allow for rapid progress in coupling the molecular to the continuum.
3. Routine access to computer allocations of 50 M core hours.	We already have efficient electronic structure codes to compute potentials of mean force (PMFs) on > 1,000-atom systems. More routine access to larger allocations to exploit the natural parallelism of statistical mechanics will be necessary to obtain benchmark calculations on using electronic structure methods to describe liquid-solid interfaces.

Impede	Why?
1. Workforce development	We do not have a sufficient supply of early-career scientists trained in high-performance computing to accelerate the development of the domain specific codes needed.
2. Focusing on scaling over time-to-solution for access to large allocations	Time to solution and generation of converged results is what the scientific community should focus on. It is necessary to have a better balance between demonstrating the importance of scaling to the computer vendor and producing useful, state-of-the-art results.

References

1. From Quanta to the Continuum: Opportunities for Mesoscale Science, BESAC Report, 2012.
2. Remsing, Richard C.; Baer, Marcel D.; Schenter, Gregory K.; Weeks, JD "The Role of Broken Symmetry in Solvation of a Spherical Cavity in Classical and Quantum Water Models," Journal of Physical Chemistry Letters 5, 2767–2774 (2014).
3. Chun, Jaehun; Mundy, Christopher J.; Schenter, Gregory K, "The Role of Solvent Heterogeneity in Determining the Dispersion Interaction between Nanoassemblies," *Journal of Physical Chemistry B* **119**, 5873–5881 (2015).

BES White Paper – John Pask (Lawrence Livermore National Laboratory)

1. **Please specify the current science drivers for your field of research.**

Where we are today? Be sure to include broad impact, DOE interest, ties between experiment/theory, etc.

The primary science driver for our SciDAC project, "Discontinuous methods for accurate, massively parallel quantum molecular dynamics: Lithium ion interface dynamics from first principles," is efficient, reliable, and safe energy storage. Such storage is critical to move from fossil fuels to clean, renewable alternatives such as solar and wind, which have intrinsic variability. Our particular focus is on the understanding and improvement of Li-ion battery cells, which have revolutionized consumer electronics and promise to do the same for transportation and electrical distribution.

A key factor in the performance, lifetime, and safety of Li-ion cells is the solid-electrolyte interphase (**SEI**) layer between the anode and electrolyte, a product of electrolyte decomposition. In or work, we seek to understand the chemistry and dynamics of the SEI layer by developing and applying first-principles quantum mechanical electronic structure methods to reach for the first time the length- and time-scales necessary to understand this critical interface.

The most widely used method to carry out such quantum molecular dynamics (**QMD**) simulations is the planewave (**PW**) method, as implemented in VASP, ABINIT, and Quantum Espresso codes, among many others. However, PW methods are not able to reach the necessary length- and time-scales to model the chemistry and dynamics of interest due to the global nature of the basis and cubic computational scaling with system size of the underlying orbital-based Kohn-Sham solution. In order to reach the necessary length- and time-scales, we proceed along two parallel paths: (1) development and application of the massively parallel Qbox PW code, specifically targeted for leadership-class computational platforms, for studies of systems of up to 2,000 atoms; and (2) development and application of new discontinuous Galerkin (**DG**) and Pole Expansion/Selected Inversion (**PEXSI**) electronic structure methodologies to reach sizes of 10,000 atoms and more. These new methods substantially reduce the need for interprocessor communications in parallel implementations by recasting the solution of the required quantum mechanical equations in a strictly local, systematically improvable, discontinuous basis, while improving scaling with system size by eliminating the need for matrix diagonalization entirely through pole expansion of the Fermi operator in the complex plane. The success of these new methodologies hinges on the ability to leverage large numbers of computing cores on leadership class machines. The success going forward will depend critically on the ability to fully leverage the more heterogeneous architectures to come.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What will probably be solved in the next 5–10 years? Why is this important to the field?

In the next 5–10 years, leveraging such methodologies as developed in the present work on the largest existing computational platforms, the basic chemistry and dynamics of the SEI layer in Li-ion cells will be clarified, including the initial phases of SEI formation and Li intercalation into the anode, with various combinations of anodes, solvents, ions, and counter-ions. This will require numerous QMD simulations of systems containing 10,000 atoms or more for 50 ps or so each. The understanding, predictions, and guidance to experiment resulting from these simulations will then inform higher-level models, which will in turn enable fundamental advances in battery design.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What might not necessarily be solved in the next 5–10 years? Again, what is the importance?

In the next 5–10 years, using the largest existing computational platforms, the larger-length-scale, longer-time aspects of the chemistry and dynamics of the SEI layer in Li-ion cells will *not* likely be clarified, due to insufficient computing cores and memory available. To eliminate finite-size effects, reduce statistical uncertainties at practical ionic concentrations with relatively slow diffusion processes, investigate processes beyond initial formation, and follow longer-time processes such as intercalation, systems sizes considerably larger than 10,000 atoms and simulation times exceeding 1,000 ps will be required. Improved physics will also require more advanced exchange-correlation functionals, including van der Waals interactions. Such simulations will require orders of magnitude more computational cores and memory than currently available and algorithms and code which can effectively leverage them.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Density functional theory (DFT) algorithms applicable to metals and insulators that scale linearly (or at worst quadratically) with system size and can leverage very large numbers of computational cores and total system memory.	Current DFT algorithms applicable to metals and insulators scale cubically with system size and are limited in the number of computational cores they can effectively leverage due to the need for substantial interprocessor communications (Fig. 1). This severely limits the number of atoms and simulation times attainable, leaving no choice but to use much less certain empirical approaches for numerous practical Li-ion interface systems of utmost interest.
2. Compilers, math libraries, and parallelization libraries which can effectively leverage heterogeneous computational nodes with large numbers of computational cores of different types and speeds (e.g., CPU, GPU) and highly nonuniform memory. The ideal is that math and parallelization libraries would be optimized for the hardware such that well written MPI+OpenMP codes using standard libraries such as BLAS/LAPACK/ScaLAPACK/FFTW could run as efficiently on the coming machines as on existing ones, with minimal modification. If the computer scientists can make this happen, the domain scientists can do the rest.	Well-optimized parallel communications (e.g., MPI, OpenMP) and math (e.g., BLAS, LAPACK, ScaLAPACK, FFTW) libraries are critical to electronic structure code performance. But current libraries, such as ScaLAPACK, do not efficiently manage local memory on heterogeneous architectures, e.g., repeatedly moving the same data back and forth between CPU and GPU for each subroutine call within a larger routine, rather than keeping the data local to cores that process it, thus incurring substantial latency costs.
3. Hardware resources that allow access to many cores (e.g., 10^6 or more) with memory per core at least that of present IBM BG/Q machines.	QMD simulations require tens to hundreds of thousands of MD steps to complete, placing extreme pressure on wall clock time per step. These times can be brought down only by

	leveraging a sufficient number of computational cores. In addition, DFT simulations require substantial memory per core. At the level of BG/Q machines, for example, it is often necessary already to distribute to more computational nodes than needed for the number of cores in order to access the needed memory. A reduction of memory per core below current levels will result in still more cores being effectively idle.

Impede	Why?
1. Absence of all three "accelerate" items above.	

Image:

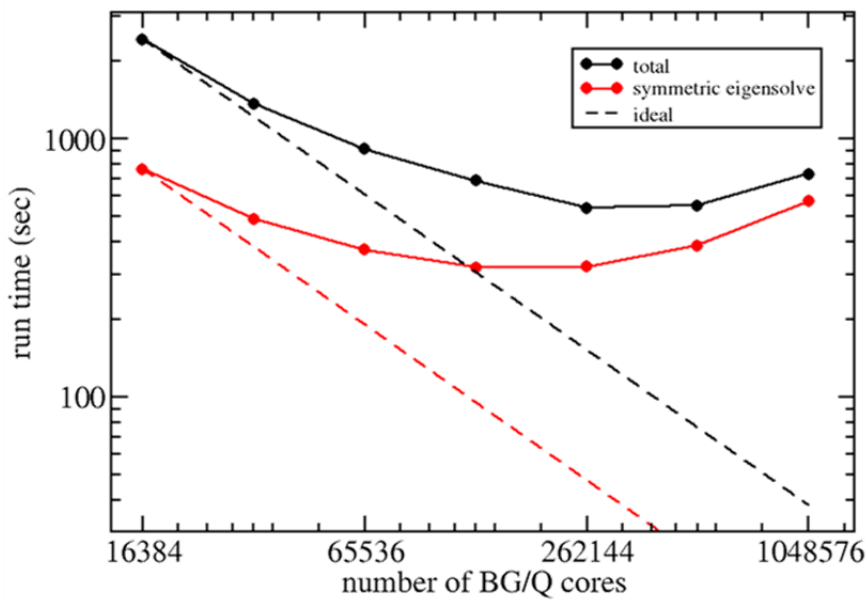


***Figure 1**: Relative time spent in the symmetric eigensolve compared with total run time as a function of number of Blue Gene/Q cores for a Qbox simulation of 1,600 gold atoms.*

BES Review White Paper – J. Ilja Siepmann (University of Minnesota)

Preamble: This white paper focuses on computing ecosystem requirements pertinent to research carried out in the Siepmann group and does not address all of the divergent requirements of the Nanoporous Materials Genome Center.

1. **Please specify the current science drivers for your field of research.**

The scientific motivation is to enhance the capabilities and to use predictive modeling (particularly, molecular simulations using force fields or Kohn-Sham DFT to describe the interacting particles) to provide accurate thermophysical properties and molecular-level understanding that aids in the design of improved separation processes and materials. Most chemical separations currently rely on highly energy-intensive processes (e.g., distillation) and improved processes involving lower energy consumption and less harmful solvents are essential ingredients for the path toward sustainability. Predictive modeling is most beneficial for high-throughput screening, for experimentally challenging conditions (e.g., high temperature, high pressure, or toxicity of compounds), and when molecular-level insight is needed to understand separation mechanisms. The research relies on a synergy between experiment and theory (e.g., data needed for force field development or materials leads needed for directed synthesis).

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Discovery of crystalline sorbents and liquid extraction solvents for numerous separation problems that currently are carried out by energy-intensive traditional separation processes. Over the next 5–10 years, predictive modeling will be able to tackle high-throughput screening even when the sorption or extraction process leads to structural changes of the host material. In addition, predictive modeling will be able to address transport properties of medium-sized molecules in micro-structured environments. Furthermore, propagation of uncertainties (beyond statistical uncertainties) will become a routine endeavor.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Accurate simulations of materials synthesis (e.g., the hydrothermal synthesis of zeolites or high-temperature synthesis of carbonaceous materials with desired pore size distribution and chemical functionality) cannot be tackled with current computational resources because the complete synthesis requires predictive modeling of highly-interdependent processes that occur over large ranges of length and time scales. Moving from materials discovery to application requires synthesis of the actual materials. Predictive modeling of phenomena involving microstructured and porous environment where length scales need to extend to hundreds of nanometer cannot be solved with current computational resources; examples include the stability of surfactant aggregates, the formation of meso-scale hydrates and asphaltene aggregates, or the flow properties of enhanced oil recovery formulations in realistic rock pores. Precise and accurate first principles simulations of phase, sorption, and reaction equilibria; using a first principles description of the interactions and treatment of nuclear quantum effects increases the computational requirements by factors of 10^6 to 10^{10} compared to force-field based simulations sampling from the classical mechanical phase space. Using a first principles description will extend the types of materials and processes amenable to predictive modeling.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Models and algorithms	Predictive modeling using particle-based simulations relies on suitable models (that determine the accuracy of the predictions) and on efficient algorithms (that determine the precision of the predictions). Significant advances have been made in these aspects over the past 5–10 years and will hopefully continue.
2. Hardware resources	As predictive modeling is applied to ever more complex systems (with respect to length scales and complexity of the models used to describe the interacting system), the need for CPU time, memory, and I/O will increase exponentially. It is expected that hardware resources will continue to grow over the next 5–10 years.
3. Application codes (development)	Availability of application codes for molecular dynamics simulations and electronic structure calculations will improve.

Impede	Why?
1. Data workflow	Particle-based simulations generate ever more data, but broadly applicable approaches for data compression (e.g., finding uncorrelated configurations), transmission, and archiving are lacking.
2. Application codes (portability)	It appears that codes are again becoming tailored to specific computational hardware and, hence, portability decreases.
3. Hardware resources	Many science problems require long calculations to reach equilibrium that cannot be replaced by large numbers of independent simulations nor simulations for larger system sizes. Hardware resources for jobs with run times in excess of 1,000 hours are lacking.

5. **(Optional) Image (next page)**

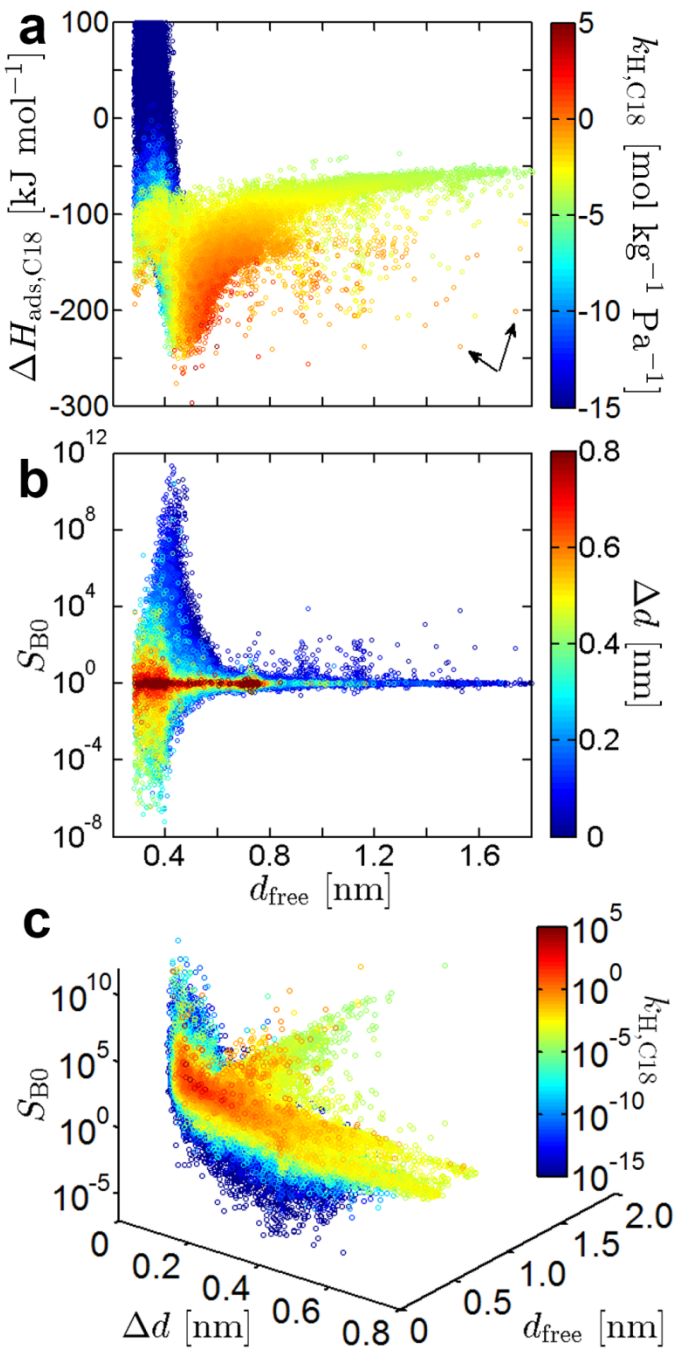


Figure 1. Adsorption properties for six hydrocarbons (with 18 to 30 carbon atoms) at the infinite-dilution limit computed for 330,000 zeolite structures. (**a**) Scatter plot of the adsorption enthalpy for *n*-octadecane ($\Delta H_{ads,C18}$) versus the free pore diameter (d_{free}) with the color scheme indicating Henry's constant ($k_{H,C18}$). (**b**) Scatter plot of linear-versus branched selectivity (S_{B0}) versus d_{free} with the color scheme indicating the pore bumpiness (Δd). **(c)** Scatter plot of S_{B0} versus Δd and d_{free} with the color scheme indicating $k_{H,C18}$. Taken from P. Bai, M.Y. Jeon, L.M. Ren, C. Knight, M.W. Deem, M. Tsapatsis, and J.I. Siepmann, "Discovery of optimal zeolites for challenging separations and chemical transformations using predictive materials modeling," *Nat. Commun*. **6**, art. no. 5912 (2015).

BES White Paper – Directing Matter to Deliver Functional Materials:
Bobby G. Sumpter (Oak Ridge National Laboratory)

Please specify the current science and operational drivers for your User Facility

The Center for Nanophase Materials Sciences (CNMS) is one of five Nanoscale Science Research Centers (NSRCs) that was established as part of the Department of Energy's (DOE) Office of Science contribution to the U.S. Government National Nanotechnology Initiative (NNI). The CNMS provides a diverse user community with access to state-of-the-art nanoscience research capabilities, expertise, and equipment, and applies these resources to execute a cutting-edge science program with emphasis in theory and simulation, nanofabrication, macromolecular synthesis and characterization, and understanding of structure, dynamics and functionality in nanostructured materials using scanning probe microscopy, electron microcopies, neutron scattering, optical spectroscopy, helium ion microscopy, and atom probe tomography. Motivating the research at the CNMS is the realization that the broadest range of energy applications relies on materials that are enhanced or even enabled by nanoscale phenomena. As described below, this applies in particular to batteries, supercapacitors, photovoltaics, catalysts, thermoelectrics, and many additional functional materials as well as structural materials. Therefore, the central motivation for the work at the CNMS can be summarized as our desire to **harness energy through nanoscience**. To this end, the mission of the CNMS is twofold: to enable the external scientific community to carry out high-impact nanoscience research through an open, peer-reviewed user program, and to conduct in-house research to understand and control the complexity of electronic, ionic and molecular behavior at the nanoscale to enable the design of new functional nanomaterials. These two aspects of the CNMS' mission are closely linked and mutually benefit from each other. In particular, the partnering of key groups of users who bring outside expertise with the sustained scientific in-house effort allows the center to be a leading effort in the development of new tools and methods for nanoscience, including synthesis, theory/modeling, and characterization.

As a case study, our experimental tools in imaging provide picometer resolved images at a fast rate and when coupled with a modern camera are capable of providing several hundreds of frames a second, it pushes the data size into the several hundreds of terabytes per experiment for a single microscope [1,2]. What is needed is the ability to analyze this data during time of the experiment; this would provide rapid feedback to-and-from models and simulations to both inform and validate the simulations/model that can subsequently provide information that is missing or guidance toward manipulation of the materials to achieve new or improved properties. This is key to address **three major gaps** in the current paradigm of materials design and discovery that typically is not tightly integrated and proceeds via synthesis ==> characterization ==> theory. **First**, there is the need for reliable computational techniques that accurately (and rapidly) address the increasingly complex functionalities required for today's technological applications, that provide the precision necessary for discriminating between closely competing behaviors, and that are capable of achieving the length scales necessary to bridge across features such as domain walls, grain boundaries, and gradients in composition. True materials are far more complicated than the simple structures often studied in small periodic unit cells by electronic structure calculations. The large length and time scales as well as finite temperatures, makes even density functional theory calculations for investigating materials under device-relevant conditions prohibitively expensive. Grain boundaries, extended defects and complex heterostructures further complicate this issue. **Second**, there is a need to take full advantage of all of the information contained in experimental data to provide input into computational models to predict and understand materials across scales. This includes integrating data efficiently from different characterization techniques to provide a more complete perspective on materials structure and function. In summary, we need more work focused toward filling these critical gaps by delivering a set of open source petascale multiscale simulation, data assimilation, and data analysis tools for functional materials design, in an approach that includes uncertainty quantification and experimental validation. Finally, pathways for actually making materials need to be established. While in general pathways for making materials are least amenable to theoretical exploration

due to the daunting dimensionality (plethora of metastable states and pathways accessing those) and primarily rely on the expertise of individual researchers, exascale computing capacities and big-deep data approaches offer new possibilities to bridge this gap. Future breakthroughs in materials design will depend on the availability of high-end simulation software and data analytics that can effectively utilize the gamut of computational resources, especially the leadership class.

Describe the broad computational and data challenges expected to be faced in the 2020–2025 timeframe.

Materials and chemistry are increasingly dependent on the ability to efficiently capture, integrate (to models and simulation), analyze (during the time of the experiment), and steward large volumes of diverse data. For example, understanding and ultimately designing new materials with complex properties suitable for energy applications will require the ability to integrate and analyze data from multiple instruments designed to probe different ranges of length, time, and energy. Chemical imaging brings other challenges as many of the imaging techniques are destructive. These techniques therefore require near real time analysis to determine structural evolution properties of the material while there is still significant material left to image. These, and many other scientific pursuits, require data science and network resources that are often distinct from, but complementary to, the computational science resources provided by traditional HPC facilities. Lossless data compression will be important.

Describe the broad computational and data challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

The size of simulation and experimental data sets will dramatically increase. This will be result from both the deployment of enhanced computational resources and from improved scalable algorithms to utilize those resources, which will permit the simulation of considerably larger nanostructured systems and phenomena on longer time scales. In addition, there will be a commensurate enhanced demand for *remote access, visualization and control, and analysis of simulation and experimental data* that will require greater network abilities. Our experimental facilities will provide considerably higher intensities to individual beamlines and coupled with corresponding advances in detector technology, this will result in unprecedented rates of data collection. The sheer volume and velocity of data from individual experiments and calculations will also increase, as scientists are able to probe increasingly complex questions that probe more subtle properties and phenomena. Additionally it will become more common to find simultaneous users at multiple experimental or computational facilities and thus the need to coordinate improved capabilities for theory/simulation guided experiments, data movement, fusion, and efficient analysis. In the future, bringing all types of disparate data to bear on a particular discovery process or mission outcome could be a new frontier of science and technology.

An aspect of theoretical materials sciences that might be *beyond exascale* is enabling full many-body theory for treating strongly correlated materials.

What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. An efficient workflow for data, modeling & simulation (this would include aspects of visualization and analysis)	This is largely lacking and it slows progress tremendously. Additionally only a small fraction of the actual data stream is captured/analyzed and that is rather slow.
2. A robust integrated feedback loop between experiment, models, and simulations	Feedback that allows in situ modifications (either to the experiment or to the model/simulation) are still largely lacking. This needs to change to enable materials discovery and design.
3. Application codes that scale, that are reasonably portable, and that are **open source**	Many traditional codes do not scale well and are not open source. This needs to be mediated to the greatest extent possible to enable large scale science exploration and potential breakthrough discoveries in a time effective fashion.
4. Workforce development	Students need to be exposed to big-deep data analytics, high performance computing, and computational physics/chemistry along with the traditional core courses of traditional science disciplines.

Impede	Why?
1. Code development	Requires substantial resources and is often not conducive to publications.
2. Resistance to new paradigms	"If it isn't broke don't fix it" attitude.
3. Discipline Language barriers	Computer science, mathematics, physics, chemistry, biology, materials science, and engineering all have their own basic language which doesn't enable easy communication of problems in the require multidisciplinary ecosystem.

References

[1] S.V. Kalinin, B.G. Sumpter, R. K. Archibald, Big-Deep-Smart Data in Imaging for Guiding Materials Design, *Nature Materials*, **14**, 973–980 (2015). DOI: 10.1038/NMAT4395

[2] Bobby G Sumpter, Rama K Vasudevan, Thomas Potok, Sergei V Kalinin, A bridge for accelerating materials by design.

Image:

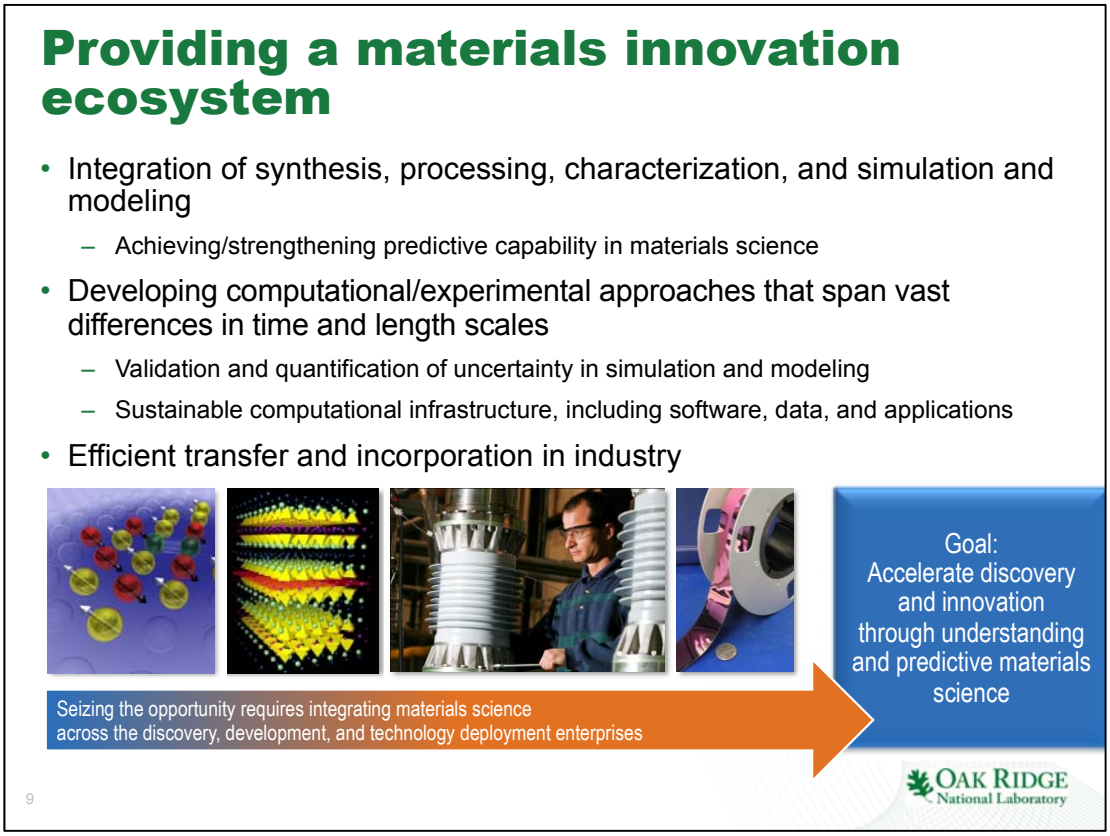


Figure: Concept of a *materials innovation ecosystem*; Next generation integration of experimental synthesis, characterization with simulations and modeling to drive materials discovery and innovation towards deployment.

C.4 White Paper Addressing Soft Matter

BES White Paper – Priya Vashishta (University of Southern California, in collaboration with Rajiv K. Kalia and Aiichiro Nakano)

Current science drivers for your field of research: Layered Materials Genomics (LMG)

The recent confluence of advancements in experimental synthesis, characterization by ultrafast X-ray laser experiments, and scalable simulation algorithms on multi-petaflop/s (10^{15} floating-point operations per second) parallel supercomputers have, for the first time, brought "first-principles-based computational synthesis" into the realm of reality. A notable prototype is "layered materials genome (LMG)." Functional layered materials (LMs) will dominate nanomaterials science in this century [1]. The attractiveness of LMs lies not only in their outstanding properties, but also in the possibility of tuning these properties in desired ways. Scientists can tune electronic, optical, magnetic, and chemical characteristics by inducing defects or dopants in LMs; by initiating interaction with external materials, such as small molecules; or by building van der Waal heterostructures with various LMs. Another attractive possibility is the introduction of multiple properties on different locations of a single nanosheet, which may lead to a completely new method of device fabrication and architectures. LMG has broad applications in electronics, as well as energy generation, harvesting, and storage.

LMG software will aid the synthesis of stacked LMs by chemical vapor deposition (CVD), exfoliation, and intercalation. The software will provide function-property-structure relationships in LMs functionalized by initiated CVD processes and will be sufficiently general to support the synthesis and characterization of other functional nanomaterials. The software suite will include plug-ins for a wide range of properties and processes, including band structures and carrier effective masses, oscillator strengths and optical absorption/emission spectra, dielectric function, electric carrier mobility and radiative recombination time, excitonic properties, electrical conductivity, heat and mass transport, and various methods for free-energy calculation.

The software will play a particularly important role in the analysis of ultrafast X-ray laser experiments on LMs. Function-property-structure relationships in stacked LMs span a wide range of length and time scales. Catalysis, plasmonics, charge separation and recombination, multielectronic excitation and exciton formation in LMs can best be studied through a joint experimental-computational approach. Experimentally, THz, optical, and UV pumps can be used for X-ray laser pump-probe to study electronic and phononic processes. In conjunction, quantum dynamics simulations should be performed that can describe nonadiabatic processes to interpret pump-probe experiments at LCLS.

Science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems: Merger of HPC and BDA and Emergence of Discovery Informatics

As articulated in the recent Executive Order by the U.S. President on National Strategic Computing Initiative (NSCI) [2], the next 10 years will see the convergence of high-performance computing (HPC) at exaflop/s (10^{15} floating-point operations per second) and Big Data analytics (BDA) at exabytes scale [3]. Until now, HPC and BDA have formed nearly disjointed computing ecosystems. The new HPC-BDA merger will enable a new generation of computational synthesis, in which large-scale quantum molecular dynamics, reactive molecular dynamics, and mesoscale simulations boosted by accelerated dynamics [4,5] will be coupled *in situ* with emerging "materials discovery informatics" tools [6] to acquire new scientific knowledge on the fly. In scientific discovery, many human activities are a bottleneck for progress, and artificial-intelligence-based systems implemented in the HPC-BDA computing ecosystem will augment human cognition.

LMG software should be scalable and extensible, featuring synthesis and characterization modules; new, community-developed plug-ins; and high-throughput and time-to-solution algorithms. The

software should run on all platforms, from desktops to petascale and exascale architectures. This will require simulation and data analysis/visualization/understanding algorithms that will continue to scale on continuously evolving computing architectures. Several "metascalable" algorithmic frameworks such as divide-conquer-recombine (DCR) have been proposed to design linear-scaling algorithms, in which the computational complexity is linear in the problem size [4,5]. In addition, fault resiliency should be built in the software to autonomously bypass unexpected failures on multimillion core architectures.

Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

The computational and experimental techniques developed in the next 10 years for targeted materials systems such as LMs are expected to be applied to broader and novel materials, such as optical and mechanical metamaterials with, for example, negative refractive index and compressibility.

