ENVIRONMENTAL
MOLECULAR SCIENCES
LABORATORY

EMSL

Workshop
Report

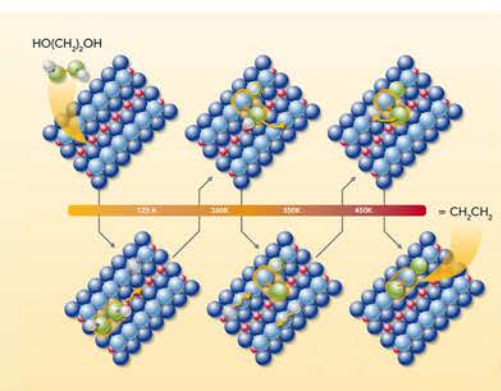# Multiscale Computation

Needs and Opportunities for BER Science

Workshop Held
August 26, 2014

Timothy Scheibe, Pacific Northwest National Laboratory
Jeremy Smith, Oak Ridge National Laboratory

U.S. DEPARTMENT OF
ENERGY
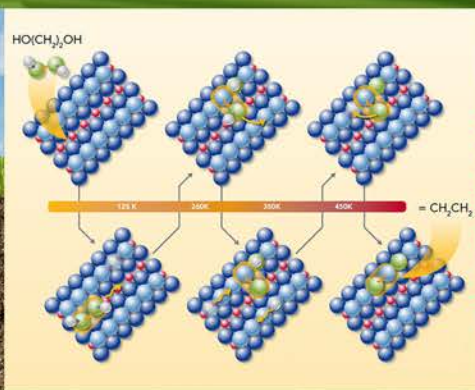
About the cover

The advancement of a robust predictive understanding of complex environmental and biological systems, one of the primary objectives for U.S. Department of Energy, Office of Biological and Environmental Research (BER), requires significant advances in our ability to connect fundamental understanding of processes at molecular, pore, and cellular scales to predictions of system behaviors at laboratory, field, and ecosystem scales.

High-performance computing software frameworks coupled with new mathematical algorithms are needed to support development of multiscale linkages across the BER code suite.

Environmental molecular research is accelerated when combined with resident expertise, leading-edge hardware, efficient parallel software, accurate and predictive theories and visualization capabilities.

Multiscale Computation: Needs and Opportunities for BER Science

Report on the Workshop held August 26, 2014

Workshop Co-chairs
TD Scheibe, Pacific Northwest National Laboratory
JC Smith, Oak Ridge National Laboratory

January 2015

Pacific Northwest National Laboratory
Richland, Washington 99352

Executive Summary

The Environmental Molecular Sciences Laboratory (EMSL), a scientific user facility managed by Pacific Northwest National Laboratory for the U.S. Department of Energy, Office of Biological and Environmental Research (BER), conducted a one-day workshop on August 26, 2014 on the topic of "Multiscale Computation: Needs and Opportunities for BER Science." Twenty invited participants, from various computational disciplines within the BER program research areas, were charged with the following objectives:

- Identify BER-relevant models and their potential cross-scale linkages that could be exploited to better connect molecular-scale research to BER research at larger scales.

- Identify critical science directions that will motivate EMSL decisions regarding future computational (hardware and software) architectures.

The workshop was organized into two breakout sessions and each session focused on one of the objectives above. Participants were also invited to contribute position papers on a topic of their choice related to multiscale simulation and a detailed description of one or more codes they use in their research.

Thirteen position papers and twenty-three detailed code descriptions were received and discussed during the breakout sessions. Based on this discussion, a number of opportunities for multiscale linkages among the various codes were identified and prioritized. Challenges and potential roadblocks to implementation, and potential means of addressing them, were also identified and discussed. These opportunities will form the basis for EMSL strategic development of a topical computing capability for multiscale modeling, focused on BER science applications and beginning with molecular scales linking upward to larger time and space scales.

This report documents outcomes of the workshop and provides recommendations that will guide strategic planning and capability development. The diversity of the applications of interest to BER scientists, and the diverse nature of the codes discussed, drive a need for consideration of hybrid architectures for future EMSL computing hardware. Specifically, a mixture of compute-intensive and data-intensive architectures is likely to best meet the needs of the scientific community. Design of future computational systems will draw from the results of this workshop, and continued interaction and partnership with scientists from selected application areas will lead to enhanced scientific impact.

The advancement of a robust predictive understanding of complex environmental and biological systems is one of the primary objectives of the BER program. It requires significant advances in our ability to connect fundamental understanding of processes at molecular, pore and cellular scales to predictions of system behaviors at laboratory, field, and ecosystem scales. This workshop is the first step in development of a strategic computing resource for BER scientists at EMSL that will focus on this problem.

Acronyms and Abbreviations

AMDF	Accelerated Molecular Dynamics Framework
BER	Office of Biological and Environmental Research
BSSD	Biological Systems Sciences Division
CESD	Climate and Environmental Sciences Division of BER
CG	coarse graining
CLM	community land model
CPU	Central Processing Unit
DPD	Dissipative Particle Dynamics
DHSVM	Distributed Hydrology Soil Vegetation Model
DOE	U. S. Department of Energy
EMSL	Environmental Molecular Sciences Laboratory
GCM	Global Climate Model
KBase	DOE Systems Biology Knowledgebase
LBM	Lattice Boltzmann Method
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator
MD	molecular dynamics
MM	molecular mechanics
PNNL	Pacific Northwest National Laboratory
QM	quantum mechanics
SBR	Subsurface Biogeochemical Research program of BER/CESD
SPH	Smoothed Particle Hydrodynamics
TES	Terrestrial Ecosystems Sciences program of BER/CESD
UCG-MD	Ultra-Coarse-Grained Molecular Dynamics

Contents

Figures

1.0   Background and Purpose

The U.S. Department of Energy (DOE) Office of Biological and Environmental Research (BER) advances world-class biological and environmental research programs and scientific user facilities to support DOE's energy, environment, and basic research missions. Both the Climate and Environmental Sciences Division (CESD) and Biological Systems Science Division (BSSD) aim to gain a predictive understanding of complex multiscale phenomena critical to BER.

CESD issued a strategic plan in 2012 for the research programs and user facilities within the division. That plan states the CESD Mission as follows:

> *To advance a robust predictive understanding of Earth's climate and environmental systems and to inform the development of sustainable solutions to the Nation's energy and environmental challenges.*

This mission is embodied in five goals that focus on developing fundamental process understanding of subsurface, terrestrial and atmospheric environments, and integrating that knowledge into earth systems simulators ("next-generation, integrated models of the human-Earth system").

Similarly, BER's Biological Systems Sciences Division (BSSD) "seeks to gain a predictive understanding of living systems, from microbes and microbial communities to plants and other whole organisms. This foundational knowledge serves as the basis for the confident redesign of microbes and plants for sustainable biofuel production, improved carbon storage, and contaminant remediation."

The desired transition to "predictive understanding" requires computational simulation models, as simpler, analytical theoretical approaches can rarely incorporate the level of complexity and heterogeneity needed to understand relevant systems. This trend to simulation, which is occurring in many branches of science and engineering, turns to computational science to integrate disparate detailed information into working models capable of providing causative explanations of the behavior of complex systems. Simulations are used to predict, given initial conditions and guiding principles enshrined in a model, how a system of interest evolves. The process of simulation furnishes an integrated detailed understanding of the dynamic nature of biological and environmental systems, enabling scientists to perform "virtual experiments", the results of which can be used to guide new real experiments to further test and refine models, in an iterative process.

Computer modeling and simulation can rationalize experimental information across scales; this integration was, for example, pivotal in the recent discovery of the mechanism of microbial mercury methylation (Parks et al. 2013). Modeling, simulation, and 3-D time-resolved reconstruction of biological events are logical next steps in this procedure. Indeed, computer modeling and simulation is the only mechanism for integrating large amounts of complex, interrelated data in a manner that provides a causative explanation of the behavior of complex systems. Computational science has thus been increasing in importance in the last decades. In the next 10 years we expect computer simulation to provide the framework for interpreting and integrating experimental biological and environmental information at multiple time and length scales.

Understanding the seamless connections inherent in biological systems is hampered by current scientific thrusts on particular network scales and on tools that examine only narrow windows of spatial, temporal or chemical information. The result is a fragmented view of biological systems that prevents understanding of the linkages and interplay between molecular scale events and observations that ultimately impact DOE's mission goals in bioenergy research, element cycling and environmental stewardship. Critical to these advances will be the ability to model and simulate interactions across hierarchies and dimensions, enabling understanding of the connections between systems functioning at diverse

length and time scales. Multiscale simulation, meaning the simultaneous simulation of processes linking scales in length and time, is of particular relevance to BER. BER research ranges from atomic to global scales in length, and femtoseconds to decades in time. However, for the purposes of the present report we confine our discussion to multiscale phenomena on the lower spatio-temporal scales, and specifically, how to link molecular-scale research to higher-level phenomena. For biological systems this requires simulation models of the dynamic interaction of components coupling molecular, cellular and organismal functions.

The Environmental Molecular Sciences Laboratory (EMSL), a DOE scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington, provides premier experimental and modeling/simulation capabilities for molecular-level research on energy and environmental needs facing DOE and the nation. By periodically engaging experts from the scientific community, EMSL identifies opportunities for future investments in capabilities that will lead to impactful scientific results in targeted research areas to address programmatic priorities within BER. EMSL comprises a unique set of experimental and computational capabilities (including scientific expertise as well as instruments) that can be enhanced to better address these goals in the context of serving the BER and broader scientific communities. A critical determinant of the success of EMSL in meeting BER research objectives is our ability to effectively utilize molecular-scale process understanding to inform and improve numerical simulators at larger scales. Accordingly, multiscale science and modeling is an important element of EMSL's research strategy. The EMSL Strategic Plan 2014 states:

> *We also intend to develop new paradigms for multiscale prediction and modeling, including development of new computational approaches to utilize molecular-scale information in models at larger spatial and temporal scales through mesoscale and multiscale modeling methods. This work will emphasize development of predictive modeling capabilities for climate, subsurface, and biological science areas, and will support the integration of EMSL experimental and modeling capabilities for scientific impact. (Section 3.3 Transforming Scientific Capabilities for the Future)*

As an initial step in development of a new multiscale modeling capability for the scientific community, this workshop had the following objectives:

- Identify BER-relevant models and their potential cross-scale linkages that could be exploited to better connect molecular-scale research to BER research at larger scales.

- Identify critical science directions that will motivate EMSL decisions regarding future computational (hardware and software) architectures.

Twenty scientists from academia and national laboratories, representing computational aspects of research in both BER divisions (CESD and BSSD), met August 26, 2014 in Washington, DC. Given the molecular-level focus of EMSL, the participants were selected to represent a range of earth system process scales from molecular up to the local land surface. Global land surface and atmospheric scales were outside the scope of this workshop. Other participants included BER program managers from both CESD and BSSD, and observers from DOE's Advanced Scientific Computing Research program. A list of invited participants and observers is provided in Appendix A.

The workshop agenda (Appendix B) was designed around the two objectives stated above. The meeting co-chairs and BER program manager for EMSL gave introductory presentations at the beginning of the meeting. The presentations provided motivation, background material, and context for this diverse scientific group. The bulk of the meeting time was spent in two breakout sessions, with each session focused on one of the two objectives. For each breakout session, participants were divided randomly into two groups and each group received the same charge questions. In this way, we provided an opportunity to obtain diverse input from the participants, optimize the groups' sizes for discussion, and

enhance interdisciplinary interactions. The co-chairs moderated the breakout sessions and a report-back period followed in which the entire group discussed the ideas generated.

Prior to the workshop, participants were asked to provide background materials in two forms: 1) completed questionnaires describing one or more simulation codes they use in their research, and 2) an optional two-page position paper describing a particular multiscale simulation problem (and potential solution approach) of interest. Code descriptions were also solicited from a selected set of non-participants prior to the workshop to obtain background information on codes important to BER programs, but not specifically represented by a workshop participant. Thirteen position papers (Appendix C) and twenty-three completed code questionnaires (Appendix D) were received.

The group was comprised approximately one half from CESD areas of research (Subsurface Biogeochemistry Research (SBR) and Terrestrial Ecosystems Science (TES) programs), and the other half from BSSD areas of research (Genomic Sciences Program). BER program managers commented this was the first time to their knowledge that representatives from these two computationally-oriented communities met to discuss opportunities for collaboration and linkages among their various modeling efforts.

2.0 Multiscale Science and Modeling

Extensive research efforts have been invested in reducing model errors to improve the predictive ability of biological and earth systems simulators. While the bulk of this research has focused on improving model parameterizations in the face of observational limitations, the more challenging type of model error/uncertainty to identify and quantify is model structural error. Such error arises from incorrect mathematical representations of (or failure to consider) important physical, chemical, or biological processes, properties, or system states in model formulations. While improved process understanding can be achieved through scientific study, such understanding is usually developed at small scales. The manifestation of small-scale processes at system prediction scales is confounded by non-linear interactions and heterogeneity that lead to *emergent behavior* in complex systems – behavior that cannot be predicted based on understanding the individual small-scale processes alone. Large-scale process-based models greatly oversimplify this complexity, leading to a potentially large error in the model structure and conceptualization. Unfortunately, this type of error is often obscured by parameter error/uncertainty, because these models typically include sufficient degrees of parametric freedom to allow successful reproduction of historical observations (calibration) even when the underlying model is incorrect. This gives a false sense of confidence in model predictions, which may be inaccurate even though the historical match is excellent. The potential importance of model structural error in land surface modeling (as an example) is emphasized by the study of Abramowitz et al. (2007), who performed numerical experiments to evaluate prediction errors in a synthetic system. Their results showed "as much as 45% of…error in these flux outputs is due to systematic problems in those model processes insensitive to changes in vegetation parameters." They concluded "efforts to improve the representation of fundamental processes in land surface models, rather than parameter optimization, are the key to development of land surface model ability."

Process-based numerical models are typically designed for a particular characteristic length and time scale; except for the most fundamental models (e.g., quantum chemistry), it is necessary to introduce approximations to describe complex systems or processes. Even for molecular chemistry, it is often necessary to make empirical simplifications because of computational constraints. This approach has been the best available to date because of 1) limited computational power, 2) limited understanding of processes and how they are coupled, and 3) a lack of rigorous frameworks for upscaling processes. However, computational power is increasing such that national laboratories and federal agencies are already considering applications and designs for future exascale computers. At the same time, our understanding of biological and environmental processes at fundamental scales has been accelerated, but using this information to advance our knowledge of the larger system behavior requires the development of multiscale simulators. The lack of rigorous frameworks for upscaling undermines the use of simulation to test the impact and coupling of chemical, physical, and biological processes. In the future realm of exascale computing and beyond, simulators will be based on first principles and accurately describe processes at a predictive scale; however, this cannot be achieved unless we advance our capabilities in multiscale techniques.

2.1 Multiscale Modeling Approaches

Recognizing this need and opportunity, focus on multiscale modeling has dramatically increased in the last decade. For example, the journals *Multiscale Modeling & Simulation* and *International Journal for Multiscale Computational Engineering* both published their first issue in 2003. Of course, scientists and engineers were not unaware of scale issues prior to that, but because of limitations in computational capability most attempts to bridge scales were limited to a broad class of methods known as upscaling. Upscaling approaches posit a microscale model, then invoke some type of formal averaging procedure (e.g., volume averaging, homogenization, mixture methods) to derive an effective macroscale model that preserves selected averaged quantities. Once the transformation of equations and parameters from the microscale to

the macroscale is prescribed, single-scale simulation is subsequently performed only at the macroscale without further reference to microscale processes or properties. Wood (2009) points out upscaling requires imposition of one or more "scaling laws", also known as closure approximations. A scaling law is an axiomatic statement about the essential character of the microscale system that allows reduction of the number of degrees of freedom and closure of the macroscopic equations. Typical scaling laws include assumptions about the statistical structure of microscale variables (e.g., statistical homogeneity, stationarity, and ergodicity), separation of scales, the magnitude of local fluctuations, and the nature of boundary conditions (e.g., infinite or periodic). When necessary closure approximations are valid, the microscale and macroscale models can be completely decoupled, and valid macroscale models and parameters can be defined that eliminate the need for explicit microscale knowledge. However, for complex systems involving multiple non-linear coupled microscale processes and strongly heterogeneous system properties, available closure approximations are often overly simplistic and not applicable.

As numerical simulation capabilities advanced, driven by continued increases in available computational power, a second class of multiscale simulation methods emerged known as multiresolution methods. These methods, referred to as "traditional" multiscale methods by E et al. (2003), provide computationally efficient ways of obtaining a complete solution to a microscale model. These methods are multiscale in that they use approximate upscaled (macroscale) models in intermediate steps to facilitate efficient computation of the microscale solution. Examples of multiresolution methods include multigrid solvers and preconditioners (e.g., Wesseling 1992; Trottenberg et al. 2001), multiscale finite element methods (e.g., Hou and Wu 1997; Jenny et al. 2003; Aarnes et al. 2005), and multiscale mimetic methods (e.g., Lipnikov et al. 2008).

However, in many problems it is simply not feasible, even with multiresolution approaches, to solve the microscale problem of interest. These problems have recently led to a third class of multiscale methods in which microscale and macroscale models are explicitly coupled in some manner. The simplest form of coupling is the use of a microscale simulation to define effective parameters for a macroscale simulation. These methods, referred to as parameterization, coarse-graining (CG) or numerical upscaling methods, assume microscale and macroscale simulations can be decoupled and information flows only in one direction (from the fine scale to the coarse scale). Recognizing that in many cases the effective behavior of a microscale system is dependent on macroscopic states, the most complex multiscale approaches invoke a two-way coupling of microscale and macroscale models. These methods, called hybrid multiscale methods, are those that combine two or more models defined at fundamentally different physical length scales within the same overall model spatial and temporal domain. In most cases, the models also have fundamentally different ways of representing the physical, chemical, and biological processes. For example, several models in the materials science literature couple molecular dynamics (MD) simulations at the molecular scale to continuum mechanics simulations at larger scales.

Scheibe et al. (2015) provide a more detailed review of the multiscale modeling approaches described above, including examples from several scientific domains. They also propose a classification scheme for methods and an approach for identifying which class of methods is best suited to a particular problem. Keyes et al. (2013) provide a thorough review and discussion of the relationship between multiscale and multiphysics models, and numerical methods for model coupling.

Workshop participants provided information on multiscale modeling methods they are pursuing in their research, both in position papers submitted and workshop discussions. Details and examples are provided in the position papers. (Appendix C)

3.0 Biological and Environmental Research Opportunities

The morning breakout session focused on the first objective: *Identify BER-relevant models and their potential cross-scale linkages that could be exploited to better connect molecular-scale research to BER research at larger scales.* Accordingly, the breakout session groups were assigned the following tasks:

- Catalog a suite of existing codes relevant to the BER science scope, identifying their critical features, existing linkages if any, and areas for potential new linkages up or down scale.

- Identify and prioritize linkages that could be made between existing codes within short-, intermediate-, and long-term time frames.

Both groups initially struggled to answer the charge; two preliminary hurdles needed to be overcome before the groups could dig deeper into specific codes and linkages. First, it was noted that focusing on codes was premature without considering basic science issues. Second, because participants had diverse backgrounds, some discussion was needed to get everyone on the same page (understanding different jargon and perspectives). As a result, some discussion around the charge continued during the first part of the second breakout session. A summary follows with key discussion points and recommendations of both breakout groups related to the breakout session one charge, whether those came from the first or second breakout session.

3.1 Science Challenges

Several overarching domain-oriented scientific challenges that motivate and/or impact multiscale model integration were identified:

- For reactive transport models, fundamental reaction rates cannot be used to predict at larger scales where mixing is incomplete. Subsurface systems are characterized by evolving (nested) scales of heterogeneity for which there often exists no clear scale separation. As a result, simplifying assumptions needed to obtain macroscopic phenomenological models are often violated and there is a need to explicitly consider finer scales.

- Strong disparities exist in time scales of various models, which can be a practical hurdle when trying to model multiple physical length scales simultaneously.

- For metabolic modeling, there are too many reactions within a cell, each with associated variable reaction rates. It is impossible to measure or infer all these rates and they are not static.

- For microbial community modeling, a daunting challenge is how to combine the effects of large numbers of different organisms within a natural system (e.g., a soil).

- We often do not know for a given model application the conditions under which coarse-scale models are inadequate, and what fine-scale information is specifically needed. When does coupling need to be tight vs. loose, two-way dynamic vs. one-way static?

- In some cases we are unable to measure variables that we can simulate at the microscale. A coarse (low-dimensionality) observable cannot be used to validate a fine model output.

- In many cases, atomistic simulations require extensive simulation of "typical" conditions before something "interesting" happens. Identifying rare events and developing ways to minimize computation of routine conditions is an important challenge.

3.2 Code Base

Twenty-two detailed code descriptions were provided by the workshop participants (and by selected owners of codes significant to BER research, but not represented by a workshop attendee) in response to a questionnaire from the workshop organizers. The detailed descriptions are provided in Appendix D. These codes fall into four broad categories:

1. Atomistic to coarse-grained scale computational chemistry/biochemistry codes:

   - AMBER: molecular mechanics (MM) and quantum mechanics (QM)/MM MD.

   - Accelerated Molecular Dynamics Framework (AMDF): Long time-scale atomistic simulations.

   - CHARMM: Biomolecular simulation.

   - GROMACS: MM/MD.

   - NAMD: MD.

   - GAMES: Quantum chemistry.

   - GAUSSIAN: Quantum chemistry.

   - NWChem: Computational chemistry. Simulates chemical reactions; electronic structure of molecular assemblies, biomolecules, nanostructures, and solid-state; ground- and excited-state chemical processes; molecular properties and relativistic effects.

   - protoMD: Self assembly, structural transition, and chemical reactions in nanoscale supramolecular assemblies.

   - RAPTOR: Rapid Approach for Proton Transfer and Other Reactions (add-on to LAMMPS). Performs reactive MD simulations.

   - REACH: Coarse-Grained Molecular Dynamics.

   - UCG-MD: Performs highly scalable coarse-grained MD simulations, especially "ultra" coarse-grained models.

2. Cell-scale metabolic modeling codes:

   - Boltzmann: Simulation of sets of coupled reactions using statistical thermodynamics/fluctuation theory. Applications in cell metabolic modeling as well as nanotechnology and materials.

   - COBRA: Metabolic network modeling. Simulates intracellular metabolic reaction rates, cellular growth, cellular consumption/production rates. Based on genome-scale reaction networks and flux balance analysis methods.

   - PLAS: Biological systems analysis. Simulates biological processes, typically metabolic pathway systems.

3. Pore-scale fluid flow, solute transport, and biogeochemical reaction simulators:

   - Unnamed code: Pore network model for pore-scale simulation coupled to continuum-scale simulators for single-phase fluid flow (Newtonian and non-Newtonian fluids), transport, and reactions.

   - ChomboCrunch: Pore-scale single-phase fluid flow and reactive transport. Note: This code was mentioned in the description of CrunchFlow/EcoTrait (see below) and a separate description was not provided.

   - Reactive transport Lattice Boltzmann Method (LBM): Simulates pore-scale multiphase flow and reactive transport, including liquid/gas flow with phase transition, heat transfer, and mineral precipitation/dissolution.

- ○ Smoothed Particle Hydrodynamics (SPH) and Dissipative Particle Dynamics (DPD) modules in LAMMPS: Multiscale, multiphase, and multiphysics modeling using SPH and DPD methods. Simulates flow of complex fluids (non-Newtonian fluids, plastic materials, and suspensions), soft matters (polymers, colloids, and microbial cells), and moving interfaces and boundaries.

- ○ Taxila LBM: Pore-scale multiphase fluid flow with limited transport and reaction capability.

- ○ Transient Energy Transport HYdrodynamics Simulator (TETHYS): Simulates laminar and turbulent incompressible fluid flow, energy transport, and solute transport. Multicomponent reaction capability under development.

4. Continuum-scale fluid flow, solute transport, and biogeochemical reaction simulators:

- ○ Community Land Model (CLM)/PFLOTRAN: Land surface model that includes 2-D surface flow and 3-D subsurface flow and reactive transport processes.

- ○ CrunchFlow: Simulates single-phase fluid flow, transport, and (bio)geochemical reactions at multiple scales (pore to field scales). See also ChomboCrunch description above under pore-scale simulators.

- ○ CrunchFlow/EcoTrait: Microbial trait-based reactive transport. Simulates variable density flow, advective-diffusive-dispersive transport, and multicomponent reactions including microbially mediated reactions with thermodynamic considerations.

- ○ Distributed Hydrology Soil Vegetation Model (DHSVM): Distributed, physics-based approach to solve coupled energy and water balance equations describing watershed processes at high spatial (10 to 90-m) and temporal (hourly) resolution.

- ○ eSTOMP: Extreme-scale subsurface transport over multiple phases. Simulates nonisothermal multiphase flow and multicomponent reactive transport, coupled geomechanical deformation.

- ○ MOSART: Model for scale-adaptive river transport. Simulates reach-scale water flow in rivers, including heat transport and biogeochemical fluxes.

- ○ PFLOTRAN: Subsurface reactive multiphase flow, heat transfer, multicomponent solute transport and biogeochemical reactions, geomechanics, and multiple interacting continua.

3.3   Opportunities for Multiscale Linkages

The breakout group participants identified a number of opportunities for linkage of the above codes (and others) in BER-relevant science applications. One breakout group went so far as to identify and prioritize seven concepts:

1. Bacterial organization within a cell, including mesoscale phenomena and compartmentalization, bacterial membranes, flagellar motors, cytoskeleton, and other organelles. Fine-scale methods could include CG molecular-scale methods such as implicit solvent/brownian dynamics, linked to mesoscale models. It was noted that mesoscale codes currently are in their infancy, although experimental data are rich (e.g., BSSD mesoscale imaging pilot program). The proposed approach would connect models to experiments across scales (tomography and imaging), and consider self-organization of function (not structure only). An example that was discussed more extensively in the second breakout session is the exchange of genetic information in yeast cells by a process initiated by pheromone reception. This process involves restructuring the cytoskeleton that changes the dynamics in a way that cannot be effectively modeled by a standard reaction/diffusion process. For this example, the approach would use an agent-based structural model of filaments composing the cytoskeletal structure to inform a coarser reaction/diffusion model in a dynamic two-way multiscale approach. Another example discussed in the

second breakout was bacterial cell wall polymorphism and bacterial outer membrane transport. For this example, doing a complete simulation with MD resolution is computationally impossible, but simplification could be achieved by coupling transport to a diffusional model at coarser resolution.

2. Metabolic pathway models informed with kinetics/functionality defined from free energy calculations (fluctuation theorem) and chemical potential values (at steady state). The proposed approach would implement a multi-agent simulation (e.g., Biocellion code), where each agent represents an idealized model of metabolism (e.g., Boltzmann code), with free energy calculations performed by a computational chemistry code (e.g., NWChem).

3. Simulate multiscale dynamics of metabolic pathways (regulation, non-steady-state). The proposed approach would perform cellular-scale simulations informed by measurements of finer-scale processes (specifically, time series of metabolite levels) derived from stimulus/response experiments.

4. Extend rare event models to longer time scales. It was noted that at the atomistic scale, milliseconds are considered long time-scales, but a similar approach could be applied to problems with other length and time scales. The proposed approach has been previously applied to fracture propagation in materials. Some BER-relevant extensions were identified including diffusion in clays and precipitation/dissolution reactions, possibly carbon sequestration. To implement this approach, we need to identify a critical spatial gap, i.e., a rare event.

5. Coupled 3-D reactive transport/land surface modeling. This is currently being done on the Next-Generation Ecosystem Experiment (NGEE-Arctic) and Accelerated Climate Modeling for Energy (ACME) projects, but over the longer term could be generalized and broadened. In order to refine existing approaches, we need to know what difference inclusion of 3-D reactive transport models makes in flux predictions, and under what contexts and for which variables?

6. Dynamical non-linear systems in which an infinitesimal difference in initial conditions leads to divergent behaviors (chaotic systems). The proposed multiscale approach is equivalent to statistical sampling of an ensemble of MD simulation trajectories.

7. Coupling pore-scale simulators with reservoir- or aquifer-scale models. A critical area that was identified for this was better representation of pore-scale geometry modifications coupled to geochemical reactions and geomechanical processes, but other applications could benefit as well.

The group provided an initial prioritization of these seven concepts in the form of a plot shown in Figure 3.1. In that figure, the seven items are ranked in terms of perceived potential impact (vertical axis) and perceived difficulty (cost/time required to achieve progress – horizontal axis).

The other breakout group raised three concepts, but did not discuss them in detail:

- Development of a protein homology model database to provide ligand binding information to plug into metabolic models.

- Hydrology/plant/microbial model interfaces.

- Scalable CG algorithms including load balancing.

Additional ideas were introduced and discussed during the second (afternoon) breakout session, but are included here to maintain
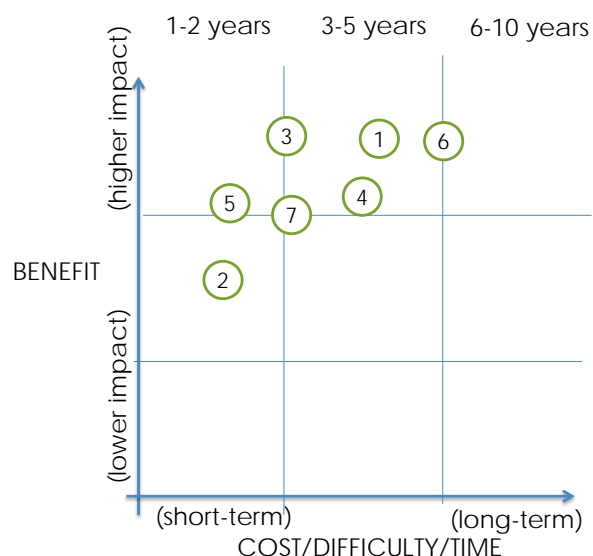
Figure 3.1 Diagram indicating prioritization of the seven target multiscale concepts/problems.

clearer organization:

- Linkage of molecular-scale fluid model and CFD simulation to evaluate critical hydrodynamic conditions (such as the assumption of no-slip boundary condition in pore-scale models).

- Modeling flux through a biochemical pathway (e.g., series of enzymes mediating a conversion). Pre-compute MD and/or CG simulation results for each individual enzyme in the pathway. To model interaction of two enzymes, use Brownian dynamics with a master equation approach based on the tabulated properties.

- Modeling dynamics of intracellular molecules (sub-nanometer up to subcellular scale) to simulate the environment for intracellular reactions. This requires methods to decouple models of reacting molecules at atomic-level detail with CG models of the intracellular environment (crowded or dilute) through calculation of potential mean force of reacting molecules in the cellular context. For example, one might treat proteins as rigid body molecules and simulate movement, then model interaction between two proteins in the cellular context as one step in a metabolic pathway.

- Modeling viruses and/or ribosomal systems. The proposed approach would use Newtonian mechanics (slow processes) to constrain atomistic simulations (e.g., MD). Using methods similar to equation-free modeling, one would obtain the rate of change of macroscopic variables from microscale simulations and then advance them in time, iterating between the two scales.

Significant time was spent in the morning discussing specific codes among the multidisciplinary participants to support improved understanding of potential linkages. Selected proposed linkages provided by code owners in the code questionnaires follow (see Appendix D for full descriptions):

- CLM/PFLOTRAN:
  - Investigate impact of explicit representation of 3-D subsurface flow and transport processes on land surface fluxes of water, energy, and carbon.
  - Aid in the development of reduced order models for representation of subsurface processes (Pau et al. 2014).

- AMDF:
  - Linkage to mesoscale using diffusive boundary conditions to exchange mass.
  - Linkage to mesoscale using CG MD approaches. Roughly corresponds to coupling to an FEM code for far-field elasticity.
  - Linkage to mesoscale by providing parameters such as equations of states, or catalogue of nanoscale processes and associated rates.
  - Linkage to mesoscale using equation-free/HMM multiscale methods. MD becomes a computational kernel for these higher scale methods.
  - Linkage to quantum calculation by doing *ab initio* MD (forces computed from quantum mechanics, classical equations of motion) or path integral MD (effective quantum equations of motion).

- Pore-Scale/Continuum-Scale Coupling:
  - Tightly coupled hybrid simulations have been performed (Mehmani et al. 2012, Mehmani and Balhoff 2014).
  - Hierarchical upscaling of macroscopic parameters (e.g., reaction rates, dispersion coefficients, permeability) is also possible.

- Boltzmann:
  - Potential linkage to 10^{-6} to 10^{-3} scale – QM-MM enzymatic reactions.
  - Linkage to simulation at $>10^{-6}$ scale – nutrient flow/changing environmental conditions.

- NWChem:
  - Hybrid QM/MM formalism: all types of *ab initio* methods can be used for describing quantum region including many-body perturbation theory, coupled cluster, density functional theory, and time-dependent density functional theory formulations for excited state processes.
  - Novel algorithms to extend temporal scales.
  - Extending length scales through continuum models (Classical DFT/RISM).
  - Interfaces to large-scale MD codes (Amber/LAMMPS).

- Reactive Transport LBM:
  - Has been coupled in hierarchical fashion to continuum reactive transport simulations (Lichtner and Kang, 2007).
  - Has been coupled in concurrent fashion to continuum reactive transport simulations (Coon et al., 2012).
  - Has been coupled in concurrent fashion to finite volume method (Navier-Stokes Equation) for reactive transport in micro reactors (Chen et al., 2013).
  - Has been coupled in hierarchical fashion to density functional theory calculation to study effect of chemical evolution on radionuclide leaching (Zhang et al., 2013).

- Taxila LBM:
  - Can be used to characterize retention curves for multiphase flows.
  - Can be used to inform interfacial area upscaling strategies.

- AMBER:
  - Connect QM with MM simulations, both still at atomic resolution (QM adds electrons) but allow atoms to move smoothly between the QM and MM region in a way that obeys detailed balance, conserves energy and angular momentum.
  - Linkage to CG and mesoscale simulations would be very interesting if the interaction between the two regions can be accurately defined and particles can move between the various scales in a way that obeys detailed balance, and conserves energy and angular momentum.

- CrunchFlow-EcoTrait:
  - CrunchFlow has been linked to Chombo for use as a geochemical engine that leads to the largest ever reactive pore scale simulations (Molins et al., 2012; Steefel et al., 2013; Molins et al., 2014) used to evaluate continuum scale (upscaled) rates. Chemical process models can be applied at different scales with conceptual adjustments (e.g., surface area as a bulk parameter at continuum scale, or as an explicit reacting boundary at pore scale).
  - Micro-continuum modeling (continuum modeling at micron scale) to understand evolution of reactive surface area (Noiriel et al., 2012), or evolution of porosity and diffusivity (Navarre-Sitchler et al., 2009). In both instances, parameters are applicable at larger scales.

- ○ The EcoTrait biological module can be applied across multiple scales when coupled to a geochemical engine like CrunchFlow, which can then couple to continuum flow and transport codes (e.g., Amanzi, Parflow).

- eSTOMP:

  - ○ The reaction module has been coupled with genome-scale metabolic models by exchanging fluxes between models at different scales.

  - ○ It can be coupled to pore-scale simulation to resolve mixing-controlled reactions at a scale where well-mixed reaction rate formulations are appropriate. For example, effective reaction rates simulated at pore scale can be provided to eSTOMP to simulate continuum-scale phenomena (e.g., Scheibe et al. 2014).

- SPH/DPD in LAMMPS:

  - ○ By analogy to MD, a broad class of particle methods facilitate concurrent coupling of complex physics and chemistry across multiple scales, i.e., from atomic to continuum scales.

- PFLOTRAN:

  - ○ Has been coupled with Lattice Boltzmann for pore-scale simulation (experimental).

  - ○ Has been coupled with CLM to study impact of 3-D subsurface flow and transport processes on land surface fluxes of water, energy, and carbon.

  - ○ Linkage to fine scale nuclear waste package degradation models under development

  - ○ Embedded multi-continuum model for resolving fine scale heat transfer and reactive transport.

- COBRA:

  - ○ Input connections: Extracellular concentrations derived from macroscale transport models (Tartakovsky et al., 2013); kinetic parameters derived from protein models.

  - ○ Output connections: Fluxes can be used in transport models (estimating biotic transformation rates); can be used in bioprocess models (e.g., reactors); fluxes can be used to model interactions in microbial communities (e.g., photosynthetic mats).

- TETHYS:

  - ○ Linkage to simulation at larger spatial scales is typically accomplished by providing boundary conditions from the larger scale model to inflow or outflow boundaries. Linkage to watershed, river basin scales is possible in this manner.

  - ○ Linkage to atmospheric forcing models can be accomplished through one-way coupling. Loose coupling could also be done via file sharing methods.

- CrunchFlow:

  - ○ Currently we are developing reactive transport modules to link with the hydrology and land-surface interaction code FLUX-PIHM at the watershed scale. There is interest in connecting subsurface biogeochemical processes to continental scale or global scale.

- DHSVM:

  - ○ Has been coupled with detailed hydrodynamic models to provide the influx of water and sediment to river system.

- ○ Has been coupled with regional climate and weather models for hydrologic forecasting and to evaluate the impacts of climate change on hydrologic processes.

  - ○ Has been coupled with detailed groundwater models to provide topographically-driven surface boundary conditions.

  - ○ Can be coupled with dynamic vegetation models.

  - ○ Can be used to estimate parameters for macroscale land surface models (e.g., GCMs).

- protoMD:

  - ○ Self-consistent use of field variables as the CG descriptors allows seamless and rigorous coupling to reaction-transport-mechanical continuum codes.

  - ○ Multiscale algorithms can be developed to naturally integrate simulation with a variety of nanocharacterization experimental data techniques.

- LAMMPS-Raptor:

  - ○ Robust multistate reactive models (i.e., force field parameters) can be derived from condensed phase *ab initio* simulations at smaller scales, (10 Å)3 and 10–100 picoseconds.

  - ○ Atomistic reactive simulations can be used to derive input for large-scale CG models (e.g., position-dependent diffusion constants in smoothed-particle hydrodynamics) to access larger length and time scales (mesoscopic).

- UCG-MD:

  - ○ Information from more detailed simulations (ab initio, all-atom) is typically used as the basis for CG molecular structures and force fields.

  - ○ The Voth group is actively exploring rigorous approaches to couple CG-MD simulation results to mesoscale representations (smoothed particle applied mechanics) and field theoretical approaches.

3.4 Critical Issues in Multiscale Modeling

In the process of discussing potential opportunities for multiscale linkages, the breakout groups also identified several important model-focused issues/challenges that cross-cut various application domains.

- How can multiscale models be tested and validated? A set of standards are needed for error propagation and uncertainty quantification to make multiscale modeling more systematic and as verifiable as possible. The models must make predictions that can be tested experimentally (or in some cases against other computational methods).

- A process is needed for deciding which model(s) and multiscale approaches are appropriate for a given problem. The example given was KBase, in which various pieces of data or models can be linked together in different ways. KBase has a narrative (like a decision tree) that guides users to appropriate methods.

- What is the appropriate level of complexity for multiscale models? In general it is desirable to make a model as simple as possible to achieve desired predictive power, but no simpler. Formal mathematical methods can provide insight into when more or less complex couplings are required.

- Loose coupling methods commonly used to link models non-invasively are difficult to analyze numerically. In some cases, it may be difficult to prove a model system will converge to a solution and to determine what conditions should be imposed for stability and accuracy (grid sizes, time steps).

4.0   Computational and Data Needs

The afternoon breakout session focused on the second objective: *Identify critical science directions that will motivate EMSL decisions regarding future computational (hardware and software) architectures.* Accordingly, the breakout session groups were assigned the following tasks:

- Identify new mathematical algorithms and software frameworks needed to support development of multiscale linkages across the BER code suite.

- Identify critical data needed to support integrated multiscale simulations.

- Identify key attributes of computational hardware needed to perform integrated multiscale simulations.

4.1   Algorithmic Considerations

Several important potential approaches and associated algorithm considerations were identified by the breakout groups:

- Interdisciplinary communication: Multiscale coupling often involves integrating models developed by different scientific disciplines (e.g., fluid dynamics and microbial metabolism). Scientists know their own algorithms and codes (and their limitations) well, but communication to scientists from other disciplines is often difficult.

- Sharing large files between different codes can lead to inefficiency in computational algorithms. Therefore, the definition of standard interfaces is a critical element of new algorithms (e.g., LAMMPS interface). It would also be very useful if application codes and computational architectures could address this issue. For example, CLM/PFLOTRAN coupling uses virtual files stored in RAM rather than disk files, which is much more efficient but requires a wrapper around both codes. Domains may be decomposed or discretized differently in various models, creating another challenge when sharing large datasets.

- Many alternative codes exist that solve the same or similar problems. Therefore, interoperable exchange of codes in a multiscale framework is a key attribute. Model interfaces must be defined as code-independently as possible to facilitate flexibility for long-term solutions. Model interfaces (i.e., APIs) should also have flexibility to accommodate multiple alternative conceptual models at each scale of interest, which again requires model-agnostic interfaces.

- In general, multiscale science is not yet mature. Some seemingly obvious (trivial) couplings exist, but are typically ad hoc and not general and physically rigorous. For most problems we do not know the conditions under which fine-scale information is specifically needed (or not), and when coupling needs to be tight or loose, two-way dynamic or one-way static. In the broader domain modeling communities, multiscale modeling methods are not well known or understood.

- In some cases, two-way multiscale simulation could be accomplished without tight coupling of codes performing linked simulation in real time. An approach was suggested in which tight coupling would be replaced with loose coupling through a lookup table method. In this approach, a set of coarse-scale conditions is prescribed for which fine-scale behaviors are pre-computed and stored for later use in coarse-scale simulations.

4.2   Data Needs

It was noted by both breakout groups that there is an important need for multiscale experimentation to provide a basis to evaluate and validate multiscale simulations. Strong communication between experimental and computational researchers is required to ensure the right variables are measured at the correct scales. Transient and dynamic measurements are likely

to provide much greater value than steady-state observations. In some cases, "data" could also take the form of high-resolution simulations performed to provide a numerical standard (benchmark) for testing alternative multiscale approaches. In this context, the discussion of data needs in the second breakout session took the form of a brainstorming session in which participants highlighted a number of data needs specific to their application areas of interest:

- Fluorescent labeling of proteins and imaging localization within intracellular environment.