What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Merger of HPC and BDA	Will allow peta- to exascale simulation data to generate scientific knowledge.
2. Exascalable algorithms	Increasingly large systems with interfacial and defect structures can be simulated.
3. *In situ* data analytics with high-throughput, time-to-solution approaches	Large spatiotemporal-scale simulation data can be coupled with discovery informatics.

Impede	Why?
1. Exaresilience	If not addressed, increasing mean time to failure will impede productivity.
2. *In situ* data analytics	Conventional file-based scientific workflows will not scale in exascale HPC-BDA platforms.
3. Uncertainty quantification	A lack of reliability will severely limit the predictive power.

Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.

Layered materials genome will combine exabyte simulation and experimental data produced at the X-ray laser facility at Stanford, as well as materials data produced at other remote locations. High-bandwidth, wide area networks, efficient data transmission protocols, and data compression schemes (e.g., compressed sensing) are essential requirements.

1. References

1. A. K. Geim and I. V. Grigorieva, "Van der Waals heterostructures," *Nature* **499**, 419 ('13).
2. Barak Obama, President, "Executive Order—Creating a National Strategic Computing Initiative," (July 29, 2015).
3. D. A. Reed and J. Dongarra, "Exascale computing and big data," *Commun. ACM* **58**, 56 ('15).
4. F. Shimojo *et al*., "A divide-conquer-recombine algorithmic paradigm for large spatiotemporal quantum molecular dynamics simulations," *J. Chem. Phys.* **140**, 18A529 ('14).
5. N. A. Romero *et al*., "Quantum molecular dynamics in the post-petaflop/s era," *IEEE Computer*, in press ('15).
6. Y. Gil *et al*., "Amplify scientific discovery with artificial intelligence," Science **346**, 171 ('14).

Images:

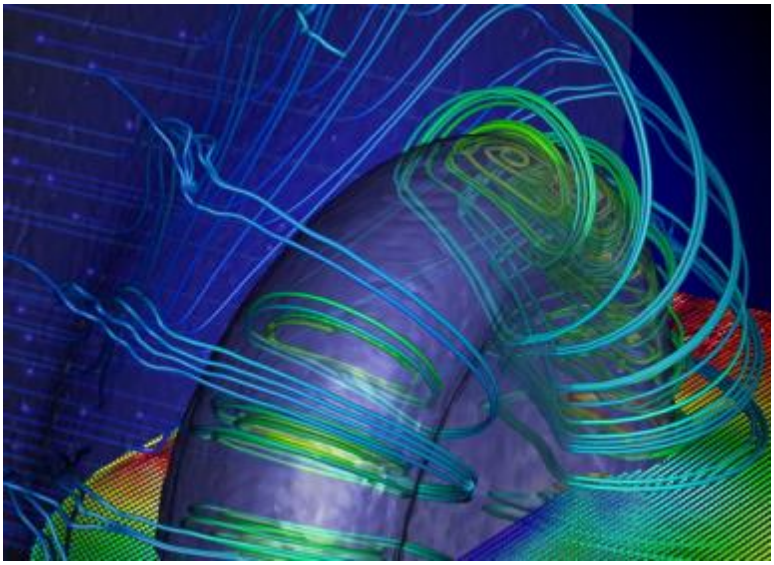


Fig. 1: Billion-atom reactive molecular dynamics simulation of cavitation bubble collapse in water [A. Shekhar *et al*., *Physical Review Letters,* **111**, 184503 (2013)].

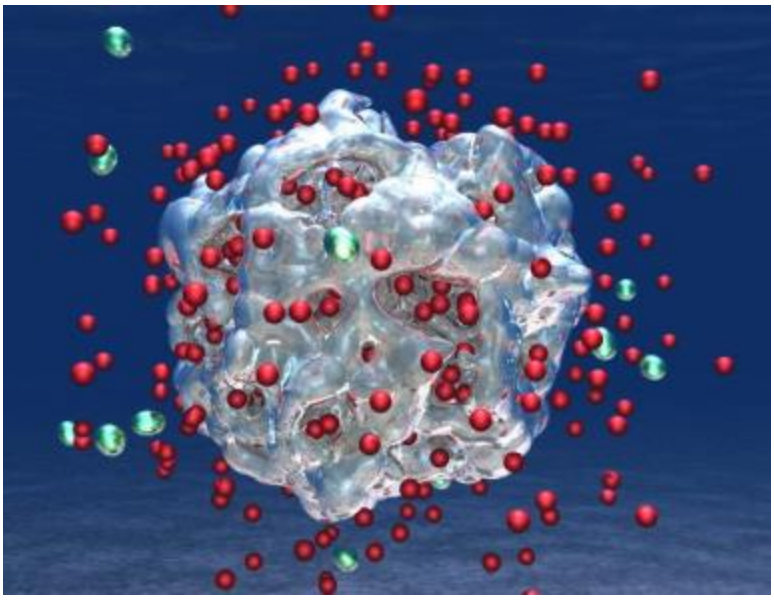


Fig. 2: A 16,611-atom quantum molecular dynamics simulation of H_2 production from water using a LiAl-alloy particle. The valance electron density (silver isosurface) is centered around Al atoms, whereas some of the Li atoms (red spheres) are dissolved into water. Produced H_2 molecules are represented by green ellipsoids. Water molecules are not shown for clarity [K. Shimamura *et al*., *Nano Letters* 14, 4090 (2014)].

BES White Paper – Thanos Panagiotopoulos (Princeton University)

1. Please specify the current science drivers for your field of research.

My field of research is the development and application of statistical-mechanics-based simulation algorithms for thermophysical properties of electrolyte solutions, surfactants, and polymers – suitable for discussion within the "Materials and Chemical Discovery" and "Soft Matter, Biochemistry, Bioinspired Materials" breakout sessions. The main science drivers in this field are the expansion of the length and time scales that can be studied with classical molecular dynamics and Monte Carlo "force-field" methods because of the availability of better hardware and also improved sampling algorithms. In addition, an important role is played by the development of rigorous methods for transfer of information across scales (e.g., using electronic structure calculations for informing the selection of force field parameters, or transferring data from atomistic-level simulations to coarse-grained ones).

2. Describe the science challenges expected to be solved in the 2020–2025 timeframe using computing ecosystems.

In the next 5–10 years, we are likely to see the development of a new generation of atomistic force fields that can accurately predict thermodynamic and transport properties of water and aqueous solutions to unprecedented accuracy. These force fields are likely to contain explicit polarizability and thus will require development of appropriate computational ecosystems to allow reaching the time and length scales we are currently able to sample using non-polarizable force fields. Being able to do this will enable accurate calculations for self-assembly in aqueous solutions, important for pharmaceutical, consumer product and biophysical applications. Accurate models will also be important for modeling novel CO_2 separation and carbon sequestration processes.

3. Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

A key issue still likely to be unresolved in the next 5–10 years are multiscale/multiphysics models that can successfully pass both equilibrium and dynamic information across scales. Such methods are important in the modeling of chemically reactive systems as well as systems with experience mechanical failure or aging (e.g., glassy polymers).

4. What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1. Models and algorithms	Improved models and algorithms are likely to be developed in the next 5–10 years.
2. Hardware resources	The planned expansion in hardware resources for BES computational science over the next 5–10 years appear supportive of accelerated progress in my research area.

Impede	Why?
1. Application codes	Many of the common application code packages for MD (e.g., LAMMPS, Gromacs) do not support polarizable force fields or suffer significantly in performance when they do. Some newer packages optimized for GPUs (e.g., HOOMD) have limited capabilities. Monte Carlo codes with multiparticle moves necessary for handling polarizable models are not generally available.

5. (Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data. For example:

The field of force-field based modeling for solutions and self-assembly is not driven primarily by data transmission, analysis, or processing. However, it could benefit from improved availability of simulation data of controlled quality for specific model systems (e.g., similar to the database for properties of the LJ system that NIST has made available).

6. References

Opportunities for Discovery: Theory and Computation in Basic Energy Sciences, http://science.energy.gov/~/media/bes/pdf/reports/files/od_rpt.pdf.

Computational Materials Science and Chemistry: Accelerating Discovery and Innovation through Simulation-Based Engineering and Science, http://science.energy.gov/~/media/bes/pdf/reports/files/cmsc_rpt.pdf.

From Quanta to the Continuum: Opportunities for Mesoscale Science, http://science.energy.gov/~/media/bes/pdf/reports/files/OFMS_rpt.pdf.

A. Jusufi and A. Z. Panagiotopoulos, "Explicit- and Implicit-Solvent Simulations of Micellization in Surfactant Solutions," Invited Feature Article, *Langmuir*, **31**: 3283–92 (2015). DOI: 10.1021/la502227v.

7. Images

I have submitted separately an image that appeared on the cover of *Langmuir*, vol. 31, issue 11 (2015), showing the self-assembly process in a system of ionic surfactants, in connection to the invited feature article by A. Jusufi and A. Z. Panagiotopoulos given above (under "references").

C.5 White Papers Addressing Advances in Algorithms for Quantum Systems

BES White Paper — Emily Carter (Princeton University)

1. **Please specify the current science drivers for your field of research.**

Where we are today? Be sure to include broad impact, DOE interest, ties between experiment/theory, etc.

Here and in the discussion below, we will focus on one of our research projects, the one must suitable to exploit exascale-class computing resources. Other methods are limited in their scalability and therefore not capable of massively parallel installations. However, we want to generally note that these methods are important (e.g., for accuracy reasons) and will continue to play an important role in the future of our field. Where insights from such projects are discussed, we will mark them.

The orbital-free density functional theory (OFDFT) method [1] is a first-principles method that is based on quantum mechanics that scales quasi-linearly with system size. Therefore, the OFDFT method has the power to understand and predict large-scale materials' fundamental properties. The OFDFT methods developed in the Carter group have been used to study mechanical properties of pure metals and metal alloys, which are determined by mesoscale defects such as dislocations, in Al [2], Mg [3], and Mg-Li [4]. We also used OFDFT to study liquid properties of Li [5] to gain a better understanding of its potential use in plasma-facing-components in fusion reactors. All our simulations not only well match the known experimental data but also provide fundamental insights of materials' properties on the atomic level. Recent exciting developments of OFDFT allow it to exploit petascale installations for simulations of over one million atoms by using more than 100,000 cores.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What will probably be solved in the next 5–10 years? Why is this important to the field?

We envision developing and applying OFDFT to study important material properties in the years ahead. Examples include the mechanisms of natural aging in alloys such as Al-Mg-Si alloys, stress corrosion cracking phenomena in Al alloys, and formation of Li dendrites in Li-air batteries. Routinely solving these scientific questions will require huge computational efforts and further algorithmic developments. The impact of understanding these effects on the atomic level goes far beyond our field and has direct influence on product development and engineering.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What might not necessarily be solved in the next 5–10 years? Again, what is the importance?

OFDFT is (currently) not suitable to study some properties of materials, for example, excited states of materials. OFDFT is also not suitable for studying reactions such as those occurring in catalytic or combustion processes. We consider method developments enabling such studies to be unlikely in the above-mentioned timeframe. Hence, other simulation tools will be required, such as (embedded) correlated wavefunction methods and/or periodic Kohn-Sham density functional theory-based

calculations. However, as discussed above, these codes have a limited scalability and will likely not be able to profit from extant computing ecosystems. Their importance for solving the scientific challenges ahead mandates that they remain feasible, supported tools for researchers in the field.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Large number of cores	OFDFT can perform petascale-class computations on more than 100,000 cores. Anticipating algorithmic progress, running OFDFT on a large number of cores will accelerate simulations.
2. Available co-processor computational resources	(Part of) our OFDFT code should be suitable to exploit co-processors such as GPUs.
3. External libraries and toolchain	Availability of well-optimized external libraries for core functions (BLAS/LAPACK/FFT) will enable us to achieve best performance for all of our codes. A developer-friendly toolchain (compilers, debuggers, profilers) will allow us to develop our codes to optimally exploit exascale-class installations.

Impede	Why?
1. Storage system (data workflow)	OFDFT-based molecular dynamics simulations need to save as many configurations (millions of atoms) as possible for analysis. The data from a single run can be in excess of several TB.
2. Memory (hardware resources)	A lot of data should be saved on-the-fly in OFDFT simulations.
3. Restrictions on running time	Typical OFDFT-based molecular dynamics simulations require several days up to months to generate a long enough trajectory.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.** For example:
   - Classify the data as simulation, experimental, both, or something else.
   - Characterize these data using 3 V's (Velocity, Volume, and Variety). How much data will be generated/stored within the next 5–10 years? What will the data rates be? Will the data be multi-modal somehow?

- Describe current or planned solutions for any of these data challenges, e.g., in situ analysis.
- Note any particular data security or privacy requirements.

As mentioned above, large volumes of simulation data will need to be stored for post-processing from molecular dynamics simulations. Other than that, the at-runtime and post-processing of data are limited in velocity and volume. A large variety in data exists but is not suitable for automation.

6. References

[1] Y. A. Wang and E. A. Carter, in *Theoretical Methods in Condensed Phase Chemistry*, edited by S. D. Schwartz (Kluwer, 2000), pp. 117.

[2] I. Shin and E. A. Carter, Physical Review B **88**, 064106 (2013).

[3] I. Shin and E. A. Carter, Int J Plasticity **60**, 58 (2014).

[4] I. Shin and E. A. Carter, Acta Mater **64**, 198 (2014).

[5] M. Chen, L. Hung, C. Huang, J. C. Xia, and E. A. Carter, Molecular Physics **111**, 3448 (2013).

BES White Paper – Mark Jarrell (Louisiana State University)

1. **Please specify the current science drivers for your field of research.**

Where we are today? Be sure to include broad impact, DOE interest, ties between experiment/theory, etc.

Some of the most interesting and important materials have strong electronic correlations, frustration, or disorder [1]. Unfortunately, their simulations often scale exponentially. As a result, even the exascale computer power expected by the end of the decade is not likely to enable new discovery using conventional approaches. To address these problems, several groups have developed multiscale methods including our own multiscale many-body approach [2], which treats the strong correlations at short length scales exactly, intermediate length scales with diagrammatic methods, and the longest length scales in a (dynamical) mean field (Fig. 1). These methods are roughly equivalent, but Rubtsov's dual Fermion (DF) approach [3,4,5] is the most elegant and extensible formulation. As illustrated in Fig. 2, the DF method is remarkably accurate when compared to the present state of the art (DCA). In addition, it scales algebraically, so that the effects of multiple bands, disorder, and interactions needed to describe real materials may be included.

2. **Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What will probably be solved in the next 5–10 years? Why is this important to the field?

We envision using the DF approach to solve complex models obtained from first principles (DFT downfolding and unfolding). The short length scale solver will remain some form of quantum Monte Carlo, and intermediate length scale solver will be a multiband fluctuation-exchange approximation (FLEX). The latter dominates the calculation, with memory scaling like Nb^4 * Nf^3 * Nc (Nb number of bands, Nf frequencies, Nc cluster size) and the computation like Nc * Nb^6 * Nf^4. So, for an eight band model, the requisite roughly 210 TB of memory would require that the calculations be spread over roughly 51,000 cores. This can be reduced by calculating the vertices on the fly at the expense of increased computational cost. This method would enable first-principles simulations of important materials [1], including the effect of non-local correlations (beyond DMFT), and produce results directly relevant to experiment.

3. **Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What might not necessarily be solved in the next 5–10 years? Again, what is the importance?

Two problems with this method appear to be difficult or intractable in the next 5–10 years. First, there is a double-counting problem since the DFT used to construct the models already contains some interaction effects. To ensure that the many-body methods do not over count these interactions, they are subtracted from the model. However, there are multiple schemes for doing this which generate different results. One possible way to address this is to use a Hartree Fock calculation to replace the DFT. Here, the diagrams describing the correlations are known so they may be explicitly subtracted off. Second, for realistic materials, to improve the accuracy achieved using the multiscale many-body approach mentioned above, one could use the parquet approximation (PA) to replace the FLEX. The PA maintains the self-consistency at both single- and two-particle levels. Our group has been working on the PA for years and achieved significant success. However, using PA for realistic materials will impose a

grand challenge on both the memory and the computational cost. While the former scales as Nc^3 * Nb^4 * Nf^3, the latter scales as Nc^4 * Nb^6 * Nf^4. Neither of these are achievable now, or in the near future, for the 8-band model described above.

4. **What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. More cores	A larger number of cores will allow us to study more complex models (more bands, different types of disorder, stronger correlations, etc.) at the expense of making the latency problems worse.
2. Better interconnects	Our problems are network bound due to the need to rotate and contract very large rank-3 tensors (the vertices associated with two-particle scattering).
3. Run-time support for better parallel efficiency	Current systems (MPI, OpenMP) make it very difficult to overlap communication with computation in a performance-portable way. A better run time system will allow us to move our parallel codes from machine to machine without major revisions.

Impede	Why?
1. I/O	Our calculations generate so much distributed data that presently it is difficult to checkpoint or drain this data from the system.
2. Memory	This is the main bottleneck for these methods. The needed memory can be reduced by calculating the vertices on the fly at the expense of increased computational cost, at least for the FLEX. This is less likely for the PA.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field of research involve the transmission, analysis (including real-time analysis), or processing of data.** For example:
   - Classify the data as simulation, experimental, both, or something else.
   - Characterize these data using 3 V's (Velocity, Volume, and Variety). How much data will be generated/stored within the next 5–10 years? What will the data rates be? Will the data be multi-modal somehow?
   - Describe current or planned solutions for any of these data challenges, e.g., in situ analysis.
   - Note any particular data security or privacy requirements.

6. References:

[1] Computational Materials Science and Chemistry. Accelerating Discovery and Innovation through Simulation-Based Engineering and Science, http://science.energy.gov/~/media/bes/pdf/reports/files/cmsc_rpt.pdf.
[2] C. Slezak, M. Jarrell, Th. Maier, J. Deisz Multi-scale Extensions to Quantum Cluster Methods for Strongly Correlated Electron Systems, JOURNAL OF PHYSICS-CONDENSED MATTER, 21, 435604. arXiv:cond-mat/0603421.
[3] A. Rubtsov, M. Katsnelson, and A. Lichtenstein, "Dual fermion approach to nonlocal correlations in the Hubbard model," Phys. Rev. B, vol. 77, no. 3, pp. 033101/1–9, 2008.
[4] S.-X. Yang, H. Terletska, Z. Y. Meng, J. Moreno, M. Jarrell Mean-field embedding of the dual fermion approach for correlated electron systems , Phys. Rev. E 88, 063306 (2013), arXiv:1305.0207.
[5] S.-X. Yang, P. Haase, H. Terletska, Z. Y. Meng, T. Pruschke, J. Moreno, M. Jarrell Dual-fermion approach to interacting disordered fermion systems , Phys. Rev. B 89, 195116, arXiv:1310.6762.

7. Images:

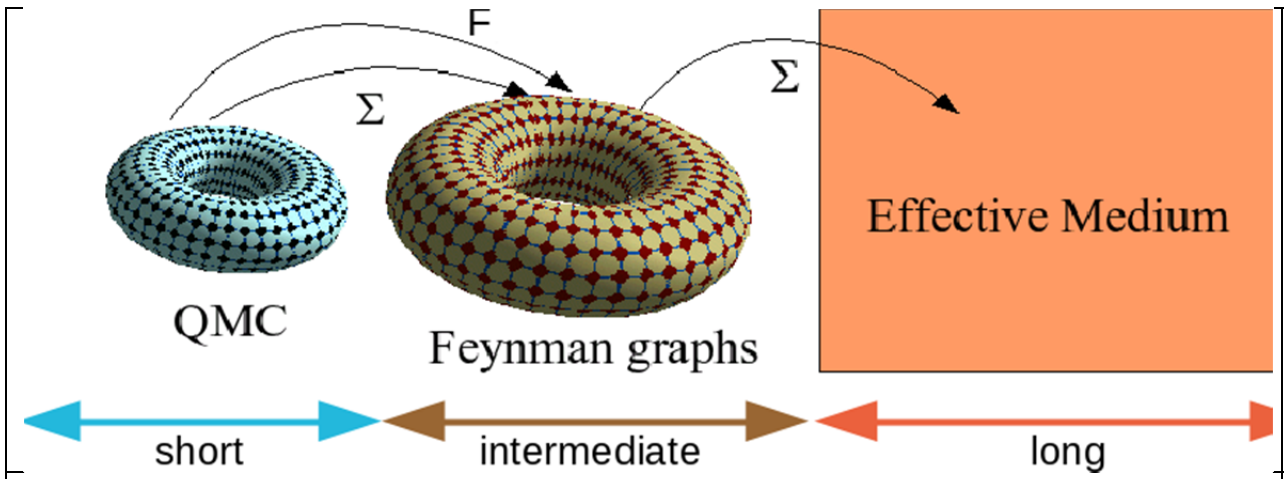


Fig. 1. Elements of the multi-scale many body approaches. Quantum Monte Carlo (QMC) is used to treat the short length scales, Feynman graphs the intermediate length scales, and a mean field the longest length scales.

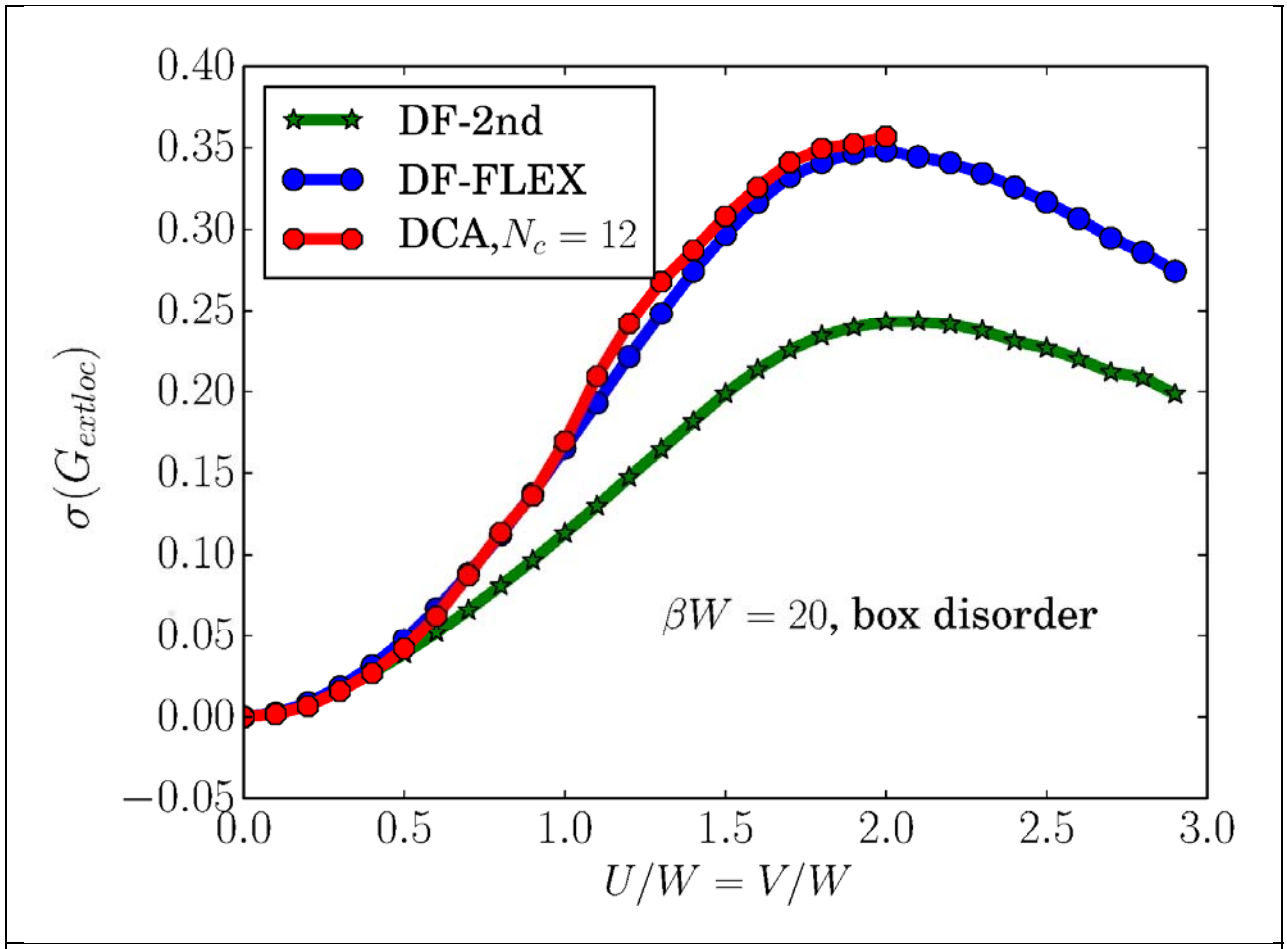


Fig. 2. Conductivity versus the interaction and disorder strength (equal) for the 1D Anderson-Hubbard model. The DCA results are essentially fully converged at this temperature. As the intermediate length scale solver improves from second order to FLEX, nearly exact agreement with the DCA results is obtained.

BES White Paper – Exascale Computing Requirements for Modeling Atmospheric Aerosols: Karol Kowalski (Pacific Northwest National Laboratory)

Please specify the current science drivers for your field of research.

There is a significant interest in several vital areas of research and development where computational chemistry can be used to address scientific questions, including catalysis, solar energy conversion, energy storage, materials, biology, and atmospheric sciences. Although the scientific questions can be different for each area, there are many common themes that drive the need for new computational capabilities to describe collective phenomena in chemical processes. As experiments become more complex, de-convoluting the intricacies of the processes involved is a daunting task. In addition, modeling these systems accurately poses new challenges. Several limitations associated with modeling chemical and materials systems in complex inhomogeneous environments are imposed by the high computational cost associated with calculations on realistic systems and with the complexity of the theoretical approaches involved, which require ever-growing computational resources to exceed system-size/accuracy-level limitations. Moreover, the contemporary challenges in the computational chemistry usually do not fit into one methodological category. By their nature, these problems are extraordinarily complex and require simulation models that span vast differences in length and time scales.

Understanding how clouds, aerosols, and dynamics couple as a function of scale and how clouds and precipitations couple with surface properties are the most pressing challenges in atmospheric chemistry. Aerosols play an important role in boundary layer and tropospheric processes by influencing cloud microphysics and radiation. The main questions address understanding the chemistry, physics, and molecular-scale dynamics of aerosols for model parameterization to improve the accuracy of climate model simulations and develop a predictive understanding of the climate:

- What are the formation mechanism and growth rates for aerosol particles?
- How do aerosol optical properties relate to particle size, morphology, and composition?
- What is the role of excited-state reactions in forming secondary organic aerosol?

In order to address these challenges, the development of a molecular-scale understanding of the processes that enhance controllable biogenic organic aerosol formation and determine their radiative properties to improve the accuracy of climate model simulations are necessary. In particular, the utilization of high-accuracy methodologies capable of capturing electron correlation effects is necessary to achieve a desired level of accuracy in calculations for optical properties and excited-state processes.

Describe the science challenges expected to be solved in the 2020–2025 timeframe using extant computing ecosystems.

NWChem offers several capabilities that contribute to the synergy between theory and experiment by enabling high-accuracy modeling tools for (1) calculating potential energy surfaces for various types of excited states (valence and core-level), (2) modeling photoelectron spectroscopy (PES), and (3) calculating nonlinear static and frequency-dependent nonlinear optical properties. Given the progress in the development of efficient parallel implementations capable of taking advantage of various computational platforms, today's capabilities in high-accuracy computational chemistry provide the means to describe excitation energies, potential energy surfaces, and static/frequency-dependent nonlinear optical properties for small aerosol particles composed of 100–200 atoms. The electronic

structure methods can also be seamlessly integrated into multi-physics/multi-scale models that are necessary to describe photochemical reactions and the effect of hydroxyl/hydrocarbon radicals on the aerosol formation.

Describe the science challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

Advancing our understanding of the physics and chemistry of the aerosol growth process and morphology will require large-scale atomistic simulations in realistic environments that span a broad range of length and time scales coupled with state-of-the-art experimentation and spectroscopies. We need to incorporate the theoretical and computational high-accuracy methodologies to simulate ultrafast transformations, adiabatic and non-adiabatic excited-state dynamics, and collective response of aerosol particles to external stimuli. We currently do not have the computational tools that will allow us to describe aerosol formation processes at larger and more realistic scales. Sizes of molecular systems required to capture complex interplay between chemical/physical processes that determine the properties of aerosol particles and their growth are far beyond what we can describe computationally today.

What *top three* computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	Why?
1.Algorithms for reducing numerical cost of high-accuracy methods	New algorithms capable of utilizing the sparsity of first-principle formulations are needed to push the envelope for system-size tractable by high-order methods. While these methods are available for ground-state formulations, there is no clear path-forward for excited-state and linear-response formulations.
2.Internal /external libraries	Novel libraries are needed to incorporate new programming models to deal with the increase in intra-node parallelism; data localization and reduction in intra-node communication; efficient utilization of runtime systems, development of topology-aware algorithms, and utilization of deeper memory hierarchies.
3.Models and algorithms	Mathematically rigorous integration of various representations of quantum mechanics to build efficient embedding schemes.
4. Hardware resources	Larger computational resources (at all scales) are needed to develop reliable multi-scale frameworks based on the accurate many-body solvers.

Impede	Why?
1.Workforce development	Strong computer science background will be required to tackle new challenges.
2.Insufficient funding	Lack of sustainable support for multi-year effort.
3.Portability	Lack of unique programming models for hybrid parallel systems.

References

1.) From Quanta to the Continuum: Opportunities for Mesoscale Science, BESAC Report, 2012.
2.) E. Apra, M. Klemm, K. Kowalski, “Efficient implementation of many-body quantum chemical methods on the Intel Xeon Phi co-processor,” Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 674–684 (2014).