- Cryo-EM and other microscopies to obtain low-resolution shapes of molecular complexes (quaternary structure).

- Adaptive bio-imaging (various spectroscopic methods) to determine localization of proteins in the cell membrane.

- HR mass spectrometry to determine the spatial distribution of individual molecules (e.g., metabolites).

- Improved tomographic imaging of pore-space geometry of soils and aquifer materials.

- Information about the structure of other porous materials (e.g., cell membranes).

- Measurements of relative uptake of metabolites by different members of a microbial community (intermediates between organisms hard to observe as they are taken up rapidly).

- Measurements needed to infer regulatory patterns, such as metabolic time series data (e.g., NMR methods). Additional spectral libraries are needed to identify unknown compounds.

- Measurements of soil carbon (dissolved and solid phase).

- Improved and more extensive measurements of CO_2 fluxes.

4.3 Hardware Needs and Recommendations

The final part of the second breakout session considered alternative computational architectures to support multiscale simulation. The discussion focused around three primary points:

1. The scientific computations relevant to BER science are highly diverse and a "one size fits all" approach to computational architecture is not likely to be effective. For example, some problems are very well suited to hybrid CPU-accelerator architectures (e.g., MD), others require large memory, while yet others need architecture to support coupled simulation and visualization of large datasets. Given this consideration, it was strongly recommended that EMSL consider purchasing a computational system that incorporates multiple node types focused on different application needs.

2. More of our applications are requiring advanced analysis of large datasets in conjunction with compute-intensive applications. An example is the generation of time-dependent probability distributions from large numbers (e.g., 10,000) of computational chemistry trajectory simulations. It would be advantageous to perform data-intensive post-processing "where the data are." This consideration again points to the need for a heterogeneous system, in this case with a portion of the system designed for compute-intensive and part designed for data-intensive (e.g., Hadoop) computations. For the example discussed, instead of writing all the trajectories to disk and transporting data to another system for subsequent post-processing, the trajectories could be communicated to distributed disk or memory on the data-intensive portion of the system and immediately analyzed using a map-reduce algorithm.

3. Since many problems of interest are biological in nature, the question was posed: "How does biology do it?" That is, can we design and build computational architectures that reflect the nature of the problem we are trying to solve? Purpose-built systems for computational chemistry have been developed (e.g., D.E. Shaw), but are more expensive than multipurpose architectures based on commodity components. Some consideration should be given to potential implications of multiscale modeling needs in this context.

5.0   Future Strategy

5.1   Biological Systems Science

Many important processes in the energy and environmental sciences involve biological organisms, particularly microbes. Research on microbial systems has led to a massive influx of "big" data produced by high-throughput experimental approaches. These data, as corralled by the DOE Systems Biology Knowledgebase (KBase), have yielded much insight into biological systems. However, whereas systems biology efforts until now have largely been in the realm of large-scale cataloguing efforts (e.g., genomics, proteomics, metabolomics, etc.), the need has now transformed to understanding how these components work together i.e., constructing working physical models from the "parts list."

In BSSD, scientists are using genome sequence and other comprehensive datasets (molecular, spatial, and temporal data) to build models of signaling networks, gene regulatory circuits, and metabolic pathways. However, static systems models will need to be assimilated into multiscale simulation engines capable of exploring hypotheses relating micro- and macro-scale system function and dynamics, understanding how processes on smaller scales lead to larger-scale phenomena (Figure 5.1). The challenge is to feed biological big data into simulation models that incorporate spatiotemporal elements. One way of thinking of this is a transformation of omics into 3-D models evolving with time.

Systems biology goes beyond the consideration of single macromolecules to obtain holistic information about interacting biological systems, such as metabolic networks, genomics and proteomics. However, an important element missing from these descriptions is a 3-D picture of the systems involved. This 3-D view requires knowledge of macromolecular structure and dynamics, the positions of sub-cellular structures in the cell, the positions of macromolecules with respect to these expressed structures, and the distribution of smaller species such as solvent, metabolites, and ions. Once a molecular-level 3-D model of the cell is established it will be necessary to use simulation to follow the evolution of this system with time.

This thrust will need to be pursued in a concerted, multidisciplinary effort with tight coupling between experiment and simulation. DOE is ideally placed for multiscale research on the nanoscale through the millimeter scale, by virtue of its large-scale facilities for molecular research, including next-generation synchrotron radiation, neutron scattering and supercomputing. The optimal use of these facilities in the service of multiscale simulation will be of increasing value.

Creating simulation models that correctly predict cellular system response to external perturbations requires many quantitative detailed inputs such as protein structures, kinetic constants, enzyme activities, and dynamic metabolic processes. Some of these will themselves need to be derived computationally. The simulation goal presents a formidable challenge, but emerging computational research approaches

Figure 5.1 Alcohol resistance in a mutant of Clostridium thermocellum was traced to just two amino-acid residues in the enzyme alcohol dehydrogenase (Brown et al. 2011). Understanding the cell-scale repercussions of molecular scale perturbations is a goal of multiscale simulation.

provide new opportunities to bridge the knowledge gap of complex systems and facilitates scaling concepts and data across multiple levels of biological organization. Interfacing simulation methods at different scales by informing larger-scale, coarse-graining (CG) methods with results from finer-detail computations will lead to a self-consistent description of mechanisms in and between cells.

The scaling and use of multiscale codes on DOE supercomputers will be integral to the success of this approach. The march of supercomputing towards the exascale will furnish major opportunities for detailed simulation. The major challenge, then, is to design methods that incorporate detailed molecular, biochemical, physiological, and structural information into accurate systems-level biological models and simulations while using the variable and sometimes esoteric architectures of modern-day high-performance computing.

Computer modeling and simulation can rationalize experimental information across scales. Essential aspects of simulation development will be 1) integrating available experimental information obtained from experimental imaging techniques and omics-based systems models, into simulation models, 2) testing and improving simulation methods by comparison with additional obtained experimental information, and 3) interfacing of simulation methods at different scales, by informing larger-scale, CG methods with results from finer-detail computations. The hoped-for result will be a self-consistent, description of mechanisms in and between cells. The scaling and use of these codes on supercomputers will be integral to the success of this approach.

Multiscale simulation in biological systems science extends upwards from quantum chemistry. Indeed, the 2013 Nobel Prize in Chemistry was awarded for the development of multiscale computational techniques that couple quantum and molecular mechanical systems (Smith and Roux 2013) and these mixed, QM/MM techniques have been applied to study systems of special interest in the energy biosciences (Parks et al. 2009). Enzyme reactions are critical to systems biology. The QM/MM techniques for computing reaction mechanisms and the associated energetics need to be improved and accelerated to the point where they provide missing information in metabolic pathway maps and guide rational engineering of enzymes. Indeed, a major challenge for the future will be how to accurately incorporate *chemical reactions*, which occur on the femtosecond timescale into cell-level models evolving on timescales many decades slower.

The majority of molecular-scale calculations in recent years have been performed with the MD technique, using empirical force fields. Force fields for these simulations need to be improved. Improvement in computer power and algorithmic simulation efficiency is of no use if interactions are inaccurate. This requires the comparison of computed quantities with experimental and quantum-chemical data obtained on smaller scales in a self-consistent manner.

Atomistic simulations provide information on the response of macromolecules to ligand binding, and on the diffusion and transport of metabolites and proteins in crowded cellular environments. An important foundation for multiscale systems simulation is the knowledge of macromolecular structure and dynamics, the positions of sub-cellular structures in the cell, the positions of macromolecules with respect to these expressed structures, and the distribution of metabolites and ions. Supercomputing now permits the simulation at atomic detail of systems up to 100M atoms in size and on the microsecond timescale. Hence, atomistic MD has moved well beyond the single-molecule level to permit systems-level simulation of hundreds of interacting biological macromolecules such as those involving transport of chemical signals across the cellular membrane. Indeed, extrapolation of current performance at the petascale to the exascale indicates we would ultimately perform MD simulations of systems of $\sim 10^{11}$ explicit interacting atoms, i.e., approximately the number of atoms in a bacterial cell for about 10 microseconds. Existing genomic and structural data, together with spatially-resolved data provided by experimental techniques proposed above, will provide information on starting configurations for large-scale simulations of subcellular systems.

However, serious limitations to MD exist, including a severe scaling problem limiting system sizes and timescales. Work is needed to overcome this obstacle. Approaches that will need to be developed to permit scaling up of MD include methods to *improve sampling* to more efficiently cover the phase space of systems of interest. For this, methodological and algorithmic improvements aimed at 'ensemble-based' simulation of sub-cellular systems will be required, including the rapid calculation of electrostatic interactions. Some of these approaches will involve independent sampling of regions of phase space followed by intelligent recombination. Furthermore, some of these approaches are inherently scalable and may be developed for use on exascale machines.

Improved, multiscale atomistic MD sampling methods are needed to understand the mechanisms of function of enzymes and macromolecular machines critical to the energy biosciences. These methods and algorithms will need to efficiently use the full available computational power to understand folding, allostery, binding, and reaction. Thermodynamics and energetics at the single-protein level are also critical to understanding cellular energy flow. We need to master structural controls of redox potentials of electron donors and acceptors, and to refine methods for calculating free energy profiles of enzymatic reactions. We need to develop methods for reliably predicting protein structures (in the absence of a close structural homolog), protein:protein and protein:ligand associations. We also need to develop methods for reliably predicting protein:protein and protein:ligand associations in a high-throughput manner. This type of development parallels needs in health sciences and can utilize supercomputer capabilities for massive-scale computational screening. Elucidating these mechanisms will permit deeper understanding and derivation of principles of biological functions, from the molecular level up to the whole microbe level, permitting their reliable engineering using site-directed mutagenesis.

Beyond improving sampling in atomistic MD, there is much current activity in CG MD. (Figure 5.2) Events in the cell that are on the millisecond timescale or longer, and system sizes beyond 100M atoms, call for simpler simulation methods than atomistic MD, averaging out the unimportant degrees of freedom to preserve long time-scale properties. Important processes occurring in the millisecond-second time window include protein folding and dynamics, macromolecular associations, receptor activation, lateral diffusion and phase separation in membranes, and the catalytic cycles of transcriptional and translational machinery.

The CG methods can scale efficiently on a variety of supercomputers (Grime and Voth 2014) and will permit cell-scale simulations on timescales up to one second or more allowing tracing the diffusion of macromolecules and metabolites across the cell, including the crowded cellular environment, providing information on system-dependent diffusion constants and associations between multiple molecules in the cytoplasm and at membranes. The CG challenge is to filter phenomena on short time and length scales that have mesoscopic consequences, so as to not 'throw the baby out with the bathwater' when conducting CG while maintaining rigorous self-consistency. Brownian dynamics simulation is among the methods requiring further development. This simulation removes explicit solvent and permits simulation of systems $100\times$ larger or longer than those accessible to MD. Methods need to be developed to merge atomistic MD with Brownian Dynamics simulations and transfer results of these calculations to whole cell descriptions.
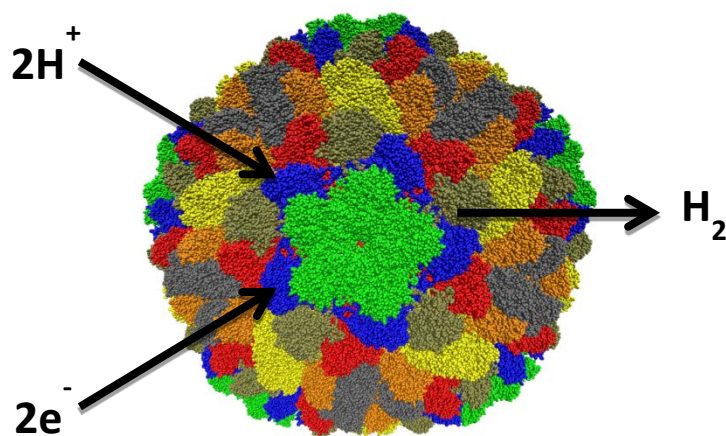

Figure 5.2 Supramillion atom structures of interest in materials engineering, environmental remediation, and renewable resource generation are being efficiently designed and analyzed with a Multiscale MD Toolkit.

The CG approaches typically first choose a functional form that is then parameterized by iteratively adjusting required parameters (for example, based on the potentials of mean force or on reverse Monte Carlo methods) to reproduce the average atomic-detail structure or thermodynamic or fluctuation data. Another approach matches effective forces on CG particles to those from atomistic simulation (Izvekov and Voth 2005). The CG parameters in the above methods are determined a posteriori, often with an iterative fitting procedure to reproduce target data. A different approach, aimed at probing collective motions, derives CG force constants directly from the variance−covariance matrix calculated from all-atom MD (Moritsugu and Smith, 2007); this kind of technique is a direct mapping, requiring no iterative fitting and no experimental input data.

The CG approaches taken can be based on the co-evolution of atomistic and CG model systems (e.g., Figure 5.2). The CG variables may be determined based on microscopic variables, and algorithms for their co-evolution using rigorous mathematical approaches. Because the CG and microscale systems are co-evolved over time, atomistic structural and chemical processes can both respond to and impact their microenvironments (Somogyi et al., 2014). Other CG models are not derived directly from a fine-grained model but are instead inferred from a combination of physical principles and empirical observations, with the model iteratively refined through guided fine-scale (all-atom) simulations combined with further experimental observations (e.g., Ayton and Voth, 2010).

Incorporating large-scale biological data into MD and CG, and scaling these up to whole-cell processes now looms as a critical and largely undeveloped area of systems biology. To handle spatial inhomogeneity and efficiently simulate time scales on the order of the cell cycle (minutes to hours), permitting *whole cell and colony simulations*, requires stochastic modeling techniques for systems of biochemical reactions inside a cell. These techniques should allow us to address complex events such as signaling cascades, transcription, translation and degradation, biofilm formation and cell division. Finally, moving beyond the level of individual organisms, new methods are needed to address biological variables at the community scale and understand evolving interactions with external signals from the environment. These may involve concepts from complexity theory, including the simulation of non-ergodic, non-stationary systems well out of equilibrium. Indeed, such concepts permeate the living world, from single proteins up to earth systems.

5.2   Climate and Environmental Sciences

Critical problems in energy-related climate and environmental sciences, whether in the atmosphere, the oceans, or the subsurface, revolve around the mass transport and transformation of fluids (gases and liquids) and their subsidiary components (e.g., dissolved chemicals, aerosols, microorganisms). These processes are often mediated by mineral and/or biological agents (e.g., solid mineral grains and surfaces, microorganisms, plants, fungi).

The range of scales involved in problems relevant to the climate and environmental sciences is daunting. Bacteria that mediate critical subsurface reactions respond to their local environment, often a single pore only a few ten of microns in diameter. Bulk observations that attempt to relate environmental conditions to microbial function typically average over millions of such pores, often rendering interpretation extremely difficult. Biogeochemical reactions, often occurring at solid-fluid interfaces, are often limited by the transport of aqueous substrates to the interface that is dominated by the relatively slow process of molecular diffusion. At the other extreme, climate sciences often model the cumulative effects of biogeochemical reactions in the oceans, land surface, and atmosphere at resolutions of hundreds of kilometers, scales as many as 40 orders of magnitude larger than the fundamental scales of the reactions (molecules, cells and grain surfaces).

Some of the most important scale transitions in the environmental and climate sciences include:

- Molecules to biological cells

- Cells to pore spaces

- Pore spaces to effective continua (porous media)

- Porous media to local field plots (including effects of plants and surface hydrologic processes)

- Local field plots to watersheds

- Watersheds to regions

- Regions to the globe.

Each of these scale transitions requires careful attention to the effects of reducing the resolution or fidelity of processes acting at smaller scales, and whether these effects can be adequately represented by upscaled (averaged) models or require a more sophisticated approach. Although all of these scales have relevance to BER, EMSL's mission focus on molecular science directs our attention to the lower end of these scales, considering molecules to local field plots and the intermediate scale transitions.

The focus of the SBR and TES programs within BER is on processes from bedrock to the top of the vegetative canopy. Many of these critical processes occur in porous media comprised of solid grains and pore spaces filled with various fluids (e.g., air, water, non-aqueous-phase liquids, methane gas). There has been strong recent interest in connecting pore-scale models (that explicitly simulate solid and fluid phases) with continuum-scale models, which treat the system as a continuous porous medium. From the standpoint of fluid flow and solute transport, pore-scale models are attractive since available process models (i.e., Navier-Stokes equations for flow and advection-diffusion equations for transport) can be considered valid in most cases of interest. Because diffusion is relatively fast at sub-pore resolution, assumptions of mixing inherent to most reaction models are also usually valid. At the continuum scale, however, many available process models require assumptions that are commonly invalidated. Therefore, connecting pore- and continuum-scale models offers the potential to provide a more fundamental underpinning for larger-scale simulations. Several position papers (Appendix C) address this issue in various ways. Scheibe points out fundamental reaction models (and associated rates) cannot be directly applied at continuum scales because of violated assumptions regarding reactant mixing. By coupling and co-evolving a pore-scale and a continuum-scale model, Tartakovsky and Scheibe (2011) demonstrated the feasibility of a hybrid multiscale approach to mixing-controlled reactive transport (Figure 5.3). Battiato proposes a related approach to simulation of changes in porous media constitutive properties (e.g., porosity, permeability) coupled to reactive transport. Valocchi's position paper poses critical questions regarding the viability of hybrid multiscale for large problems (and possible solutions), and regarding uncertainty quantification and model verification/validation.

$t_d = 27.5$ $t_d = 223.3$ $t_d = 1112.5$
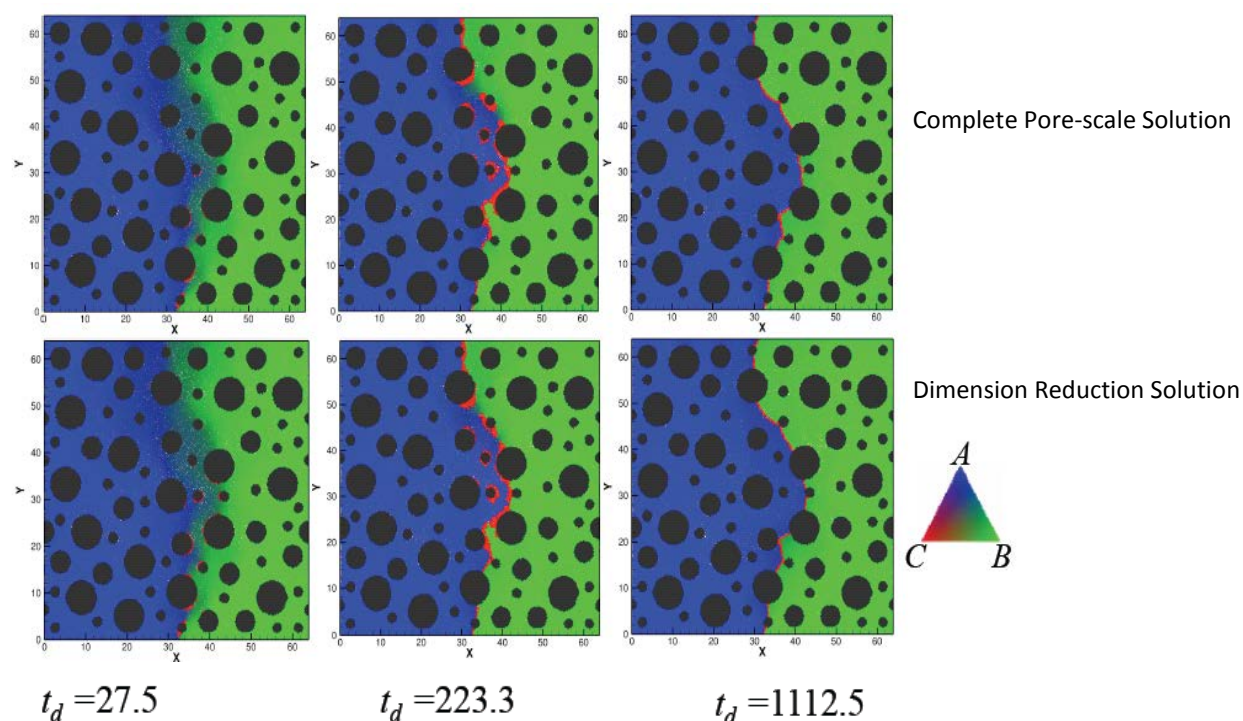
Figure 5.3 Simulation of two reacting solutes (blue, green) forming a mineral reaction product (red). Top row is a pore-scale simulation result at three selected times; bottom row is the result for a comparable hybrid multiscale simulation at the same three times.

Many environmental chemistry applications can use high-accuracy computational chemistry methods (e.g., to accurately predict reaction rates and dynamics). However, these applications are often limited by short time frames over which simulations can be performed within reasonable wall clock time. For example, high-level *ab initio* MD codes, even when run at maximum scalability on massively parallel computing platforms, may require several months to run simulations of picosecond duration, which is too short to represent important chemical processes. Multiscale modeling techniques may be employed to increase the degree of concurrency of solutions by parallelizing in time (Bylaska et al., 2013), thus reducing time to solution and enabling simulation of longer time periods.