3.) K. Kowalski, S. Krishnamoorthy, R.M. Olson, V. Tipparaju, E. Apra, "Scalable implementations of accurate excited-state coupled cluster theories: Application of high-level methods to porphyrin-based systems," *High Performance Computing, Networking, Storage and Analysis (SC)*, 2011 International Conference, pp. 1–10 (2011). *[Also the source for Figure 1.]*

Image:

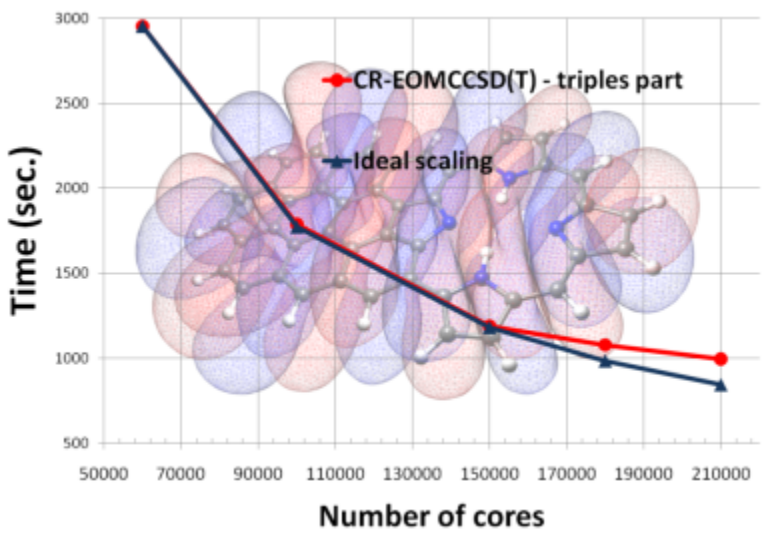


Figure 3 Scalability of the triples part of the excited-state CR-EOMCCSD(T) approach for the functionalized form of the free base porphyrin (FBP) system. Timings were determined from calculations on the Jaguar XT4 computer system.

C.6 White Papers Addressing Computing and Data Challenges @ BES Facilities

BES White Paper – Maria Chan, Pierre Darancet, Stephen Gray, Ian McNulty, Subramanian Sankaranarayanan, and Michael Sternberg (Argonne National Laboratory)

1. Please specify the current science and operational drivers for your User Facility (e.g., detector performance, real-time data analysis).

The Center for Nanoscale Materials (CNM) at Argonne is a user facility providing expertise, instrumentation, and infrastructure for interdisciplinary nanoscience and nanotechnology research. Academic, industrial, and international researchers access the center through its user program for both nonproprietary and proprietary research. The center's goal is to perform basic research and instrumentation development that explores ways to tailor nanoscale interactions by creating, visualizing, and assembling hybrid nanomaterial architectures for energy-related research and development programs. High-impact staff and user science is accommodated within the primary cross-cutting theme of "Energy and Information Transduction at the Nanoscale" that addresses grand challenges in energy and information conversion and transport (transduction), while furthering the DOE missions of energy generation, storage, and efficiency.

2. Describe the broad computational and data challenges expected to be faced in the 2020–2025 timeframe.

A broad computational challenge that might be solved in the next ten years is to bridge a wide range of length and time scales so that phenomena captured in atomistic simulations can be modeled at the nano- to the meso-scales [1,2]. Such multiscale modeling implies the use of modern algorithms running on agile high-performance computer architectures that are able to accommodate and efficiently handle the full gamut of electronic structure, atomistic, coarse-grained, and continuum modeling tasks.

Valuable mature methods have been developed and optimized in past decades by scientists of different disciplines, perspectives, and on different software platforms. Progress in multiscale modeling depends on this work and is thus contingent on computer systems that support a wide variety of algorithms by having generous characteristics for CPU, memory, and interconnect bandwidth, and, not least, have a broad set of state-of-the-art software libraries and applications. We believe that agile computing ecosystems will foster the development of integrated multiscale, multidisciplinary methods that are able to model systems at time- and length-scales beyond the reach of atomistic methods with a controlled loss in accuracy.

One data challenge that might be solved in the next ten years is the establishment of materials databases that collect, store, meaningfully index, and disseminate large volumes of unprocessed as well as distilled first-principles data. This will be important for the eventual success of the materials genome initiative, i.e., to create a new era of policy, resources, and infrastructure that supports U.S. institutions in the effort to discover, manufacture, and deploy advanced materials faster and at a fraction of the cost of current approaches.

Another broad challenge on both the computational and data front is the direct and real-time integration of materials modeling with *in situ* and *in operando* experimental measurements, through the use of high-throughput atomistic simulations and machine learning. In order to accurately predict atomistic structures and corresponding experimental measurements, either first principles approaches involving quantum mechanical calculations or large-scale polarizable or reactive molecular dynamics

calculations are necessary. A large number of parameters, each involving a distinct configuration and corresponding experimental measurements, would have to be sampled in parallel processes. On both the modeling and experimental measurement fronts, finding solutions to problems of data (pre)processing, reduction, and pattern matching is paramount.

The National User Facilities such as the Nanoscale Science Research Centers and Light Sources also have a significant "big data" problem: a growing fraction of the prodigious amount of data they produce is not getting analyzed and puts a strain on data storage systems. Site-to-site data transfers are becoming less feasible as transfer bandwidth cannot keep up with advances in local storage technologies. The primary sources of this flood of data are multispectral, multi-dimensional, high-resolution, and high frame-rate detectors (up to 10 Kframes/s), substantially brighter (100–200 ×) X-ray sources, and increasingly larger-scale (up to petascale) computational capabilities being developed to analyze and simulate algorithmically and numerically more complex problems. From a facility utilization standpoint it is highly inefficient to take days to months to analyze a data set acquired in minutes. Guiding the course of a research program or making critical decisions on the time scale of the data acquisition becomes impossible under these conditions. Solutions to the Analysis Bottleneck—new analysis approaches, paradigms and infrastructure for keeping up with the data being collected—will have the highest impact on our community, and provide the best possible utilization of the considerable investment in our National User Facilities such as the CNM.

3. Describe the broad computational and data challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.

It is uncertain if the "big data" problem noted in the last part of sec. 2. above would be solved with extant computing ecosystems.

The challenges listed above likely cannot be solved using existing computational resources, infrastructure, and algorithms. On the integration of atomistic modeling with experimental measurements, each calculation requires at least ~10^3 cores in order to model materials systems of typical complexity, and in order to adequately sample the different possibilities corresponding to a realistic material sample or process, 10^3–10^6 such calculations are often needed. Data processing and matching, as well as communications, between computational and large-scale experimental facilities will have to be integrated. Existing computational resources, infrastructure, and codes are unlikely to meet such requirements.

4. What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

Accelerate	**Why?**
1. Visualization and analysis resources	These are especially important for emerging "big data" problems.
2. Models and algorithms	New models and algorithms almost always accelerate problem solving.
3. Hardware	Faster computers with more memory almost always accelerate problem solving.

Impede	Why?
1. Workforce development	Lack of HPC expertise is a big bottleneck to implementing many theory and modeling projects.
2. Data workflow	Transferring data between applications and between institutes can impede progress.
3. Application codes	Many codes, particularly for electronic structure are difficult to parallelize efficiently for HPC use.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field involve the transmission, analysis (including real-time analysis), or processing of data.** For example:
    - Classify the data as simulation, experimental, both, or something else.
    - Characterize these data using 3 V's (Velocity, Volume, and Variety). How much data will be generated/stored within the next 5–10 years? What will the data rates be? Will the data be multi-modal somehow?
    - Describe current or planned solutions for any of these data challenges, e.g., in situ analysis.
    - Note any particular data security or privacy requirements.

6. **References**

[1] "Opportunities for Discovery: Theory and Modeling in Basic Energy Sciences," BES Advisory Committee, US Department of Energy (2005).
[2] "Theory and Modeling in Nanoscience," BES Advisory Committee, U.S. Department of Energy (2002).

BES White Paper —User Facilities: Alexander Hexemer and Dula Parkinson, Advanced Light Source (Lawrence Berkeley National Laboratory)

1. Current science and operational drivers.
More and more users are proposing experiments that are time-resolved, combinatorial, or high throughput. To meet user needs, synchrotrons are pushing to increase their brightness; to develop fast, high resolution, and high efficiency detectors; and to develop more automation in sample handling and beamline operation. But to maximize both the quantity and quality of light source science output, all users must also have access to the best in class with respect to beamline controls, applied mathematics, algorithms, software, networks, data management, workflow, and computing resources. These computational aspects are an integral part of a feedback loop with the experiments: data must be packaged, transferred, and in many cases reduced, analyzed, and compared with simulations or other results in real time in order to provide feedback to the user or directly to the instrument control system during a time-resolved experiment. Data are often then subsequently transferred for longer-term storage and sharing, or prepared for further, more computationally intensive offline analysis.

To cite one example of computing integrated into light sources, at the Advanced Light Source, soft X-ray ptychography is a technique which allows for very high-resolution (nano-scale) imaging with chemical sensitivity. For this experiment, an inverse problem must be solved before the user can see even a 2D image, so it is essential that this be done in real time. In the next few years, multiple additional ptychography beamlines will be commissioned; based on extrapolation from current computational needs, these ptychography beamlines, when combined with requirements from just two other currently existing high-data-rate beamlines, will need several petaflops' worth of compute power to process all the acquired data in real time.

2. Broad computational and data challenges, 2020–2025.
Current prototypes must greatly expand and become more robust. SPOT Suite, one such prototype system, has been a collaborative effort between the Berkeley Lab Computational Research Division, NERSC, ESnet, and the Advanced Light Source. Underlying its functionality is a workflow system to drive both the automated, near-real-time processing and user-triggered actions; a metadata and provenance management system and database; and a suite of easy-to-use, web-based visualization tools. The Center for Advanced Mathematics for Energy Research Applications (CAMERA) was launched recently to develop cutting edge algorithms and analysis tools in collaboration with facilities like the ALS. The algorithms produced by CAMERA can be used on their own, but are also integrated into SPOT Suite.

As the desired speed of time-resolved experiments (and the accompanying data rate) increases, it is more and more important to have some kind of online, streaming analysis that allows steering of the data collection. This is already being implemented at some beamlines. The challenge is that as the complexity of experiments increases, the accompanying computing is more complex and requires more computational resources than can be available at the synchrotron. So the concept of streaming analysis will have to be extended for use at computing facilities that have not run in real-time mode, much less in a streaming mode. For some beamlines, real-time feedback is not a priority, but offline analysis is computationally VERY expensive—reaching easily in the petaflop regime—so facilitating easy access for users to computing power is essential.

Along with the increasing complexity of experiments, there is an increasing need for users who collect data to share this data with collaborators in analysis and theory. There is also an increasing mandate to make federally funded data publicly available, where the community can take advantage of the data for

re-analysis, to extract additional information, or to compare with other data of the same or of other techniques. Sharing data is a challenge both because of data size but also because of the lack of metadata that would be necessary to make the data useful beyond the scientist who collected it. Ongoing research in such areas as machine learning is starting to point to ways of classifying data sets and finding features in data sets to make them more searchable and useful.

We hope for the spread of federated identity which will simplify access to different resources. While there are a wide variety of file formats in existence, there seems to be consensus in the synchrotron community that hdf5 is an appropriate file format, and various sub-communities are slowly but surely converging towards more standardized file formats.

3. Broad computational and data challenges that cannot be solved, 2020–2025, with extant systems.
As mentioned above, some of the biggest challenges have to do with sharing data and doing analysis across many different samples or across multiple modes of data. We believe that although there is work being done in these areas, there will still be much more work to do 10 years from now. A major reason for this is that data are not perfect: there is missing metadata or missing portions of data; there are systematic errors that are different across different data types; there can be very low signal to noise; and there are many other problems in the data. Finding ways to take large amounts of bad data from multiple techniques and finding ways to extract relevant information will be increasingly important.

Despite work on classification and feature extraction, we also expect that there will still be significant work to do 10 years from now in automatically adding missing scientifically meaningful metadata to data, which is needed to help researchers search for and find the data most relevant to them.

4. Top three computing ecosystem aspects to accelerate or impede progress, 5–10 years?

Accelerate	Why?
1. Development of distributed, scalable software infrastructure incorporating workflow, data management, and resource management.	To enable the BES facility goal of delivering science knowledge (rather than data) requires the development of the Super-Facility, using ASCR resources.
2. Access to guaranteed Network QoS (perhaps with SDN) and the ability to co-schedule network and compute resources with beamline experiments.	In-situ, time-resolved experiments needing real-time feedback have burst needs for network, compute, storage, and other resources.
3. Inter-facility federated ID and single sign-on capabilities.	As BES facilities build partnerships with multiple ASCR facilities, disjoint trust domains and procedures are barriers.
4. Low-overhead and high-bandwidth data I/O capable systems (e.g., burst-buffer or RDMA-like I/O).	Many processing and analysis systems and applications for experimental and observational science are very data I/O intensive.

5. Data ecosystem aspects.
The ALS has 40 beamlines with a wide variety of x-ray science techniques, scientific objectives, file formats, software, and data themes. Many ALS beamlines see experimenters bringing in experiment-specific instruments to augment the default instruments of the beamline - bringing variety and multi-modality. The ALS currently produces petabytes' worth of data per year—much of it not centrally managed. This data rate is increasing faster than Moore's law curves. Instantaneous data rates for certain beamlines will approach >10Gbps within the next 1–3 years. ALS beamlines produce raw experimental data, metadata, and provenance. Reduced data and subsequent "downstream" data are

often compared with the results of simulations and/or used as input or as a parameter or constraint for the input to a simulation. In general, there is no single system currently in use that directly captures experimental data from the light source and simulation data for curation. Conceptually, ALS data and metadata are, by default, viewable only by the PI of the experiment which generates it with capability to share with select collaborators. Subsequent release of the data to the wider scientific community and public is envisioned.

6. References:

1. *A.Hexemer, D.Kumar, S.Venkatakrishnan, A.Sarje, S.Patton, S.Li, J.Deslippe, C.Tull, E.Dart, F.Liu, T.Russell, E.Gomez, C.Zhu, E.Schaible, P.Stewart, "Fast Analysis of Time-Resolved Scattering Data", Bulletin of the American Physical Society, APS March Meeting 2015, Volume 60, Number 1, March 2, 2015, Accepted.*
2. *J.Blair, R.S.Canon, J.Deslippe, A.Essiari, A.Hexemer, A.A.MacDowell, D.Y.Parkinson, S.J.Patton, L.Ramakrishnan, N.Tamura, B.L.Tierney, C.E.Tull, "High performance data management and analysis for tomography", Proc. SPIE 9212, Developments in X-Ray Tomography IX, 92121G (September 12, 2014); doi:10.1117/12.2069862.*
3. *K.Troutmana, E.Arenholza, D.Parkinsona, A.Hexemera, D.Shapiroa, A.MacDowella, C.Tull, W.Yanga, C.Ralstonc, S.Guptac, J.Bohond, "ALS User Meeting Highlights Accomplishments and Challenges", Synchrotron Radiation News, Volume 28, Issue 1, 2015, DOI: 10.1080/08940886.2015.991645.*
4. *J.Deslippe, A.Essiari, S.J.Patton, T.Samak, C.E.Tull, A.Hexemer, D.Kumar, D.Parkinson, "Workflow management for real-time analysis of lightsource experiments", Proceedings of the 9th Workshop on Workflows in Support of Large-Scale Science, Pages 31–40, IEEE Press Piscataway, NJ, USA ©2014, ISBN: 978-1-4799-7067-4 doi>10.1109/WORKS.2014.9.*
5. *S.Patton, T.Samak, C.E.Tull, C.Mackenzie, "Spade: Decentralized Orchestration of Data Movement and Warehousing for Physics Experiments", Proceedings IFIP/IEEE Integrated Network Management Symposium (IM 2015), May 11, 2015.*

BES White Paper – LCLS Requirements for Exascale:
Amedeo Perazzo (Stanford Linear Accelerator Center)

1. **Please specify the current science and operational drivers for your User Facility (e.g., detector performance, real-time data analysis).**

The first X-ray FEL to generate hard X-rays, the Linac Coherent Light Source (LCLS), began operation in 2009 and has dramatically exceeded performance expectations. This facility, which was created using an existing electron accelerator which limits its pulse rate to 120 Hz, generates X-rays by amplifying spontaneous noise in the electron beam. LCLS has already had a significant impact on many areas of science, including: resolving the structures of macromolecular protein complexes that were previously inaccessible; capturing bond formation in the elusive transition-state of a chemical reaction; revealing the behavior of atoms and molecules in the presence of strong fields; and probing extreme states of matter.

The LCLS data acquisition systems read out a variety of 1D and 2D optical and X-ray cameras and various digitizers. Each of the seven LCLS instruments has a dedicated DAQ system and adopts a common framework and architecture. Most instruments can acquire 5GB/s (CXI was designed to read up to 10GB/s since it can operate two experiments concurrently). All instruments share the same data management system. Experimenters can analyze the data at three different stages with increasing complexity: online monitoring in real-time (< 1s), feature extraction in quasi real-time (< 10s), and full analysis offline.

2. **Describe the broad computational and data challenges expected to be faced in the 2020–2025 timeframe.**

LCLS will go through a major upgrade in 2020 called LCLS-II, which will operate at a much higher repetition rate (up to 1MHz) and, above all, with a higher data throughput (~100GB/s).

3. **Describe the broad computational and data challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

Real time analysis, event building, fast feedback storage: The interval between pulses will decrease from 8ms to 1us. On-the-fly event building will become too fragile. Instead we envision writing the data to a fast, flash-based, buffer where multiple clients can access them. This buffer would need to handle read/writes at hundreds of GB/s.

Data storage and archiver: The SLAC tape archive system is approaching limits in overall storage capacity and throughput. Based on the current storage requirements and the estimated increase in the amount of acquired data, it is expected that LCLS-II will store tens of PB per year.

Data Management: LCLS has developed a powerful data management system that handles both the automatic workflows of the data through the various storage layers (e.g., long-term data archival) and the users' requests through a web portal (e.g., restoring data from tape). Some aspects of the current system, such as checksum calculations, HPSS interface, and lack of prioritization, will become limitations at higher data volumes and will need to be upgraded.

Data Processing: Based on the current computing requirements and the estimated increase in the amount of data to process, it is expected that LCLS-II will require between 200 teraflops and 1 petaflop. As with data-storage, deploying and maintaining a very large processing capacity at SLAC would require

a significant increase in the capabilities of the existing LCLS and/or SLAC IT groups. A more effective solution would be offloading part of the LCLS-II data processing to larger computing facilities like NERSC.

Data Transfer: SLAC recently upgraded its connection to ESNET from 10Gb/s to 100Gb/s to gain the ability to offload part of the LCLS science data processing to NERSC. This link will not be enough for LCLS-II, and terabit capabilities will be required if LCLS relies on NERSC for processing LCLS data.

Veto and Compression: A veto signal could be delivered to the front-end electronics, to the readout nodes, to the online cache or in the fast feedback layer. In general, a veto in the front-end electronics reduces the throughput requirements on the DAQ components, while a veto in the following layers provides cheaper/larger buffers and more time to reach a decision. Advanced generic lossy compression or physics based features extraction will also be considered to reduce the amount of data stored.

Data format: The LCLS DAQ is currently writing the raw data in XTC format. Users can request that their data be translated to HDF5. The translation step will become a bottleneck in the future and LCLS-II should adopt a single data format. HDF5 is becoming the de-facto standard for storing science data at light source facilities, but in order to effectively replace XTC in LCLS, a couple of critical features are required, namely the ability to read while writing and the ability to consolidate multiple writers into a consistent virtual data set.

4. **What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?**

Accelerate	Why?
1. Detector performance	Detectors that operate at a fraction of the machine repetition rate will require proportionally more time to provide the same statistics.
2. Algorithms	Being able to run sophisticated algorithms on-the-fly will improve the chances of success of an experiment and reduce the time to completion.
3. Data management	Automatically managing the data will help the scientists to focus on the experiment and maximize the amount and quality of the acquired data.

Impede	Why?
1. Real-time data analysis and visualization	LCLS scientists rely on fast feedback information to steer the experiment.
2. Data reduction	LCLS-II may be unable to acquire and store a significant fraction of the data without a veto and/or compression mechanism.
3. Data storage	Ability to write and read hundreds of GB/s will be a challenge in LCLS-II.

5. **(Optional) Characterize the data ecosystem aspects if the primary drivers for your field involve the transmission, analysis (including real-time analysis), or processing of data.**

6. References

[1] LCLS-II upgrade and its associated science opportunities:
https://portal.slac.stanford.edu/sites/lcls_public/Documents/LCLS-IIScienceOpportunities_final.pdf.

7. (Optional) Images

See attachments.

BES White Paper – User Facilities: Mark Stevens (Sandia National Laboratories)

1. **Please specify the current science and operational drivers for your User Facility (e.g., detector performance, real-time data analysis).**

Where we are today? Be sure to include broad impact, DOE interest, ties between experiment/theory

The Center for Integrated Nanotechnologies (CINT) plays a leadership role in the area of integration science through its function as a DOE/SC national user facility. By creating a collaborative community of diverse users matched to expert facility scientists with advanced capabilities, CINT fosters high-impact nanoscience discoveries, leads next-generation technique development, and advances the frontiers of knowledge beyond that which is achievable by individual researchers or any single institution. The unique properties of nanostructured materials cannot be fully exploited without a predictive understanding of the underlying phenomena. This requires a spectrum of theory/simulation techniques developed and optimized not only for the nanomaterial component itself but also for its interactions with surrounding components and materials. Hence, the CINT Theory & Simulation of Nanoscale Phenomena Thrust enables studies of nanomaterials assembly, interfacial interactions, and emergent properties of nanoscale systems, including their electronic, magnetic, and optical properties.

2. **Describe the broad computational and data challenges expected to be faced in the 2020–2025 timeframe.**

What will probably be solved in the next 5–10 years? Why is this important?

Typically, increased computational speed is used to treat dynamical phenomena outside the range of previous capability and/or to improve the accuracy of calculations. In soft matter, the range of time scales for the underlying molecular phenomena covers many orders of magnitude. Fig. 1 shows time scales for solvents, proteins, and polymers. In the next decade, multiple notable advances will occur as atomistic simulations in the μs to ms range become more common.[1]

Particularly for macromolecules, we are finally reaching time scales involving critical physical phenomena (e.g., reptation, ion transport) that both atomistic simulations *and* measurements can treat, enabling direct comparison between theory and experiment, which will substantially accelerate progress. Given that many materials properties of polymers depend strongly on the reptation dynamics, this will be a major advance.

Coarse-grained models are an important tool in soft matter simulations that enables a wider range of time and length scales simulated.[5] A new stage has begun, in which well-defined coarse-grained models are developed from atomistic models and retaining key chemical details while providing large increases in computational speed.[5,6] These advances are enabling simulations of phenomena such as surfactant self-assembly, polymer dynamics, and nanoparticle assembly that have not been previously possible.

Atomistic simulations of the interactions between nanoparticles with small molecule coatings of nanoparticles have recently started and in the next decade should provide an understanding of simple nanoparticle assemblies. A related area that is ripe for advancement is treatment of systems containing both hard and soft components. The major development will be in treating interactions at hard/soft interfaces, which are common in nanoparticle systems. Being able to calculate the influence of the surrounding solvent on the inherent optical or electronic properties of nanoparticles are fundamental to designing new materials with the unique characteristics that nanoparticles bring.

3. **Describe the broad computational and data challenges that cannot be solved in the 2020–2025 timeframe using extant computing ecosystems.**

What might not necessarily be solved in the next 5–10 years? Again, what is the importance?

As shown in Fig. 1, there are many phenomena at time scales beyond what can be reached in the next 5–10 years. Many of these phenomena are critical determiners of polymeric material behavior

particularly involving processing of polymers. With respect to industrial application, they are essential issues to be treated. Soft matter systems are often complex assemblies of many parts. Most simulations in the near future will only be able to treat individual parts or one level of a hierarchy. We presently do relatively simple systems with identical components, e.g., a single polymer type or a single nanoparticle type. Combining multiple elements adds important new material properties, but also adds longer time scales involving the mixing and interactions among the components that remain beyond future resources.

4. What top three computing ecosystem aspects will accelerate or impede your progress in the next 5–10 years? Why?

One clear need is improved use of GPUs and other coprocessors. In particular, the efficient parallelization of codes using these coprocessors can substantially increase the computational power, since on a single processor they already have had a large impact. Efforts such as the KOKKOS[7] library are attempting a more general approach and it is being incorporated in the MD package LAMMPS. The extent of this development will determine how well new architecture accelerates future progress.

While there is demand for accurate calculations, direct funding for the broad force-field development necessary to obtain accurate simulations is absent. In particular, much better water potentials for materials applications have been developed in the last decade,[8] but the corresponding force-field for the solute molecules (surfactant, polymers) still needs to be done. Polarizable[2] and reactive potentials[3] are other examples, which do have some development, but are lacking for many materials systems and are critical for many phenomena. With respect to simulations of systems with hard and soft components, which have force-fields with distinct functional forms, code development needs to be done as well as the development of force-fields for the cross interactions.

Coarse graining inherently changes the local energy barriers. Methods need to be developed to determine the effective time scales in coarse-grained models and the coupling between different dynamic processes. In addition, a framework to understand and address transferability to different state conditions is a challenge to be addressed in the next decade.

The application of time acceleration and rare event algorithms to soft matter will also greatly help.[4] Given the importance of nonequilibrium dynamical processes in soft matter, a strong foundation of the statistical physics for nonequilibrium simulations is critical, but presently not a focus.

MD simulations of materials often requires week-long runs. Most HPC centers impede the use of such simulations with very short time run times allowed. The needs of materials simulations needs to be incorporated in the queuing systems at the DOE HPC centers.

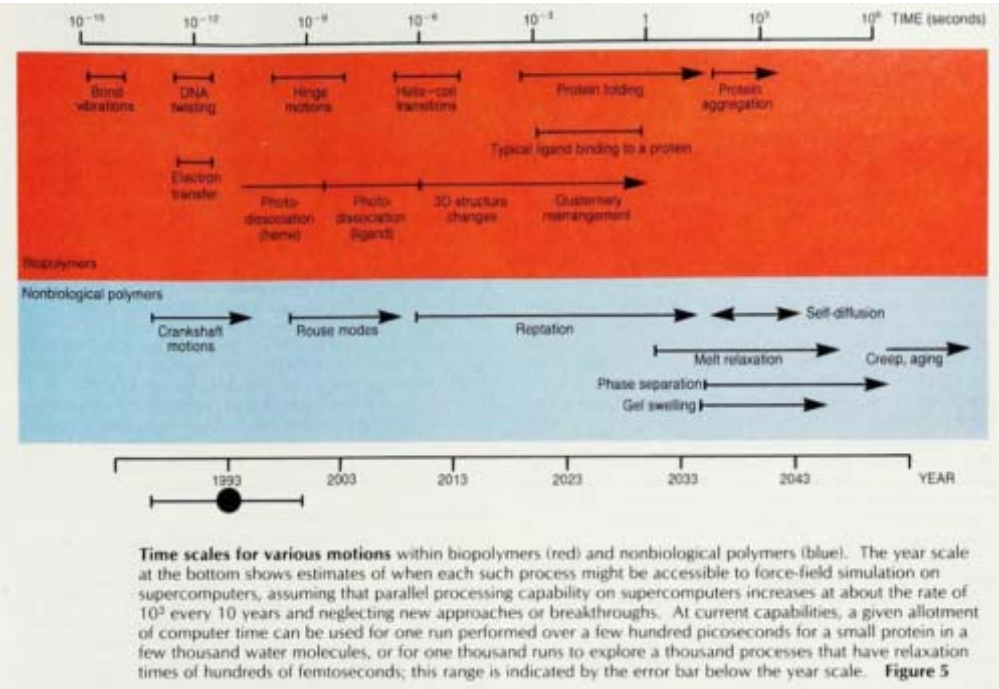
Presently, simulations of protein folding can reach ms time scales,[9] which is far beyond what polymer simulations can do. Polymer and other soft matter simulations are distinct from protein folding in that the systems typically have many polymers, not a single protein. However, there is much that can be learned from biomolecular simulations, but is being impeded by lack of transferring knowledge and code.

5. References *:*

1. “Opportunities in Theoretical and Computational Polymeric Materials and Soft Matter,” NSF Workshop, Santa Barbara, CA 2013.
2. Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Journal of Chemical Theory and Computation 2005, 1, 153- 168.
3. Liang, T. et al., Materials Science and Engineering: Reports 2013, 74, 255–279.
4. P. Tiwary and M. Parrinello, Phys. Rev. Lett. 111, 230602 (2013); J-M. Bello-Rivas and R. Elber, J. Comp. Chem. (2015) online.

5. G. Voth, Coarse-Graining of Condensed Phase and Biomolecular Systems (CRC Press/Taylor and Francis Group, Boca Raton, FL, 2009).
6. W. Noid, J. Chem. Phys. 139, 090901 (2013).
7. KOKKOS https://github.com/kokkos/kokkos; http://lammps.sandia.gov/doc/accelerate_kokkos.html.
8. T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande, Current Opinion in Structural Biology 23, 58 (2013).
9. J. Abascal and C. Vega, J. Chem. Phys. 123, 234505 (2005).

6. Images:



Time scales for various motions within biopolymers (red) and nonbiological polymers (blue). The year scale at the bottom shows estimates of when each such process might be accessible to force-field simulation on supercomputers, assuming that parallel processing capability on supercomputers increases at about the rate of 10^3 every 10 years and neglecting new approaches or breakthroughs. At current capabilities, a given allotment of computer time can be used for one run performed over a few hundred picoseconds for a small protein in a few thousand water molecules, or for one thousand runs to explore a thousand processes that have relaxation times of hundreds of femtoseconds; this range is indicated by the error bar below the year scale. **Figure 5**

Figure 1. From Hue Sun Chan and Ken Dill, Physics Today **46**(2), 24 (1993).

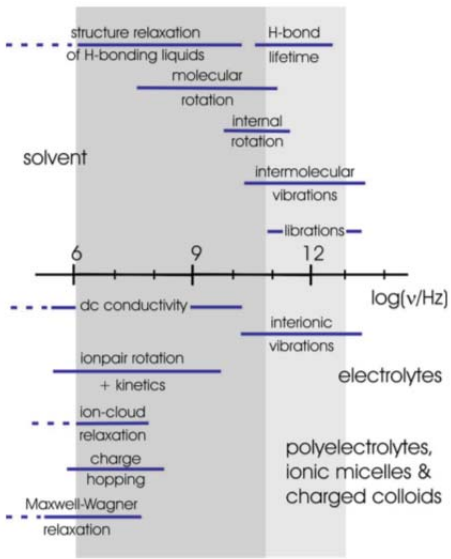


Fig. 2 Characteristic frequencies of dynamical processes contributing to the total dielectric response, $\hat{\eta}(\nu)$, of electrolyte systems at ambient temperatures. Modes arising from possible cross-correlations of solute and solvent motions are neglected for simplicity. The shaded area indicates the frequency range currently covered by the authors, partly in collaboration with other groups (light shading).

From R. Buchner and G. Hefter, Phys. Chem. Chem. Phys. 11, 8984 (2009).