Parallel-in-time methods (Figure 5.4) utilize a coarse-scale propagator to establish initial conditions on sub-periods of time, thus breaking time dependency and allowing independent parallel solution of the microscale sub-problems. Iteration between the coarse propagator and the microscale model is performed until convergence is achieved. Bylaska et al. (2013) demonstrated the application of parallel-in-time methods by using Python scripts to make calls to the precompiled quantum chemistry package NWChem (Valiev et al., 2010). For a simulation that reached its maximum possible speedup through standard parallelization of the electronic structure calculation,
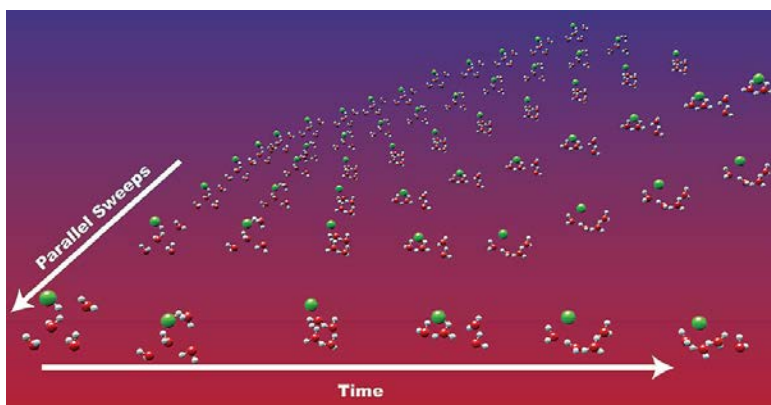
Figure 5.4 Schematic diagram illustrating parallel sweeps (iterations) where enhanced concurrency is achieved by breaking time into multiple sub-periods.

they observed an additional speedup from 32 s/time step to 6.9 s/time step through parallelization in time.

Future strategies in the environmental and climate area will build on approaches described in these examples, with an effort to generalize them (Scheibe et al., 2015) and enable efficient implementation on advanced computational platforms.

5.3   Multiscale Modeling Topical Computing Capability for BER Science

The 2014 EMSL Strategic Plan makes the following statement regarding the EMSL strategy for multiscale modeling:

We also intend to develop new paradigms for multiscale prediction and modeling, including development of new computational approaches to utilize molecular-scale information in models at larger spatial and temporal scales through mesoscale and multiscale modeling methods. This work will emphasize development of predictive modeling capabilities for climate, subsurface, and biological science areas, and will support the integration of EMSL experimental and modeling capabilities for scientific impact. (Section 3.3 Transforming Scientific Capabilities for the Future)

A critical part of this strategy is the development of a topical computing capability in multiscale modeling aimed at critical BER science targets, utilizing the computational facilities (existing and planned) at EMSL to support the needs of the BER science community. Implementation of the EMSL strategy, shown schematically in Figure 5.5, has three primary elements:

The EMSL strategy has three primary elements:

1. Establish a suite of BER-relevant codes on EMSL high-performance computing systems (initially Cascade), including resident expertise for user support of each code.

Figure 5.5 Schematic diagram of the EMSL strategy for a multiscale computing capability for BER. On the left are examples of "at-scale" codes commonly used in BER research; these and/or similar codes will be installed at EMSL with resident expertise for user support. Arrows denote multiscale couplings to be developed between codes at different scales using a script-based loose coupling approach. The right side denotes potential connections to experimental data from facilities and research projects ranging from molecular and subcellular to laboratory and local field scales.

Currently, BER investigators who want to incorporate modeling into their research scope must allocate significant funding and effort for code development/enhancement and installation and compilation on a host computer, and apply separately for a computational allocation at EMSL, NERSC, or other facility. By investing in installation, maintenance, and support of the most widely used BER-relevant codes (e.g., selecting from but not limited to those codes described in Appendix D), EMSL will provide a resource to the BER community that will encourage greater use of EMSL computational resources and enhance opportunities for integration with experimental capabilities embodied in EMSL. This element will include outreach to investigators in the BER program and other programs that support BER-relevant science, to attract a targeted user base to EMSL and enhance use of EMSL computational resources in other areas in addition to EMSL's traditional focus on computational chemistry. Particular attention will be given to modifying (where appropriate) existing codes to take advantage of the hybrid CPU/accelerator architecture of the current EMSL supercomputer, Cascade.

2. Develop a platform for multiscale code linkages among the user codes.

   This element is a unique aspect of the proposed topical computing resource. Whereas other computational facilities emphasize large-scale (i.e., petascale moving toward exascale) simulations to address selected at-scale problems, EMSL topical computing will emphasize integration of moderately-sized simulation capabilities across multiple length and time scales of interest. This will provide solutions to the fragmentation challenge called out in a recent DOE report on the "Virtual Laboratory":

   > "…[the] fragmentation of science, technologies, and predictive capabilities among disciplines and the focus on studying mostly individual, scale-based system components…leads to fundamental uncertainties about how coupled subsystems interact with each other and respond to environmental changes across different space and time scales. The lack of sufficient science-based capabilities to predict these interactions and responses hinders the ability to sustainably manage and mitigate energy and environmental problems." (BERAC, 2013)

   This platform will initially provide tools for loose coupling of resident codes using script-based workflow tools similar to that described in Scheibe et al. (2014), implementing alternative multiscale coupling methods drawn from various multiscale modeling motifs defined in Scheibe et al. (2015). Our goal is to implement as generally as possible methodologies such as the Heterogeneous Multiscale Method (HMM, see E et al. 2003), using scripted workflows that will require limited customization for adaptation to specific application codes. In the case of HMM, for example, the customization will take the form of definitions of variables and variable transformations to be used in "lifting" and "restriction" steps of the algorithm (Figure 5.6). We envision initial efforts will focus on relatively small scales, starting with molecular scales central to the EMSL mission and developing linkages to pore and cellular scales, with eventual adaptation to larger scales of interest to BER.

3. Construct data linkages to instruments, other user facilities, and project data.

   Integration of models and data across scales is essential for advancing predictive capability of models and improving fundamental process understanding. The multiscale modeling capability at EMSL will provide, as appropriate, linkages to available data systems such as KBase, MyEMSL, and other developing systems. These linkages will take the form of standardized interfaces and scripted codes to convert data between formats required for various modeling applications. This effort will involve coordination with developing BER plans for integrated data management across various research programs.
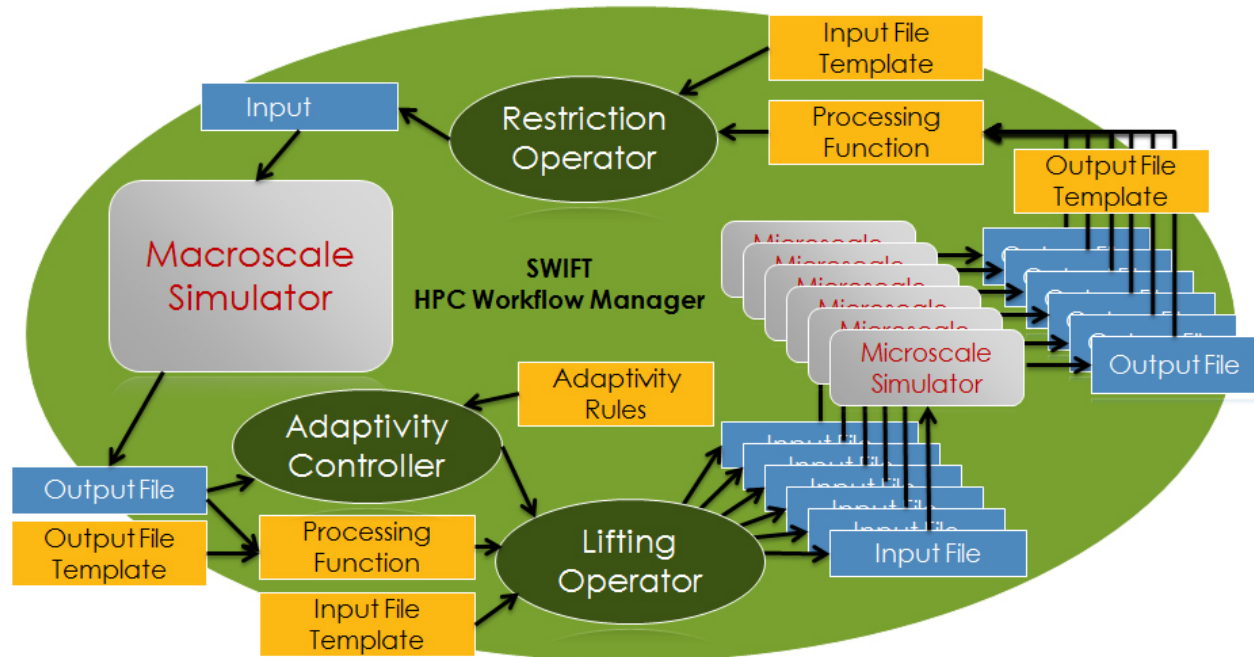
Figure 5.6 Schematic diagram of the EMSL strategy for a multiscale computing capability for BER. The left side denotes a suite of "at-scale" codes commonly used in BER research, to be installed and maintained at EMSL with resident expertise for user support. Schematic diagram of generic implementation of the Heterogeneous Multiscale Method for any pair of codes operating at different scales. Gray boxes denote instances of the microscale and macroscale codes; yellow boxes are user-provided templates, rules and functions that define information transfer between the two scales; green ovals are script elements that perform the data transfer; and blue boxes are script-generated input files and simulator outputs.

5.4   First Science on Future Hardware Systems

The EMSL computing leadership is engaging the user community to identify critical science drivers that will guide the architectural design for EMSL's future supercomputing systems. This workshop represents one important form of that engagement. As part of the design and procurement process, EMSL has dedicated funds to implementing "First Science" on future systems. This is aimed at preparing key users and codes for efficient utilization of new systems as soon as they become available, to maximize the potential scientific impact. As an initial step, a small number of target users and their associated codes will be invited to participate in a working meeting early in 2015. The participants of this workshop (and codes identified in Appendix D) will form the primary selection pool of target users. The objective of the working meeting will be to obtain further details on the architectural needs of candidate codes and identify final participants for First Science. Participating teams will partner with EMSL scientists and computational experts to adapt and/or develop their codes in ways that will 1) facilitate utilization on the current system (Cascade), in particular making use of the Intel MIC accelerators in Cascade, and 2) prepare the codes for efficient and early use on the next-generation EMSL supercomputing system.

6.0   Concluding Remarks

The workshop on "Multiscale Computation: Needs and Opportunities for BER Science," sponsored by EMSL and conducted on August 26, 2014 in Washington, DC, was a unique opportunity. The BER scientific communities met and discussed critical challenges in multiscale simulation that are shared across the wide range of scientific disciplines represented within BER programs. Participants found great value in the initial dialogue, particularly between computational biologists and computational environmental scientists.

This report summarizes the discussions and outcomes of the workshop, and provides objective input that will guide EMSL investments in development of software tools and design of future hardware systems that will optimize their benefit to and impact on the EMSL user community. Several strategies were identified, together with concrete steps to implement those strategies. EMSL has and will continue to invest in the development of a unique multiscale computational capability, drawing on input from this workshop, which will meet the needs identified here and enable transformational advances in mechanistically-based predictive models of complex earth and biological systems.

There is increasing need for the ''virtual reality'' provided by multiphysics and multiscale simulations to quantitatively match, with reasonable accuracy, what happens in the real world. Thus, even if the simulations are not perfect, we would like to be sure the ''correct'' answer falls reliably and predictably within some interval around the computational result. Achieving such reliability is critically important for consolidating the usefulness of simulation in the energy and environmental biosciences. Confidence that simulations provide genuine information about the system under study has important implications. We will need to reach a stage where if the result of a computation does not match some experimental measurement, then one should be able to conclude it is not the calculation, but some underlying hypothesis about the system that is wrong. At this point simulation will furnish the 'predictive understanding' required for future BER science.

7.0   References

Aarnes, J.E., V. Kippe, and K.A. Lie, "Mixed multiscale finite elements and streamline methods for reservoir simulation of large geomodels," *Advances in Water Resources* 28(3): 257–271, 2005.

Abramowitz, G., A. Pitman, H. Gupta, E. Kowalczyk, and Y. Wang, "Systematic bias in land surface models," *J. Hydrometeor.* 8: 989-1001, doi:10.1175/JHM628.1, 2007.

Ayton, G. S. and G. A. Voth, "Multiscale computer simulation of the immature HIV-1 virion," Biophysical Journal 99:2757-2765, DOI:10.1016/j.bpj.2010.08.018, 2010.

BERAC (Biological and Environmental Research Advisory Committee), *BER Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges*, Report DOE/SC-0156, U.S. Department of Energy, February, 2013.

Brown, S. D., A. M. Guss, T. V. Karpinets, J. M. Parks, N. Smolin, S. Yang, M. L. Land, D. M. Klingeman, A. Bhandiwad, M. Rodriquez, B. Raman, X. Shao, J. R. Mielenz, J. C. Smith, M. Keller, and L. R. Lynd, "Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*," *Proceedings of the National Academy of Sciences* 108(33): 13752-7, 2011.

Bylaska, E. J., J. Q. Weare, and J. H. Weare, "Extending molecular simulation time scales: Parallel in time integrations for high-level quantum chemistry and complex force representations," *Journal of Chemical Physics* 139, 074114, 2013.

Chen, L., Y. He, Q. Kang and W. Tao, "Coupled numerical approach combining finite volume and lattice Boltzmann methods for multi-scale multi-physicochemical processes," *Journal of Computational Physics*, 255: 83-105, 2013.

Coon, T., M. L. Porter, Q. Kang, J. D. Moulton, and P. C. Lichtner, "Coupling lattice Boltzmann and continuum equations for flow and reactive transport in porous media," in *Proceedings*, *Computational Methods in Water Resources* (CMWR), Champaign-Urbana, IL, 2012.

Dama J. F., A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth, "The theory of ultra-coarse-graining. 1. General principles," *J. Chem. Theory Comput.* 9:2466-2480, 2013.

E, W., B. Engquist, and Z.Y. Huang, "Heterogeneous multiscale method: A general methodology for multiscale modeling," *Physical Review B* 67(9): 1–3, 092101, 2003.

Grime, J. M. A. and G. A. Voth, "Highly scalable and memory efficient ultra-coarse-grained molecular dynamics simulations," *J. Chem. Theory Comput.* 10: 423-431, 2014.

Hou, T.Y., and X.H. Wu, "A multiscale finite element method for elliptic problems in composite materials and porous media," *Journal of Computational Physics* 134(1): 169–189, 1997.

Izvekov, S. and G. A. Voth, "A multiscale coarse-graining method for biomolecular systems," *The Journal of Physical Chemistry B* 109(7):2469-2473, 2005.

Jenny, P., S.H. Lee, and H.A. Tchelepi, "Multi-scale finite-volume method for elliptic problems in subsurface flow simulation," *Journal of Computational Physics* 187(1): 47–67, 2003.

Joshi, H., A. Singharoy, Y. V. Sereda, S. C. Cheluvaraja, and P. J. Ortoleva, "Multiscale simulation of microbe structure and dynamics," *Prog. Biophys. Mol. Biol.* 107(1): 200-217, doi: 10.1016/j.pbiomolbio.2011.07.006, 2011.

Keyes, D.E., L.C. McInnes, C. Woodward, W. Gropp, E. Myra, M. Pernice, J. Bell, J. Brown, A. Clo, J. Connors, E. Constantinescu, D. Estep, K. Evans, C. Farhat, A. Hakim, G. Hammond, G. Hansen, J. Hill, T. Isaac, X.M. Jiao, K. Jordan, D. Kaushik, E. Kaxiras, A. Koniges, K. Lee, A. Lott, Q.M. Lu, J. Magerlein, R. Maxwell, M. McCourt, M.

Mehl, R. Pawlowski, A.P. Randles, D. Reynolds, B. Riviere, U. Rude, T. Scheibe, J. Shadid, B. Sheehan, M. Shephard, A. Siegel, B. Smith, X.Z. Tang, C.Wilson, and B. Wohlmuth, "Multiphysics simulations: Challenges and opportunities," *International Journal of High Performance Computing Applications* 27(1): 4–83, 2013.

Lichtner, P. C. and Q. Kang, "Upscaling pore-scale reactive transport equations using a multi-scale continuum formulation," *Water Resources Research* 43, W12S15, 2007.

Lipnikov, K., J.D. Moulton, and D. Svyatskiy, "A multilevel multiscale mimetic (M-3) method for two-phase flows in porous media," *Journal of Computational Physics* 227(14): 6727–6753, 2008.

Mehmani, Y., T. Sun, M. T. Balhoff, P. Eichhubl and S. Bryant, "Multiblock pore-scale modeling and upscaling of reactive transport: Application to carbon sequestration," *Transport in Porous Media*, 95(2): 305-326, 2012.

Mehmani, Y. and M. T. Balhoff, "Bridging from pore to continuum: A hybrid mortar domain decomposition framework for subsurface flow and transport," *Multiscale Modeling & Simulation*, 12(2): 667-693, 2014.

Molins, S., D. Trebotich, L. Yang, J. B. Ajo-Franklin, T. J. Ligocki, C. Shen, and C. I. Steefel, "Pore-scale controls on calcite dissolution rates from flow-through laboratory and numerical experiments," *Environ. Sci. Technol.* 48(13): 7453-7460, 2014.

Molins, S., D. Trebotich, C. I. Steefel, and C. Shen, "An investigation of the effect of pore scale flow on average geochemical reaction rates using direct numerical simulation," *Water Resour. Res.*, 48(3), 2012.

Moritsugu, K. and J. C. Smith, "Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian," *Biophysical Journal*, 93(10): 3460-3469, 2007.

Navarre-Sitchler, A., C. I. Steefel, L. Yang, L. Tomutsa and S. L. Brantley, "Evolution of porosity and diffusivity associated with chemical weathering of a basalt clast," *Journal of Geophysical Research*, 114, F02016, 2009.

Noiriel, C., C.I. Steefel, L. Yang and J.B. Ajo-Franklin, "Upscaling calcium carbonate precipitation rates from pore to continuum scale," *Chemical Geology*, 318-319, 2012.

Parks, J., H. Guo, C. Momany, L. Liang, S. Miller, A. Summers, and J. C. Smith, "Mechanism of Hg-C protonolysis in the organomercurial lyase MerB," *Journal of the American Chemical Society* 131: 13278-13285, 2009.

Parks, J. M., A. Johs, M. Podar, R. Bridou, R. A. Hurt, S. D. Smith, S. J. Tomanicek, Y. Qian, S. D. Brown, C. C. Brandt, A. V. Palumbo, J. C. Smith, J. D. Wall, D. A. Elias, and L. Liang, "The genetic basis for bacterial mercury methylation," *Science* 339(6125): 1332-1335, 2013.

Pau, G., G. Bisht, and W. J. Riley, "A reduced order modeling approach to represent subgrid-scale hydrological dynamics for land-surface simulations: Application in a polygonal tundra landscape," *Geosci. Model Dev.* 7(5): 2091-2105, 2014.

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, "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, btt055, 2013.

Scheibe, T. D., X. Yang, K. Schuchardt, J. Chase, B. Palmer, and A. Tartakovsky, "A many-task parallel approach for multiscale simulations of subsurface flow and reactive transport," In *Proceedings, 7th Workshop on Many-Task Computing on Clouds, Grids and Supercomputers (MTAGS) 2014*, New Orleans, LA, Nov. 16, 2014.

Scheibe, T. D., E. M. Murphy, X. Chen, A. K. Rice, K. C. Carroll, B. J. Palmer, A. M. Tartakovsky, I. Battiato, and B. D. Wood, "An analysis platform for multiscale hydrogeologic modeling with emphasis on hybrid multiscale methods," *Groundwater*, 53(1): 38-56, doi: 10.1111/gwat.12179, 2015.

Somogyi, E., A. Abi Mansour, and P. J. Ortoleva. *ProtoMD: A Prototyping Toolkit for Multiscale Molecular Dynamics*. arXiv:1309.6397, 2014.

(Available from: https://github.com/Andrew-AbiMansour/proto_md).

Smith, J. C. and B. Roux, "Eppur si muove: The 2013 Nobel Prize in Chemistry," *Structure* 21(12): 2102-2105, 2013.

Steefel, C. I., S. Molins, and D. Trebotich, "Pore scale processes associated with subsurface CO2 injection and sequestration," *Rev. Mineral. Geochem.*, 77: 259-303, 2013.

Tartakovsky, A. M. and T. D. Scheibe, "Dimension reduction method for advection-diffusion-reaction systems," *Advances in Water Resources*, 34(12):1616-1626, doi:10.1016/j.advwatres.2011.07.011, 2011.

Tartakovsky, G. D., A. M. Tartakovsky, T. D. Scheibe, Y. Fang, R. Mahadevan and D. R. Lovley, "Pore-scale simulation of microbial growth using a genome-scale metabolic model: Implications for Darcy-scale reactive transport", *Advances in Water Resources* 59: 256-270, 2013.

Trottenberg, U., C.W. Oosterlee, and A. Schuller, *Multigrid: Basics, Paralellism and Adaptivity*, London, UK: Academic Press, 2001.

Valiev, M., E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Want, J. Nieplocha, E. Apra, T. L. Windus, and W. A. De Jong, "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations," *Computer Physics Communications* 181(9):1477–1489, doi:10.1016/j.cpc.2010.04.018, 2010.

Wesseling, P., *An Introduction to Multigrid Methods*, Stafford BC, Queensland: John Wiley & Sons, 1992.

Wood, B. D., "The role of scaling laws in upscaling," *Advances in Water Resources* 32(5): 723-736, 2009.

Zhang, F., N. A. Marks, J. D. Gale, Q. Kang, B. P. Uberuaga, C. R. Stanek, and N. Henson, "The effect of chemical evolution in 137CsCl on radionuclide leaching," abstract in Goldschmidt Conference, DOI:10.1180/minmag.2013.077.5.26, 2013.

Appendix A Workshop Attendees

Appendix A

Co-Chairs:

- Scheibe, Timothy, Pacific Northwest National Laboratory (tim.scheibe@pnnl.gov)

- Smith, Jeremy, Oak Ridge National Laboratory (smithjc@ornl.gov)

U. S. Department of Energy Program Managers:

- Paul Bayer (Office of Biological and Environmental Research, CESD)

- Daniel Drell (Office of Biological and Environmental Research, BSSD)

- Thomas Ndousse-Fetter (Office of Advanced Scientific Computing Research)

Observers:

- David Cowley (Pacific Northwest National Laboratory)

- Lois McInnes (Argonne National Laboratory)

Workshop Participants		
Balhoff, Matthew	University of Texas at Austin	balhoff@mail.utexas.edu
Battiato, Ilenia	San Diego State University	ibattia@mail.sdsu.edu
Bisht, Guatam	Lawrence Berkeley National Laboratory	gbisht@lbl.gov
Brodie, Eoin	Lawrence Berkeley National Laboratory	ELBrodie@lbl.gov
Cannon, William	Pacific Northwest National Laboratory	william.cannon@pnnl.gov
Cardon, Zoe	Marine Biological Laboratory	zcardon@mbl.edu
Cheng, Xiaolin	Oak Ridge National Laboratory	chengx@ornl.gov
Kang, Qinjun	Los Alamos National Laboratory	qkang@lanl.gov
Li, Li	Pennsylvania State University	lili@eme.psu.edu
Ortoleva, Peter	Indiana University	ortoleva@indiana.edu
Parks, Jerry	Oak Ridge National Laboratory	parksjm@ornl.gov
Perez, Danny	Los Alamos National Laboratory	danny_perez@lanl.gov
Reed, Jennifer	University of Wisconsin	reed@engr.wisc.edu
Roberts, Elijah	Johns Hopkins University	eroberts@jhu.edu
Roux, Benoit	University of Chicago	roux.uchicago@gmail.com
Valocchi, Albert	University of Illinois	valocchi@illinois.edu
Voit, Eberhard	Georgia Institute of Technology	eberhard.voit@bme.gatech.edu
Voth, Gregory	University of Chicago	gavoth@uchicago.edu
Walker, Ross	San Diego Supercomputing Center	ross@rosswalker.co.uk
Zhou, Huan-Xiang	Florida State University	hzhou4@fsu.edu

Appendix B Workshop Agenda

Appendix B

MULTISCALE COMPUTATION: NEEDS AND OPPORTUNITIES FOR BER SCIENCE
HOSTED BY THE ENVIRONMENTAL MOLECULAR SCIENCES LABORATORY (EMSL)

AUGUST 26, 2014
WASHINGTON, D.C.
BATTELLE WASHINGTON OFFICE
901 D ST. SW #900
TEL: 202-479-0500

WORKSHOP OBJECTIVES:

- Identify BER-relevant models and their potential cross-scale linkages that could be exploited to better connect molecular-scale research to BER research at larger scales.
- Identify critical science directions that will motivate EMSL decisions regarding future computational (hardware and software) architectures.

AGENDA:

8:30 am **Welcome and Logistics** .... Tim Scheibe[a] / Jeremy Smith[b]
[a]*EMSL Lead Scientist, Multiscale Modeling and High-Performance Computing*
[b]*Director, UT/ORNL Center for Molecular Biophysics*

8:45 am **Role of Scientific Simulation in BER's Research Strategy** .... Paul Bayer
DOE Program Manager for EMSL

9:00 am **Science Challenges in Multiscale Simulation for**
Biological and Environmental Research ....Tim Scheibe / Jeremy Smith

9:30 am **Charge Instructions for Breakout Session One and Transition Time / Break**

10:00 am **Breakout Session One – Components of a BER Multiscale Simulation Framework**

12:00 **Lunch – Food Court (No Host)**

1:00 pm **Report Back – Breakout Session One**

1:30 pm **Charge Instructions for Breakout Session Two and Transition Time / Break**

2:00 pm **Breakout Session Two – Computational and Data Needs**

4:00 pm **Report Back – Breakout Session Two**

4:30 pm **Group Discussion and Closing Remarks**

5:00 pm **Adjourn**

Appendix C Position Papers

Appendix C

Battiato, Ilenia (San Diego State University): Adaptive stochastic hybrid modeling of highly reactive fronts in subsurface geologic systems.

Brodie, Eoin (Lawrence Berkeley National Laboratory): Maximizing new information on microbial function to produce a more accurate and dynamic representation of microbial processes in biogeochemical models.

Cannon, Bill (Pacific Northwest National Laboratory): Simulating microbial metabolism with statistical thermodynamics.

Cardon, Zoe (Marine Biological Laboratory): 3-D reality check: Data mining with structural support.

Kang, Qinjun (Los Alamos National Laboratory): Multiscale simulation of reactive transport in porous media.

Li, Li (Penn State University): Understanding cross-scale behavior.

Ortoleva, Peter (Indiana University): Multiscale theory and computation for microbes and nanomaterials.

Roberts, Elijah (Johns Hopkins University): Multiscale modeling of reaction-diffusion and cytoskeletal dynamics during gradient signaling.

Scheibe, Tim (Pacific Northwest National Laboratory): Hybrid multiscale simulation of mixing-controlled biogeochemical reactions.

Smith, Jeremy (Oak Ridge National Laboratory): Multiscale biological computer simulation at ORNL.

Valocchi, Al (University of Illinois): Concurrent hybrid multiscale methods for reactive transport in porous media: Some challenges.

Voit, Eberhard (Georgia Institute of Technology): Of mice and men… and microbes.

Voth, Gregory (The University of Chicago): (Untitled).

"Adaptive stochastic hybrid modeling of highly reactive fronts in subsurface geologic systems"
Ilenia Battiato
San Diego State University

Motivation: Ultra-long time predictions of transport of highly reactive fluids in the subsurface challenge our current modeling efforts, and their uncertainty undermines our ability to predict the future impact of, and the risks associated to, anthropogenic injection and storage of CO_2 and nuclear wastes. *We propose to develop an organizing framework (Fig. 1) to describe nonlinear reactive transport in geologic systems over many length and time scales, and corresponding physics-based adaptive multi-scale hybrid models. We will also provide computational tools for robust uncertainty quantification in hybrid simulations. Our theoretical and computational models will be grounded on laboratory experiments.* Since our physical understanding of processes is scale-dependent through the models we employ to decipher and interpret reality (and measurements) at a particular (observation) scale, such a theoretical and computational framework is needed i) to control modeling errors in inherently multi-scale geological systems and ii) to improve our predictive capabilities. Both grand challenges are relevant to DOE and office of Science missions.
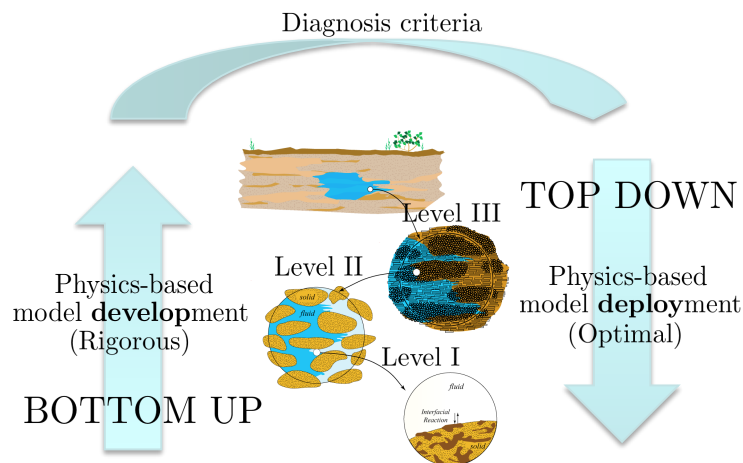
Fig. 1: Conceptualization of the connection between *develop*ment and *deploy*ment of physics-based multi-scale models. The missing link between model development and deployment is generally the identification of diagnosis criteria to identify suitable models and suitable modeling scales at which continuum-scale quantities and parameters are well defined. (Battiato I., 2013 DOE Young Investigator Proposal)

Challenge: Subsurface flow and transport pervasively exhibit non-linear dynamics and lack of (temporal and spatial) separation of scales, i.e. physical and bio-geochemical phenomena on one scale (e.g., a pore scale) affect, and are coupled to, phenomena on a vastly different scale (e.g., a field scale). For example, pore-scale molecular diffusion fundamentally affects field-scale mixing of (bio)chemically reacting solutes, and differential dissolution of rock minerals (over time scales spanning many orders of magnitude) can lead to drastic permeability changes due to granular coating development, successive liquefaction and jamming at pore/fracture throats. Common features of such subsurface phenomena are their high localization (e.g., propagation of reactive fronts and biofilm growth) and/or strong nonlinear coupling between the processes involved (e.g., dynamic changes in porosity and permeability due to dissolution or precipitation). Current modeling approaches can be subdivided into Darcy-scale (continuum), pore-scale, and multi-scale models. The theoretical validity, and predictive power, of Darcy-scale (one-point closure) formulations totter, when confronted with these highly coupled, nonlinear systems. Fine-scale models (e.g., Lattice-Boltzmann or Navier-Stokes equations),

which describe these phenomena with a high degree of fidelity, do not represent a viable alternative, as they are computationally prohibitive and impractical on the field scale. Current multi-scale models share many of the limitations of continuum-scale representations as they are often based on empirical closures, upscaling methods and/or assumed macroscopic behavior of microscopic variables. Additionally, there exist no adaptive criteria either to determine *a priori* the error associated to a specific (continuum-, multi-scale) model, or to select the optimal representation scale in terms of accuracy and computational burden. Further, structural and parametric uncertainty call for quantification of predictive uncertainty of multi-algorithm approaches.

Approach: The primary objective of the proposed research is to provide a framework to modeling multi-scale coupled non-linear reactive transport in geologic systems through the development of an upscaling framework where stochastic hybrid models are employed adaptively. While *a priori* suitable to many physical systems, **we will apply such a framework to abrupt permeability changes in carbonate rocks formations due anthropogenic injection of highly reactive fluids**.

Permeability and porosity changes in carbonate rocks (e.g., argillaceous limestone) due to injection of acidic waters are determined by a complex interplay between flow and nonlinear transport processes. Abrupt permeability/porosity changes due to dissolution and/or clogging represent an outstanding example of tipping point events, where the system state can undergo a 'sudden' transition (compared to the temporal scales of the transport processes involved) driven by temporally and spatially localized impulses (fluctuations). Some additional challenging features of these systems are: i) nonlinear coupling between scales, ii) strong out-of-equilibrium dynamics under anthropogenic loading, iii) time- and spatial-scales spanning many orders of magnitude, iv) heterogeneity at different scales, v) system sensitivity to fluctuations, vi) long range interactions (e.g. force networks in clay particles coating). The former characteristics make this physical phenomenon highly suitable to stochastic hybrid approaches.

The proposed approach consists of the following steps:
1. We will develop a generalized multi-scale theoretical framework for non-linear multi-component reactive transport based on *spatial* and *temporal* **homogenization of dimensionless (fine-scale) equations**. The temporal homogenization allows one to include explicitly the effect of temporal fluctuations of, e.g., driving forces and boundary conditions. This formulation, in its unclosed form, *a priori* will account for a lack of (temporal and spatial) scale separation, and will explicitly depend upon dimensionless (e.g., Peclét and Damköhler) numbers, which describe the dynamics of the physical system under consideration. The degree of coupling between scales can be represented in a phase diagram in terms of the adimensional quantities previously identified. Evaluating the order of magnitude of such parameters in the 'real' system provides one with a quantitative estimate of the degree of coupling between its characteristic scales. When scales are not fully coupled, this leads to a simplification of the original formulation through the choice of the most appropriate closure scheme. In this case, the closed equation resulting from the previous homogenization step may be used as the new fine-scale model in the following upscaling step. This procedure guarantees that the coarser scale model is employed, wherever and whenever possible. If, on the other hand, there is no separation of scales, coupling conditions must be formulated.
2. We will develop hybrid formulations for systems with different degrees of coupling. We will combine continuum-scale with fine-scale models to describe the dynamics of highly correlated systems. For example, we will employ (fine-scale) force-network models and pdf methods to determine the probability of rupture of granular clay coating under hydrodynamic shear.
3. We will develop diagnosis criteria for algorithm refinement. The criteria for adaptive spatial and temporal hybridization will be formulated in terms of coarse-scale

functionals, e.g. space-time dependent dimensionless numbers. The spatio-temporal evolution of appropriately defined adimensional quantities will identify an orbit in the phase-diagram. Each point of the orbit will be characterized by a given degree of scale coupling/separation. This information will be used to select the most appropriate hybrid scheme at any fixed instance in time and space.

4. We will generalize our former approaches to fine-scale (deterministic) equations in random domains to account for unknown pore geometry. This approach will lead to an equivalent formulation in terms of fine-scale stochastic equations in deterministic domains. Finally, the hybrid coupling will be generalized to stochastic PDEs of nonlinear reactive transport, and will allow one to propagate parametric uncertainty of fine-scale quantities (e.g. geometry) through a modeling process.

The proposed research would develop: 1) a framework to describe nonlinear reactive transport in geologic systems over many length and time scales, and corresponding physics-based adaptive multi-scale hybrid models, with specific application to permeability/porosity changes in fractured and porous rocks, and 2) methods to propagate uncertainty between scales. Both outcomes have a strong transformational potential in modeling of geological (or, more generally, physical) nonlinear systems, which lack, completely or partially, scale separation. Such organizing framework would also lay a solid basis for quantitative risk analysis of anthropogenic injection and storage of highly reactive substances.

Position Paper for "**Multiscale Computation: Needs and Opportunities for BER Science**"

Maximizing new information on microbial function to produce a more accurate and dynamic representation of microbial processes in biogeochemical models

Eoin Brodie – Lawrence Berkeley National Laboratory

Motivation

Our understanding of the extent of microbial diversity and function in natural systems has been transformed through the use of new genomic and computational approaches. From metagenomic datasets, terabases (and terabytes) of sequence information are becoming routine, and their reconstruction into complete genomes is accelerating. These genomes bear the signatures of environmental selection, the imprint of historical conditions within a biome, effectively turning the microbial community into a network of sensors with a memory for past conditions, deployed at the pore scale. This microbial dark matter is gradually being illuminated, revealing metabolic potential that has yet to be cultivated and studied in the laboratory much less considered in biogeochemical models.
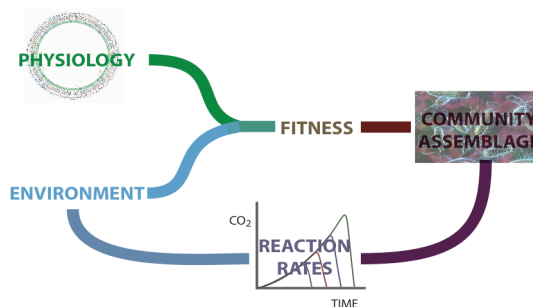
Fig. 1. Trait-based models of microbial function informed by genomics with emergent populations resulting in dynamic reaction rates.

Challenge

How do we move forward, assimilating and condensing this information such that it can be used to inform biogeochemical reaction networks and constrain parameter values? Can we reduce a complex network of biochemical reactions within a microbial cell to a few key functions/traits that determine its fitness, can we reduce organisms with similar functions/trait space to 'guilds' and should we associate guild members with specific reactions within our biogeochemical reaction networks? Can we make generalizations that can be applied beyond the systems where we have such data, and when and where is this information useful and how much of it do we really need?

Approach

To answer these questions requires that we begin to incorporate biological complexity into models that represent life at the microbial (pore) scale. Reactive transport models offer an attractive solution with flow and transport constraints regulating the supply of substrates to microbes where reactions are represented by a Michaelis-Menten type function with a thermodynamic driver (e.g. CrunchFlow; Steefel, 2009). By sequencing metagenomes from a series of samples taken over what we consider to be biogeochemically relevant locations or times in the system of interest, we can assemble

genomes (using genome properties and abundance profiles) (Handley et al., 2013; Kantor et al., 2013) and with well-curated databases of key functional genes/pathways we can ascribe a functional role to those genomes (e.g. ggkbase.berkeley.edu). We can possibly go further a derive proxies for kinetic parameters, estimating theoretical maximum growth rate (Vieira-Silva & Rocha, 2010) and optimal growth temperatures (Zeldovich et al. 2007), all without growing any organisms in the laboratory. If suitably validated this approach would replace the current practice of using literature parameters derived under irrelevant conditions for irrelevant organisms, with ecosystem specific information or a solid basis with which to derive meaningful distributions of parameters with some relationship to important drivers.

Key References

Steefel, C.I. (2009) CrunchFlow- Software for Modeling Multicomponent Reactive Flow and Transport. http://www.csteefel.com/WebCrunch.html

Handley, K. M., Verberkmoes, N. C., Steefel, C. I., Williams, K. H., Sharon, I., Miller, C. S., … Banfield, J. F. (2013). Biostimulation induces syntrophic interactions that impact C, S and N cycling in a sediment microbial community. *The ISME Journal*, *7*(4), 800–816. doi:10.1038/ismej.2012.148

Kantor, R. S., Wrighton, K. C., & Handley, K. M., Sharon, I., Hug, L.A., Castelle, C.J., Thomas, B.C., Banfield J.F. (2013). Small Genomes and Sparse Metabolisms of Sediment-Associated Bacteria from Four Candidate Phyla. mBio, 4: 5 e00708-13 doi:10.1128/mBio.00708-13.

Vieira-Silva, S., & Rocha, E. P. C. (2010). The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genetics*, *6*(1), e1000808. doi:10.1371/journal.pgen.1000808

Zeldovich, K. B., Berezovsky, I. N., & Shakhnovich, E. I. (2007). Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Computational Biology*, *3*(1), e5. doi:10.1371/journal.pcbi.0030005

Bill Cannon – Pacific Northwest National Laboratory

Simulating Microbial Metabolism with Statistical Thermodynamics

Motivation. Ideally, models of metabolism should predict metabolite levels, characterize the thermodynamic requirements of pathways and processes, be testable with experimental data, and provide physical insight into the principles of emergence of cellular function and self-organization. Simulations based on the law of mass action, such as kinetic simulations, can in principle meet these requirements. However, these simulations require knowledge of the thousands of rate constants involved in the reactions. The measurement of rate constants is very labor intensive, and hence rate constants for most enzymatic reactions are not available. Moreover, the same *prima facie* enzymes (ortholog) from different species, or even different strains, can have vastly different rate constants. If one were to model the metabolism of an organism using kinetic simulations, the rate constants for each enzyme would first need to be measured. This is an overwhelming task that is not likely to be accomplished for any single organism in the near or distant future.

Challenge. The challenge is to model coupled reactions on a large-scale using methods that do not ignore the physics of the system. Without modeling the physics of the system, the emergence of function and self-organization can only be fit to data and not understood from principles. Such fitting does not allow for prediction of new properties or function, and does not provide insight necessary for design of new functionality.

Approach. Instead of setting rate constants and sampling for steady-state concentrations and fluxes, we set chemical potentials and sample for rates, fluxes and concentrations. Statistical thermodynamics provides us with much of the mathematical framework to accomplish this, with the exception of the ability to model time-dependent non-equilibrium processes [1, 2]. However, development of fluctuation theories over the past 20 years are promising and in many cases, accurately allow us to model the time-dependence of coupled reactions across large time scales with the same accuracy as mass action-based kinetic simulations.
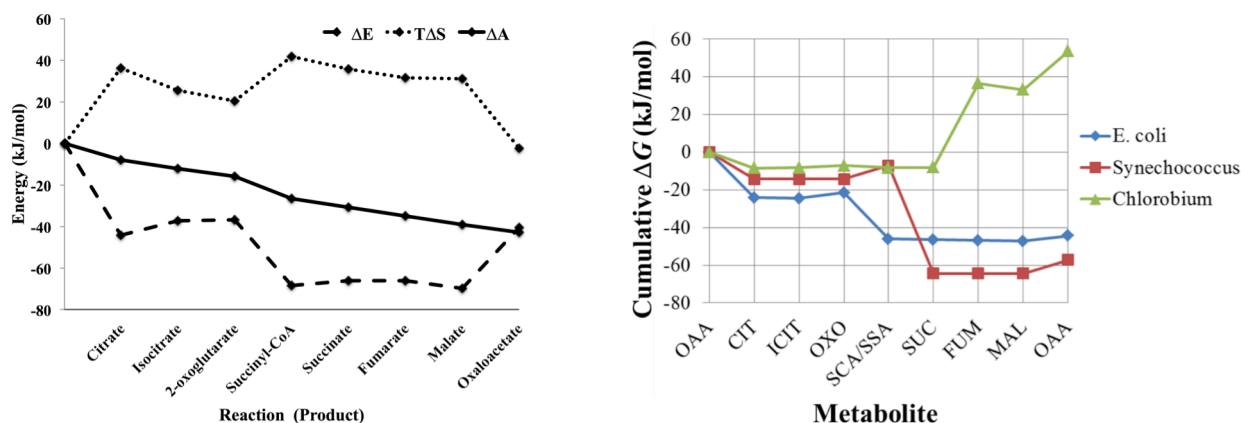

Figure 1. (Left) Free energy profile determined from metabolomics data and statistical thermodynamics simulations. (Right) Free energy profiles of E. coli, a cyanobacterim and a green sulfur bacterium under similar conditions.