BES White Paper – Building an APS Computing Ecosystem: Brian Toby (Argonne National Laboratory)

1. Science and Operation Drivers for the APS

The operation of the APS is in the middle of a revolution, due to improvements in X-ray area detection. Techniques that can utilize area detectors, which include the vast majority of APS beamlines, have seen 10- to 1,000-fold improvements in data rates and/or sensitivity in the past decade or less. Further, X-ray detector technology continues to improve at rates far faster than Moore's Law. As one example, in X-ray photon correlation spectroscopy, this year we will see a 10-fold expansion in data rates and expect at least another order of magnitude from detectors in development.

These rapid data rates, combined with types of measurements that require computation for even basic interpretation, provide a quality assurance problem to beamlines, in that real-time diagnostics and visualization are needed to verify that an experiment is progressing properly. Likewise, the APS needs to incorporate machine-learning techniques into beamline operations, so that instruments spend more time on collecting data from regions of interest and only survey less important areas. With today's instruments, this will improve efficiency, but in the future, as measurement speeds improve, humans will not be fast enough to assimilate and respond to the experimental data stream.

Further, the great success of X-ray techniques has caused the APS user community to evolve from a small cadre of X-ray scattering experts to a highly disparate range of domain scientists, who wish to apply APS measurements as one of many characterization techniques for samples generated within their research. This newer community needs easier-to-use beamlines and data analysis software, since they cannot devote years to mastering techniques, as did the previous generation of users. Frequently, analysis stalls when users find they lack expertise to complete their analysis.

The future of the APS is dominated by the planned Upgrade to an MBA lattice, which will give multiple orders-of-magnitude gains in brightness and coherence. This will act as a multiplier of data rates beyond the current detector revolution. Likewise, instruments will conduct multiple types of measurements simultaneously, providing multi-terabyte, >>4 dimensional, datasets each hour.

2. Solvable Computational Challenges for 2020–2025

Routine beamline use of leadership-scale computing is rare. The challenge is how to design facilities that can perform large-scale but rapid computations in near real time, while the large fraction of remaining compute cycles is utilized for the long-duration computations (typically simulations) that are the routine work for large machines. Development of large-scale machines that can spin up partitions, perform rapid task switching, and accommodate on-demand scheduling of high-priority tasks has not been done, but should not present any significant problems.

Data management hardware and software at the APS has also not caught up to the existing challenges. However, the technology needed for these problems is also readily available.

3. Insolvable Computational Challenges for 2020–2025

The challenge the facility cannot expect to solve is how to create the software needed to operate instruments with software as modern as the measurement techniques they offer, and to develop, maintain, and support user-friendly data analysis software, operating on platforms desired by users

(laptops through HPC). Also of great importance will be to implement automated workflows for data management and reduction to reduce staffing demands. For the X-ray Science Division, with close to 40 beamlines to support, all unique and many offering more than one type of measurement, the needs for high-quality data collection and data analysis software could not be met even if the entire scientific staff were applied to only software development. The APS hopes to begin use of a new MBA storage ring in 2022, which will improve the performance of many beamlines by as much as three more orders of magnitude. This will put even more pressure on software development.

4. Top accelerators and blockers for computing progress

Accelerate:

1. Well-developed and supported data standards, with documentation, APIs and basic tools. ***Why***? While a number of data standards have been proposed and partly developed, none are easy to implement or provide any compelling crosscutting applications to entice use. The lack of this leads to a "Tower of Babel" with respect to incompatible software development.
2. Tools that allow domain scientists without HPC skills to easily develop parallelizable code in programming environments they prefer, most commonly at present Python. ***Why***? Many scientists are very comfortable translating their data analysis concepts to computer code, but are much more comfortable doing that in a script language environment like Python, Matlab or R, but not in Java or C++. At present, significant effort from HPC experts is needed to adapt such codes to make effective use of even the multicore processors found in laptops.
3. New approaches to data set management/storage. ***Why***? Conceptually, one thinks of data as a single or perhaps a hierarchical set of files in a single place. In practice, at our facility a data set is accumulated as a set of files and sometimes database entries that are dispersed amongst multiple computers, as they are collected. After the experiment and during data reduction and analysis, these files are migrated to other locations, and additional intermediate results are added. Further, if one considers the frequent occurrence that multiple researchers work on analysis with periodic restarts and overlapping approaches, a data set looks more like a distributed github project than a hierarchical directory.
4. Multimodal data analysis tools. ***Why***? At present we are unprepared to reconstruct from multimodal data with self-consistent corrections for experimental artifacts, to sift through large data series to identify regions where the most significant changes are found, or to interpret very high dimensional data streams.

Impede:

1. Lack of inter-lab coordinated software development. ***Why***? Each SUF addresses its own most strategic development needs, but development targets are not coordinated across labs; there are no DOE-wide plans for project prioritization, scope, staffing continuity, etc.
2. The need to have long-running tasks in order to make use of large-scale computing facilities. ***Why***? In general, leadership computers do not task-switch quickly. For integration into a beamline analysis system, it might well be optimal to use a very large number of cores to complete a computation in minutes, where hours would be required on a small cluster, but if it takes 5 minutes to spin-up the analysis code and then another 5 minutes to start the next task, plus delays for data transport and queues, local computing becomes preferable.

APPENDIX D: BASIC ENERGY SCIENCES CASE STUDIES

The following case studies were submitted by the authors listed below soon after the Exascale Requirements Review to guide development of BES requirements and the text of this report.

D.1 Case Studies Addressing Novel Quantum Materials and Chemicals

D-3 Hai-Ping Cheng (University of Florida)

D-7 Aurora E. Clark (Washington State University), Christine Isborn (University of California-Merced), and Thomas Markland (Stanford University)

D-13 James Freericks (Georgetown University), Hulikal Krishnamurthy (Indian Institute of Science), and Tom Devereaux (Stanford Linear Accelerator Center)

D-17 Mark Gordon (Ames Laboratory)

D-20 Thomas Maier and Michael Summers (Oak Ridge National Laboratory)

D-24 A.J. Millis (Columbia University)

D.2 Case Studies Addressing Catalysis, Photosynthesis and Light Harvesting, and Combustion

D-28 W.A. de Jong, J. Brabec, and C. Yang (Lawrence Berkeley National Laboratory)

D-32 Laura Gagliardi (University of Minnesota)

D-35 Stephen J. Klippenstein (Argonne National Laboratory)

D-39 Xiaosong Li (University of Washington)

D-42 Roger Rousseau and Vassiliki-Alexandra Glezakou (Pacific Northwest National Laboratory)

D-45 Lyudmila V. Slipchenko (Purdue University)

D.3 Case Studies Addressing Materials and Chemical Discovery

D-49 Giulia Galli (University of Chicago/Argonne National Laboratory) and Francois Gygi (University of California-Davis)

D-52 Christopher J. Mundy and Gregory K. Schenter (Pacific Northwest National Laboratory)

D-56 J. Ilja Siepmann (University of Minnesota)

D-61 Mark Stevens (Sandia National Laboratories)

D.5 Case Studies Addressing Advances in Algorithms for Quantum Systems

D-65 Karol Kowalski (Pacific Northwest National Laboratory)

D-70 A.M.N. Niklasson (Los Alamos National Laboratory)

D.6 Case Studies Addressing Computing and Data Challenges @ BES Facilities

D-73 Dilworth Y. Parkinson, Alexander Hexemer, and Craig E. Tull (Lawrence Berkeley National Laboratory)

D-80 Jana Thayer and Amedeo Perazzo (Stanford Linear Accelerator Center)

D.7 Case Study Addressing Mathematics and Computer Science Transforming BES Science

D-86 J. Donatelli, A. Hexemer, D. Kumar, R. Pandolfi, D. Parkinson, V. Venkatakrishnan, P.H. Zwart, and J.A. Sethian (all of Lawrence Berkeley National Laboratory)

Next-Generation Workforce

No case studies addressed only this topic.

D.1 Case Studies Addressing Novel Quantum Materials and Chemicals

Case Study Title: Atomistic Simulations of Physical and Chemical Nanoscale Processes

Lead Author: Hai-Ping Cheng (University of Florida)

1. Description of Research

1.1 Overview and Context:

We study nanoscale physical and chemical processes based on atomistic simulations. Projects in my group include electron and spin transport in 2D nano-junctions, properties of magnetic nanoparticles, electron relaxation in quantum dots, interfacial phenomena at water-oxide interfaces, etc. Specifically, (1) we use first-principles density functional theory (DFT) method and theory beyond DFT in conjunction with the nonequilibrium Green's function method, effective screen medium, and Boltzmann theory to investigate graphene-2dCrystal-graphene heterogeneous structures (2dCrystal=BN, WS_2, MOS_2 etc.); (2) we use DFT and theory beyond DFT to study the electronic, magnetic, and spin-transport properties of $[Mn]_n$ nanocrystals with ligands, as well as to search for magnets with desired properties (a material genome project); (3) we use DFT in conjunction with the density matrix method based on quantum mechanical models of electron relaxation to study the excited state dynamics of molecules related to organic solar cells; state-of-the-art electronic structure calculations are performed to describe electronic behavior in these systems, pushing the limits of current memory requirements and requiring efficient parallelization over many cores; (4) we use classical molecular dynamics modeling to study the mechanical loss and structural properties of amorphous oxides, which are important materials for gravitational wave detection and optical measurements; high-performance computing and storage are necessary for running long simulations of large amorphous systems to thoroughly search potential energy landscapes; and (5) we use DFT and theory beyond to simulate $LaMnO_3$, a catalyst for alkaline fuel cells, and its interaction with adsorbates. Better understanding the catalytic efficiency of $LaMnO_3$ is essential to improving fuel cell applications. Our calculations require a large number of processors. These outputs help us look at energetic costs for form defects, the binding energy of interactions with intermediates of fuel cell interactions, and also explore the details of chemical binding on the surface.

1.2 Research Objectives for the Next Decade:

Our scientific goals for the next decade in this area of research include:

- Running first-principles simulations and design of realistic electronic devices.
- Simulating electron and spin transport using DFT with inclusion of strong correlation effects.
- Improving the parallelization and speed of potential energy landscape searches used with classical molecular dynamics simulations.
- Developing an electron relaxation method that applies to periodic systems using approximations while taking into account many-body electronic effects (GW and BSE).

2. Computational and Data Strategies

2.1 Approach:

First-principles DFT-based methods are hindered by low efficiency in parallel computation beyond 10^3 atoms. Kohn-Sham equations are converted to an eigen-problem or linear equation problem. In the next decade, the emerging many-body extensions may become the routine procedure and replace the role of the mean-field density functional theory.

2.2 Codes and Algorithms:

Our current codes and workflows and the algorithms that characterize them are as follows:

- Gaussian – A quantum-mechanical code implementing density functional theory to calculate the electronic structure of molecules and other finite systems.
- VASP and PWSCF – Plane-wave quantum-mechanical codes implementing density functional theory to study large periodic systems. VASP has better pseudo-potentials, but PWSCF has more added modules for investigating physical properties.
- Exciting+ (modified ELK) – All-electron, quantum-mechanical codes implementing density functional theory to study large periodic systems. This is a modern (best) version of full potential LAPW code for high-precision calculations, with many implementations.
- TranSiesta – Atomic-orbital quantum-mechanical codes implementing density functional theory and nonequilibrium Green's functions to electronic transport. It solves the Green's function by transforming the Schrodinger equation into linear equations.
- DL_POLY and LAMMP – A classical molecular dynamics code.
- In-house code to study charge and spin transport using a hybrid DFT-Boltzmann approach.
- In-house code to calculate excited electron dynamics in finite systems that is used in conjunction with output from Gaussian software.

3. Current and Future HPC Needs

3.1 Computational Hours:

We project needing the following to reach our scientific goals:

- Interfaces, junctions, and tunneling field transistors: 500,000 core-hours per system, for a total of 5,000,000.
- Magnetic nanoparticles: 100,000 per system, for a total of 5,000,000 (high throughput).
- LIGO – 200,000 core-hours required per amorphous sample studied, for a total of 2,000,000.
- Electron Relaxation – 100,000 core-hours required per system studied, for a total of 1,000,000.
- Response functions in oxides – 200,000 per system, for a total of 1,000,000.
- CNT in aqueous environment study (VASP) – 2,000,000 core-hours per system, for a total of 8,000,000.

These are calculated based on a one-year period.

3.2 Parallelism:

Our use is as follows:

- Classical MD codes DL_poly and LAMMP – Both multi-node and on-node parallelization (OpenMPI).
- Electron Relaxation – On-node parallelization.
- DFT codes VASP/PWSCF/TranSiesta/Exciting+ – Both multi-node and on-node parallelization.

3.3 Memory:

N/A.

3.4 Scratch Data and I/O:

N/A.

3.5 Long-term and Shared Online Data:

N/A.

3.6 Archival Data Storage:

N/A.

3.7 Workflows:

N/A.

3.8 Many-Core and/or GPU Readiness:

Our current codes are not yet ready for a transition to "lightweight" cores and/or hardware accelerators with deepening memory hierarchies. To convert in-house codes, it would be helpful to receive training and online documentation about how to make the transition.

4. Requirements Summary Worksheet

Table 1 shows our projected HPC requirements.

Table 1. Requirements

Code:	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	22,000,000	3×	5×
Computational node hours (Homogeneous many-core)[b]			
Computational node hours (w/GPU or accelerator)[c]	N/A	N/A	N/A
Memory per node	10–30 GB	20–50 GB	50 GB
Aggregate memory	1 TB	2 TB	4 TB
Data read and written per run	0.3 TB	1 TB	2 TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O	5	5	5
Scratch file system space needed	100 TB	300 TB	100 TB
Permanent online data storage	3 TB	10 TB	30 TB
Archival data storage needed	1,000 TB	3,000 TB	10,000 TB

[a] "Core hours" are used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" are used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" are used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

Case Study Title: Realistic Solution-Phase Chemistry: From Multi-component Non-ideal Mixtures, to Ground and Excited State Charge and Energy Transfer Processes

Lead Authors: Aurora E. Clark (Washington State University), Christine Isborn (University of California-Merced), and Thomas Markland (Stanford University)

1. Description of Research

1.1 Overview and Context: Complex multi-component liquids are non-ideal and contain solutes whose reactivity may encompass multiple equilibria, and be driven by external perturbations including evolution on excited states. Consider that electron transfer and proton-coupled electron transfer reactions are an essential feature of energy conversion/catalytic systems, but are also endemic to the chemistry of complex waste liquids found in DOE environmental management and waste processing sites. Fundamental software and model developments for realistic solution conditions and the reactions therein (including the interplay between photo-excitation, energy, proton, and electron transfer processes) must be implemented in a highly scalable approach that leverages analytics that can identify these reactive processes in real time.

1.2 Research Objectives for the Next Decade: Accurate knowledge gained across length and timescales is mandated for understanding the collective behavior of complex solutions. Higher accuracy means not only better mathematical descriptions of the physics, but also realistic depictions of the system and timescales long enough to observe unanticipated phenomena. Improvements to force fields, molecular dynamics algorithms, mixed QM/MM, and sampling strategies can all expand traditional classical approaches. However, major algorithm developments will take place in the area of large-scale QM calculations. This includes accelerating 2 e- integrals and maximizing calls to libraries on GPUs (e.g., NVIDIA CUBLAS library). This approach has recently been enabled within TeraChem for excited state calculations, which lead to a speedup of over 300× compared to standard CPU code (Figure 1).

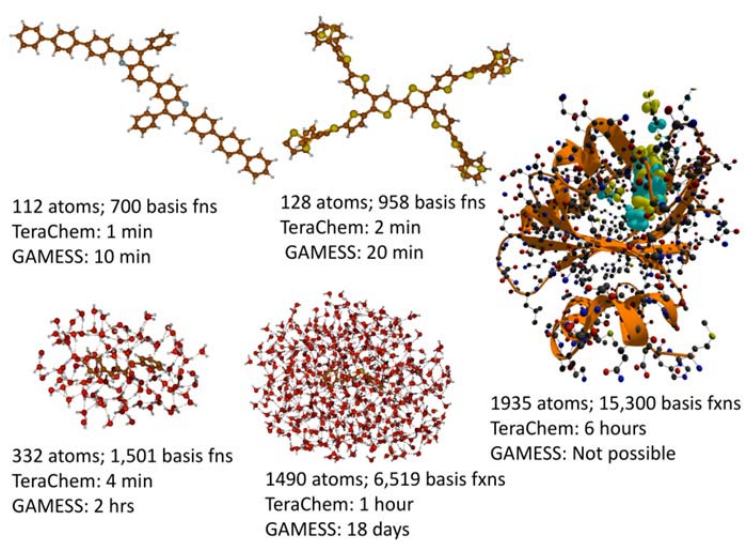


Figure 1. Timings for an excited state computation with TERACHEM vs. the CPU-based GAMESS package on an Intel Xeon X5570 with 8 Tesla C1060 GPUs. The use of GPUs with TERACHEM dramatically accelerates quantum chemistry calculations, enabling computation of systems of unprecedented size. This hardware and software technology will be used to compute the electronic excitations of solutes in the condensed phase.

GPU acceleration must be extended beyond electronic structure codes to include quantum dynamics algorithms. Nonadiabatic semiclassical quantum dynamics techniques can retain a high level of accuracy by evolving the reduced density matrix using a semiclassical approximation within the generalized quantum master equation framework. This approach (i) takes advantage of the short time accuracy of semiclassical trajectories, (ii) reduces the cost of each trajectory, (iii) allows for massive parallelization, and (iv) avoids the need to use a specific form of the interactions permitting its use with

force fields and ab initio excited state surfaces. Real-time data analytics of the perturbations in chemical interactions can identify both known and unknown reaction pathways as they happen such that total simulation time can be optimized, and algorithms that enable steered dynamics or rare event sampling can be implemented. This will be based upon Intermolecular Network Theory, which creates graphs of all relevant interactions in the system and then borrows from contemporary network analysis methods within Computer Science, such that the complex local and global changes that occur within the network can be dissected.

2. Computational and Data Strategies

2.1 Approach: Generally, when dynamics of molecular interactions at either the classical or quantum mechanical (QM) level of theory are performed; large amounts of data are generated and must be saved in the form of molecular and solvent configurations. Configurations are then either analyzed for trends in those molecular interactions or used to perform a higher-level QM calculation. Real-time analysis is one mechanism around this data obstacle. Path-integral molecular dynamics itself requires relatively small resources, but with ab initio interactions, many separate electronic structure calculations are required to obtain the forces that drive the dynamics.

2.2 Codes and Algorithms: Classical molecular dynamics codes include LAMMPS and DL_POLY, while electronic structure codes include TeraChem and NWChem, with the quantum dynamics algorithms CP2K (Markland implementation) and data analytics done by *ChemNetworks*.

3. Current and Future HPC Needs

3.1 Computational Hours: See Tables 1, 2, and 3.

3.2 Parallelism: TeraChem and LAMMPS are GPU accelerated codes. Both run well on conventional multi-core processors with multiple GPUs on a single node. CP2K is an electronic structure and ab initio molecular dynamics program parallelized using MPI and optionally OpenMP, with experimental GPU acceleration. It is highly optimized and runs on tens to thousands of cores, depending on the size of the system. Further, for path integral simulations, several (units to tens) separate instances are needed for different replicas of the system, offering another level of parallelism.

3.3 Memory: TeraChem has large memory requirements for many of the specialized jobs (256 GB per node). CP2K does not have large memory requirements, fitting under 1 GB of RAM per core for common types of simulations.

3.4 Scratch Data and I/O: TeraChem I/O is fairly minimal. CP2K does not have any notable I/O requirements.

3.5 Long-term and Shared Online Data: See Tables 1, 2, and 3 below.

3.6 Archival Data Storage: Archival data are generally not a requirement; however, we acknowledge that with future analytics capabilities, re-analysis of archived data may become attractive unless major advances in simulation algorithms are achieved that make it easier to

perform re-running the simulations. Therein, data management tools that reproduce the original simulation protocols must be employed.

3.7 Workflows: Current workflows rely on post-production analysis of all simulation data. Real-time analysis represents a major change to the workflow environment and has the possibility of dramatically increasing the chemical space that can be explored. By 2025, performing data analytics in real time should be routine. Trivially it can be used to optimize simulation time, but its largest benefit will be as a real-time chemical discovery tool. Changes in the patterns of chemical reactions can lead to the discovery of new correlations and reaction pathways that can subsequently be sampled at increased rates if analytics is paired with steering algorithms during the simulation.

3.8 Many-Core and/or GPU Readiness: TeraChem already takes advantage of GPUs, and shows improved performance with the latest generation of NVIDIA cards (Maxwell, with larger dedicated shared memory). CP2K is already highly optimized for shared and distributed memory architectures. New support for GPU acceleration of MPI parallel runs is now available and under further active development.

4. Requirements Summary Worksheets

Tables 1, 2, and 3 present estimates of required exascale resources for the TERACHEM, CP2K, and LAMMPS codes, respectively.

Table 1. TERACHEM Requirements

Code: **TERACHEM** GPU accelerated electronic structure code. Runs well on conventional multi-core with multiple (4–12) GPUs on a single node.	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	1,900 k/yr	10×	50×
Computational node hours (Homogeneous many-core)[b]			
Computational node hours (w/GPU or accelerator)[c]	120 k/yr	10×	50×
Memory per node	256 GB	500 GB	500 GB
Aggregate memory	TB	TB	TB
Data read and written per run	TB	TB	TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O			
Scratch file system space needed	TB	TB	TB
Permanent online data storage	TB	TB	TB
Archival data storage needed	TB	TB	TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.
[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.
[c] "Node hours" is used for "GPU or accelerator" usage.
[d] For example, 32 × column 1.

Table 2. CP2K Requirements

Code: CP2K, MPI parallel electronic structure, experimental GPU acceleration.	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)******	**Future Usage: 2025 2 (As a factor of column 1)******
Computational core hours (Conventional)*	2,300 k/yr	10×	50×
Computational node hours (Homogeneous many-core)**			
Computational node hours (w/GPU or accelerator)***	0 k/yr	140 k/yr	1,000 k/yr
Memory per node	64 GB	4×	4×
Aggregate memory	64–384 TB	4×	4×
Data read and written per run	50 GB	10×	10×
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O	0%		
Scratch file system space needed	0 TB	TB	TB
Permanent online data storage	10 TB	TB	TB
Archival data storage needed	20 TB	TB	TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.
[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.
[c] "Node hours" is used for "GPU or accelerator" usage.
[d] For example, 32 × column 1.

Table 3. LAMMPS Requirements

Code: **LAMMPS** GPU accelerated classical molecular dynamics code. Runs well on conventional multi-core with multiple (4–12) GPUs on a single node.	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)******	**Future Usage: 2025 (As a factor of column 1)******
Computational core hours (Conventional)*	100,000 k/yr	10x	50x
Computational node hours (Homogeneous many-core)**			
Computational node hours (w/GPU or accelerator)***	100,000 k/yr	10x	50x
Memory per node	32GB	32GB	32GB
Aggregate memory	TB	TB	TB
Data read and written per run	TB	TB	TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O			
Scratch file system space needed	TB	TB	TB
Permanent online data storage	TB	TB	TB
Archival data storage needed	TB	TB	TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.
[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.
[c] "Node hours" is used for "GPU or accelerator" usage.
[d] For example, 32 × column 1.

Case Study Title: Numerical Renormalization Group–based Methods for Nonequilibrium Many-Body Physics

Lead Authors: James Freericks (Georgetown University), Hulikal Krishnamurthy (Indian Institute of Science), and Tom Devereaux (Stanford Linear Accelerator Center)

1. Description of Research

1.1 Overview and Context: Nearly all electronics, including energy-relevant electronics, employ nonlinear properties of the materials in making the device work. There is a widespread interest in employing so-called strongly correlated materials into future devices, because they have the potential for being highly tunable and working with low power. Unfortunately, the theory to describe such systems remains in its infancy. Over the past decade, significant work to solve these problems has advanced with a technique called dynamical mean-field theory. Key to the success of this approach is the development of a nonequilibrium impurity solver. This case study will focus on the development of a new solver and its implementation. Note that this tool has not been developed yet, so this case study will focus on what we plan over the next decade.

1.2 Research Objectives for the Next Decade:

The research objectives are to be able to use numerical calculations to predict new properties of strongly correlated materials that are driven into nonequilibrium. Current technology employs perturbation theory to describe nearly all systems with the associated problems of convergence and accuracy. While some more exact methods, like quantum Monte Carlo, have been developed, they tend to suffer from the difficulty of not being able to run long enough to describe interesting behavior. Our goal is to develop a different type of algorithm, which will allow for accurate solution of these problems. With such a solver in hand, we will be able to investigate the nonequilibrium behavior of pump/probe experiments, witching experiments, ccur only in nonequilibrium, and so on. There may even be new kinds of behavior not yet dreamed of that await discovery. The computational effort is critical to the success of this approach. First, we have to have an exact algorithm that can be employed, and second, we need to have it run efficiently to be able to analyze the most interesting scientific cases.

2. Computational and Data Strategies

2.1 Approach: The approach is based on something called the numerical renormalization group, developed by Ken Wilson in 1975 to tackle one of the most complex many-body physics problems of that time, the so-called Kondo problem. In nonequilibrium, we formulate the

problem in the time domain rather than the frequency domain, but we still involve a series of "bath" states that the impurity interacts with in order to create the fluctuating field that determines the complex quantum mechanical behavior.

2.2 Codes and Algorithms:

We do not have any current codes available, because we are developing this approach right now. But the basic idea is to evaluate the Green's function as a trace over relevant states by properly treating the time evolution of the states. This approach works best for low-temperature systems, because the number of states that contribute is small. The parallelism will involve sending the evolution of different states to different processors. This part of the algorithm is highly parallelizable. The critical elements in this approach are first finding an efficient way to represent the wavefunction at each time step that allows us to calculate how it evolves in the next time step, and to truncate the representation of the wavefunction at each time step to keep the size small enough so that it remains representable in the machine. These ideas are similar in many respects to density matrix renormalization group techniques (and the tensor product extensions of that method) but with a different way to encode the wavefunction.

3. Current and Future HPC Needs

3.1 Computational Hours:

Because we do not currently have functioning codes, we will be making estimates of these numbers for codes in the future. The algorithm involves two steps, one which is essentially a sparse matrix vector multiply and one which is a vector projection to truncate the growth of the number of nonzero elements in the vector. As with many codes of this type, the ability to effectively use accelerator units will rely on the ability to manage the memory handling of the data as the calculations proceed. We suspect accelerators will be able to be used to some extent, but are not clear on how well they will perform. The problem size grows with the number of bath states that are employed and with the number of time steps. As the field matures, we will want to run for longer and longer times to find the interesting behavior. Larger numbers of bath states will be required to give more precise answers. Our estimates of computational hours for these codes is given in Table 1.

3.2 Parallelism: Because most operations involve different linear algebra steps, it is likely this algorithm will be able to employ both coarse and fine-grained parallelism. But, because it requires the states and the matrices to be stored in an ultra sparse format, it is likely we will need to write all of the codes for these parts of the algorithm ourselves because common approaches from lapack and blas will not be directly applicable.

3.3 Memory: These codes are almost certainly going to be memory limited and will benefit by running on machines with the highest memory per core that can be found. Even though we have an ultra sparse system, the dimension of the vector space we work in is so large, one could never imagine storing the entire vector on any computer. Thus, figuring how to properly manage the memory and the different hierarchies will be particularly important for this work.

3.4 Scratch Data and I/O: The program does not need significant I/O because the output involves dense matrices of moderate dimensions. Checkpointing, on the other hand, can require large amounts of data if we have to record the intermediate vectors and matrices to be able to restart the calculation. We have not yet decided precisely how we will deal with the checkpointing of the code.

3.5 Long-term and Shared Online Data: We will need to store all of the output data from given runs for a substantial period of time. This is because we will need to hold onto the raw data used for future analysis, and we will need to store it to meet the data management policies of DOE for its sponsored research. In any case, however, we do not anticipate the stored data to be huge. On the order of a few 10s to 100s of terabytes is probably all we will need.

3.6 Archival Data Storage: Similar to the on-line storage, this will be used primarily for data in projects that have already been completed.

3.7 Workflows: The typical workflow for these types of calculations is that we employ HPC resources to generate the Green's functions of the electrons (and phonons) for the particular drive fields of the given experiment. These results are then stored for further post processing (which often is done on smaller computers). The post processing includes calculating measurable quantities for a given experiment or generating graphics to display the final results.

3.8 Many-Core and/or GPU Readiness: While we will try to design codes with an eye toward different hierarchical schemes for memory and processing, it is too early to tell how this will be done. However, if robust middleware is developed that will allow developers to program these schemes more efficiently for the newer machines, we would definitely employ such software in our project.

4. Requirements Summary Worksheet

Table 1. Requirements for Nonequilibrium Wilson

Code: _Nonequilibrium Wilson _	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	5 million	5×	25×
Computational node hours (Homogeneous many-core)[b]	5 million	5×	25×
Computational node hours (w/GPU or accelerator)[c]	?	?	?
Memory per node	64 GB	256 GB	1,024 GB
Aggregate memory	20 TB	80 TB	320 TB
Data read and written per run	2 TB	2 TB	2 TB
Maximum I/O bandwidth needed	100 GB/sec	100 GB/sec	100 GB/sec
Percent of runtime for I/O	5	5	
Scratch file system space needed	10 TB	10 TB	10 TB
Permanent online data storage	50 TB	100 TB	200 TB
Archival data storage needed	500 TB	1,000 TB	2,000 TB

[a] "Core hours" is used for "conventional" processors" (i.e., node-hours * cores_per_node). Intel's "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

Case Study Title: LiquidPhenom INCITE

Lead Author: Mark Gordon (Ames Laboratory)

1. Description of Research

1.1 Overview and Context: The focus of this project, just starting its second INCITE year, is the study of liquid behavior using high-level electronic structure theory. In particular, the project involves the simulations of room-temperature ionic liquids that are important in the separation process of heavy elements in critical materials and the simulations of aqueous solvation of anions such as halides and nitrate. The primary method used for the project is the fragment molecular orbital (FMO) method. The FMO method scales well to the petascale.

1.2 Research Objectives for the Next Decade: The goal is to be able to perform FMO3 molecular dynamics (MD) simulations at the level of second-order perturbation theory (MP2) with less than 60 seconds (preferably less than 30 seconds)/time step. The FMO3 variation includes explicit three-body interactions, which have been shown to be essential for simulations that involve water. Fully analytic FMO3 gradients have recently been developed and implemented for open and closed shell Hartree-Fock (HF), density functional theory (DFT), and MP2, and excellent scalability has been demonstrated.

2. Computational and Data Strategies

2.1 Approach: The methods used in this project are electronic structure theory, a subset of computational chemistry, as implemented in GAMESS (General Atomic and Molecular Electronic Structure System). At the HF and DFT levels of theory, the calculations are dominated by the computation of large numbers of 2-electron integrals. Several 2-electron integral methods are employed, tailored to the specifics of the integral components (basis functions). The MP2 calculations are dominated by DGEMMs. Traditional algorithms have been written in Fortran; however, the advent of novel computer architectures, such as GP-GPU, Intel Phi, ARM, and others, have prompted a rethinking of code development in terms of object-oriented C++ and associated software, such as templates. As new architectures become available and feasible, code development will have to adjust accordingly.

2.2 Codes and Algorithms: The primary code is GAMESS, which is primarily written in Fortran. A secondary code is libcchem, written in object-oriented C++. Libcchem contains a small subset of the current GAMESS functionality. At present, libcchem contains HF, MP2, and coupled cluster (CC) methods for closed shells. Open shell variants (for molecules with unpaired electrons) are in progress. Over the next several years, other functionalities will be

implemented for libcchem. These include analytic gradients for all functionalities, Hessians for some functionalities, and multireference methods.

3. Current and Future HPC Needs

I am opting to put answers in each section in narrative form.

3.1 Computational Hours: This is variable, depending on the application and the method used. GAMESS is a multifunctional suite of programs. The INCITE project uses Hartree-Fock (HF), density functional theory (DFT), and second order perturbation theory (MP2), combined with the fragment molecular orbital (FMO) method. The FMO method scales nearly linearly with the size of the system, but with a significant prefactor. FMO/HF and FMO/DFT still scale very close to linearly, whereas FMO/MP2 is a little worse but still better than quadratically. In a recent study of water clusters using FMO/HF on the BG/Q at Argonne National Laboratory, the scaling was shown to be very good up to ~260,000 cores. This excellent scaling was accomplished using the Generalized Distributed Data Interface (GDDI) combined with a novel approach that requires no I/O beyond input and output files. Over the next 5–10 years, it will become possible to make effective use of one million cores or more. These advances will enable molecular dynamics (MD) simulations with high-quality ab initio electronic structure theory, including periodic boundary conditions.

3.2 Parallelism: GDDI makes use of multilevel parallelism. The FMO method is ideally suited for GDDI, because the method divides a molecular system into fragments, each of which can be calculated independently on a separate node. Many of the GAMESS functionalities (e.g., MP2) additionally have parallel algorithms and are therefore able to take advantage of fine-grain parallelism among the cores within a node. At present, GAMESS is not thread aware. The addition of thread capability is expected to further improve the GAMESS parallelism and scalability.

3.3 Memory: We generally need at least 2 GB/core. For more demanding calculations, this will increase to up to 8GB/core.

3.4 Scratch Data and I/O: The required online scratch space depends on the problem and can range from very little to tens or hundreds of GBs. The latest development of the FMO method is virtually I/O-free, in order to improve scalability. For the most demanding coupled cluster calculations, our new C++ CCSD(T) code uses well under 10% I/O, because the most demanding part of such a calculation is the triples (T), which are not I/O bound.

3.5 Long-term and Shared Online Data: Virtually none.

3.6 Archival Data Storage: Virtually none.

3.7 Workflows: If I understand this question (which is pretty vague), input file → internal data preparation → computation → output file. There are no unusual requirements.

3.8 Many-Core and/or GPU Readiness: We have been in the forefront of the development of codes for lightweight, low-power architectures for electronic structure theory codes. We have developed a C++ library (libcchem) for such codes for both CPUs and GPUs and continue to expand the methods that are available in libcchem. We are also in the forefront of the development of codes for ARM architectures (with two papers already in print). Our collaborators are also investigating FPGA architectures. The strategy is to secure funding to keep this process going (so far, so good) and to collaborate with the hardware vendors (e.g., NVIDIA, Microsoft).

3.9 Software Applications, Libraries, and Tools: This topic does not apply.

3.10 HPC Services: This topic does not apply.

3.11 Additional Needs: What the entire community needs are more collaborative efforts among the application developers, applied mathematicians, and computer scientists.

Case Study Title: Dynamic Cluster Quantum Monte Carlo Simulations of High-Temperature Superconductors

Lead Authors: Thomas Maier and Michael Summers (Oak Ridge National Laboratory)

1. Description of Research

1.1 Overview and Context: The scientific goal of this research is to understand, predict, and ultimately optimize the complex behavior of high-temperature superconductors, a class of materials that can transport electricity with perfect efficiency below a critical temperature. Harnessing these materials can have important technological benefits in energy-related areas, such as power transmission and generation, but optimization is necessary to unleash their full potential. The lack of a "small parameter" in the physics of these materials has eluded controlled analytical treatments, and high-end computing is necessary to handle the quantum many-body problem underlying their complex behavior.

1.2 Research Objectives for the Next Decade: Remarkable progress has been made in understanding the qualitative physics of these systems by simulating generic, often paradigmatic models that do not account for the chemical details of a certain material. An important goal for the next decade is to obtain a quantitative understanding of these systems, such as the variation of the critical temperature Tc across a certain class of superconductors and the factors that determine Tc, and thereby obtain information on how to optimize these systems. Computationally, the required simulations of more complex models, which incorporate the chemical details and thus can distinguish different materials in a class, imply a rapid increase in numerical complexity that can only be handled by the most advanced HPC systems.

2. Computational and Data Strategies

2.1 Approach: The quantum many-body simulations are based on a dynamic cluster quantum Monte Carlo (QMC) approximation in which a finite-size cluster of atoms is self consistently embedded in a mean-field designed to describe the rest of the bulk system. The cluster quantum many-body problem is then solved exactly by a QMC algorithm with numerical complexity that scales as the cube of the number of degrees of freedom (number of atoms, electronic orbitals per atom) on the cluster.

2.2 Codes and Algorithms: The DCA++ code implements a dynamic cluster continuous-time auxiliary-field QMC algorithm using C++ and generic programming models on distributed multi-core and hybrid CPU-GPU systems. It uses MPI for internode parallelization, pthreads for multi-threading on the nodes, and CUDA for the hybrid implementation.

3. Current and Future HPC Needs

3.1 Computational Hours: In 2015, our simulations used around 50 M CPU hours on ORNL's Titan or roughly 3 M node hours. As a specific example, the calculation of the superconducting transition temperature Tc in a generic model with only a single orbital degree of freedom uses on the order of 0.1 M node hours on Titan. Simulations of materials-specific models with N orbital degrees of freedom will be a factor N3 more expensive and thus will use for N = 3—5 of the order of 3—13 M node hours for the simulation of a single material. Based on this, an estimate of 100 M node hours seems reasonable for achieving the scientific goals listed under Section 1.2.

3.2 Parallelism: The DCA++ code uses MPI for internode communication and a task-based multi-threading model for fine-grained intra-node communication. Using this model, the DCA++ code scales with almost perfect parallel efficiency to 18,600 nodes on Titan with a sustained performance of up to 15.4 petaflops. In order to further increase this parallelism, the HPC system must: 1) permit GPU kernels to be able to launch other kernels and call GPU libraries (Kepler Dynamic Parallelism), 2) put the GPUs and CPUs in the same NUMA memory space and 3) provide CPU-to-GPU inter-process coordination mechanisms (e.g., semaphores and atomics). Titan's PCI 2 bus ruled out such GPU to CPU coordination techniques. In the future, we plan to 1) take advantage of Kepler's dynamic parallelism to run executive routines on the GPU, and 2) explore the development of a task-processing executive kernel that will use GPU-CPU semaphores enabling the GPU to pull tasks from a task pool.

3.3 Memory: Most current DCA++ simulations fit in the memory available on a node (32 + 6 GB on Titan). The memory requirements scale as N4 with the number N of orbitals, so that future simulations will require more main memory on each node than is available. In this case, the code must be modular enough to reconfigure itself for distributed memory operations, even though this will be at the expense of performance. We look forward to systems that provide nodes with many terabytes of memory per node. Titan provides some capability to control allocation of threads to NUMA domains, but the ability to do so dynamically under program control is poorly supported. The ability to do this well is very important to the highly threaded, dynamically adaptive DCA++ code. It will be all the more important in HPC systems that provide nodes with terabyte memories. Tools that support fine-grained NUMA tuning of a multi-threaded, multi-MPI-process applications will also be important.

3.4 Scratch Data and I/O: The DCA++ code minimizes the data volume written to disk and I/O by performing much of the analysis on the fly during the DCA++ simulation. Restart data are

minimal, in the 10- to 100-MB range. Written data are more significant but usually less than 50 GB per run. I/O requirements are therefore marginal. A scratch storage space of 1 TB easily accommodates the data generated by a simulation of a single material. The numbers will scale to more significant values for the runs planned in the future, but an on-the-fly analysis model will be able to keep data volumes and I/O requirements manageable.

3.5 Long-term and Shared Online Data: A long-term storage of about 5 TB that can be shared among and accessed by all project members accommodates our current needs, including collaborations between different institutions. In the future, this will likely increase by a factor of 5–10.

3.6 Archival Data Storage: We currently have about 3 TB of data stored in archival storage. We expect this number to increase by a factor of 5–10 in the 2020–2025 timeframe.

3.7 Workflows: A typical workflow consists of running the DCA++ code on Titan, generating data, which are then analyzed by a separate analysis code that is usually run on a local GPU-enabled cluster using the latest MAGMA library for the diagonalization of large matrices. This workflow will likely change, due to the increased data size of future simulations. A possible model to accommodate the increase in data volume will be to perform much of the analysis on the fly, on a dedicated part of the HPC system, while the main QMC simulation is running.

3.8 Many-Core and/or GPU Readiness: The DCA++ code has GPU support built in, and the architectural approach is being adapted to provide the modularity and flexibility needed to adapt to various computing platforms and evolving science requirements. At the top level, the DCA++ code runs as a collection of interacting MPI applications. Each such application spans multiple nodes and uses MPI intra-communicators to coordinate their computations. These applications primarily interact with each other via inter-communicators that are used to send and receive task requests and their associated data. The system also requires the use of inter-process, shared-memory pools for zero-copy transfers. Within each MPI application, task-based multi-threading is used to further decompose the computations and to control computations being performed on the GPU. Supporting MPI dynamic process creation would not only be advantageous from a software engineering perspective but would also enable the code to dynamically configure itself to respond to the demands of the current run.

4. Requirements Summary Worksheet

Table 1 shows our projected HPC requirements.

Table 1. DCA++ Requirements

Code: DCA++	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	NA	NA	NA
Computational node hours (Homogeneous many-core)[b]	NA	NA	NA
Computational node hours (w/GPU or accelerator)[c]	~ 3 M / year	~ 15×	~ 30×
Memory per node	10–40 GB	~ 5×	~ 10×
Aggregate memory	10–40 GB	~ 16×	~ 81×
Data read and written per run	100 MB (read), 10–40 GB (written)	~ 16×	~ 81×
Maximum I/O bandwidth needed	100 GB/sec	500 GB/sec	1,000 GB/sec
Percent of runtime for I/O	< 1%	< 5%	< 10%
Scratch file system space needed	1 TB	10×	20×
Permanent online data storage	5 TB	5×	10×
Archival data storage needed	3 TB	5×	10×

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

Case Study Title: Modern Materials Theory: Collective Electronic Properties

Lead Author: A.J. Millis (Columbia University)

1. Description of Research

1.1 Overview and Context:

The goal of materials theory is to predict and control collective electronic properties of molecules and solids such as superconductivity and magnetism, and to understand their response to changes in chemistry, strain, applied fields, and ultrafast (high-intensity laser) fields. Addressing these issues requires solving the quantum mechanical equations describing many interacting electrons. The combination of quantum entanglement (anti-symmetry of electron wave function) and electron-electron interactions means that the general solution of the many-electron problem is hard (indeed, it is generally believed to be NP-hard), and until recently, a useful solution was believed to be out of reach. New ideas involving both novel intelligently crafted approximations and large-scale numerics are now enabling significant progress**.**

1.2 Research Objectives for the Next Decade:

Research objectives for the next decade include completing the solution of key model systems (e.g., the ``Hubbard model'' believed to describe properties of the high-temperature copper-oxide superconductors); improving the interface between model system and real materials studies (solving the double-counting problem in density functional plus dynamical mean field theory and developing the GW+DMFT methodology); developing diagrammatic Monte Carlo and tensor network/matrix product state methods to the point where they provide robust approaches to studies of important materials; and developing, implementing, and validating the methods needed to understand the properties of materials under extreme non-equilibrium conditions, including intense laser irradiation and high current flow. The result will be a general computational framework for understanding and optimizing properties of other classes of materials, including thermoelectrics, battery materials, catalysts, light harvesters, and Mott transition systems.

2. Computational and Data Strategies

2.1 Approach:

The key HPC issues in this research pertain to computing; the storage and networking issues are manageable. The crucial tasks involve matrices (determinants, eigenvalues, and eigenfunctions), which are large and typically not sparse, and stochastic evaluation of a variety of different kinds of expressions. These tasks require cycles and also local memory. A wide variety of pre and post-processing steps are also required, including solution of mean field (density functional) electronic structures, computation of large numbers of matrix elements, fast Fourier transforms of non-equispaced data, and fast and multipole computations of Columb integrals, etc.

2.2 Codes and Algorithms:

The current codes used in my group are continuous time quantum Monte Carlos (ALPS and TRIQS libraries) and Lanczos diagonalization and home-made diagonalization involving adaptively chosen bases. These are wrapped with standard density functional and applied math codes.

3. Current and Future HPC Needs

3.1 Computational Hours: See Table 1.

3.2 Parallelism: We use only local parallelism because the programming overhead to parallelize matrix operations across many cores is too great. This is unlikely to change.

3.3 Memory: See Table 1. On-node memory is very important to our work. Having less on-node memory in the future would be very bad for our research. We need more.

3.4 Scratch Data and I/O: I/O is not a serious issue for our operations.

3.5 Long-term and Shared Online Data: See Table 1.

3.6 Archival Data Storage: See Table 1. Data storage is not likely to be a major bottleneck.

3.7 Workflows: Key to progress in this field is the ability to run a large number of different and relatively small-scale simulations. We all operate in a “burst-y” mode, with periods of algorithm development and testing alternating with intense periods of calculation.

3.8 Many-Core and/or GPU Readiness: We do not currently use GPUs or other hardware accelerators. Our codes are not ready for this. We do not currently have a strategy for exploiting these technologies. Optimizing code for new generations of hardware accelerators is beyond the resources of our group and of most groups doing similar work. Also, the hardware accelerators have, to date, had little impact on the important problems of materials theory, and this situation is likely to continue. The physics of the many-electron problem means that all approaches face an exponential barrier of computational difficulty. Progress has come from developing and implementing creative algorithms (e.g., tensor networks, diagrammatic Monte Carlos, or the continuous time QMC used in DMFT calculations), and this situation will continue. What is needed above all else is a flexible and well-supported computing environment that enables easy development of sophisticated codes. Libraries and other resources that compile and function on all relevant machines are essential, as is a high degree of user support.

3.9 Software Applications, Libraries, and Tools: This topic does not apply.

3.10 HPC Services: As noted above, libraries, advanced training, and user support are essential now and will become more important. The advances will come from developing new algorithms, and the materials theory community will need to be able to implement these efficiently.

4. Requirements Summary Worksheet

Table 1. Modern Materials Theory: Collective Electronic Properties Code Requirements

Code: _____	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 2 (As a factor of column 1[d]**
Computational core hours (Conventional)[a]	2,000,000	3	9
Computational node hours (Homogeneous many-core)[b]	N/A		
Computational node hours (w/GPU or accelerator)[c]	N/A		
Memory per node	32 GB	128 GB	1 TB
Aggregate memory	1 TB	5 TB	20 TB
Data read and written per run	0.1 TB	0.5 TB	1 TB
Maximum I/O bandwidth needed	Not a constraint	GB/sec	GB/sec
Percent of runtime for I/O	Small		
Scratch file system space needed	1 TB	3 TB	10 TB
Permanent online data storage	TB	TB	TB
Archival data storage needed	10 TB	50 TB	200 TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

D.2 Case Studies Addressing Catalysis, Photosynthesis and Light Harvesting, and Combustion

Case Study Title: In-Silico Design of Catalysts for Production of Sustainable Chemicals and Fuels

Lead Authors: W.A. de Jong, J. Brabec, and C. Yang (Lawrence Berkeley National Laboratory)

1. Description of Research

1.1 Overview and Context: Development of sustainable sources of essential energy, chemicals, and materials as an alternative for, and supplement to, fossil fuels is crucial to meet the increasing demands for consumables in our growing society. The key scientific and engineering challenge in the production of sustainable products is to develop efficient, environmentally friendly, and cost-effective methods at industrial scale. Catalysts are central to overcoming the engineering and scientific barriers to economically feasible routes for the conversion of biomass-derived and solar-mediated fuel and chemicals into usable products. An example is the conversion of cellulose into sugars and bio-oils, which through catalytic processes can be converted into biofuels or building blocks for industrial applications such as plastics. In silico design utilizing high-performance computing resources is critical to accelerating the development of new catalysts, chemical reactions, and transformation processes. Accurate simulations of the kinetics and thermodynamics of chemical transformations enable scientists to discover new and novel ways to predict, control, and design optimal – industrially viable – catalytic activity and selectivity by rapidly scanning the large design space.

1.2 Research Objectives for the Next Decade: The first target for biomass catalysts at the exascale computing systems would be to obtain accurate thermodynamic and kinetic properties for the chemical transformation of two widely used bio-oil products: free fatty acids (50 atoms) or triglycerides reacting (160 atoms) with methanol and a catalyst for biodiesel. To model the catalytic process (transition states, thermodynamics), one will have to include at least the active part of the catalytic material (which could be a zeolite or clay-like material).

2. Computational and Data Strategies

2.1 Approach: To be predictive requires the capability to model chemical reaction landscapes with very high accuracy to determine and discover dynamic reaction pathways that can exhibit complex behavior over a wide range of time scales. Utilizing a petascale machine, thermodynamics and kinetics of organic molecules of less than 100 atoms are feasible. Considering the O(N6-N7) scaling of coupled cluster methods, molecular systems that are 2.5 to

3 times larger will be feasible, more if advances in reducing the computational complexity due to sparsity can be achieved.

2.2 Codes and Algorithms: The main algorithm that will utilize the majority of the compute cycles will be the coupled cluster algorithm as it has been implemented in the NWChem computational chemistry software. The algorithm relies heavily on local DGEMM operations and remote put-and-get data movement operations. It has been shown to perform on accelerator platforms.

3. Current and Future HPC Needs

3.1 Computational Hours: Computational cycles in coupled cluster methodologies are mainly driven by DGEMM-like operations. Both conventional and accelerator compute cores can be utilized for this class of computation. With accelerators, the biggest challenge has been the slow transfer of data to the card. Next-generation architectures are addressing this challenge, further improving the platform suitability for this class of algorithms.

3.2 Parallelism: The highly scalable, open-source NWChem software suite can perform large-scale, coupled cluster-based simulations. Coupled cluster algorithms have a computational complexity of O(N6-N7) and use block-sparse DGEMM-like kernels. In NWChem, these complex-to-code kernels are computer-generated with the Tensor Contraction Engine software. Data movement is done using one-sided put, get, and accumulate operations using the Global Arrays Toolkit or LBNL's GASNet. Single-level parallelism is achieved using a global task counter as a queue. The software has been demonstrated to scale to 210,000 CPU cores on the ORNL Cray XT5 Jaguar leadership platform to determine the energetics of a 68-atom organic molecule with 240 electrons and 780 basis functions, achieving over 80% of peak performance for the O(N7) part of the algorithm. The O(N6) component does not achieve this level of scalability due to increased data movement and fewer available parallel tasks. Hand-coded multilevel parallelism for selected algorithms provides some extra scalability, computing different terms of the equations in parallel on predefined groups of processors or multiple loosely coupled terms in multi-reference coupled cluster calculations in parallel.

3.3 Memory: Memory is utilized to store a large configuration space vector in a distributed fashion. Necessary data are obtained locally through one-sided put-and-get operations. Generally, our algorithms perform best if the peak DGEMM performance can be achieved. Reduction in the memory footprint available to store local data for these DGEMM operations will affect the performance. On-chip fast memory could be exploited through its use as a pre-fetch buffer, or if the memory and bandwidth are fast enough as a streaming data source into the caches.

3.4 Scratch Data and I/O: The increase in the problem size accessible on the exascale platforms will inherently lead to increased data of discretized wave functions generated by the simulations. These data are already of petabyte size on current petascale platforms and are stored for restart purposes only requiring using efficient parallel I/O libraries or emerging technologies such as burst buffers. If the MTBF can be sustained at current levels one may be able to limit restart needs.

3.5 Long-term and Shared Online Data: Most of the essential simulation results volume tends to be small, although more information may need to be stored in online databases.

3.6 Archival Data Storage: Most of the essential simulation results volume tends to be small.

3.7 Workflows: Truly accelerating scientific discovery requires the generation of large numbers of simulation data of realistic models and a seamless integration with analysis and deep learning in a reasonably short timeframe. Researchers are building homegrown databases for their data that are hard to integrate for advanced knowledge discovery. Federated database infrastructures that contain both computational and experimental data, combined with transparent data movement, will be essential to accelerate the design of new catalysts and industrial catalytic processes.

3.8 Many-Core and/or GPU Readiness: In recent years, the O(N7) compute-intensive kernels have been hand-coded in CUDA for GPU, and manually optimized and extended with OpenMP directives to achieve performance on multicore platforms such as the Xeon Phi. Next-generation codes and algorithms need to be developed that can exploit the drastically increasing concurrency and can handle increasingly expensive and more dynamical data movement. Programming languages and runtime schedulers that enable developers to express work concurrency and data movement in a system-agnostic way will be instrumental in developing efficient and scalable codes.

3.9 Software Applications, Libraries, and Tools: The efficiency and scalability of computational chemistry and materials applications are driven by the ability to move data into the cache of the processor. Processing ability is increasing in next-generation architectures, but the ability to store and move data is progressing much more slowly, which will affect application performance. Programming languages and runtime schedulers that enable developers to express work concurrency and data movement in a system-agnostic way will be instrumental in developing efficient and scalable applications. Algorithms need to be developed that can exploit the drastically increasing concurrency and can handle increasingly expensive and more dynamical data movement.

3.10 HPC Services: An impediment is a focus on hero runs and the development of hero applications that have limited science impact and a small user base. It is amazing to me that all the computing centers within the DOE complex have not been able to put a unified sign-on system in place for the open-access compute resources.

3.11 Additional Needs: Co-design of the applications developers, computer scientists, and applied mathematicians is critical in identifying the most suitable discretization and solver techniques, in addition to developing simulation software that will take optimal advantage of the extant platforms in the 2020–2025 timeframe to enable new scientific discoveries.

Table 1 presents estimates of required exascale resources for the NWChem code.

4. Requirements Summary Worksheet

Table 1. NWChem Requirements

Code: NWChem, not accounting for new novel algorithms to reduce computational cost or memory footprint.	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	600 K per run	40×	100×
Computational node hours (Homogeneous many-core)[b]	20 K per run	40×	100×
Computational node hours (w/GPU or accelerator)[c]	30 K per run	40×	100×
Memory per node	32 GB	1× is target	1× is target
Aggregate memory	600 TB	2×	4×
Data read and written per run	1 TB to disk	2×	4×
Maximum I/O bandwidth needed	500 GB/sec	2×	2×
Percent of runtime for I/O	<5%	1×	1×
Scratch file system space needed	1–2 TB	2×	4×
Permanent online data storage	1 TB	10×	10×
Archival data storage needed	1 TB	10×	10×

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

Case Study Title: Next-Generation Catalysts for Natural Gas Conversion

Lead Author: Laura Gagliardi (University of Minnesota)

1. Description of Research

1.1 Overview and Context:

We are performing computational research aimed at designing the next generation of catalysts for natural gas conversion. We are focusing on well-defined, supported subnanometer cluster catalysts with controlled cluster size, shape, composition, and positioning of atoms/ions of specific metals (Figure 1). We study how these various factors affect adsorption properties, catalytic rates/selectivities, and stabilities. The potential number of candidate cluster catalysts is enormous, making it impractical to synthesize even a modest fraction of them, much less to characterize their structures, physical properties, and catalytic efficacy. Therefore, we guide the selection of synthesis targets computationally. An ultimate goal of the computations is high-speed (high-throughput) predictive characterization of putative cluster structures, stabilities, and catalytic competency to yield a database of potential catalysts that is unparalleled in size and chemical diversity. Powerful, newly developed quantum chemical methods have predictive accuracy for complex catalytic problems that were not amenable to reliable theoretical predictions as recently as ten years ago. Data mining for these hypothetical structures, together with strategic implementation of high-level theory for specific hypothetical and real examples, are indispensable. We wish to explore potential energy surfaces describing the detailed steps occurring in the catalytic reactions of interest. High-end computing allows us to perform quantum mechanical calculations to understand these processes. Storage of data and networking are important because a lot of information needs to be collected and shared among different computational efforts.

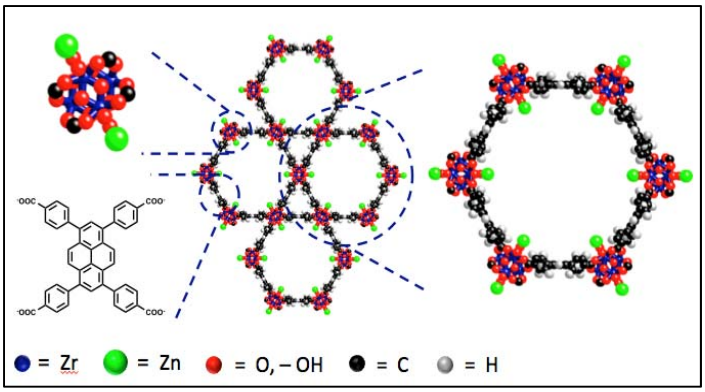


Figure 1. Robust mesoporous scaffold featuring catalytic Zn(II) sites on nanoscale zirconia-like nodes. The siting of the Zn(II) atoms can be determined via computational modeling.

References:

Yang, D., S.O. Odoh, T.C. Wang, O.K. Farha, J.T. Hupp, C.J. Cramer, L. Gagliardi, and B.C. Gates, "Metal-organic framework nodes as nearly ideal supports for molecular catalysts: NU-1000-

and UiO-66-supported iridium complexes," *J. Am. Chem. Soc.*, 137 (23), 7391–7396 (2015). DOI: 10.1021/jacs.5b02956

Carlson, R.K., G. Li Manni, A.L. Sonnenberger, D.G. Truhlar, and L. Gagliardi, "Multiconfiguration Pair-Density Functional Theory: Barrier Heights and Main Group and Transition Metal Energetics," *J. Chem. Theory Comput.* 11 (1), 82–90 (2015). DOI:10.1021/ct5008235

1.2 Research Objectives for the Next Decade:

We would like to automatize potential energy surface explorations for these complex reactions. Some of these processes will involve systems in their electronically excited states. This means that multiple potential energy surfaces will have to be explored at the same time. This will require new methods to study efficiently the ground and excited states. Various groups are currently developing these methods and the relevant software. Once the software becomes available, more powerful hardware resources will be needed with large memory capabilities for each single job and fast communication among the processors.

2. Computational and Data Strategies

2.1 Approach:

We currently use state-of-the art quantum chemistry software to study the systems described in Section 1. We moderately benefit from parallel architectures. We benefit from data storage facilities. Our calculations become particularly efficient when all the data can be stored in memory. If we want to create large data-bases of information data storage and easy accessibility becomes very important.

2.2 Codes and Algorithms:

We use commercial packages like Gaussian but also modules developed in our group for multireference calculations, which are implemented in MOLCAS and NWChem.

3. Current and Future HPC Needs

3.1 Computational Hours:

The average duration of our calculations is of the order of magnitude of a few days to one week. We expect our requirements to increase by one order of magnitude. Would it be possible to run a calculation for 10 days consecutively, with the total availability of the requested resources, both in terms of memory, disk-speed, and storage?

3.2 Parallelism:

This topic does not apply.

3.3 Memory:

At the moment, we request at most 500 Gb for one single calculation. Would it be possible to have 1 Tb for one calculation?

Case Study Title: Computational Combustion Chemistry

Lead Author: Stephen J. Klippenstein (Argonne National Laboratory)

1. Description of Research

1.1 Overview and Context: Fuel economy, emissions requirements, and environmental and natural resource pressures are demanding more efficient engines fueled by alternate fuels. Numerical simulations of internal combustion engines, which couple chemical models for conversion of the fuel into combustion products with computational fluid dynamics, are beginning to aid in the engine design process.1,2 Recent efforts to meet technological demands focus on low-temperature combustion schemes in which the chemical aspects of the simulation are particularly important. Global chemical models typically consist of thermochemical and transport properties for hundreds of species, together with rate coefficients for the thousands of reactions that describe the conversion of the fuel into oxidation products and the formation of various pollutants.

1.2 Research Objectives for the Next Decade:

The fidelity of the underlying chemical models needs to improve in order for engine simulations to be truly effective in reducing the number of expensive and time-consuming engine prototypes that need to be built. Furthermore, the development of high-fidelity chemical mechanisms for new fuels must occur more rapidly. Ultimately, what is needed is a fully automated, high-accuracy procedure for predicting the combustion chemistry of arbitrary fuels.

2. Computational and Data Strategies

2.1 Approach: The recent transformation of theoretical chemical kinetics from an empirical to a predictive science has led to its widespread utility in the determination of high-accuracy thermochemical kinetics parameters for combustion. Currently, domain scientists generally consider one reaction at a time. Over the next decade, we envision the development of an overarching simulation code that automatically predicts the chemical reaction kinetics from first principles for the full set of reactions involved in the combustion of an arbitrary new fuel. Such a development will require a major investment in workflow software that couples and improves the codes of a wide variety of domain-level scientists. The effective use of such a global combustion chemistry code will require HPC resources.

2.2 Codes and Algorithms: The accurate prediction of the kinetics for a single reaction currently requires a sequence of (1) potential energy surface exploration through electronic structure calculations; (2) chemical dynamics simulations of the microscopic dynamics of energy

[1] Reitz, R.D., 2013, "Directions in Internal Combustion Engine Research," *Comb. Flame* 160:1–8.