Recent Results.

We have recently published an outline of the statistics of modeling coupled reactions from a stochastic thermodynamic perspective [1]. The principles of the approach were demonstrated using a simple set of coupled reactions, and then the system was characterized with respect to the changes in energy, entropy, free energy, and entropy production. Finally, the physical and biochemical insights that this approach provides for metabolism was demonstrated by application to the tricarboxylic acid (TCA) cycle of Escherichia coli.

The use of the chemical potential in modeling metabolic reactions provides a natural integration platform for experimental metabolomics data. The assumption inherent in the use of the standard chemical potential for modeling reactions is that each change of state occurs with a frequency proportional to the thermodynamic driving force for the respective reaction. In regions of state space where the system responds linearly, this is an excellent assumption. We have found that 80% of the possible state space can be modeled in this way.
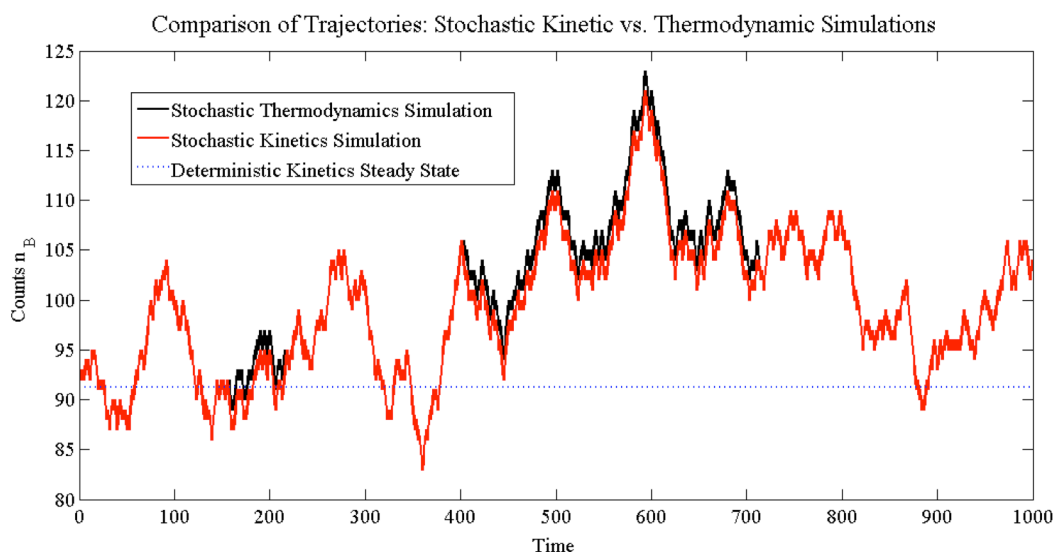
Figure 2 Comparison of simulation trajectories from a stochastic kinetic simulation and a stochastic thermodynamic simulation using fluctuation theory.

Key References

[1]   W. R. Cannon, "Simulating metabolism with statistical thermodynamics," *PLoS One,* vol. 9, p. e103582, 2014.http://www.ncbi.nlm.nih.gov/pubmed/25089525
[2]   T. Schmiedl and U. Seifert, "Stochastic thermodynamics of chemical reaction networks," *Journal of Chemical Physics,* vol. 126, Jan 28 2007.<Go to ISI>://WOS:000243891100005

3D Reality Check: Data Mining with Structural Support
Zoe G. Cardon, *Marine Biological Laboratory, Woods Hole, MA*

THE CHALLENGE: The microbial activities on which humanity and ecosystems depend are determined by both microbial community membership and the resources and conditions in the microbes' local environments. In order to effectively mine "omics" (metatranscriptomic, metagenomic, proteomic, and amplicon-based membership) datasets for insight into controls over microbial community structure and function, a pervasive disconnect must be addressed: The length of a typical bacterium is approximately one micrometer, yet environmental sampling meant to provide contextual insight into environmental conditions surrounding, and resources available to, microbial communities is often (by necessity) carried out at much larger scales. "Metadata" (pH, nutrient concentrations, salinity, availability of organic matter etc.) are measured from liters of well-mixed seawater, or from homogenized, coffee-mug-size cores of sediment or soil. From such sampling, it is impossible to infer the localized, heterogeneous, micron-scale environmental conditions the microbes are actually reacting to, and influencing, at the time of sampling.

The importance of 3D microheterogeneity for microbial life is not restricted to particular environments. Many bacteria in open oceans, nearly all in estuaries, and microbes in sediments and soils are associated with particles, whether on or in marine snow, organic matter, or inorganic substrate. Even in open oceans, free-living microbes in water can not be considered to be surrounded by a well-mixed, homogeneous, aqueous environment; plumes of resources appear and disappear as particles pass and solutes diffuse. Though many computational approaches in use today aim to discern key environmental conditions associated with particular microbial processes or community structure, they do so by developing extensive correlation networks among environmental variables (measured from homogenized large samples) and the gene expression or community patterns (produced from millions of individual microbes each sensing its actual local environment). Gene expression or community patterns that don't correlate strongly with the large-scale contextual data are ignored as noise (e.g. how could strictly aerobic processes be occurring in an anaerobic, saturated sediment?). But, the diversity of process and membership in "omic" datasets is itself information, information that can and should be mined for insight into how microenvironmental heterogeneity enables functional and community microbial diversity on Earth. This insight is essential if we are to use existing "omics" datasets for scaling up from e.g. genome-scale metabolic models or gene regulatory network structure and function (in well-defined communities), or thermodynamic predictions of likely biogeochemistries (in extremely diverse environmental microbial samples), to understand controls over, and predict, larger-scale processes of interest in bioremediation, food and fuel production, and decomposition and nutrient cycling across ecosystems.

Goals: The basic mis-match in scales of contextual environmental sampling and microbial response hinders maximization of information gained from existing "omics" data. We will tackle this problem by:
- developing a computational framework that explicitly embeds microbial communities within high resolution 3D microenvironmental structure (see page 2);
- predicting diagnostic "fingerprints" of the resulting small-scale heterogeneity in function and niche expected to appear in "omics" datasets, e.g. caused by diffusion limitations, fluid flow patterns, and/or abiotic and biotic reactions within the 3D microstructure;
- examining our and others' existing "omics" datasets for such fingerprints, first targeting sets from environments with highly contrasting structure, and exploring patterns inexplicable based on associated large-scale metadata (e.g. simultaneous aerobic and anaerobic processes; operation of multiple nitrate reduction pathways though one appears thermodynamically favored based on large-scale metadata);
- predicting the larger-scale manifestation of the millions of small-scale processes operating simultaneously. This larger-scale manifestation would represent the process rate measured from e.g. contaminated soil, or a bottle of ocean water, or a reactor to cleanse water of pollutant nitrate.

FUTURE RESEARCH DIRECTION: Our development of a new computational framework embedding diverse microbial communities within high resolution 3D structure requires melding two existing models. (1) Scheibe and colleagues (e.g. Tartakovsky et al., 2013, *Advances in Water Resources* 59:256) have developed a 3D modeling framework capturing pore-scale environmental heterogeneity in a saturated or unsaturated volume structured by particles. They use the Smoothed Particle Hydrodynamics numerical

method to simulate fluid flow (Navier-Stokes equations), solute transport (advection and diffusion), and biotic and abiotic reactions. Pore geometry is specified randomly or from known 3D structure. Scheibe is located at the Environmental Molecular Sciences Laboratory, which makes available to outside scientists the programming expertise and supercomputer access necessary for computationally-intensive projects such as ours. For example, Scheibe and colleagues have folded a genome-scale metabolic network for the well-characterized bacterium *Geobacter* (Tartakovsky et al. 2013) into the 3D framework, to explore how microheterogeneity of conditions caused by 3D physical structure, and influenced by microbial activity embedded in that structure, affects *Geobacter*'s ability to reduce contaminant uranium in groundwater. They showed that an assumption of homogeneous resources and conditions for all microbes (i.e. equivalent to gleaning contextual environmental information from large, mixed samples) overpredicted reduction rates, compared to the more realistic case with microbes embedded in the 3D matrix.

(2)   In most environmental communities, however, community membership and processes of interest are not dominated by one or just a few well-characterized, sequenced microbes. A complementary method is needed to predict localized microbial function, one that does not require genome-scale networks for every microbe present. <u>Vallino and colleagues (e.g. Vallino, 2010, *Phil. Trans. R. Soc. B* 365:1417; Algar and Vallino, 2014, *Aquatic Microbial Ecology*, 71:223) provide a promising thermodynamically-based framework</u>. Though transformations of environmental compounds are orchestrated by microbes housing complex intracellular networks, a distributed metabolic network, based on extracellular metabolites in the microbes' environment, can be substituted as a first order approximation. Vallino's approach is different from optimization of selected, individual metabolic reactions at each point in space and time, as currently implemented by Gawande in Scheibe's framework (in prep); it instead optimizes energy dissipation through multiple pathways over the entire time frame of interest. Vallino et al.'s distributed metabolic network will be folded into Scheibe's 3D pore-scale model, capturing biotic-abiotic functional interconnections and microenvironmental diversity over time, in 3D space. The new framework underlies goals for "omics" predictions, dataset exploration, and system implications noted pg 1.

Payoff for data-mining within existing and future datasets: Development of this quantitative, broadly-applicable framework enables a new kind of data-mining within environmental "omics" datasets, mining that not only recognizes the biotic detail available for perusal (curious community structures, unusual genes, or novel genome-scale regulatory networks, e.g. Konstantinidis et al., 2009, PNAS 106:15909), but also recognizes the "omics" biotic information as a signal resulting from structured interaction between the microbiotic and the abiotic at microscales. Recognizing this interaction, and its potential implications at larger scales, will support greater understanding of how the remarkable microbial functional and community diversity in natural ecosystems persist, and how essential ecosystem functions, and applied microbial processes important for humanity, are sustained.

Multiscale Simulation of Reactive Transport in Porous Media

Motivation
Multiphase flow and reactive transport processes in natural and man-made porous media are ubiquitous in Earth and energy systems. Examples include development of petroleum and geothermal reservoirs, geological storage of carbon dioxide and nuclear wastes, fate and transport of underground contaminants, fuel cells and batteries[3], fixed/fluidized bed reactors[4], and micro power plants. In these systems, although the key processes of fluid mobility and associated physical, chemical, and biological processes are ultimately governed by pore-scale interfacial phenomena, it is virtually impossible to solve the pore-scale processes over any desired macro length scales because of the wide disparity in scales ranging from pore and molecular to field. Consequently, a continuum formulation of reactive flow relying on averaging system properties over a representative elementary volume (REV) has been generally employed. The unresolved pore-scale phenomena, together with the empirical constitutive parameters, will likely lead to significant uncertainties and deviations in reactive transport modeling at the larger scale[1]. For example, the scaling effect may be a significant cause of the discrepancy between lab-measured and field-derived rate constants. Therefore, to better predict reactive transport processes in porous media and to reduce uncertainties in numerical modeling at scales of interest, it is imperative to greatly enhance our understanding and capability to simulate multiphase flow, advection, diffusion, and reaction processes at the pore-scale, and to develop the ability to incorporate these processes in the continuum scale[5].

Challenge
In the past years, reactive transport processes at the pore scale involving strongly coupled flow, transport, and precipitation/dissolution reactions in realistic geochemical systems have been explicitly included in numerical models[6-10]. These studies have provided the basis for significantly improved understanding of reactive transport in porous media at the pore scale. However, it remains challenging to directly use the pore-scale results to improve quantitative predictions based on large-scale simulations for all but the most simplified systems. This is due to a lack of fundamental understanding of the impact of pore-scale phenomena at the continuum scale and direct integration across scales in the simulations. Multiscale capabilities are essential for numerical models to accurately predict multiscale processes and multiscale heterogeneities in geological systems. Yet the effectiveness and limitations of existing multi-scale modeling approaches have not been rigorously evaluated.

Fig. 1: Comparison of concentration distribution for upscaled LBM and various continuum model simulations of reactive transport in a structured medium[1].

Approach
In cases where the systems are not far from equilibrium so that the nonlinear effects are negligible, and various length scales are sufficiently different for processes at two or more consecutive scales to be separable, a dual (Fig. 1) or multiple continuum approach[1, 11, 12] can be used to more accurately model the impact of sub-grid scale variability for reactive multiphase flow and transport. In this approach, the pore fluid is partitioned into flowing and stagnant domains. This approach has been
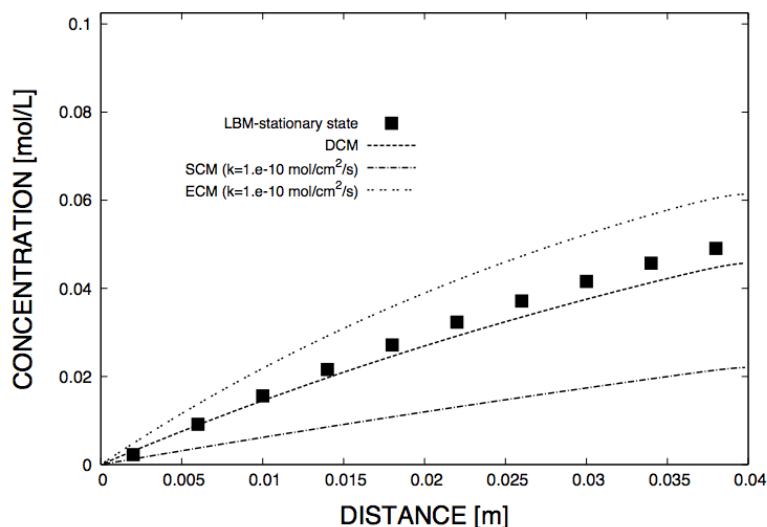
used to model chemical processes in soils and fractured media. Some success with multi-rate models has been achieved at the Hanford 300 Area site for modeling leaching of uranium from Hanford sediments. In many cases, however, there are no clear distinctions among length scales, and the reaction terms are generally nonlinear. For these cases, the hybrid multiscale approach combining a pore-scale model in a small fraction of the overall domain and a continuum-scale model that solves Darcy-scale governing equations at the REV scale over the remainder of the domain can be used[2] (Fig. 2).
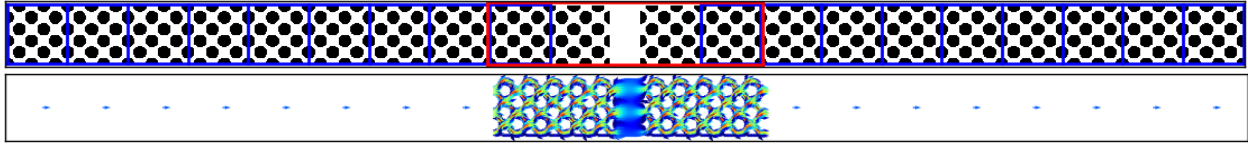

Fig. 2: Hybrid model simulation for flow across a fracture in a fractured medium. Top: simulation domain consisting of **continuum** and **pore-scale** subdomains; Bottom: velocity field, with a single value for each continuum cell and well-resolved flow within the pore-scale region[2].

Key References
1.   Lichtner, P. C.; Kang, Q. J., Upscaling pore-scale reactive transport equations using a multiscale continuum formulation. *Water Resources Research* **2007,** *43*, (12).
2.   Coon, E.; Porter, M. L.; Kang, Q. J.; Lichtner, P. C.; Moulton, J. D., Coupling lattice Boltzmann and continuum equations for the solution of multiscale flow and reactive transport in porous media. In *Computational Methods in Water Resources (CMWR) XIX International Conference*, Urbana-Champaign, IL, 2012.
3.   Mukherjee, P. P.; Kang, Q. J.; Wang, C. Y., Pore-scale modeling of two-phase transport in polymer electrolyte fuel cells-progress and perspective. *Energy & Environmental Science* **2011,** *4*, (2), 346-369.
4.   Chen, L.; Kang, Q. J.; He, Y.-L.; Tao, W.-Q., Pore-scale simulation of coupled multiple physicochemical thermal processes in micro reactor for hydrogen production using lattice Boltzmann method. *International Journal of Hydrogen Energy* **2012,** *37*, (19), 13943-13957.
5.   Pore-scale modeling challenge & workshop. In EMSL at Pacific Northwest National Laboratory, Richland, Washington., August 9-10, 2011, http://pore-scale.emsl.pnl.gov.
6.   Kang, Q. J.; Lichtner, P. C.; Zhang, D. X., Lattice Boltzmann pore-scale model for multicomponent reactive transport in porous media. *Journal of Geophysical Research-Solid Earth* **2006,** *111*.
7.   Kang, Q. J.; Lichtner, P. C.; Zhang, D. X., An improved lattice Boltzmann model for multicomponent reactive transport in porous media at the pore scale. *Water Resources Research* **2007,** *43*, (12).
8.   Li, L.; Peters, C. A.; Celia, M. A., Upscaling geochemical reaction rates using pore-scale network modeling. *Advances in Water Resources* **2006,** *29*, (9), 1351-1370.
9.   Tartakovsky, A. M.; Meakin, P.; Scheibe, T. D.; Wood, B. D., A smoothed particle hydrodynamics model for reactive transport and mineral precipitation in porous and fractured porous media. *Water Resources Research* **2007,** *43*, (5), 18.
10.   Kang, Q. J.; Lichtner, P. C.; Viswanathan, H. S.; Abdel-Fattah, A. I., Pore Scale Modeling of Reactive Transport Involved in Geologic CO2 Sequestration. *Transport in Porous Media* **2010,** *82*, (1), 197-213.
11.   Liu, C. X.; Zachara, J. M.; Qafoku, N. P.; Wang, Z. M., Scale-dependent desorption of uranium from contaminated subsurface sediments. *Water Resources Research* **2008,** *44*, (8), 13.
12.   Zyvoloski, G. A.; Robinson, B. A.; Viswanathan, H. S., Generalized dual porosity: A numerical method for representing spatially variable sub-grid scale processes. *Advances in Water Resources* **2008,** *31*, (3), 535-544.

Understanding cross-scale behavior

RTM has been extensively used in the past 3 decades to understand complex subsurface processes in many applications. Up until now most applications have focused on processes at spatial scales ranging from single pore or single cell scale (microns, tens of microns) to field scale (10s of meter) (Li et al., 2011), with a few exception at the watershed or catchment scale (100s meter) (Maher and Chamberlain, 2014). With recent advances in microbial techniques, there is a need to develop models that describe processes at scales within individual cells to understand controls of microbial community structure and their response to changing biogeochemical conditions.

On the other hand, with the recent emphasis in climate change, there is a significant need to develop models at the watershed, continent, and global scales to understand and quantify the contribution of subsurface processes to global processes such as elemental cycling (C or N). Subsurface processes have in general been ignored in existing models that describe large-scale carbon processes (De Kauwe et al., 2014; Walker et al., 2014). Models that couple subsurface processes with surface hydrology, land surface interactions, meteorological and climatic processes are essentially lacking, although vigorous interests have developed in recent years.

In my opinion, downscaling to smaller scales with finer details or upscaling to larger scale with much larger spatial domain is not merely a matter of increasing computational cost. It is a matter of developing conceptual models that can adequately capture the dominant processes at the scales of interest. When downscaling, we inevitably need to know more details, while upscaling means inevitably ignoring some finer scales details. Mechanistic understanding and conceptual models that can describe data and phenomena at the scale of interests require understanding of cross-scale behavior.

With a few exceptions (Bao et al., 2014; Battiato et al., 2009; Battiato et al., 2011), most existing work using RTM have focused on individual scales without linking processes and phenomena across scales. As an example, the following upscaling question has remained largely unanswered:

- *How does scale transition occur? What fine scale processes and parameters are essentially in representing large scale phenomena?*

For example, for carbon decomposition at the watershed scale, the following upscaling question is relevant:

- *What are the key controls at finer scales that need to be carried over to represent soil carbon decomposition at the watershed scale? What is the form of the governing equation?*

References:

Bao, C., Wu, H., Li, L., Long, P.E., Newcomer, D., Williams, K.H., 2014. Uranium Bioreduction Rates Across Scales: Biogeochemical Hot Moments and Hot Spots During a Biostimulation Experiment at Rifle, Colorado. Environmental Science & Technology, in press.

Battiato, I., Tartakovsky, D.M., Tartakovsky, A.M., Scheibe, T., 2009. On breakdown of macroscopic models of mixing-controlled heterogeneous reactions in porous media. Advances in Water Resources 32, 1664-1673.