[2] Virtual Engine Research Institute and Fuels Initiative, http://verfi.anl.gov.

transfer, chemical reactions, and product state distributions; and (3) master equation simulations to treat the coupling of energy transfer with reaction dynamics. The codes used in this process range from highly developed commercial codes, such as MOLPRO or GAUSSIAN, to the transitory working codes of individual domain scientists. The automated development of a full chemical mechanism requires kinetic predictions for a large number of elementary reactions as directed by a mechanism generation code. Uncertainty analyses can direct the level of computational effort applied to each of the individual rate predictions. The current major shortcoming is the absence of overarching software that couples together the various domain-level codes and that deals effectively with the inevitable code failures.

3. Current and Future HPC Needs

3.1 Computational Hours: Please see Table 1.

3.2 Parallelism: Each of the various codes I use currently makes effective use of fine-grained parallelism for up to ~10 nodes. Coarse-grained parallelism is embarrassingly simple for many aspects of the calculations, such as the generalization to many reactions, the propagation of many trajectories, and the evaluation of energies for large numbers of geometries. The codes that exist generally make use of this coarse-grained parallelism, but much effort is needed to tie all aspects together in one effective program.

3.3 Memory: The high-level electronic structure calculations require large amounts of memory (e.g., ~100 GB/node). The remaining calculations generally involve only modest amounts of memory (e.g., ~1–10 GB/node). My understanding is that the development of lower memory, more well-parallelized electronic structure codes is a focus of other research efforts.

3.4 Scratch Data and I/O: Currently, our highest-level electronic structure calculations use up to 2 TB/node. This should not change much with time, but we will likely be performing such calculations on many more nodes. To be effective, it should be possible to have an I/O bandwidth that correlates with writing this amount of data in a few hours. I think devoting 5–10% of the runtime to I/O will be a reasonable limit. To some extent, the calculations can be scaled according to the machine capabilities, but rapid I/O is an integral part of current high-level electronic structure calculations.

3.5 Long-term and Shared Online Data: Currently, our long-term storage is limited to about 100 GB. I expect that would increase by about a factor of 100 for the project envisioned here.

3.6 Archival Data Storage: I do not currently have any data archived and I do not foresee a need for that.

3.7 Workflows: My current workflow mostly involves human-level couplings of the various codes described in Section 2.2. For this project to be effective, considerable resources must be devoted to workflow/software development, so that the machines are effectively used and the programs run smoothly through to completion.

3.8 Many-Core and/or GPU Readiness: I do not really know the level of readiness for GPU accelerators. I think the electronic structure codes are far from ready, but I suspect the development of such codes is already considered in other projects. I suspect that things like trajectories and statistical sampling of geometries are readily accelerated with GPUs.

4. Requirements Summary Worksheet

I have tried to fill this worksheet out the best I can, but there are a lot of qualifiers that should really be described. Most problematic is filling out things like memory, disk, and data I/O requirements, due to wide variability across the various aspects of the calculations. Furthermore, the calculations can be performed with various computational requirements, with higher accuracy requiring more resources. For simplicity, I have noted my yearly usage and presumed that a fully automated code would consider about 1,000 times more reactions than I currently consider. Whether this is feasible and happens depends a lot on whether we make a significant investment in software/workflow development.

Table 1. Computational Combustion Chemistry Code Requirements

Code: Automated Combustion Chemistry	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]			
Computational node hours (Homogeneous many-core)[b]	1×10^7	100	1,000
Computational node hours (w/GPU or accelerator)[c]			
Memory per node	100 GB	200 GB	200 GB
Aggregate memory	TB	TB	TB
Data read and written per run	2 TB/node for 10% of the nodes	2 TB/node for 10% of the nodes	2 TB/node for 10% of the nodes
Maximum I/O bandwidth needed	1 GB/sec	1 GB/sec	1 GB/sec
Percent of runtime for I/O	5%	5%	5%
Scratch file system space needed	2 TB/node for 10% of the nodes	2 TB/node for 10% of the nodes	2 TB/node for 10% of the nodes
Permanent online data storage	0.1 TB	1 TB	10 TB
Archival data storage needed	0 TB	0 TB	0 TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

Case Study Title: Time-Dependent Electronic Structure Theory and Quantum Electronic Dynamics

Lead Author: Xiaosong Li (University of Washington)

1. Description of Research

1.1 Overview and Context: My research focuses on developing time-dependent electronic structure theories for studying photochemical/photophysical properties and dynamics in large systems. New methods are developed and distributed through an open-source software development effort. The possibility of applying time-dependent, many-body theory to study quantum electronic dynamics in realistic material systems is hinged on the ability of software that can take the full advantage of the state-of-the-art computer and software architecture.

1.2 Research Objectives for the Next Decade: Because many-electron quantum dynamics are foundational to numerous advanced technologies, most notably in the areas of photonics and spintronics, the products of this research field will result in the development of novel materials with new or enhanced photophysical properties for application in a variety of scientific contexts from fundamental research to energy conversion. In order to achieve this goal, highly efficient, multidimensional data analysis and dimension-reduction techniques will be developed to resolve complex, many-electron and multi-photon processes to probe for insights into materials/ photochemical/photophysical properties.

2. Computational and Data Strategies

2.1 Approach: Time-dependent electronic structure theories explore physical processes in a high-dimensional domain including both space and time. As a result, the HPC aspect of algorithm development is intrinsically heterogeneous because the spatial degrees of freedom are parallel computing–compatible but the integration over time is simply a serial computing task. In addition, a single picosecond's worth of electronic dynamics with output including density, dipole, Kohn-Sham matrices at every time step can easily reach TB size, posing challenges to data storage and analysis.

2.2 Codes and Algorithms: The current algorithm developed in the Li group separates the time-propagation of the electronic degrees of freedom from that for nuclei via a multi-split operator approach. The software development employs a heterogeneous strategy, i.e., while the nuclear integrator is based on the minimal FLOP strategy while the electronic integrator utilizes a minimal MOP approach and is ideal for single instruction multiple data (SIMD) parallel architecture.

3. Current and Future HPC Needs

3.1 Computational Hours: Currently, a single 1-picosecond TDDFT quantum electronic dynamic simulation for a relatively realistic material system (e.g., 350 electrons and 80 nuclei) will take ~150k hours on a conventional[3] core (no accelerators). To achieve the scientific goals, the computational time will need to be reduced by a factor of 103 to 104 so that a statistically meaningful ensemble of dynamics can be computed.

3.2 Parallelism: A hybrid parallelism is implemented in our codes. Both SMP using OpenMP and MPI parallelism are deployed. Currently, we are planning to acquire a computer cluster with Intel mini-core (Knights Landing) infrastructure and will explore mini-core–based parallelism.

3.3 Memory: For real-time TDDFT electronic dynamics, the memory requirement scales as N4 where N is the number of basis functions (proportional to the number of electrons for a given basis set). It is unlikely that all integrals for a realistic material system can be stored in memory. We are currently planning on developing a "fusion" type memory strategy that can seamless integrate fast RAM and slower SSD.

3.4 Scratch Data and I/O: For a DFT-based electronic structure method, scratch data during simulation are minimal. The main I/O cost comes from the runtime output, which is estimated at approximately 5% of the total runtime. However, for time-dependent post-SCF methods, such as coupled cluster–based methods, the scratch data can be significant if on-disk integral storage is used.

3.5 Long-term and Shared Online Data: Currently, we do not use online data storage except for code distribution via Github. We plan to explore options to share processed simulation data online.

3.6 Archival Data Storage: Currently, our data have been compressed and saved in a 50-TB RAID 5 NAS server. The amount of data is likely to grow by a factor of 10 in 2020 and 2025.

3.7 Workflows: Currently, Github provides a very convenient way for our software development effort with multiple developers. Communications between developers and between developers and users were supported by Github. Software version control and distribution are also supported by Github. We currently use public domain open-source licensing.

[3] Conventional = current multicore CPUs like Intel's "Ivy Bridge."

3.8 Many-Core and/or GPU Readiness: Although our current code is not currently tuned to take advantage of "lightweight" cores and/or hardware accelerators (e.g., Knights Landing mini-cores, GPUs) with deepening memory hierarchies, we have development plans to explore the new computer infrastructures. To facilitate transition to them and tune the software to its optimal performance, we will need direct access to mid-size HPCs with the latest hardware architectures and opportunities to collaborate with software engineering from hardware vendors.

4. Requirements Summary Worksheet

Table 1 shows our projected HPC requirements.

Table 1. Chronus Quantum Requirements

Code: _Chronus Quantum_____	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	150 K per simulation	100×	500×
Computational node hours (Homogeneous many-core)[b]	10 K per simulation	100×	500×
Computational node hours (w/GPU or accelerator)[c]	0	0	0
Memory per node	32 GB	1,000 GB	2,000 GB
Aggregate memory	1 TB	10 TB	20 TB
Data read and written per run	0.5 TB	5 TB	10 TB
Maximum I/O bandwidth needed	0.5 GB/sec	3 GB/sec	6 GB/sec
Percent of runtime for I/O	5%	20%	20%
Scratch file system space needed	0.5 TB	10 TB	20 TB
Permanent online data storage	0 TB	10 TB	20 TB
Archival data storage needed	20 TB	200 TB	500 TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

Case Study Title: Coupled Reactivity of Supported Metal Nanoparticles on Reducible Oxide Catalysts

Lead Authors: Roger Rousseau and Vassiliki-Alexandra Glezakou (Pacific Northwest National Laboratory)

1. Description of Research

1.1 Overview and Context: Reducible oxides like CeO2, TiO2, and RuO2 are ubiquitously encountered in thermal-, photo-, and electrocatalysis as either the catalytic support or the actual catalyst. These oxides promote redox reactions with surface-bound adsorbates4 because of the presence of metal atoms with accessible redox states (e.g., Ce3+/Ce4+, Ti3+/Ti4+, and Ru3+/Ru4+) and intrinsic defects, such as interstitial ad-atoms or vacancies, which create mobile electrons that can travel to the surface to perform chemistry. As support materials, these oxides exhibit strong interactions with supported metal nanoparticles, which lead to unique structural and chemical properties that are distinct from other redox-inactive supports and have confounded catalysis science for decades. In the past few years, ab initio molecular dynamics (AIMD), which account for the thermal motion of nuclei and the response of electrons to this motion, have provided invaluable insights into: (i) how mobile charge carriers move through these oxides and perform surface redox chemistry; (ii) how crystal defects determine surface redox properties; (iii) how supported metal particles interact with the support oxide to synergistically catalyze reactions and determine the catalyst structure under operating conditions5 (see Figure 1). An emerging paradigm out of these studies is that the redox properties of the supported metal nanoparticles, the nature of the catalytic active sites, and the chemical make-up of the adsorbed species are coupled. The resulting nonlocality of the interactions places stringent demands on modern theory and simulations. In essence, this requires simulations of hundreds to thousands of atoms and their associated electrons combined with the appropriate statistical mechanical sampling of free energetics to properly study reactivity. This

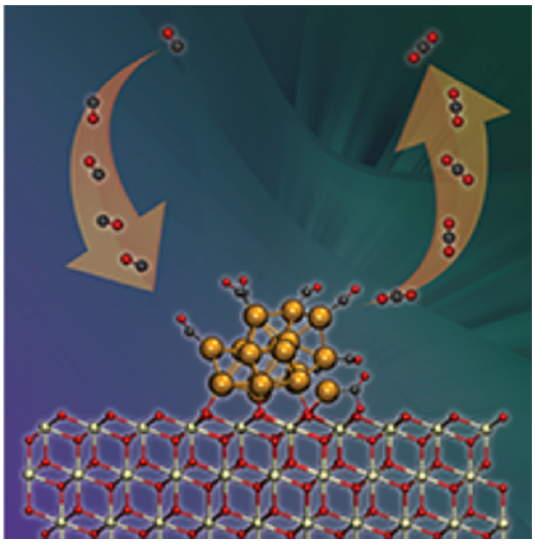


Figure 1. Dynamic formation of single Au catalytic site on CeO2 that facilitates the CO oxidation. After CO2 is formed, the Au atom returns to the cluster.

[4] Y Yoon, YG Wang, R Rousseau, VA Glezakou, *ACS Catalysis* 5 (3), 1764–1771 (2015); R Mu, DC Cantu, X Lin, VA Glezakou, Z Wang, I Lyubinetsky, R Rousseau, Z Dohnalek, *Journal of Physical Chemistry Letters* 5 (19), 3445–3450 (2014); YG Wang, D Mei, J Li, R Rousseau, *Journal of Physical Chemistry* C 117 (44), 23082–23089 (2014); Z Dohnálek, I Lybinetsky, R Rousseau, *Progress in Surface Science* 85 (5), 161–205 (2010).

[5] YG Wang, Y Yoon, VA Glezakou, J Li, R Rousseau, *Journal of the American Chemical Society* 135 (29), 10673–10683 (2013); YG Wang, D Mei, VA Glezakou, J Li, R Rousseau, *Nature Communications* 6 6411 (2015).

information is distilled into realistic micro-kinetic models, taking into account these couplings and relating the computed energetics of individual chemical reactions to the macroscopically observed global rate phenomena.

1.2 Research Objectives for the Next Decade: Over the next decade, a comprehensive understanding of the coupling between the system components needs to be evaluated to derive a more systematic understanding of these phenomena, which transcends the simulations. Key questions that need to be resolved are: (i) how supports affect multicomponent nanoparticles and the nature of the active sites; (ii) how the redox state of the support materials influences the catalyst structure's active sites; (iii) how reactions occur on the oxide support or the nanoparticle; and (iv) how the presence of a liquid state affects reactivity. The ultimate goal will be to ask if we can identify the key control parameters that allow us to predict catalyst performance a priori and design novel catalytic processes with enhanced activity, selectivity, and durability.

2. Computational and Data Strategies

2.1 Approach: Current computational methods involve large-scale ab initio molecular dynamics using a linear scaling Density Functional Theory to account for system models on the order of ~103 atoms at finite temperature. This information allows for the assessment of a single chemical reaction at a time and its dependence on the support's redox state. The models of reactivity are then used to construct rate equations that describe the kinetics of chemical reactivity for an entire catalytic process, including side reactions and catalyst deactivation. As the understanding of these coupling phenomena grows, it is expected that system size will need to be enhanced by at least 1 order of magnitude and that sampling of configuration space will need to grow by 2–3 orders of magnitude to capture currently observed phenomena.

2.2 Codes and Algorithms: Computations employ CP2K (or equivalent software) for large-scale AIMD using a linear-scaling Density Functional Theory, which exploits multiple levels of parallelization to perform efficient electronic structure calculations (strongly coupled parallelization) and sampling based on multiple system replicas (loosely coupled parallelization). Data sets generated from these trajectories are relatively small (in tens of GB), and the code is designed to minimize I/O. Kinetic models are at least an order of magnitude simpler to run (often requiring only a single to few tens of processors) and are run separately, with the HPC challenge residing in obtaining reaction-free energetics.

3. Current and Future HPC Needs

3.1 Computational Hours: See response in Table 1.

3.2 Parallelism: See response in Table 1.

3.3 Memory: See response in Table 1.

3.4 Scratch Data and I/O: See response in Table 1.

3.5 Long-term and Shared Online Data: See response in Table 1.

3.6 Archival Data Storage: See response in Table 1.

3.7 Workflows: See response in Table 1.

3.8 Many-Core and/or GPU Readiness: See response in Table 1.

4. Requirements Summary Worksheet

Table 1. CP2K Requirements

Code: CP2K	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	250,000[e]	10×	100×
Computational node hours (Homogeneous many-core)[b]	--	--	--
Computational node hours (w/GPU or accelerator)[c]	--	--	--
Memory per node	4 GB	10 GB	20 GB
Aggregate memory[f]	?? TB	?? TB	??TB
Data read and written per run	0.02 TB	0.2 TB	2 TB
Maximum I/O bandwidth needed	10–14 GB/sec	25 GB/sec	50 GB/sec
Percent of runtime for I/O			
Scratch file system space needed	0 TB	0 TB	0 TB
Permanent online data storage[g]	1 TB	5 TB	10 TB
Archival data storage needed[g]	0.1 TB	0.5 TB	1 TB

[a] "Core hours" are used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.
[b] "Node hours" are used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.
[c] "Node hours" are used for "GPU or accelerator" usage.
[d] For example, 32 × column 1.
[e] This number is based on computations with ~500 atoms, ~2,300 electrons, and ~5,800 atomic bases for one AIMD trajectory on an X86 64-core machine. Typically, we need 10 to 100 trajectories to complete a project.
[f] Current mixed and shared memory algorithms are under development.
[g] Estimate per project requiring ~ 10–50 trajectories. For future demands, we assume compressed data form.

Case Study Title: Electronic Structure and Dynamics of Photochemical Processes in Condensed Phases

Lead Author: Lyudmila V. Slipchenko (Purdue University)

1. Description of Research

1.1 Overview and Context:

The focus of my research program is on the development of theoretical and computational approaches targeting the electronic structure of extended systems, such as photosynthetic and fluorescent proteins, molecular solids, polymers, and bulk liquids. A special focus in my research is on the development of polarizable QM/MM methods, fragmentation techniques, and vibronic models. We use the developed techniques to investigate fundamental aspects of non-covalent interactions and the effect of the environment on electronic structure and dynamics. We maintain collaborations with several experimental groups that provide us with motivation and exposure to real-life chemistry and biophysics problems.

High-end computing, storage, and networking are essential for achieving our research goals. Complexity of considerand the level of detail attained today are beyond the reach of computational modelers years ago. For example, ten years ago, a good publication would discuss simulations of the excited states of a model chromophore in gas phase, often at a single geometry. Now, we routinely perform simulations of photoactive proteins, i.e., including the chromophore, polypeptide, solvent, and configurational sampling. Obviously, such tasks became practical due to development of new algorithms and approaches, as well as powerful computational resources, in terms of high-end computing, data storage, and networking.

1.2 Research Objectives for the Next Decade:

The goal for the next decade is to perform *in silico* modeling of systems and phenomena that can be directly observed experimentally, rather than simulating simplified model systems. This involves employing model Hamiltonians of higher complexity and extending simulation times and length scales. Thus, we will be simulating more complex phenomena in larger systems for a longer time. Development of parallelization algorithms is crucial toward this end. Additionally, development of workflows that allow automation of repetitive steps of simulation protocols and unification of data formats will be necessary. Algorithms for data mining and analysis will be of high demand as well.

2. Computational and Data Strategies

2.1 Approach:

One of the problems we face currently is a different data format between different packages. For example, our typical sequence of calculations involves molecular dynamics simulations with classical force fields (PDB format) and then hybrid QM/MM simulations (XYZ format or similar). This often requires time-consuming reformatting the data. Currently, we use a set of home-made scripts to convert data into different formats. In the future, we plan to unify the scripts and combine them into generic automated workflows.

2.2 Codes and Algorithms:

We use a combination of classical MD (GROMACS, CHARMM) and QM/MM and fragmentation methods (Q-Chem, GAMESS). Our QM/MM and fragmentation methods involve high-level correlation, which constitutes the majority of our computational needs. The correlation methods are memory and hard-drive intensive, and their parallelization is challenging. Recent developments in tensor libraries allow efficient fine-grain parallelization of correlation methods, while coarse-grain parallelization and use of accelerators is a task that should be solved in the next decade. Generally, fragmentation methods are naturally scalable, but a challenge remains, which is to achieve efficient load balancing.

3. Current and Future HPC Needs

3.1 Computational Hours:

See Table 1.

3.2 Parallelism:

Most codes we use in Q-Chem have very good fine-grained (on-node) parallelization. Some work is done (by other developers) to extend parallelization to GPUs and coarse-grained parallelism.

3.3 Memory:

Our electronic structure calculations are memory-extensive. Currently, we use nodes with 64 GB of memory, which is barely sufficient (these are single-node calculations). Hopefully, multi-node parallelization will be in place when low-memory architectures become dominant.

3.4 Scratch Data and I/O:

Our calculations can be disk-intensive, requiring large and fast scratch disks (3 TB or larger per run currently). We do not use much online scratch space today, but this might change in the future.

3.5 Long-term and Shared Online Data:

We have only a small need for online long-term storage right now. However, we anticipate more collaborative projects in the future where online long-term storage might be important.

3.6 Archival Data Storage:

None currently; not sure about future needs.

3.7 Workflows:

Very few workflows are used today, but this is the area that will certainly change in the future. We expect to develop workflows involving multiple (>100) calculations, including calculations of various types.

3.8 Many-Core and/or GPU Readiness:

The codes we use now are not ready for new architecture types. On the other hand, some codes have recently been rewritten (e.g., new tensor libraries, new integral codes) such that extensions toward accelerators and memory hierarchy might be possible. However, I currently do not have either human resources or funding to work on extending the codes to new architectures. Additional funding opportunities to adjust the codes to new architectures as well as workshops/summer schools/internships that train computational chemists in new types of programming would be very helpful.

4. Requirements Summary Worksheet

Table 1. Q-Chem Requirements

Code: Q-Chem	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	3,000,000	10×	100×
Computational node hours (Homogeneous many-core)[b]	150,000	10×	100×
Computational node hours (w/GPU or accelerator)[c]	None	?	?
Memory per node	64 GB	128 GB	?
Aggregate memory	–	–	–
Data read and written per run	3 TB	10 TB	30 TB
Maximum I/O bandwidth needed	72 GB/sec	?	?
Percent of runtime for I/O	10–90%	?	?
Scratch file system space needed	10 TB	50 TB	200 TB
Permanent online data storage	< 1 TB	3 TB	10 TB
Archival data storage needed	10 TB	50 TB	200 TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.
[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.
[c] "Node hours" for "GPU or accelerator" usage.
[d] For example, 32 × column 1.

D.3 Case Studies Addressing Materials and Chemical Discovery

Case Study Title: First-Principles Simulations of Functional Materials for Energy Conversion: Transport and Dynamics

Lead Authors: Giulia Galli (University of Chicago/Argonne National Laboratory) and Francois Gygi (University of California-Davis)

1. Description of Research

1.1 Overview and Context: This case study focuses on the properties of nanostructured materials for use in energy conversion processes, with emphasis on systems exhibiting complex structures on multiple length scales, inclusive of interfaces between nano- and meso-building blocks. The goal is to predict the dynamical and transport properties of these systems, by using a theoretical framework that combines classical and ab initio molecular dynamics with accurate electronic structure methods beyond Density Functional Theory. Molecular simulations are employed to compute ensemble averages of thermodynamic and transport properties, as well as the assembly and synthesis routes of heterogeneous systems. The same trajectories subsequently serve as input to Many Body Perturbation Theory calculations to compute electronic and transport properties. Scalable, integrated first-principles algorithms combined with the unique computational resources provided by exascale platforms will enable studies of unprecedented scope.

1.2 Research Objectives for the Next Decade: In order to accelerate the discovery of innovative functional materials, it is not sufficient to compute the properties of the end product; rather, it is critical to simulate and validate the assembly processes that occur during synthesis and fabrication. In addition, in order to design materials relevant to energy technologies, it is essential that the basic mass, charge, and energy transport phenomena involved in energy storage and conversion processes be understood. Most of these phenomena (e.g., electron transport) are inherently quantum mechanical and require a first-principles treatment. Others, such as ionic transport, occur at the molecular scale. It is therefore necessary that electronic- structure methods be coupled to appropriate dynamical descriptions of matter, thereby providing the means to capture all the relevant length and timescales of importance to a material's performance. Acquiring the ability to predict the transport and dynamical properties of heterogeneous materials across multiple length scales represents a **major research goal in materials science for the next decade.**

2. Computational and Data Strategies

2.1 Approach: The calculation of the electronic and transport properties of complex systems from first principles will be based on density-functional and many-body perturbation theory using the Qbox and WEST codes; transport coefficients obtained from first principles may serve as inputs to continuum particle codes that will predict the effect of applied fields on a material's structure and performance. Within a client server strategy, quantum (Qbox) and classical Molecular Dynamics (MD) and Monte Carlo codes (e.g., LAMMPS and HOOMD-blue) will be coupled and enhanced through a suite of advanced generalized-ensemble sampling techniques, which will in turn operate in tandem with continuum codes. In the next 4 to 5 years, this approach will enable simulations of assembly processes of nano- or meso-building blocks of arbitrary shapes with designer electronic properties. Importantly, through a marriage of forefront quantum MD, advanced sampling, and particle-continuum coupling, one will enable ab initio-based calculations of the free energy of complex materials (and its derivatives with respect to field variables), both at equilibrium and far from equilibrium.

2.2 Codes and Algorithms: Major first-principles codes to be employed in this project are Qbox: (http://qboxcode.org) and WEST (http://www.west-code.org/codes). These are codes for first-principles molecular dynamics simulations and many-body perturbation theory calculations, respectively, on which advanced sampling capabilities as well as transport property calculations will be built.

3. Current and Future HPC Needs

3.1 Computational Hours:

The goals stated in Section 1.2 may require on the order of ten billion core-hours per year.

3.2 Parallelism:

The codes currently in use can run efficiently on BG/Q on up to 128k cores (Qbox) and 512k cores (WEST). There is no current deployed implementation that uses accelerators. The current programming model is MPI+OpenMP.

3.3 Memory:

Based on current implementations, 8 GB/core can be sufficient. However, simplicity of the architecture is more important than the amount of memory itself. For example, 6 GB on a GPU

that is only accessible via a PCIe interface is difficult to use and requires considerable rewriting of the applications.

3.4 Scratch Data and I/O:

Checkpoint restart files are expected to reach the range of 1 TB in the coming 5 years. Each user of a project will routinely store 20–50 temporary restart files on scratch storage.

3.5 Long-term and Shared Online Data:

This topic does not apply.

3.6 Archival Data Storage:

This topic does not apply.

3.7 Workflows:

This topic does not apply.

3.8 Many-Core and/or GPU Readiness:

A prototype of a GPU-compatible version of Qbox has been developed, although the resulting performance indicates that a full rewrite of the application would be needed to obtain acceptable performance on even a few hundred nodes. Qbox development currently focuses on the MIC architecture.

Case Study Title: Toward the Predictive Science of Synthesis

Lead Authors: Christopher J. Mundy and Gregory K. Schenter
(Pacific Northwest National Laboratory)

1. Description of Research

1.1 Overview and Context:

The predictive science of synthesis will require both (i) the development of a fundamental comprehension of the driving forces, processes, and phenomena such as solvation, nucleation, assembly, transport, and reactions in complex condensed-phase, heterogeneous and interfacial molecular environments, and (ii) the development of theoretical and computational methods required to accelerate scientific advances in condensed-phase and interfacial molecular science.

1.2 Research Objectives for the Next Decade:

The future will require connecting theoretical and computational frameworks. No single framework will be sophisticated enough to solve problems in catalysis and synthesis. The future will be all about understanding how complexity (e.g. environment, interfaces) influences phenomena and how frameworks must be extended, modified, and redeveloped in response to complexity. Exporting information from molecular simulations to advance continuum frameworks like Born Theory, Guoy Chapman, and Derjaguin-Landau-Verwey-Overbeek (DLVO) will be a research objective.

Key examples will be to elucidate the solid-electrolyte interface to (i) understand how interfaces modify reactivity and drive new ideas in the area of catalysis, and (ii) understand how interfaces between phases couple to the solution and drive assembly and processing of new materials.

2. Computational and Data Strategies

2.1 Approach:

The approach is to advance the art of molecular simulation. We do this by combining molecular electronic structure with statistical mechanical sampling. We still struggle with the balance between efficiency and accuracy in (i) the solution of the many-body Schrodinger equation (for

electronic structure), and (ii) the effective sampling of molecular configurations to accurately understand (predict, control) emergent collective phenomena.

2.2 Codes and Algorithms:

We utilize CP2K (**www.cp2k.org**), which consists of state-of-the-art algorithms for ab initio sampling (MC and MD) that scale for large interfacial systems (>2,000 atoms) utilizing advanced statistical mechanical sampling protocols.

3. Current and Future HPC Needs

3.1 Computational Hours:

Today we are using 10 to 20 M (see Table 1). We anticipate using 10 to 100 times more core-hours in the future. GPUs are used for fast matrix multiplies for preconditioners for the SCF cycles.

3.2 Parallelism:

Code can make efficient use of MPI and OpenMP directives.

3.3 Memory:

Memory that is currently used in the CRAY XC40 is sufficient for our purposes. Having the same relative memory/core will be sufficient for future needs.

3.4 Scratch Data and I/O:

This is not currently a limitation for systems of interest.

3.5 Long-term and Shared Online Data: 10–100 TB (see Table 1).

3.6 Archival Data Storage: 10–100 TB (see Table 1).

3.7 Workflows:

Running long trajectories (approaching nanoseconds) using ab initio dynamics. Workflow will be largely ensemble-based computing and long, sustained data collection. The per run requirements will not be large, but the aggregation and accumulation of trajectories and statistics can potentially generate 10 to 100 times what is used today.

3.8 Many-Core and/or GPU Readiness:

Currently, CP2K is GPU ready. Examples include matrix multiplies using GPU libraries.

3.9 Software Applications, Libraries, and Tools:

This is not currently a limitation for systems of interest.

3.10 HPC Services:

This is not currently a limitation for systems of interest.

3.11 Additional Needs:

Additional needs include queuing and usage policies that allow for long, sustained sampling using ab initio molecular dynamics.

4. Requirements Summary Worksheet

Table 1. Predictive Science of Synthesis Code Requirements

Code: CP2K	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	10–20M	10	100
Computational node hours (Homogeneous many-core)[b]	20 M	10	100
Computational node hours (w/GPU or accelerator)[c]	30 M	10	100
Memory per node	*16 GB*	*2*	*5*
Aggregate memory	*TB*	*TB*	*TB*
Data read and written per run	0.1 TB	2	5
Maximum I/O bandwidth needed			
Percent of runtime for I/O	N/A	N/A	N/A
Scratch file system space needed	1 TB	5	10
Permanent online data storage	10 TB	10	50
Archival data storage needed	10 TB	10	50

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

Case Study Title: Nanoporous Materials Genome

Lead Author: J. Ilja Siepmann

1. Description of Research

1.1 Overview and Context:

1.2 Research Objectives for the Next Decade:

The scientific motivation for our research is to use **predictive modeling** to provide accurate thermophysical properties and molecular-level understanding that aids in the design of improved chemical separation processes and materials. Most chemical separations currently rely on highly energy-intensive processes (e.g., distillation), whereas improved processes involving lower energy consumption and less harmful solvents are essential ingredients for the path toward sustainability. Predictive modeling is most beneficial for high-throughput screening, for experimentally challenging conditions (e.g., high temperature, high pressure, or toxicity of compounds), and when molecular-level insight is needed to understand separation mechanisms. To this extent, we develop computational tools that enable predictive Monte Carlo and molecular dynamics simulations where the interactions are described by force fields or Kohn-Sham DFT, depending on chemical complexity and length/time scales. Over the next decade, researchers will be able to use predictive modeling to tackle high-throughput screening — even when the sorption or extraction process leads to structural changes of the host material — and to address transport properties in micro-structured environments. Advances in algorithms, models, and a high-end computing infrastructure are pivotal to increasing accuracy, reducing uncertainties, and modeling chemical processes and materials under realistic operating conditions.