Battiato, I., Tartakovsky, D.M., Tartakovsky, A.M., Scheibe, T.D., 2011. Hybrid models of reactive transport in porous and fractured media. Advances in Water Resources 34, 1140-1150.

De Kauwe, M.G., Medlyn, B.E., Zaehle, S., Walker, A.P., Dietze, M.C., Wang, Y.-P., Luo, Y., Jain, A.K., El-Masri, B., Hickler, T., Wårlind, D., Weng, E., Parton, W.J., Thornton, P.E., Wang, S., Prentice, I.C., Asao, S., Smith, B., McCarthy, H.R., Iversen, C.M., Hanson, P.J., Warren, J.M., Oren, R., Norby, R.J., 2014. Where does the carbon go? A model–data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO_2 enrichment sites. New Phytologist 203, 883-899.

Li, L., Gawande, N., Kowalsky, M.B., Steefel, C.I., Hubbard, S.S., 2011. Physicochemical heterogeneity controls on uranium bioreduction rates at the field scale. Environmental Science and Technology 45, 9959-9966.

Maher, K., Chamberlain, C.P., 2014. Hydrologic Regulation of Chemical Weathering and the Geologic Carbon Cycle. Science 343, 1502-1504.

Walker, A.P., Hanson, P.J., De Kauwe, M.G., Medlyn, B.E., Zaehle, S., Asao, S., Dietze, M., Hickler, T., Huntingford, C., Iversen, C.M., Jain, A., Lomas, M., Luo, Y., McCarthy, H., Parton, W.J., Prentice, I.C., Thornton, P.E., Wang, S., Wang, Y.-P., Warlind, D., Weng, E., Warren, J.M., Woodward, F.I., Oren, R., Norby, R.J., 2014. Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO_2 enrichment experiments: Model performance at ambient CO_2 concentration. Journal of Geophysical Research: Biogeosciences 119, 2013JG002553.

Position Paper for "Multiscale Computation: Needs and Opportunities for BER Science"
Peter Ortoleva – Indiana University

Multiscale Theory and Computation for Microbes and Nanomaterials

Motivation

Nanoscale materials self-assemble or undergo structural transformation in response to changes in their environments. These systems have distinguished strength, electronic, or chemical properties that make them of interest for materials engineering, energy conversion and storage, medical, and environmental applications. Since these systems typically involve millions of atoms and evolve via processes occurring on a broad range of scales in length and time, multiscale methods are needed to understand them and achieve the goal of computer-aided design. Responses to their microenvironments are driven by structural and chemical processes.

Challenge

Theoretical and computational approaches are developed to simulate these systems with atomic accuracy over long times. To enhance the reliability of multiscale modeling, a goal is to develop approaches that involve a minimal of phenomenology. Specifically, a supramillion atom system is simulated using only the interatomic force field, i.e., without the use of conjectured forms of rate equations satisfied by coarse-grained variables which are calibrated using experimental or molecular dynamics simulation. As experimental data at atomic resolution is now available for validation, and as many processes (e.g. chemical reactions) require atomic scale information, the models should simultaneously account for processes on space and time scales from atomic to course-grained. A spectrum of coarse-grained variables is introduced that are tailored to the system and phenomena of interest. Types that have been coevolved with the atom-resolved states include space-warping [1, 2], curvilinear coordinates for polymers [3], density-field variables [4], conserved quantities such as energy [5], and scaled coordinates [6, 7]. Atomic resolution is preserved by coevolving the coarse-grained and atom-resolved states [8, 9]. The multiscale coevolution approach is developed for both classical systems (i.e., molecular dynamics) as well as for quantum electronic systems.
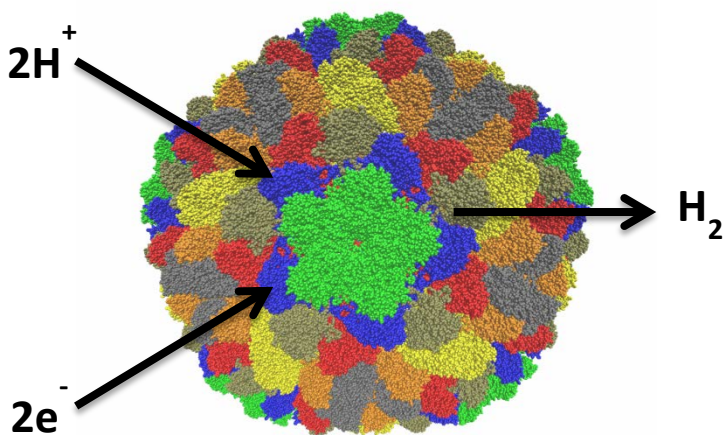

Fig. 1: Supramillion atom structures of interest in materials engineering, environmental remediation, and renewable resource generation are being efficiently designed and analyzed with our Multiscale MD Toolkit.

Approach

The coevolution methodology is deductive in character. Thus we start with Newton's or Schrodinger's equations for N particles and introduce coarse-grained variables defined in terms of microvariables. The equations for the coevolution of the micro and CG states are then derived. Techniques used to derive the coevolution algorithms include renormalized perturbation expansion [8, 10], Lie-Trotter factorization [5, 9], and variational principles [7, 11]. Results for the many-atom classical dynamics problem are implemented as an efficient computational platform that takes advantage of existing conventional MD, and similarly for the use of standard packages for quantum problems. To advance the field of multiscale theory and computation, the implementation is in two distinct forms. For the MD type calculations, the protoMD software is developed for researchers with interests in proposing and testing multiscale algorithms [12]. In contrast, the DMS package folds all the multiscale techniques directly into the GROMACS packages to achieve optimal computations efficiency by avoiding writing and reading files.

KEY REFERENCES

1. Jaqaman, K. and P.J. Ortoleva, *New space warping method for the simulation of large-scale macromolecular conformational changes.* J. Comput. Chem., 2002. **23**(4): p. 484-491.
2. Singharoy, A., S. Cheluvaraja, and P.J. Ortoleva, *Order parameters for macromolecules: Application to multiscale simulation.* J. Chem. Phys., 2011. **134**: p. 044104.
3. Shreif, Z. and P. Ortoleva, *Curvilinear All-Atom Multiscale (CAM) Theory of Macromolecular Dynamics.* J. Stat. Phys., 2008. **130**(4): p. 669-685.
4. Shreif, Z., S. Pankavich, and P.J. Ortoleva, *Liquid-crystal transitions: A first-principles multiscale approach.* Phys. Rev. E, 2009. **80**(3): p. 031703.
5. Sereda, Y.V., J.M. Espinosa-Duran, and Ortoleva, *Energy Transfer Between a Nanosystem and its Host Fluid: a Multiscale Factorization Approach.* Journal of Chemical Physics, 2014. **140**(7).
6. Shreif, Z. and P. Ortoleva, *Scaling Behavior of Quantum Nanosystems: Emergence of Quasi-particles, Collective Modes, and Mixed Exchange Symmetry States.* J. Chem. Phys., 2011. **134**(11): p. 104106.
7. Sereda, Y.V. and P.J. Ortoleva, *A multiscale variational approach to the kinetics of viscous classical liquids: The coarse-grained mean field approximation.* Journal of Chemical Physics, 2014. **140**(13).
8. Ortoleva, P.J., *Nanoparticle dynamics: A multiscale analysis of the Liouville equation.* J. Phys. Chem. B, 2005. **109**(45): p. 21258-21266.
9. Abi Mansour, A. and P.J. Ortoleva, *Multiscale Factorization Method for Simulating Mesoscopic Systems with Atomic Precision.* J. Chem. Theory Comput., 2014. **10**(2): p. 518-523.
10. Ortoleva, P.J., A. Singharoy, and S. Pankavich, *Multiscale theory of supramolecular assemblies.* Soft Matter, 2013. **9**: p. 4319-4335.
11. Sereda, Y.V. and P.J. Ortoleva, *Variational Principles and Multiscaling for Time-dependent Many-particle Classical Systems.* Physica A, 2013. **392**(4): p. 628-638.
12. Somogyi, E., A. Abi Mansour, and P.J. Ortoleva. *ProtoMD: A Prototyping Toolkit for Multiscale Molecular Dynamics*. arXiv:1309.6397 2014; Available from: https://github.com/Andrew-AbiMansour/proto_md.

Position Paper for "**Multiscale Computation: Needs and Opportunities for BER Science**"
Elijah Roberts – Johns Hopkins University

Multiscale modeling of reaction-diffusion and cytoskeletal dynamics during gradient signaling

Motivation
Many vital eukaryotic cellular processes require the cell to respond to a directional gradient of a signaling molecule. The first two steps in any chemotactic pathway are sensing and cell polarization. Many details are known regarding the biochemical details of these steps in various pathways, but much of the underlying physics of the process as a whole remains unresolved. Particularly, although it has become apparent in the last decade that cellular noise plays a large role in the accuracy of any response, the capacity for a given pathway to accurately track a chemical gradient despite cellular and environmental fluctuations is unknown. To effectively perform their functions, cells must have ways of mitigating (or taking advantage of) these fluctuations.

A fundamental omission in many existing models of eukaryotic chemotaxis is the lack of coupling between the biochemical description of the pathway and the mechanical description of the cytoskeleton. The cellular cytoskeleton may play a critical role in helping gradient sensing networks make accurate decisions. The long-term memory afforded by the slowly changing cytoskeletal structure introduces a longer time scale that can work with the fast reaction-diffusion dynamics to filter fluctuations. Thus the cytoskeletal dynamics, not just the biochemistry of the pathway, may be critical for understanding cellular gradient sensing behavior.

Challenge
To correctly model the spatial response of a eukaryotic cell one must be able to probabilistically describe both i) the reaction-diffusion dynamics of the signal transduction pathway, and ii) the vesicle and filament dynamics of the cytoskeleton, along with their interactions. The time and length scales of these processes are dramatically different and no tractable methods presently exist for simulating an integrated model.

Approach
Stochastic reaction-diffusion simulations have traditionally been considered too computationally costly for studying decision-making processes, which require long simulations. Two new approaches may break this assumption. The first is the development of algorithms with fine-grained parallelism that can take advantage of GPU accelerators. The second is the use, for the first time, of enhanced sampling (ES) methods to calculate the rare event statistics of stochastic reaction-diffusion processes. ES methods should allow agent-based models of cytoskeletal dynamics to be integrated with stochastic reaction-diffusion models using a simultaneous multiscale approach to explore the cytoskeleton's role in noise filtering. It can be anticipated that incorporating cytoskeletal dynamics will be a key innovation toward models with predictive power for

studying the chemotactic decisions of individual cells.
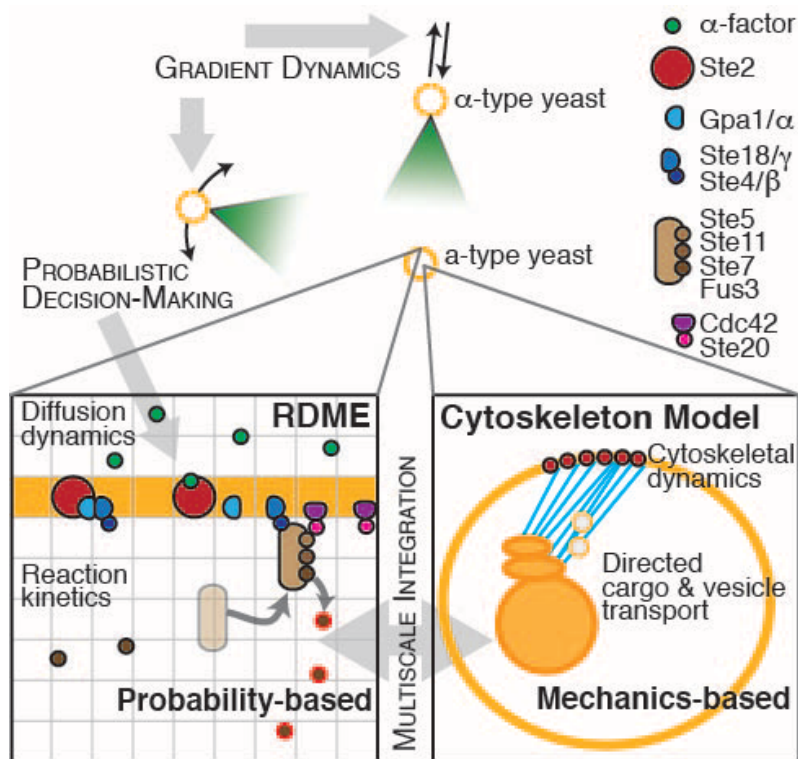
Figure 1 - Multiscale modeling of gradient signal transduction in the yeast mating response. Only a portion of the pathway is shown.

Key References

Roberts, E (2014) Cellular and molecular structure as a unifying framework for whole-cell modeling. Curr Opin Struct Biol 25:86–91.

Roberts,E, Stone, JE, Luthey-Schulten, Z (2013) Lattice Microbes: High-performance stochastic simulation method for the reaction-diffusion master equation. J Comput Chem 34:245–55.

Roberts, E, Magis, A, Ortiz, JO, Baumeister, W, Luthey-Schulten, Z (2011) Noise contributions in an inducible genetic switch: a whole-cell simulation study. PLoS Comput Biol 7:e1002010.

Hybrid Multiscale Simulation of Mixing-Controlled Biogeochemical Reactions

Motivation

It is now well known that apparent reaction rates observed in natural systems vary with observational scale, with rates generally decreasing with increasing length scales. This negatively impacts predictive ability of models, since directly measured or predicted fundamental rates cannot be used in application-scale simulations. This phenomenon is generally attributed to the apparent effect of transport limitations in heterogeneous systems at larger scales. For example, the apparent rate of microbial reduction of solid iron oxide minerals depends not only on the fundamental reaction rate at the grain surface but on the diffusion-limited rate of delivery of soluble electron donor and other substrates to the surface. This hinders the applicability of fundamental genome-scale models of microbial metabolism to field-scale reactive transport simulations (Tartakovsky et al. 2013).
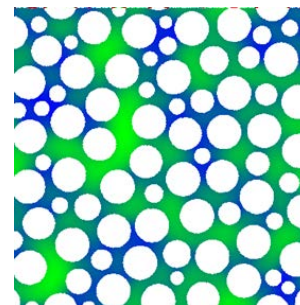

Figure 1: Simulated acetate distribution (green) in a pore-scale simulation of microbial iron reduction. Iron reduction occurs at the surface of solid grains (white). From Tartakovsky et al. 2013.

Challenge

Direct simulation of fundamental processes (microscale simulation) is computationally infeasible and lacks sufficient supporting characterization data. However, in some situations macroscopic (upscaled) models and parameters may not be valid. In such situations the ability to robustly predict system behavior from fundamental principles is called into question.

Approach

When it is possible to define (in terms of macroscopic variables) conditions under which macroscopic models are invalidated (Battiato et al. 2009), and such conditions are limited to a relatively small fraction of the model domain (in space and/or time), it may be possible to combine microscale and macroscale simulation in a single model framework, a process known as hybrid multiscale simulation (Scheibe et al. 2014). In the example discussed above, one would utilize pore-scale simulations (at which the fundamental genome-scale simulation of microbial metabolism is appropriate) where necessary, combined with upscaled models and parameters (e.g., advection-dispersion equation) where upscaling conditions are valid. Alternatively, one could use a hierarchy of models in which fundamental (e.g., pore-scale) models are run over short time periods to inform (parameterize) upscaled models with periodic updating as necessary (e.g., Tartakovsky and Scheibe 2011).
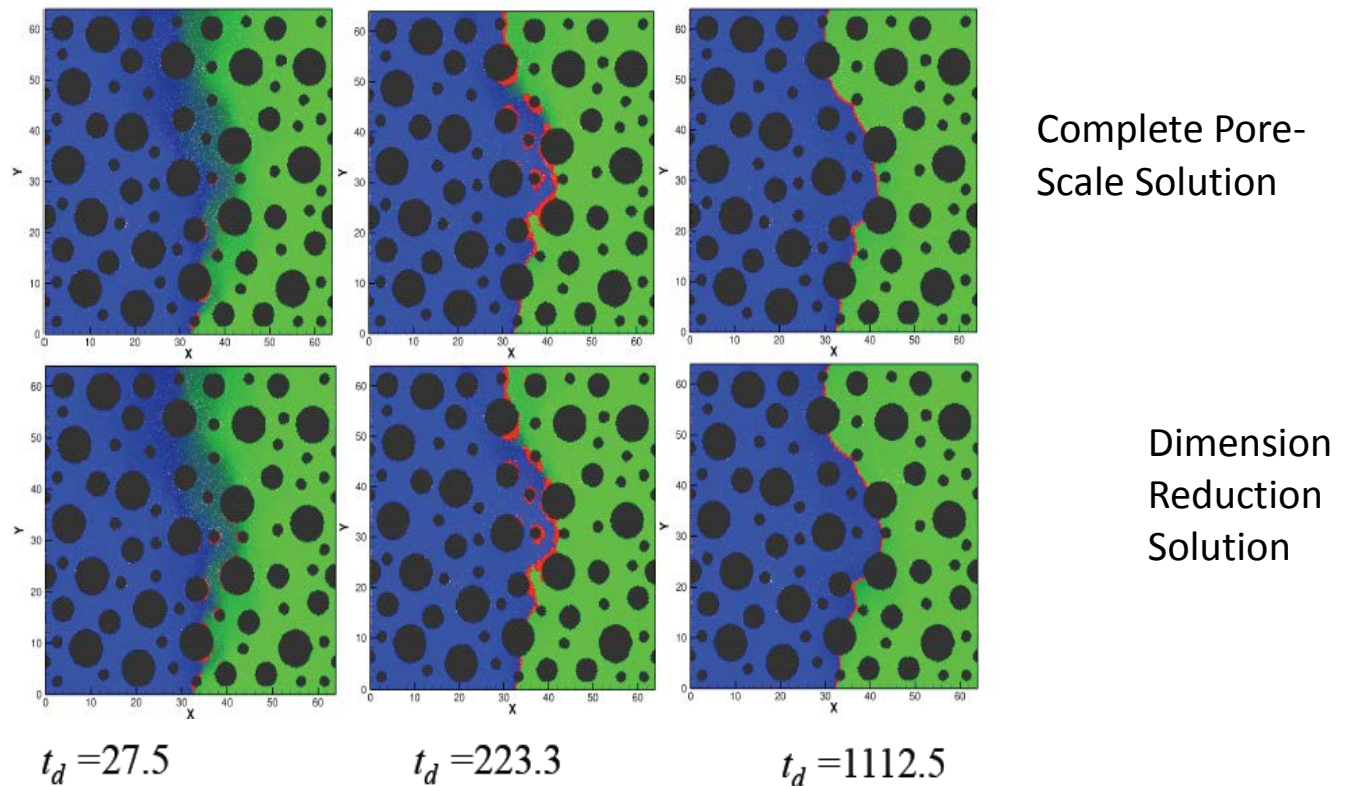
Complete Pore-Scale Solution

Dimension Reduction Solution

$t_d = 27.5$ $t_d = 223.3$ $t_d = 1112.5$

Figure 2: Simulation of two reacting solutes (blue, green) forming a mineral reaction product (red). Top row is a pore-scale simulation result at three selected times; bottom row is a comparable multiscale simulation with pore-scale processes simulated only over a portion of the time period. This result demonstrates the feasibility of a hybrid multiscale approach to a mixing-controlled reaction. From Tartakovsky and Scheibe, 2011.

Key References

Battiato, I., D. M. Tartakovsky, A. M. Tartakovsky, and T. Scheibe, "On breakdown of macroscopic models of mixing-controlled heterogeneous reactions in porous media," *Advances in Water Resources*, 32(11):1663-1673, 2009.

Scheibe, T. D., E. M. Murphy, X. Chen, K. C. Carroll, A. K. Rice, B. J. Palmer, A. M. Tartakovsky, I. Battiato, and B. D. Wood, "An analysis platform for multiscale hydrogeologic modeling with emphasis on hybrid multiscale methods," *Ground Water*, published online March 13, doi: 10.1111/gwat.12179, 2014.

Tartakovsky, A. M. and T. D. Scheibe, "Dimension reduction method for advection-diffusion-reaction systems," *Advances in Water Resources*, 34(12): 1616-1626, doi:10.1016/j.advwatres.2011.07.011, 2011.

Tartakovsky, G. D., A. M. Tartakovsky, T. D. Scheibe, Y. Fang, R. Mahadevan and D. R. Lovley, "Pore-scale simulation of microbial growth using a genome-scale metabolic model: Implications for Darcy-scale reactive transport", *Advances in Water Resources* 59: 256-270, doi:10.1016/j.advwatres.2013.05.007, 2013.

Multiscale Biological Computer Simulation at ORNL

Jeremy C. Smith, Xiaolin Cheng, Jerry Parks.

Computer modeling and simulation can rationalize experimental information across scales. Essential aspects of the planned development at ORNL are i) the integration into simulation models of available experimental information as well as –omics based systems models, ii) interfacing of simulation methods at different scales, by informing larger-scale, coarse-grained methods with results from finer-detail computations. The result will be a self-consistent description of mechanisms in and between cells. The scaling and use of these codes on DOE supercomputers, and in particular the TITAN machine and its successors at ORNL, will be integral to the success of this approach.