2. Computational and Data Strategies

2.1 Approach:

2.2 Codes and Algorithms:

Our predictive modeling of chemical separation processes and materials relies on a **hierarchical strategy** driven by the enormous chemical/materials space and the wide range of operating conditions. On the one hand, the initial screening for a sorption-based separation process may involve computation of the Henry's constant for every compound in the mixture (say, about 10) in about 10^6 porous materials. The computational cost for each of these Monte Carlo calculations ranges from minutes (small molecules, such as carbon dioxide, in rigid host material) to hours (complex hydrocarbon in flexible host material). Here, design of efficient workflows that can manage $>10^6$ processes (each running on only a few cores) and their I/O is

paramount. On the other hand, a high-fidelity prediction may involve first-principles modeling of reaction equilibria under confinement in a complex host material where the computational cost for each state point may exceed 10^7 hours and involves first-principles codes with good, strong scaling up to 10^5 cores.

3. Current and Future HPC Needs

3.1 Computational Hours:

At present, the annual allocation on *Mira* is 120,000,000 core hours with about two-thirds of the allocation devoted to first principles simulations with **CP2K** and one-third to force-field-based simulations with **MCCCS**. We expect the computational requirements to increase by about one order of magnitude every five years with the increase driven by increasing chemical system complexity, improving accuracy, and reducing uncertainty. Code and algorithm development will allow the computational load to shift from conventional cores, over homogeneous multi-core architectures, to GPU/accelerator-enabled architectures. **CP2K** can already utilize the combination of CPU/GPU on *Titan*.

3.2 Parallelism:

Both **CP2K** and **MCCCS** already utilize a combination of coarse-grained and fine-grained parallelism, but **MCCCS** is currently not able to utilize GPU/accelerator-based architectures. Thus, **MCCCS** will need to be adapted to utilize GPU/accelerator-based architecture where the GPU/accelerator handles numerous independent tasks, which is one of the advantages of Monte Carlo simulations.

3.3 Memory:

The relatively modest memory requirements for force field-based simulations will allow **MCCCS** to be tailored to utilize a complex memory hierarchy with on-chip fast memory. For many applications, 16 or 32 MB of on-chip fast memory would be desirable to greatly reduce the need to access off-chip memory. The minimum shared memory pool for **MCCCS** would be 1 GB for larger systems. For **CP2K**, the aggregate memory and time for access are often the major limitations.

3.4 Scratch Data and I/O:

The large numbers of calculations coupled together by an **MCCCS** workflow lead to relatively large I/O requirements that increase linearly with the number of calculations.

3.5 Long-term and Shared Online Data:

Only metadata accessible through the Nanoporous Explorer (currently hosted by the Materials Project) needs online long-term storage. The amount of data is relatively small.

3.6 Archival Data Storage:

For both **MCCCS-** and **CP2K**-based data, it will be possible to use mostly archival data storage for restart, trajectory, and external energy grid files. A delay in moving these data to scratch space will not hamper these projects.

3.7 Workflows:

At present, both **MCCCS** and **CP2K** use a single-level workflow where multiple simulations are bundled to generate capability type jobs. The current workflow for **MCCCS** can handle 2^{16} concurrent simulations and involves managing I/O utilization. For 2020 and 2025, we envision workflows that allow for multiple levels of simulations with on-the-fly decision making.

3.8 Many-Core and/or GPU Readiness:

Neither CP2K nor MCCCS are ready for deepening memory hierarchies. Successfully transitioning to such architectures will require dedicated one-on-one help from computational scientists familiar with such architectures.

4. Requirements Summary Worksheet

Tables 1 and 2 show our projected HPC requirements.

Table 1. Requirements with CP2K

Code: _____**CP2K**_____	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 2 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	80,000,000	2×	4×
Computational node hours (Homogeneous many-core)[b]	500,000	4×	16×
Computational node hours (w/GPU or accelerator)[c]	100,000	20×	40×
Memory per node	16 GB	16 GB	16 GB
Aggregate memory	4 TB	8 TB	16 TB
Data read and written per run	0.1 TB	0.2 TB	0.4 TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O	2	2	2
Scratch file system space needed	8 TB	16 TB	32 TB
Permanent online data storage	0 TB	0 TB	0 TB
Archival data storage needed	32 TB	64 TB	128 TB

[a] "Core hours" are used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" are used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" are used for "GPU or accelerator" usage.

[d] For example, 32× column 1.

Table 2. Requirements with MCCCS

Code: _____**MCCCS**_____	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 2 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	40,000,000	2×	4×
Computational node hours (Homogeneous many-core)[b]	50,000	20×	40×
Computational node hours (w/GPU or accelerator)[c]	5,000	200×	400×
Memory per node	16 GB	1 GB	1 GB
Aggregate memory	1 TB	2 TB	4 TB
Data read and written per run	1 TB	4 TB	16 TB
Maximum I/O bandwidth needed	1 GB/sec	4 GB/sec	16 GB/sec
Percent of runtime for I/O	10	10	10
Scratch file system space needed	16 TB	32 TB	64 TB
Permanent online data storage	1 TB	2 TB	4 TB
Archival data storage needed	128 TB	256 TB	512 TB

[a] "Core hours" are used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" are used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" are used for "GPU or accelerator" usage.

[d] For example, 32× column 1.

Case Study Title: Nanoparticle Systems

Lead Author: Mark Stevens (Sandia National Laboratories)

1. Description of Research

1.1 Overview and Context:

Nanoparticle systems are especially intriguing because of the ability to create functional materials with properties that presently do not exist. Nature has already shown us that nanoparticles such as protein can "master energy and information on the nanoscale."[7] We thus see a path using synthetic nanoparticles to develop materials rivaling living systems. By combining nanoparticles with different intrinsic properties in an active matrix, composite systems yielding new materials that are beyond the simple combination of the parts can be created. For simulations to provide computational tools to guide the synthesis, assembly, and integration of nanomaterials with desired properties, they must be able to treat systems with a large number of nanoparticles. They also must include the ligand coating that is an essential part of the properties and the surrounding medium, which strongly influences their behavior as well as providing addition material properties (e.g., mechanical). These are very large, complex systems that require high-performance computing to follow their assembly and determine their properties. In addition, the rich parameter space that characterizes these systems is an extraordinarily large parameter space to search, for which computational work is ideally situated to offer a way to tune and control the system properties.

1.2 Research Objectives for the Next Decade:

The goal is to understand how the nanoparticle constituents interact to form new types of materials. The interactions among the three main parts (nanoparticle core, ligand coating, and surrounding medium) determine the overall structure and dynamics of the system. From that, we need to develop the means to design materials based on the choices of these components. The computational goal is to treat large systems with a variety of interaction types and long relaxation times. Efforts in force-field development and efficient algorithms for calculating complex terms will be needed. Treating the slow relaxation will require development of novel sampling methods and further development of coarse-graining methods. For such large systems, post-analysis will have to be done using parallel computers. The analysis codes will have to be efficiently parallelized. These systems present a visualization challenge as they are too large to view on a single screen. Novel ideas for visualization will have to be developed.

---

[7] *Challenges at the Frontiers of Matter and Energy: Transformative Opportunities for Discovery Science*, John Hemminger et al. (2015).

2. Computational and Data Strategies

2.1 Approach:

The biggest challenge in these simulations is the need for extremely long simulation times. Because the main cost in MD is evaluation of the forces, speeding up the force calculation is paramount. Already, the LAMMPS MD code has implemented some routines to use coprocessors to speed this up. Ultimately, in order to handle memory efficiently, the main loop of the code must be rewritten, as well as implementing all of the force-fields and other related subroutines (e.g., constraints) that are part of the force loop. To date, data analysis has been inexpensive, but we are fast approaching a transition where data analysis (i.e. post-simulation run) will also require parallel computation. This will be a big challenge because there are so many distinct calculations that are also typically specific to the problem. Some of these calculations will become routines in the MD code (and potentially slow it up), but many others will be separate codes that have to be written specifically for the project at hand.

2.2 Codes and Algorithms

LAMMPS is a massively parallel molecular dynamics code that treats a variety of materials (e.g., metals, polymers, semiconductors). Parallelism is done based on domain decomposition. The basic workflow is to first build the system. For nanoparticle systems, separate codes are presently written to create the system. Presently, after the LAMMPS MD run is performed, analysis is done primarily using workstations. The analysis involves calculations of pair correlation functions, static and dynamics structure factors, viscoelastic response, and visualization.

3. Current and Future HPC Needs

3.1 Computational Hours:

The numbers provided in Table 1 are for a single simulation. A project typically would involve about 100 such simulations. For a leading edge project, this number may be reduced to as low as 10.

3.2 Parallelism:

LAMMPS already can already use NVIDIA coprocessors to speed up some of the force-field calculations. However, I do not have timings for the nanoparticle system using such a computer. Based on other cases, there should be a speed up of 3×. Further development is ongoing. In particular, the KOKKOS package is being developed with the intent that LAMMPS programmers

will not have to write for NVIDIA or Intel phi, but just write in “kokkos” and the KOKKOS library will take care of the different coprocessors.

3.3 Memory:

MD codes typically use small amounts of the available memory. However, implementations on NVIDIA coprocessors run best with as little communication to CPU or node memory as possible. Thus, putting as much data on the NVIDIA coprocessor as possible speeds up performance considerably. For parallel simulations using domain decomposition, the memory of the coprocessor then influences the size of the problem to be treated.

3.4 Scratch Data and I/O:

See Table 1.

3.5 Long-term and Shared Online Data:

Presently, we bring most data back to our local computers and do analysis here, but for such large systems, this will no longer be feasible unless bandwidth increases commensurately. In addition, the large systems will require parallel analysis codes. Much analysis will have to be performed at the computing facilities. Some of the analysis will be incorporated in the LAMMPS code. The rest will be in separate codes.

3.6 Archival Data Storage:

See Table 1.

3.7 Workflows:

See Section 2.2.

3.8 Many-Core and/or GPU Readiness:

See Section 3.2.

4. Requirements Summary Worksheet

Presumably the numbers for 2020 and 2025 will not scale by 32 each time as I have estimated. Usage of GPUs should lower use of the node hours by at least 3× and hopefully by 10×, although we will most likely use that speed up to perform different simulations (longer and/or more expensive force-fields). The amount of data dumped will probably be lower, as well, as more judicious choices of needed data are made.

Table 1. LAMMPS MD Code Requirements

Code: __LAMMPS_____	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	864,000	27,648,000	884,736,000
Computational node hours (Homogeneous many-core)[b]			
Computational node hours (w/GPU or accelerator)[c]			
Memory per node	0.05 GB	0.05 GB[e]	0.05 GB
Aggregate memory	0.50 TB	16 TB	512 TB
Data read and written per run	5 TB	160 TB	5,120 TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O	1%		
Scratch file system space needed	50 TB	1,600 TB	51,200 TB
Permanent online data storage	100 TB	3,200 TB	100,000 TB
Archival data storage needed	500 TB	16,000 TB	500,000 TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

[e] Maybe a better guess (and it is a guess) is that the 2020 node will be equivalent to X 2016 nodes and the memory per node will be X*0.05 GB. For example, the number X = the number of CUDA cores on an NVIDIA coprocessor. For TITAN, X = 64 or 192, depending on double or single precision, that is, 3–10 GB.

D.5 Case Studies Addressing Advances in Algorithms for Quantum Systems

Case Study Title: Computational Chemistry

Lead Author: Karol Kowalski (Pacific Northwest National Laboratory)

1. Description of Research

1.1 Overview and Context:

This project focuses on the development of parallel implementations of electronic structure methods to model various chemical transformations. High-end computing plays a crucial role in (1) reducing time-to-solution, (2) addressing steep numerical scaling of ab initio methodologies, and (3) pushing the systems-size limit tractable by ab initio methodologies (more complex processes can be tackled). Larger computational resources will also enable us to achieve unprecedented level of accuracy in molecular simulations.

1.2 Research Objectives for the Next Decade:

The most pressing scientific challenges/goals in the area of computational chemistry include the following topics: the rational design of molecular complexes and materials for energy conversion and storage, low-pressure/temperature catalysis, biomass degradation, and understanding metabolic processes. These challenges can be addresses by the propagation of a molecular level of understanding across spatial and temporal scales. The main computational goal is to develop parallel algorithms capable of taking advantage of exascale architectures.

2. Computational and Data Strategies

2.1 Approach:

A significant re-write of the existing codes will be required to take advantage of exascale resources to address current bottlenecks including: (1) efficient concurrency management to maximize the parallel performance, (2) auto-tuning and library-oriented design of key computational kernels, (3) network topology-aware execution for networks with hierarchical topologies, (4) efficient tools for the efficient handling of deeper memory hierarchies, and (5) dramatic loss of resilience at exascale. In particular, a significant increase in intra-node parallelism, data localization and reduction in intra-node communication, efficient utilization of runtime systems, development of topology-aware algorithms, and utilization of deeper memory

hierarchies will play crucial roles in achieving this goal. Current parallelization algorithms are based on dynamic load balancing and one-sided communication. The key strategy to address these issues is the development and use of high-level DSLs (such as the Tensor Contraction Engine) to separate the mathematical and physical models from the actual implementation on the hardware; use and development of low-level libraries that provide abstractions to choose between architecture-specific optimization choices (alignments, vectorization, low-level parallelism, etc.) and various tensor distribution and layout strategies; development of a more adaptable and flexible framework using modern languages and standards that use existing and future runtime environments (such as the asynchronous, task-based runtime environments that are under current development); and development of components, libraries, and APIs that enable easy, open-access development of new capability, multiple levels of parallelism (including multi-and many-core and heterogeneous architectures), and code transformations to new hardware as it becomes available.

2.2 Codes and Algorithms:

NWChem is an open-source, computational chemistry software code to provide users with the massively parallel and scalable computational chemistry software necessary to tackle complex questions in molecular sciences. NWChem is actively developed by a consortium of developers and maintained by the Environmental Molecular Sciences Laboratory (funded by DOE BER) located at Pacific Northwest National Laboratory. Current parallel implementations of many-body methods are based on the Global Array and MPI programming models. NWChem is composed of several modules which are defined by different design principles, computational requirements, and numerical overhead. For example, while the performance of the plane-wave DFT algorithm is mostly defined by network latency, the high-accuracy methods are generally more computationally intensive. In particular, the development strategy will be geared toward enabling computational and theoretical frameworks for connecting interaction-driven representations of many-body phenomena at various scales. This goal can only be met by assuring the proper choice of physical models that provide an environment for the proper inclusion of correlation effects, interoperability between electronic structure methods, and effective algorithms for reducing the numerical overhead of first-principle methods:

(1) State-of-the-art formulations of electronic structure methods:
    a. Multi-reference perturbative, configuration interaction, coupled-cluster methods (MRMBPT/MRCI/MRCC).
    b. FCI formulations: Density matrix renormalization group (DMRG) and Monte Carlo FCI formulations (MCFCI).
    c. Formulations utilizing inherent sparsity of quantum-mechanical formulations: local CC methods and CC formulations utilizing tensor decomposition techniques.
(2) Excited-state and linear-response methods:
    a. Equation-of-motion coupled cluster (EOMCC) methods.

b. Linear-response and Green function coupled-cluster methods for: (i) integrating various representations of many-body theory and (ii) transport phenomena.
c. Analytical gradients and dynamics with high-accuracy methods.

(3) Multi-scale/multi-physics methods:
a. Embedding methodologies: density embedding and DMFT formalisms.
b. Adaptive formulations of embedding methods: controlling the accuracy control by adaptively invoking the appropriate level of the wave function/Green function/DFT formalisms.

Achieving these goals will require close collaboration between domain scientists from various fields and will leverage existing strengths of NWChem in the areas of computational chemistry, high-performance computing (HPC), and applied mathematics.

3. Current and Future HPC Needs

3.1 Computational Hours:

Depending on the size of the problem, production runs range in scale from desktops to leadership-class computing facilities. In large-scale user facilities, runs can vary in scale from a single node to several hundred thousand cores. For example, a large 200,000-core run took around 30 minutes. To address new challenges, computational requirements may increase 10- to 100-fold or more (see Table 1).

3.2 Parallelism:

In large part, NWChem is based on dynamic load balancing where data and computations are distributed over the network. Several numerically intensive parts of NWChem can take advantage of Nvidia GPU/Intel MIC accelerators using CUDA/OpenMP programming models. A part of the current and future effort is geared toward the development of a library-oriented design of tensor contractions.

3.3 Memory:

The memory requirements of electronic structure methods are largely determined by the storage of 2-electron integrals. The storage requirements for these integrals are proportional to N^4, where N stands for the basis set size. For systems of interest, the memory requirements may vary between 100–1,000 TB (or more). When direct algorithms (integrals are recomputed in each iteration cycle) or integral-compressing techniques are used, the memory requirements are significantly reduced.

3.4 Scratch Data and I/O:

As in Section 3.3, 15–500 TB would be required in order to provide restart capabilities.

3.5 Long-term and Shared Online Data:

This topic does not apply.

3.6 Archival Data Storage:

This topic does not apply.

3.7 Workflows:

Current workflows consist of a relatively small number of applications that require tera-/petascale resources. Future workflows will engage a large number of high-end peta- and exascale applications.

3.8 Many-Core and/or GPU Readiness:

We have had considerable success with achieving good performance across multiple supercomputing platforms. This is evidenced by the fact that NWChem had the first quantum many-body codes that scaled to full machine scales and utilized NVIDIA GPUs and Intel MICs. However, we expect additional effort for the upcoming supercomputing platforms due to deeper memory hierarchies, continued increase in compute-to-communication and compute-to-memory ratios, and changes in the nature of compute resources (heterogeneity and wider SIMD units).

4. Requirements Summary Worksheet

Table 1. NWChem Code Requirements

Code: NWChem	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	20K–100K	10–100×	100–500×
Computational node hours (Homogeneous many-core)[b]	20–100 K	10–100×	100–500×
Computational node hours (w/GPU or accelerator)[c]	40K	10–100×	100–500×
Memory per node	128 GB	128–256 GB	128–256 GB
Aggregate memory	10–20 TB	100 TB	500 TB
Data read and written per run	TB	TB	TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O			
Scratch file system space needed	TB	TB	TB
Permanent online data storage	TB	TB	TB
Archival data storage needed	TB	TB	TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32× column 1.

Case Study Title: Next–Generation, Quantum-based MD

Lead Author: A.M.N. Niklasson (Los Alamos National Laboratory)

1. Description of Research

1.1 Overview and Context: Quantum-based Born-Oppenheimer molecular dynamics (QMD) simulations provide a general, powerful, but computationally expensive multidisciplinary tool to understand and design materials directly from the fundamental principles of quantum physics. Merging QMD with exascale computing therefore holds the promise of a paradigm shift in materials science, chemistry, and biology. Unfortunately, this opportunity requires a radical redesign of current approaches to QMD.

1.2 Research Objectives for the Next Decade: Our goal is to bypass the major obstacles to large-scale QMD through the development of a new, truly transformative high-speed computational framework. Recognizing that it is not possible to achieve this goal by a single invention, we have to reformulate the underlying theory to generate new equations that can be solved with new algorithms. These algorithms can be implemented with new data structures and run with high performance on new and emerging computer architectures.

2. Computational and Data Strategies

2.1 Approach: Linear scaling O(N) electronic structure theory provides a solution to the complexity problem of QMD. However, the promise of current O(N) theory has never been fully realized because of serious shortcomings. In a multidisciplinary coordinated design effort, we are developing a completely new O(N) computational framework based on graph theory that is highly scalable and allows optimizations tailored for large-scale heterogeneous platforms.

2.2 Codes and Algorithms: A fast Fermi-operator expansion scheme is used to calculate the electronic structure and the forces in QMD with a computational kernel of sparse matrix-matrix multiplications. The integration of the equations of motion is performed through a geometric integration scheme enabled by a novel extended Lagrangian formulation of Born-Oppenheimer MD, which drastically reduces the computational overhead and removes instabilities. These LANL-unique schemes have been implemented in the LATTE QMD code (and in a separate PROGRESS/BML library) based on self-consistent density functional-based tight-binding (DFTB) theory. These methods are powerful; but to reach their full potential, the new graph-based framework is needed. An unexplored opportunity is the ability to calculate quantum response properties on-the-fly to aid visualization and analysis of the massive amount of generated data.

3. Current and Future HPC Needs

3.1 Computational Hours: The LATTE QMD code used the PROGRESS and BML libraries for the calculation of the density matrix. We currently achieve an unprecedented efficiency of about 2 ps/day for QMD simulations on 104 atoms on one compute node (16 cores, no MPI). Accelerated, parallel replica MD of rare-event systems allows us to run on many MPI ranks. The development of efficient multi-node parallelism with precise error control for the density matrix build via our graph theoretical approach will enable 100–1,000 more atoms to be simulated at the same rate. Hence, in 5–10 years, we expect to be using beyond 16,000 cores.

3.2 Parallelism: We currently use OpenMP for threading over all available CPU cores and Nvidia GPUs to either accelerate dense matrix algebra operations or to perform in parallel many small matrix-matrix multiplications. We have stand-alone implementations of PROGRESS/BML for the computation of the density matrix based on a graph-theoretical distribution of the workload that combines MPI with threading over CPU and GPU cores. With further development and optimization of the graph-based electronic structure framework, the parallel efficiency can be significantly enhanced, especially in the strong scaling limit, which is crucial for QMD simulations.

3.3 Memory: We use sparse data structures heavily so that memory usage remains reasonable even for large, quantum systems. Traditional non-sparse data structures would rapidly limit the maximum size of our simulations. We do not anticipate requiring more on-node memory in the coming years.

3.4 Scratch Data and I/O: In comparison with classical large-scale MD (SPaSM, LAMMPS, etc.) that can handle 109 atoms, our QMD requires relatively little disk storage. Even writing the density matrix to file is not onerous because we use thresholded sparse data structures. Nevertheless, storage requirements will increase significantly once large-scale (>106 atom) QMD is reconciled fully with parallel replica MD. Extracting and visualizing data on-the-fly will then be required.

3.5 Long-term and Shared Online Data: None.

3.6 Archival Data Storage: Current archival storage at LANL is sufficient, and we anticipate that it will continue to be adequate for the system sizes accessible to QMD relative to those accessible to classical MD.

3.7 Workflows: Unknown.

3.8 Many-Core and/or GPU Readiness: We already use accelerators heavily (GPUs and Phis) and are well positioned to use them in the future. Vendor-supplied math libraries help tremendously (CuBLAS and MKL) but we have also developed our own optimized kernels for sparse matrix operations that often outperform commercial software.

4. Requirements Summary Worksheet

Table 1 shows projected requirements for our code.

Table 1. LATTE Requirements

Code: LATTE (data consistent with a typical parallel replica QMD simulation)	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	125,000	×100	×1,000
Computational node hours (Homogeneous many-core)[b]			
Computational node hours (w/GPU or accelerator)[c]	8,000	×100	×1,000
Memory per node	32 GB	32 GB	32 GB
Aggregate memory	TB	TB	TB
Data read and written per run	TB	TB	TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O	>1	>1	>1
Scratch file system space needed	0.1 TB	1 TB	10 TB
Permanent online data storage	0 TB	0 TB	0 TB
Archival data storage needed	10 TB	10 TB	100 TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

D.6 Case Studies Addressing Computing and Data Challenges @ BES Facilities

Case Study Title: Advanced Light Source

Lead Authors: Dilworth Y. Parkinson, Alexander Hexemer, and Craig E. Tull (Lawrence Berkeley National Laboratory)

1.1 Science Use Case

1.1.1 Present or Near Term

The Advanced Light Source (ALS), located at Lawrence Berkeley National Laboratory, is a third-generation synchrotron and national user facility that attracts scientists from around the world. The ALS has 39 beamlines as of October 2015, providing hard and soft X-rays, IR, and EUV light for imaging, scattering, and spectroscopy experiments for chemical, geological, life, material, and physical sciences.

More and more users are working on time-resolved, combinatoric, and high-throughput experiments. To meet this need experimentally, synchrotrons are pushing to provide the necessary X-ray source by increasing their brightness, to build beamlines with appropriate optics and sample environments, and to work with detector developers on fast, high-resolution, high-efficiency detectors.

But bright sources, good optics, and fast detectors are not the only developments necessary to meet the new user needs. They must be accompanied by fast networks, high-performance computers, and advanced software and algorithms. These are necessary in many cases to manage and store the large amounts of data coming at high rates, but also to reduce, process, and analyze the data to extract the useful information. Some of this computing must happen very quickly to provide feedback to users as they collect data. In other cases, more computationally intensive algorithms may be chosen–these may be slower, but they can give optimal results for subsequent analysis and publication.

The ALS has participated in two collaborations with ASCR scientists to attempt to meet users' computational needs: the Center for Advanced Mathematics for Energy Research Applications (CAMERA), which is an integrated cross-disciplinary center aimed at inventing, developing, and delivering the fundamental new mathematics required to capitalize on experimental investigations at scientific facilities; and SPOT Suite, a suite of tools developed jointly by the ALS, Berkeley Lab's Computational Research Division, the Energy Sciences Network, and the National Energy Research Scientific Computing Center (NERSC), to provide ALS users with access to best-of-breed data management, data analysis, and simulation tools. The following is an overview of four representative ALS beamlines that have been part of these initiatives.

- **Imaging (Beamline 8.3.2, hard X-ray micro-tomography).** Scans at this beamline consist of tens to thousands of 2D X-ray transmission images ("radiographs"), which are collected as a sample is rotated, generally through 180 degrees. Tomographic reconstruction yields a 3D volume with approximately 1-micron spatial resolution; in many cases, image volumes are collected every few seconds to minutes to measure dynamic processes. This beamline is used by earth scientists to study, e.g., flow-through porous media; by materials scientists to study, e.g., material failure under strain; and by biologists to study, e.g., plant and insect anatomy.

- **Scattering (Beamline 7.3.3, small- and wide- angle X-ray scattering).** Small- and wide-angle X-ray scattering (SAXS/WAXS), as well as grazing incidence X-ray scattering, are techniques where the scattering of X-rays by a sample is recorded. The pattern of scattering yields information about characteristic distances within the sample on the nanometer scale, and about the shapes and sizes of macromolecules. One characteristic experiment is on organic photovoltaic (OPV) materials. Printing these materials with a specialized printer shows promise as a less expensive, more flexible way to fabricate solar cells to convert sunlight to electricity. The ALS is one of the only facilities that has been able to print and measure these materials simultaneously. By capturing an image of the solution every second for five minutes, scientists can watch the structures crystallize during the drying process.
- **Micro-diffraction (Beamline 12.3.2).** Laue X-ray microdiffraction has been successfully used to probe the microstructure of materials at the (sub)micron scale. Quantities such as crystal orientation, strain/stress, and defect density can be extracted from the analysis of a Laue pattern. One example of a recent experiment is with single crystal nickel-based super-alloys, the material of choice for making turbine blades in the aeronautical industry because of their excellent resistance to thermal creep and hot corrosion, and their strength at high temperature. However, the cost of replacing the turbine blades when damaged can be prohibitive. Laser-assisted 3D printing is the most promising alternative for repairing worn parts, as the single crystallinity needs to be maintained for the material to retain its mechanical properties. In this project, layers of Ni-based superalloys grown on a single crystalline substrate of the same material by laser-assisted 3D printing under various conditions of laser power and speed, are investigated with Laue X-ray micro-diffraction. Finding the conditions for the appearance of the deleterious stray grains and crack formations inside the layers are of particular interest for fine-tuning the technique. Modeling and simulation are used to assess the crystal nucleation and solidification process as well as strain distribution to be directly compared with the experimental data.
- **Hybrid (COSMIC Beamline, ptychography).** Ptychography is a coherent diffractive X-ray imaging method, which enables X-ray imaging at a spatial resolution that is limited by the X-ray wavelength rather than the quality of X-ray optics. Images are reconstructed by a phase retrieval algorithm that acts on coherent diffraction data and information known a priori about the imaging geometry. It is a scanning method, so the field of view, and hence the diffraction dataset, can be arbitrarily large. X-ray ptychography is in the early stages of development but is already having a very large impact in the study of chemistry and magnetism in nanomaterials.

1.1.2 Future

Much of the computing at the ALS — especially prior to around 2013 — was with local desktop-class machines, along with serial software developed for this platform. Much of this computing infrastructure did not integrate advanced computer science solutions. Currently and in the future, ALS needs in this area will increase because of at least three changes: upgraded beamlines, new beamlines, and new approaches to the analysis and use of data after it is collected. These changes will lead to a need for adopting additional computing resources and parallelized software solutions:

- Data rate increases at existing beamlines will come from new and improved detectors, as well as from increases in flux and brightness due to upgrades in the storage ring, beamline optics, and end stations. The increase in complexity of the experiments is enabled in part by the increases in speed, but also because of the constantly improving reliability and stability of the normal beam- line components, which means that more risky and challenging experiments can be attempted.

- For the ALS, two prime examples of new data-intensive beamlines that will come online are the new ptycho-tomography beamlines and new infrared tomography beamlines. In both of these cases, detectors will be used that can approach 10 Gbps data rates. And in both cases, the processed data will result in 5D data sets: 3D volumes that contain spectral information and which are collected as a function of time.
- There will be an increasing demand to combine data from multiple sources — not just from multiple beamlines, but from beamlines and other types of experiments, including neutron, electron, optical, and other experiments. To the extent that data become more widely shared and accessible across communities, we also see in the future opportunity for a large new effort in data mining to find patterns across data. Rather than relying solely on data you collect, you can combine results from your data with data collected by many other researchers. In many cases, this will mean the use of algorithms and questions that go far beyond the original questions and conclusions made by the researchers who collected the data.

1.1.3 Data Lifecycle

We acknowledge that one barrier to collaboration between science domains and computer scientists is the lack of common terminology or representations for modeling and profiling the data lifecycle. On the other hand, it is a challenge for us to accurately portray in a simple way a characteristic data lifecycle at the ALS because it is highly experiment dependent–even for experiments at a given beamline, it can be highly variable. The following presents the data lifecycle for one characteristic experiment from each of the beamlines discussed above.

- **Imaging (Beamline 8.3.2, hard X-ray micro-Tomography).** During a scan at Beamline 8.3.2 (Figure 1), an acquisition computer saves images it receives from a camera to a Data Transfer Node, where SPOT Suite software packages each set of images (10 GB) and transfers them both to NERSC and to a local temporary storage server. At NERSC, preprocessing (normalization, phase retrieval, and other filtering) and tomographic reconstruction (fast analytic approaches based on Fourier transforms) is launched automatically for each data set, which results in reconstructed 3D image volumes. Users can also submit a limited number of data sets for tomographic reconstruction using iterative and model-based methods, which give superior results but are orders of magnitude more computationally intensive. All results are presented to users through a web portal. Users then download the reconstructed image volumes (20–50 GB each) and carry on with subsequent steps. Data on NERSC is kept on disk for a period of a few days, until it is moved to tape, and staged back to disk on demand.

  The kinds of analysis performed on the reconstructed 3D volumes is extremely diverse, but it is common to filter and then segment structures of interest (define their boundaries); the ALS has collaborated with CAMERA to develop faster and more automated and robust methods for these steps. In many cases, segmentation is a precursor to measure porosity, or to generate statistics about the size, shape, and distribution of certain features within the volume. In other cases, a reconstructed (and often segmented and then meshed) volume is used as the input to a simulation such as reactive transport, which combines fluid dynamics, structural changes, and chemistry; these simulations using an initial volume as its starting point can be compared to the measured experimental sample during in situ time-resolved experiments.
- **Scattering (Beamline 7.3.3, small- and wide- angle X-ray scattering).** For one of the OPV materials printing experiment described above, an attempt was made to illustrate a super facility concept, with seamless integration of multiple, complementary DOE Office of Science user facilities into a virtual facility offering fundamentally greater capability. The facilities were the ALS, NERSC, the Oak Ridge Leadership Computing Facility (OLCF), and

ESnet. The SPOT Suite workflow management system running at NERSC was used to create a prototype data pipeline: as data were collected from an ALS GISAXS experiment, they were sent via ESnet to the Titan Supercomputer at OLCF for analysis on 8,000 nodes using CAMERA's HipGISAXS code, a customized high-performance code that exploits advanced graphics processors and particle swarm optimization to quickly reverse engineer the sample from simulated scattering patterns based on distorted wave Born approximations. The project demonstrated the capability for researchers in organic photovoltaics to not only measure scattering patterns for their samples at the ALS and see real- time feedback on all their samples through the SPOT Suite application running on NERSC, but also to see near-real-time analysis of their samples running at the largest scale on the Titan supercomputer at OLCF. This allowed the researchers to understand their samples sufficiently during beamtime experiments to adjust the experiment to maximize their scientific results. Making a super facility available to users on a regular basis would have a large positive impact on the kind of work that could be done.

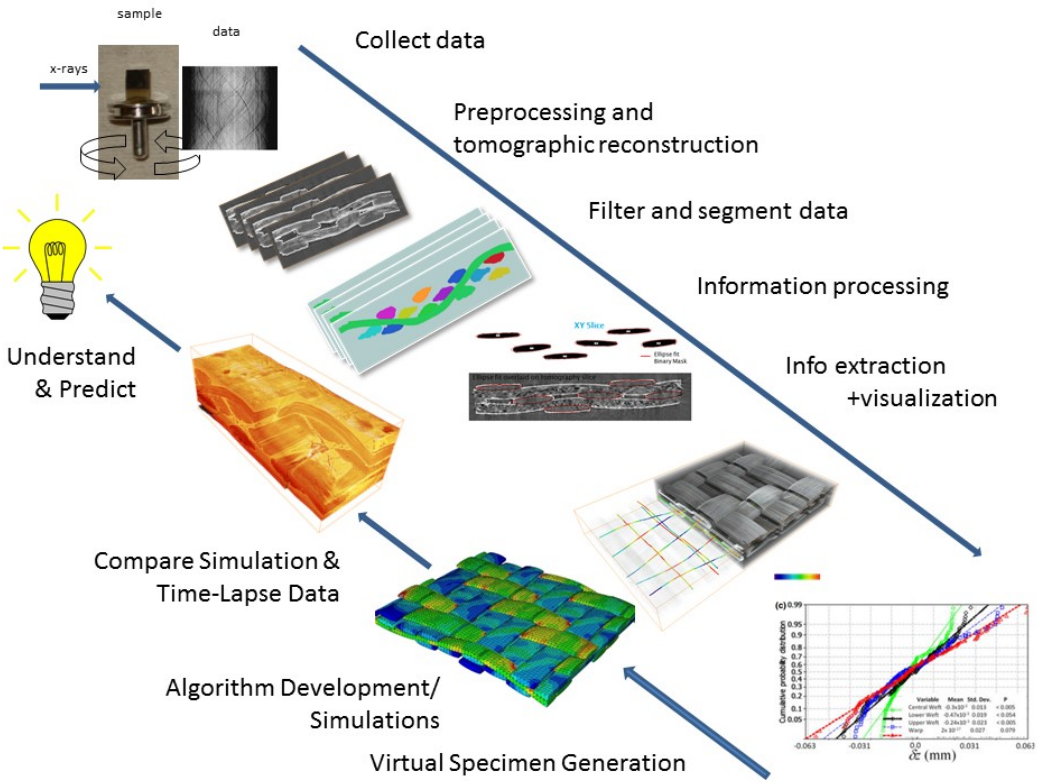


Figure 1. Data lifecycle for Beamline 8.3.2, from the perspective of a domain scientist (in other words, lacking details of where the computations occurred, or details about data sizes, software used, etc.). Images courtesy Rob Ritchie (UC Berkeley/LBNL) and Hrishi Bale (now at Zeiss)

- **Micro-diffraction (Beamline 12.3.2).** With the advent of fast and large-size X-ray detectors such as the DECTRIS Pilatus hybrid pixel array detector, it has become possible to map a large portion of a sample with micron step sizes within a few hours. The technique becomes particularly useful when the data generated can be analyzed in real time. We have written a Laue indexing and strain refinement code that can process multiple images in parallel. Data collected on beamline 12.3.2 of the Advanced Light Source can be transferred to NERSC automatically through SPOT Suite. Users can then log into a web portal, where they input their desired data processing parameters, and calculations are then launched on NERSC. Results are presented within the web portal.

Processing tens of thousands of Laue patterns, which previously took weeks on a desktop computer, can be done in just a few hours, so that users can get results during their beamtime and use the feedback to adjust their experiments. The use of high-performance computing and fast detector technology has provided the opportunity to transition from Laue X-ray micro-diffraction mapping (a few hundred data points on localized area of the sample) to a quantitative micro- structural imaging tool — 1-megapixel images showing the distribution of grain orientation, phases, strain/stress, and deformation inside a material.

- **Hybrid (COSMIC Beamline, ptychography).** Currently, diffraction data are streamed from a high-frame-rate CCD detector, through a data transfer node, and to a multi-GPU cluster during the sample scan. Data are then submitted to a preprocessing computation, which removes background, filters outliers, and ideally samples the diffraction measurements. After preprocessing, the sample image is reconstructed by a phase retrieval algorithm that acts on the full set of diffraction pat- terns. The parallel projection algorithm iteratively recovers reciprocal space phases and allows for a direct computation of the image via FFT. After image reconstruction, higher-level analysis may proceed on a set of projections that represent a tomographic or spectroscopic image data set.

1.1.4 Data-centric Requirements: Capabilities, Speeds, and Feeds

The ALS submitted a case study to the 2014 ESnet Basic Energy Sciences Network Requirements Review. This included an analysis of current and predicted data rates; those trends and predictions still hold true. Figure 2, a table that was part of that report, summarizes the results of that case study.

1	Upgrade Scenario	Max Gbps	Operating Ave. Gbps	Overall Ave. Gbps
Current	Current	7.24	0.91	0.43
Current	10G LAN/WAN	7.57	0.92	0.44
Current	Detectors 5x	13.94	0.95	0.44
Current	Exposure 5x	10.37	0.95	0.44
Current	Sample 5x	7.24	1.59	0.68
Current	All 5x+10G LAN/WAN	34.56	2.98	1.31
Current+New	Current	25.84	3.22	1.81
Current+New	All 5x + 10G LAN/WAN	81.58	8.43	4.68

Figure 2. Based on a data rate prediction tool formulated by interviews with beamline scientists at the ALS (bl832web.lbl.gov/esnet), various scenarios can be investigated to determine expected future data rates. Predicted data rates indicate that the exponential rate of growth in network traffic seen over the last 5+ years (at least) will continue at about the same pace.

1.2 Impediments, Gaps, Needs, and Challenges

There are a number of needs based on future plans at the ALS. Some of these have been mentioned in the previous sections. We will review them here. We note that this list has significant overlap with the report of the BES Facilities Computing Working Group from their May 2015 meeting.

- One overarching challenge is the number and diversity of light source experiments. Even for a given beamline, there are often tens of different types of experiments. This means

that it is possible to "solve" all the problems of one user without helping the next user at all. Even when there is a potential for a given tool to benefit other users, it often takes significantly more development time to make a tool robust and easy to use to be useful to the community rather than for a single user (documentation, testing, bug tracking, message boards, outreach), and the incentive to do this extra development is often not there. On the other hand, consolidation of software would mean better software with overall less investment. One approach could be to focus on easy libraries and languages that allow relatively easy customization, rather than full environments. It will be important for these solutions to focus on parallelism in the processing and on taking advantage of emerging hardware, while facilitating running applications on multiple platforms. Another approach could be focusing on workflow tools, which could provide the vehicle that would lead to a community catalog of software libraries.

- Usability and accessibility are key concerns. It is necessary to minimize the need for facility users to have detailed knowledge of system hardware and operating systems. The ALS has been extremely successful at expanding its user base to a wide variety of science areas as well as to industry users. Many of these users will benefit from advanced computing, but they are experts in areas other than computing — writing a script is something many of them have never done. Many users also do not have easy access to computing power beyond their laptop. Even for users who do have access to more computing power, for some of the newer beamlines and for planned beamlines, it is getting to the point where users cannot just download their data — their hard drive is not large enough, and if it was, they would not have the computing power needed to do anything with it. In these cases, a new form of instrument combining storage, compute, data, and code is required (a "super facility" or "discovery engine").
- As data rates and experiment complexity increase, it becomes more desirable to steer the data collection, and near-real-time feedback can permit qualitatively different, more interactive, and collaborative discovery modalities. One requirement for this will be automating and abstracting key analysis tasks, to better allocate the human in the loop — another way to think about it is the requirement to "mathematicize" more of the process (to formalize and quantify metrics that were previously qualitative). Of course, doing this would have applications beyond just real-time feedback. Another requirement to make this work will be workflows, which must help cross the boundaries of multiple data sources, computing resources, operating systems, and runtime environments to provide the necessary feedback. This is a challenge because, among other reasons, there is currently a lack of common scheduling across facilities, or a common language to define job pipeline operations across centers.
- With increasing data rates, ALS personnel are seeing more problems with storing data quickly enough, with how to let users access it, and with how to transfer it to users' home institutions or to their collaborators. One issue is a lack of a network infrastructure to end-users, or other issues reaching the end users — to cite one example, many users from industry have internal security policies which preclude them from using globus.org for data transfer.
- There are a number of areas in which new algorithms or analysis approaches are necessary. In many cases, "noisy" data are collected or information is missing. Some users collect large 5D data sets and would like to visualize them with low latency or collaboratively share interactive visualizations with people at multiple locations. In other cases, users would like to combine different types of data from different instruments in a way that adds value. Many tools focus on single data sets of low dimension, so these data ensembles and high-dimensional data provide a particular challenge. New visualization methods must use novel visual encoding, interactive tools for dealing with higher-dimensional data, and automatic algorithms to identify salient variables across ensembles or for dimension reduction with real-time feedback.

- Ideally, any relevant data should be made available to the scientific community after some amount of time. But more than data preservation is required—proactive data curation is necessary for the data to be really useful. This will require more detailed metadata than is currently available, and it is a challenge to find ways to have automated but customizable ways to capture metadata. Data curation would mean making the data accessible in such a way that it can be searched — not just based on the existing metadata, but also based on scientifically meaningful metadata that are filled in through machine learning or other approaches. Ideally, there would be ways for data to find interested parties as datasets are being created, rather than waiting for scientists to search for the data, which may be spread across many different archives. It might mean making the data accessible along with the software and computing infrastructure necessary to process the data. The benefit of curation would be to reduce duplication of effort in data creation, but also for reuse of data for further high-quality research. Another benefit would be that it could lead to the increased availability of algorithms and software to the community, as researchers write code that can be benchmarked and used against curated data. It is not clear who would host or pay for this data curation.

Case Study Title: Linac Coherent Light Source (LCLS)

Lead Authors: Jana Thayer and Amedeo Perazzo (Stanford Linear Accelerator Center)

1. Description of Research

1.1 Overview and Context:

The Linac Coherent Light Source (LCLS) is a Free Electron Laser which generates short, high-intensity (~4-mJ) X-ray pulses in the 250 eV–11 keV range. The LCLS's highly focused beam gives researchers the intensity needed to probe complex, ultra-small structures and the ultrafast pulses (< 5–500 fs) required to freeze atomic motions, thus shedding light on the fundamental processes of chemistry, physics, drug development, and technology. It has had a significant impact on many areas of science including resolving the structures of macromolecular complexes that were previously inaccessible.

1.2 Research Objectives for the Next Decade:

Throughout this document, we will use the quasi real-time (< 10 s) nanocrystallography pipeline (Figure 1) as the main example of the LCLS computing requirements because it takes the most CPU cycles, it provides well-known and scalable computing demands, and because it is instrumental for many LCLS experiments, which study atomic-scale structural dynamics and fluctuations in matter:

1. Complex materials (novel functional properties): Heterogeneity and fluctuations at the nano-scale, nano-particle dynamics.
2. Catalysis (efficient, selective, robust, earth-abundant): Chemical, structural, and electronic changes, nano-particles, interfacial chemistry.
3. Biological function: Protein crystallography – structure and dynamics.

The LCLS-II upgrade will increase the repetition rate of the machine from 120 Hz to up to 1 MHz. With detectors' readout rates going from today's 120 Hz to 5 kHz in 2020 and 100 KHz in 2025, the expected throughput from the front-end electronics is expected to increase by a factor of 20 in 5 years and by three orders of magnitude in 10 years when LCLS will operate three beamlines concurrently.

2. Computational and Data Strategies

2.1 Approach:

Currently, the main data challenges derive from the need to develop a large variety of advanced algorithms for FEL science and by the need to handle multi GB/s data streams and distribute them to the users for processing. Data access today is handled through the use of large (few petabytes) parallel, spindle based, file systems and multiple storage layers.

We plan to manage the large increase in throughput, which will come with LCLS-II, by adopting flash-based storage technologies and by offloading the most demanding computing requirements to large data facilities like NERSC.

2.2 Codes and Algorithms:

As data are read out from the detector, diffraction data are processed serially using the following algorithms, enabling users to distill the raw data down to meaningful content:

→ Hit finding: identify images with useful information by detecting diffracted photons;

→ Indexing: based on the peak positions, infer what the orientation of the crystal was when it was hit; this algorithm attempts to determine the orientation of the crystal by repeatedly generating a lattice, predicting the spot position and comparing it to the observed spots; and

→ Merging and post-refinement: scales each diffraction pattern to conform to the mean and feeds results into any standard phasing package suite used for automated determination of molecular structures, which produces the atomic position.

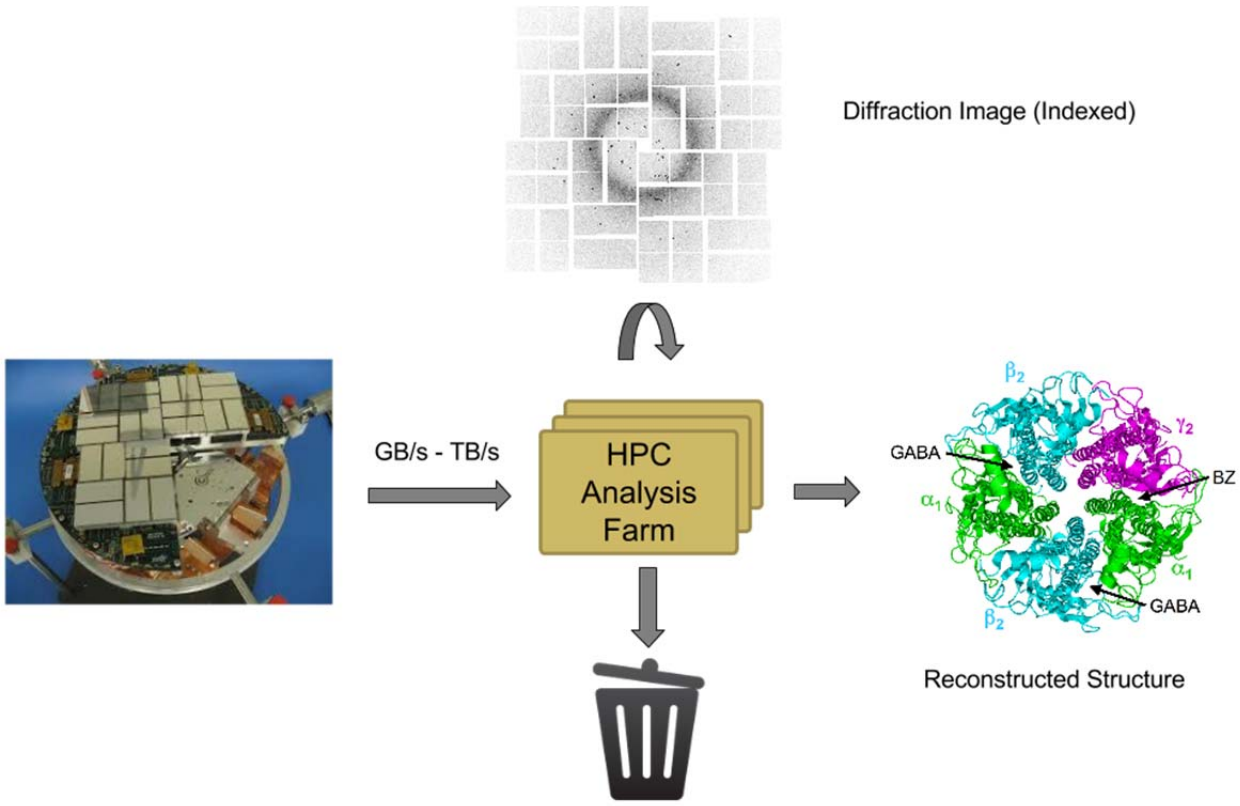


Figure 1: The nanocrystallography pipeline provides atomic-scale vision to researchers at the beamline in < 10 s. The data stream is sent from the detector to the analysis farm for indexing, classification, and reconstruction in order to provide quasi real-time response. For LCLS-I, this processing requires roughly 50 TFLOPS.

3. Current and Future HPC Needs

3.1 Computational Hours:

The nanocrystallography pipeline uses about 1280 cores, each delivering ~36.8 GFLOPS, i.e., 50 TFLOPS aggregated. Over the span of one year, assuming 50 experiments operating 60 hours, 1280 cores deliver about 4×10^6 core hours (that's an average: not all experiments require 50 TFLOPS, and there are about 100 experiments in 1 year).

3.2 Parallelism:

The LCLS psana analysis framework supports the use of MPI and OpenMP parallelization for both offline analysis and real-time analysis. Currently, we distribute jobs over tens of cores. These jobs will need to scale up by a factor 20 by 2020 and by 3 orders of magnitude by 2025. The dramatic increase in the level of parallelism will require development effort to effectively utilize all cores and will also have a profound effect on the I/O pattern of the data analysis

applications. In particular, it will require storage systems which would be effectively managing/serving many (tens, hundreds of) thousands of I/O requests concurrently.

3.3 Memory:

We currently run with 4 GB per core. It is reasonable to assume the memory requirements may scale up by a factor of 2 with the size of the images.

3.4 Scratch Data and I/O:

Currently, LCLS operates with 600 TB of scratch space, i.e., roughly 10% of the total disk storage. We expect to maintain the same ratio for scratch moving forward.

I/O examples by 2020:

→ 1 × 16 Mpixel ePix@360Hz = 12 GB/s

→ 100 K points, fast digitizers @ 100 kHz = 20 GB/s

→ 2 × 4 Mpixel ePix @ 5 kHz = 80 GB/s

→ Distributed diagnostics = 1–10 GB/s

I/O examples by 2025:

→ 3 beamlines × 2 × 4 Mpixel ePix @ 100 kHz = 4.8 TB/s

Ideally, the processing would take the same amount of time as the pre-fetch of the next event so that as soon as one event was processed, the CPU could immediately proceed to the next without any idle time in between. If the fetch took longer, then the analysis would be I/O bound and the CPU would be idle. If the processing took longer, then the analysis would gradually fall behind the data acquisition.

3.5 Long-term and Shared Online Data:

We currently operate with 5 PB of disk storage. Data retention policy is 6 months unlimited on disk, 2 years with quota, 10 years on tape. Assuming the same data retention, we will need 100 PB by 2020 and 6 EB by 2025.

3.6 Archival Data Storage:

Currently, the SLAC tape archive system has an overall storage capacity of ~20+PB. We foresee the same scaling as above (factor of 20 by 2020, 3 orders of magnitude by 2025).

3.7 Workflows

The analysis workflow for a typical crystallography analysis starts with accessing the data, doing a per-image analysis using software suites such as crystFEL or cctbx, dataset aggregation, and structural interpretation/model building (PHENIX). In the future, the workflow will look similar, but the data throughput will be much higher, and some tools for monitoring the progress of dataset aggregation will need to be developed. The workflow for non-crystallography experiments requires that users write parallel Python code to analyze their data. The LCLS portal will handle on-boarding, experiment preparation, group management, experiment logbook, data management, and post-run debrief. The data management system handles both the automatic workflows of the data from the online cache to the various storage layers as well as the users' requests, such as restoring data from tape, all through a web portal. Some aspects of the current system, such as checksum calculation, HPSS interface, and lack of prioritization, will become limitations at higher data volumes and will need to be upgraded.

3.8 Many-Core and/or GPU Readiness:

At this time, we are not envisioning use of GPUs, but we plan to be able to eventually operate on Xeon Phi.

3.9 Software Applications, Libraries, and Tools:

Ability to scale up the parallelism of our Python jobs to tens of thousands of cores will be critical.

3.10 HPC Services:

We will need to offload part of the LCLS-II data processing to larger computing facilities like NERSC. Although SLAC recently upgraded its connection to ESNET from 10 Gb/s to 100 Gb/s, this link will not be enough for LCLS-II, and terabit capabilities will be required if LCLS relies on NERSC for processing its data.

3.11 Additional Needs:

The ability to keep up the analysis with the DAQ throughput is critical. Given the dramatic increase in the amount of data coming from the future LCLS cameras, we foresee that the main challenges will come from the storage system, in terms of both size and fast access. In this sense, LCLS will need a system focused on data archiving and data throughput as much as CPU cycles.

4. Requirements Summary Worksheet

Table 1: LCLS Nanocrystallography Pipeline Requirements

Code: LCLS Nanocrystallography Pipeline	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)**[d]	**Future Usage: 2025 (As a factor of column 1)**[d]
Computational core hours (Conventional)[a]	3.8 M	20×	1,200×
Computational node hours (Homogeneous many-core)[b]	N/A	N/A	N/A
Computational node hours (w/GPU or accelerator)[c]	N/A	N/A	N/A
Memory per node	128 GB	2×	2×
Aggregate memory	5 TB	40×	2,400×
Data read and written per run	150 TB	20×	1,200×
Maximum I/O bandwidth needed	5 GB/sec	100 GB/s	4.8 TB/s
Percent of runtime for I/O	100%	100%	100%
Scratch file system space needed	600 TB	20×	1,200×
Permanent online data storage	5 PB	20×	1,200×
Archival data storage needed	20 PB	20×	1,200×

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.

[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.

[c] "Node hours" is used for "GPU or accelerator" usage.

[d] For example, 32 × column 1.

D.7 Case Studies Addressing Mathematics and Computer Science Transforming BES Science

Case Study Title: Advanced Mathematics for High-Performance Imaging for Light Sources

Lead Authors: J. Donatelli, A. Hexemer, D. Kumar, R. Pandolfi, D. Parkinson, V. Venkatakrishnan, P.H. Zwart, and J.A. Sethian (all of Lawrence Berkeley National Laboratory)

1. Description of Research

1.1 Overview and Context:

The U.S. Department of Energy supports a spectrum of experimental science at scientific facilities. CAMERA (The Center for Advanced Mathematics for Energy Research Applications) builds sophisticated, state-of-the-art mathematics to greatly accelerate progress at user facilities, accomplished through coordinated teams of applied mathematicians, computer scientists, beamline scientists, materials scientists, and computational chemists.

1.2 Research Objectives for the Next Decade:

We consider two different case studies: (1) Emerging techniques such as single particle imaging, fluctuation scattering, and cryo- EM allow researchers to transcend the limitations of traditional imaging methods. However, determining structural information from these experiments requires the development of new advanced mathematical algorithms and robust, high-performance software that scales to handle the vast amounts of data collected during these experiments. (2) The fundamental makeup of complex materials can be determined through scattering experiments (GISAXS, WAXS, SAXS) and through tomographic reconstruction. A critical need is to make the associated mathematical inversion methods fast and accurate.

2. Computational and Data Strategies

2.1 Approach:

New Reconstruction algorithms and software are needed to solve complex inverse problems to determine molecular structure, visualize molecular motion, and classify heterogeneity from a vast number of diffraction or EM images, which are expected to be collected at rates up to 1 MHz in the near future. In scattering and tomography, new methods are required for inverting data to produce structural information for highly complex materials.

2.2 Codes and Algorithms:

Our current codes use iterative reconstruction techniques utilizing optimization techniques and fast transforms to solve the associated inverse problems and extract structural information from these images. New techniques, such as applied machine learning, will assist in these reconstructions.

3. Current and Future HPC Needs

3.1 Computational Hours: ALS currently has a 40-million-hour allocation for 2016, and we expect our need for compute hours to continue to increase, though the increase will be modest compared to the increase in the need for computational hours with accelerators.

3.2 Parallelism: We have written hybrid MPI/OpenMP code, as well as GPU code. ALS currently has a 15-million-hour allocation on Titan. This does not count the hours of time on ALS-owned clusters with GPUs for ptychography, tomography, or scattering. Due to a CAMERA focus on algorithms that run on GPUs for imaging, tomography, and scattering, there will be a strong increase in demand for computational hours with accelerators.

3.3 Memory: Regarding memory per node, for tomography, due to the switch in data read orientation between the normalization step and the reconstruction step, it is useful to have the full data set in memory. Currently, raw data sets are 10 GB, but this will grow.

3.4 Scratch Data and I/O: This topic does not apply.

3.5 Long-term and Shared Online Data: The definition of "run" is ambiguous for us. If you talk about one image (for scattering), this is 10 MB. If you talk about one tomography run, one raw data set input is 10 GB, and the output is 50 GB. In the future, when we do tomographic reconstruction of multiple time points in tandem (sharing information), we will have 5 to 50 TB of raw data that will be read in at once. The scratch file system request approximately with data read/write per run.

3.6 Archival Data Storage: We currently have 1.5 PB on tape at NERSC under spot suite, and this number will continue to increase. Archival data are accessible online but may involve a delay in accessing it (e.g., data stored on HPSS tapes).

3.7 Workflows: Raw data are transferred from the light source, preprocessed, analyzed through our reconstruction codes, and then visualized.

3.8 Many-Core and/or GPU Readiness: The bulk of computation involves dense matrix operations and FFTs. If standard libraries, such as BLAS and FFTW, are optimized for these new architectures, then our code can readily take advantage of them.

4. Requirements Summary Worksheet

Tables 1 and 2 present estimates of required exascale resources for our codes.

Table 1. Iterative Reconstruction Requirements

Code: Iterative Reconstruction for SPI (single particle imaging), FXS (fluctuation scattering), and cryo-EM	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	10^6 per year	10×	100×
Computational node hours (Homogeneous many-core)[b]			
Computational node hours (w/GPU or accelerator)[c]			
Memory per node	32 GB	64 GB	128 GB
Aggregate memory	200 TB	1,000 TB	100,000 TB
Data read and written per run	50 TB	500 TB	5,000 TB
Maximum I/O bandwidth needed	GB/sec	GB/sec	GB/sec
Percent of runtime for I/O	0.1%	1%	10%
Scratch file system space needed	50 TB	200 TB	2,000 TB
Permanent online data storage	50 TB	200 TB	2,000 TB
Archival data storage needed	200 TB	500 TB	5,000 TB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.
[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.
[c] "Node hours" is used for "GPU or accelerator" usage.
[d] For example, 32 × column 1.

Table 2. Scattering and Tomography Requirements

Code: Scattering and Tomography	**Column 1: Current Usage**	**Future Usage: 2020 (As a factor of column 1)[d]**	**Future Usage: 2025 (As a factor of column 1)[d]**
Computational core hours (Conventional)[a]	40 million	10×	100×
Computational node hours (Homogeneous many-core)[b]			
Computational node hours (w/GPU or accelerator)[c]	15 million	5×	20×
Memory per node	64 GB	128 GB	256 GB
Aggregate memory			
Data read and written per run	10 MB to 60 GB	10 MB to 1 TB	10 MB to 10 TB
Maximum I/O bandwidth needed			
Percent of runtime for I/O	10–80%	<10%	<10%
Scratch file system space needed	0.5 TB	5 TB	50 TB
Permanent online data storage	400 TB	800 TB	2,048 TB
Archival data storage needed	1.5 PB	5 PB	50 PB

[a] "Core hours" is used for "conventional" processors (i.e., node-hours * cores_per_node). Intel "Ivy Bridge" is an example conventional processor.
[b] "Node hours" is used for homogenous many-core architectures. A self-hosted Intel Xeon Phi "Knights Landing" is an example.
[c] "Node hours" is used for "GPU or accelerator" usage.
[d] For example, 32 × column 1.

U.S. DEPARTMENT OF
ENERGY