Large-Scale Molecular Dynamics (MD) – Supercomputing now permits the simulation at atomic detail of systems up to 100M atoms in size and on the microsecond timescale. Hence, atomistic MD has moved well beyond the single-molecule level to permit systems-level simulation of hundreds of interacting biological macromolecules such as those involving transport of chemical signals across the cellular membrane. Extrapolation of current performance at the petascale to the exascale indicates that we would, ultimately, be able to perform MD simulations of systems of ~10^{11} explicit interacting atoms, *i.e.,* approximately the number of atoms in a bacterial cell, for about 10 microseconds. Existing genomic and structural data will serve to provide information on starting configurations for these simulations.

Cell-Level Brownian Dynamics (BD) Simulation – Events in the cell that are on the millisecond timescale or longer, and system sizes beyond 100M atoms, currently call for simulation methods more simplified than atomistic MD, averaging out the unimportant degrees of freedom so as to preserve long time-scale properties. BD removes explicit solvent, and thus will permit us to simulate systems 100× larger or longer than those accessible to MD. These methods scale efficiently on a variety of supercomputers, and will permit cell-scale simulations on timescales up to one second. Important processes occurring in the millisecond-second time window include protein folding and dynamics, macromolecular associations, receptor activation, lateral diffusion and phase separation in membranes, and the catalytic cycles of transcriptional and translational machinery.

Integration of Simulation Models across Timescales – An important aspect of the work will be the use of information from atomistic MD to derive parameters for BD simulations, the results of which are in turn transferred as parameters for whole cell reaction-diffusion description such as that pioneered by Zan Luthey-Schulten (UIUC).

References:

1. F. NOE, I. HORENKO, C. SCHÜTTE and J.C.SMITH. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states **Journal of Chemical Physics** *126* (15):155102 (2007).
2. F. NOE, S. DOOSE, I, DAIDONE, M, LOLLMANN, M. SAUER, J.D. CHODERA & J.C. SMITH. Dynamical Fingerprints: Probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. **Proceedings of the National Academy of Sciences (U.S.A.)** 108(12):4822-7 (2011).
3. K. MORITSUGU and J.C. SMITH. Coarse-Grained Biomolecular Simulation with REACH: Realistic Extension Algorithm via Covariance Hessian. **Biophysical Journal** *93*(10):3460-3469 (2007).

Positon Paper for "**Multiscale Computation: Needs and Opportunities for BER Science**"
Al Valocchi- Dept. of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

Concurrent Hybrid Multiscale Methods for Reactive Transport in Porous Media: Some Challenges

Motivation

There have been several experimental and theoretical studies demonstrating that some standard continuum-scale models can break down for environmentally relevant cases of reactive transport in porous media. One important class of problems is mixing-controlled reactions, which have been investigated intensely in recent years, including many studies by several of the workshop participants (e.g., [1], [2] [3]). When the reaction occurs on a time scale that is shorter than the time scale of advection and diffusion processes that homogenize the reactant concentrations over the scale of the REV, then classical theory breaks down and the rate of reaction depends upon localized concentrations at the pore scale. Moreover, if the reaction leads to changes in the porosity (e.g., precipitation/ dissolution, biofilm growth) then it may also be difficult to upscale the impact of localized changes in pore space morphology upon continuum-scale properties like effective diffusion, dispersion, and permeability.

Hybrid multiscale methods are one approach to address this challenge. The spatial location of the mixing zone is identified; the "micro-scale" (i.e., pore scale) model is solved in this domain, and information is exchanged between the pore and continuum scale domains across the boundary. Several papers have been published recently demonstrating this approach (e.g., [4], [3], [5]). A schematic is taken from the survey paper by [1], shown below on the left. Pore-scale equations are solved in the central mixing zone, while continuum-scale equations are solved away from the mixing zone. On the right is shown a figure from our work on modeling biofilm growth [6]. Reactants diffuse across a central mixing zone where biodegradation occurs and biofilms develop; pore-scale models are required in the central zone when the reaction rate is large.
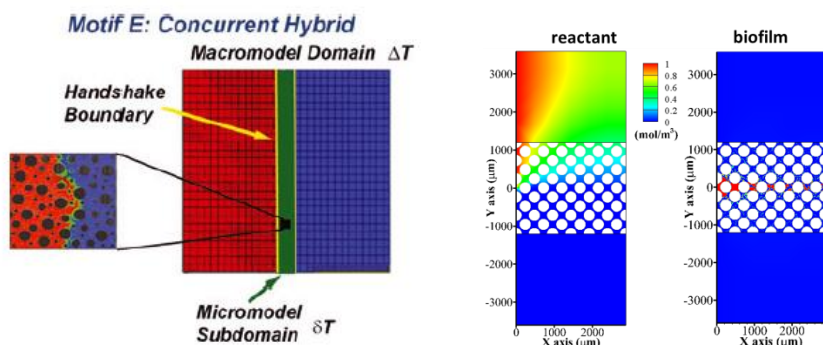
Some Challenges

I would like to raise several questions/challenges for discussion at the workshop. I believe that these issues may also pertain to other fields of application in addition to porous media transport.

Is it practical to explicitly solve the micro-scale problem within the mixing zones for realistic problems?

Field-scale problems normally encompass large spatial scales. For example, contaminant plumes span tens or hundreds of meters at hazardous waste sites, and sites of interest to BER (e.g., Hanford, Rifle). Mixing zones at the interfaces between fluids with different composition can therefore extend over these large scales, and it may not be feasible to solve pore-scale models at the millimeter to centimeter scale throughout the domain. An alternative may be to use "response-surface" approaches (e.g., [7]) where many micro-scale simulation are performed "off-line" for a wide range of parameters and conditions, the appropriate quantities of interest at the continuum-scale are computed, and then statistical (or maybe machine-learning) methods are used for upscaling. Perhaps other approaches are possible.

How do we account for uncertainty about the micro-scale?

Most published work applying hybrid models in porous media solve the pore-scale based on some assumed structure of the porous medium. Even if the porous medium is statistically homogeneous, it is not clear if another realization of the porous medium will yield the same macro-scale behavior. One approach would be to solve an ensemble of pore-scale problems, which leads to the possibility of stochastic multi-scale approaches. However, this may not be computationally feasible. In addition to uncertainty about the porous medium geometry, there will be many other sources of uncertainty regarding the pore-scale physics and biogeochemistry.
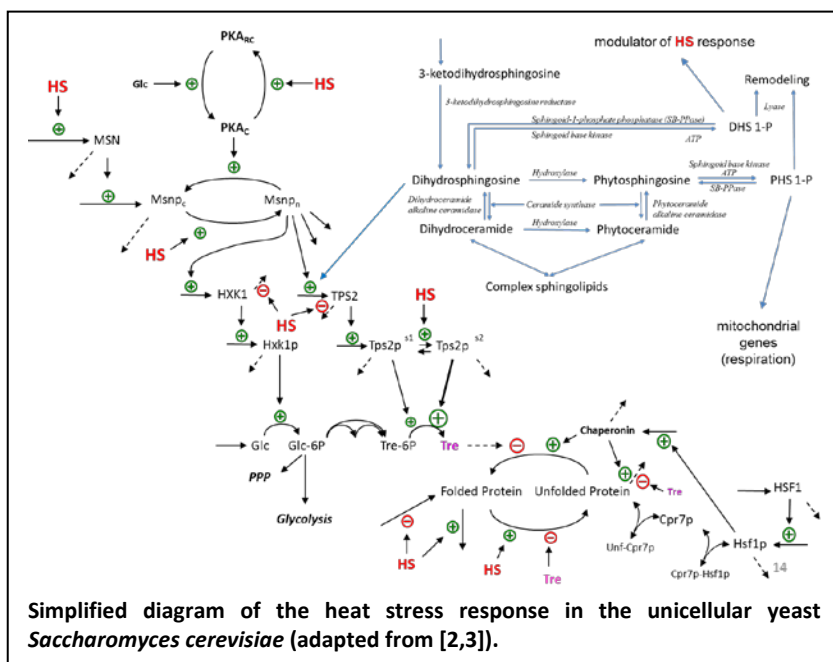
How do we evaluate (test/validate) multi-scale models?

Are there some benchmarks or tests to assess if model 'performance' is improved when a multi-scale approach is used? Will parameter calibration still be required when applying multi-scale models to real-world cases, and if so, will that complicate comparison with calibrated continuum-scale models with empirical reaction models?

References

1.   Scheibe, T.D., et al., *An Analysis Platform for Multiscale Hydrogeologic Modeling with Emphasis on Hybrid Multiscale Methods.* Groundwater, 2014: p. n/a-n/a.
2.   Battiato, I., et al., *On breakdown of macroscopic models of mixing-controlled heterogeneous reactions in porous media.* Advances in Water Resources, 2009. **32**(11): p. 1664-1673.
3.   Tartakovsky, A.M., et al., *Hybrid Simulations of Reaction-Diffusion Systems in Porous Media.* SIAM J. Sci. Comput., 2008. **30**(6): p. 2799-2816.
4.   Battiato, I., et al., *Hybrid models of reactive transport in porous and fractured media.* Advances in Water Resources, 2011. **34**(9): p. 1140-1150.
5.   Balhoff, M., S. Thomas, and M. Wheeler, *Mortar coupling and upscaling of pore-scale models.* Computational Geosciences, 2008. **12**(1): p. 15-27.
6.   Tang, Y., A.J. Valocchi, and C.J. Werth, *A hybrid pore-scale and continuum-scale model for solute diffusion, reaction and biofilm development in porous media.* Water Resour. Res., 2014. **submitted**.
7.   Varloteaux, C., S. Békri, and P.M. Adler, *Pore network modelling to determine the transport properties in presence of a reactive fluid: From pore to reservoir scale.* Advances in Water Resources, 2013. **53**(0): p. 87-100.

Of Mice and Men … and Microbes

Motivation
Biological systems are at the core of human existence. They directly determine our health and disease, are responsible for supplying food for a growing world population, and govern the dynamics of our living environments and ecosystems. Biological systems are found at a wide spectrum of scales in multiple dimensions of time, space, and functional organization. Most biological processes that ultimately govern these systems occur inside cells, at the molecular (nanometer) level and within seconds or faster. Yet, these processes operate in such large numbers and in such a synergistic and hierarchical manner that they make hearts tick, supply seven billion people and incomparably more animals with oxygen, keep the oceans viable, and are the source of most of the energy we use today. Recent advances in computing and systems biology offer the opportunity to study these complex phenomena in a quantitative, systemic manner. As an example from our lab, we currently analyze the dynamics of bacterial meta-populations in Atlanta's main water reservoir. In a very different project, we studied how slightly altered biochemical processes can lead to disease that emerges close to the end of life [1]. The vast discrepancy in scales makes the study of multiscale biological systems exciting and, at the same time, daunting.

Simplified diagram of the heat stress response in the unicellular yeast *Saccharomyces cerevisiae* (adapted from [2,3]).

Challenge
Biological systems pose a challenge that can be summarized in one word: complexity. This complexity is manifest in large numbers of components and processes, the nonlinearity of most of these processes, and threshold effects where small alterations in parameter values can trigger dramatic qualitative changes.

Large Numbers. Even simple bacteria typically possess thousands of genes, proteins, and metabolites. Our brain contains an estimated 100 billion neurons, which correspond with each other through trillions of connections in the form of synapses. Moreover, the function of the brain is supported by other cells, such as astrocytes and glia cells, at an order of 10 to 50 cells per neuron [4]. These magnitudes indicate that even bookkeeping is a difficult aspect of biological systems analysis and that functional information integration is much more challenging.

Nonlinear Processes. Essentially all processes in biology are nonlinear. In particular, we seldom find situations where operations scale linearly from small to large quantities. The nonlinearities often stem from the very rich regulatory networks that control even relatively simple systems. As a very

prevalent example, feedback inhibition keeps variables in check, but this mechanism also has the potential of creating oscillations and instabilities. The nonlinearity of processes creates significant challenges to the biomedical scientist or engineer, as it invalidates the principle of superposition, which in simple terms holds that the output to a sum of two inputs equals the sum of the outputs if the inputs are applied separately. As a consequence, biological systems cannot necessarily be taken apart and validly studied through analyses of their parts, and one must wonder to what degree the results of any *in vitro* experiments can be reliably translated into insights regarding functioning systems *in vivo*.

Threshold Phenomena. As a special case of nonlinear behaviors, many biological systems have the capacity to respond in a discontinuous, switch-like manner. Such switches have been observed in gene regulatory systems, signaling systems, and many other biological phenomena. The challenge of these thresholds is that critical values at which switches happen are rarely known, and that many potentially switching processes may coexist within the same system.

Approach

Until about the turn of millennium, mathematics beyond the high school level was essentially banned from mainstream biology, with the exception of statistics, which however was executed by specialists. The –omics revolution, with its supporting high-throughput data generating technologies, has left no doubt that this lack of interest in mathematics and computing is no longer affordable. The response to this situation has been the emergence first of bioinformatics and subsequently of systems biology. The former focuses primarily on large-scale, static molecular datasets, such as gene expression patterns and profiles of proteins and metabolites, while the latter embraces dynamic phenomena, adaptation, and the complexity sketched above. Bioinformatics and systems biology borrow from statistical methods such as machine learning, and also utilize dynamical systems and control theory, as well as numerous techniques from computer science. Systems biology furthermore draws from biomedical engineering, materials science, and robotics. While it is a young science, early practical applications are already crystalizing in the partner fields of synthetic biology and metabolic engineering. The range of potential applications is arguably limitless, spanning the spectrum from the microbial production of valuable organics to improved food production, personalized medicine, and the responsible stewardship of our environments.

Our lab uses primarily, although not exclusively, Biochemical Systems Theory (BST) [5-8] as the modeling framework for most of our biomedical applications. BST provides strong guidance for designing, diagnosing, and analyzing models, requires only minimal assumptions, and is in principle scalable to very large systems. Also, BST offers germane means for bridging temporal and organizational scales [9]. The key feature of BST is the exclusive use of products of power-law functions as process representations. While this strategy may seem restrictive, it is fundamentally advantageous for the design and analysis of models that are ill characterized. Furthermore, it was shown that BST models are rich enough to represent any nonlinearities that can be created with arbitrary systems of ordinary differential equations [10].

References
[1] Ferreira, A.E.N., A. Ponces, and E.O. Voit, *Biochem. J.* **376(1)**, 109-121, 2003.
[2] Fonseca, L.L., C. Sánchez, H. Santos, and E.O. Voit*, **7 (3),** 731 – 741, 2011.
[3] Voit, E.O.: *A First Course in Systems Biology*. Garland Science, New York, NY, 2012, xiv+445pp.
[4] http://www.livescience.com/32311-how-many-cells-are-in-the-brain.html
[5] Savageau MA (1976) Biochemical Systems Analysis. Addison-Wesley Pub. Co.
[6] Voit, E.O.: *Computational Analysis of Biochemical Systems.* Cambridge Univ. Press, Cambridge, 2000.
[7] Torres, N.V., and E.O. Voit: *Pathway Analysis and Optimization in Metabolic Engineering.* Cambridge University Press, Cambridge, U.K., 2002.

[8] Voit, E.O.: Biochemical Systems Theory: A review. *Int. Scholarly Res. Network (ISRN – Biomathematics)*, Article 897658, pp. 1-53, 2013.
[9] Savageau, M.A., *PNAS* **76**, 6023-6025, 1979.
[10] Savageau, M.A. and E.O. Voit, *Mathem. Biosci.* **87**, 83-115, 1987.

Position Paper for "Multiscale Computation: Needs and Opportunities for BER Science"
Gregory A. Voth – The University of Chicago

Motivation: There is a pressing need for a new integrated multiscale modeling methodology to study complex biomolecular systems of relevance to the DOE-BER mission (e.g., in bacteria and the environment). Applications of such an approach can also include large scale scientific computations on DOE leadership class computational resources. This effort will be unique in both its scope and its methodology, and it can provide a complementary synergy with existing DOE-funded activities in structural biology, environmental science, and bioinformatics.

Challenge: Biomolecular systems that operate at scales relevant to cellular biology collectively function on length scales of the order 1 µm and timescales of ms or longer. These phenomena are also intrinsically multiscale in nature and the molecular features of the underlying biomolecules help to define this collective behavior. Brute force computational approaches such as atomistic MD, even carried out on leading edge supercomputers, can provide only a relatively limited snapshot into the overall behavior of such systems; however, novel multiscale theoretical and computational approaches (in combination with large-scale computing) can provide a new paradigm to access the scales relevant to cellular biology. *A significant challenge therefore exists to develop and apply a transformative, integrated multiscale modeling methodology that can push the frontiers of biomolecular simulation into realms of cellular-scale biology, especially that relevant to the BER mission.*

Figure 1: Integrated Multiscale Biomolecular Simulation

(1) Identify Key Experimental Results and Targets
(2) Initial Coarse-Grained (CG) Model Development
(3) Carry Out Extensive Large Scale CG Simulations
(6) CG Model Refinement with Experimental Prediction/Verification
(5) Carry Out New All-Atom MD Simulations Constrained by CG Variables and Boundary Conditions
(4) Identify Important Interactions at the CG Scale
More Complex CG Interactions

(1) Carboxysome structure: Tanaka *et al.*, *Science* 2008, 319:1083-1086.

Approach: Developing and implementing a multiscale computational methodology to simulate key biomolecular processes requires incorporating chemical reactions, large and small-scale protein dynamics, and macromolecular self-assembly into a single unifying framework that spans length scales ranging from Ångstroms to microns and beyond. Moreover, this methodology must accurately bridge dynamical

phenomena that occur on the femto- to picosecond timescale to processes that occur on the millisecond biological timescale, and it must do so for multi-million or billion atom molecular systems. This is a multiscale challenge of the highest degree in both the size and time domains. Overcoming this challenge will require an elegant incorporation of the quantum (chemically reactive) phenomena into classical simulations, the sampling of rare events, atomically detailed and coarse-grained (CG) force fields, and coupled kinetic models all within a rigorous framework of multiscale molecular modeling and simulation to ultimately understand the mechanism of complex biomolecular processes. A key component of this overall scientific methodology can be derived from coarse-graining theory,[1] broadly expanded, designed, and implemented within an innovative "hybrid" multiscale modeling framework (cf. Fig. 1) that also extensively utilizes experimental data and "knowledge-based" computational modeling.

The proposed overall hybrid multiscale modeling methodology is schematically depicted in Fig. 1 on the previous page for the specific and very interesting example of the bacterial carboxysome, a specialized microcompartment that dramatically enhances the rate of carbon fixation in bacteria.[2] This application is just one example of the overall approach, but it illustrates the key integrated concepts of the approach. As the first step, an initial multiscale simulation is designed with direct input from experimental data, e.g., in this case cryo-electron microscopy data (upper left panel). From here an initial knowledge-based "first pass" CG model is developed (upper middle panel) that represents a substantial paradigm shift away from traditional "bottom-up" CG modeling of biomolecular systems.[1] In this CG model the system is very aggressively coarse-grained (i.e., not at the amino acid level, but at a significantly lower CG resolution). Importantly, physical principles are still utilized extensively in this level of CG modeling along with experimental data (i.e., blending physical interactions and "knowledge-based" modeling). Next, large-scale and highly scalable CG MD simulations are run (upper right panel) in order to characterize the multi-protein system and, more importantly, to identify specific regions at the CG level (lower right panel) that are then targeted for refinement and additional high resolution all-atom MD simulations (lower middle panel), which are *guided by both the CG model and experimental data*. (In this particular case, the coat proteins that control diffusion across the carboxysome shell would be studied in the CG-guided all-atom simulations.) Next, the CG models are refined, expanded, and verified based on the CG-guided all-atom MD simulations in combination with additional high resolution experimental results if available (e.g., structural information of hexamer interfaces or coat protein composition via x-ray crystallographic or NMR studies). Finally, the overall multiscale methodology either continues refinement via an iterative process (looping back to the upper left) or it terminates and provides the desired predictions/results. This overall hybrid multiscale modeling framework can provide a transformative theoretical and computational methodology that will be important to he DOE-BER mission. Preliminary results demonstrating the feasibility of certain aspects of this approach have been published by the Voth group for the case of the HIV-1 virion,[3] and key aspects of the coarse-grained model theory and algorithms published.[4,5]

Key References

1. Voth GA ed (2009) *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC Press/Taylor and Francis Group, Boca Raton, Fla.), 1st. Ed.
2. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO, Atomic-Level Models of the Bacterial Carboxysome Shell. *Science* **2008**, *319*, 1083-1086.
3. Ayton GS, Voth GA, Multiscale Computer Simulation of the Immature Hiv-1 Virion. *Biophys. J.* **2010**, *99*, 2757-2765.
4. Dama JF, Sinitskiy AV, McCullagh M, Weare J, Roux B, Dinner AR, Voth GA, The Theory of Ultra-Coarse-Graining. 1. General Principles. *J. Chem. Theory Comput.* **2013**, *9*, 2466-2480.
5. Grime JMA, Voth GA, Highly Scalable and Memory Efficient Ultra-Coarse-Grained Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 423-431.

Appendix D Code Descriptions

Code	Organization	Process Simulated	Primary Applications	User Base	Contacts
CLM-PFLOTRAN	LBNL	Land surface model that includes 2-D surface flow, 3-D subsurface flow, and reactive transport processes • 2-D non-isothermal surface flow • 3-D non-isothermal variably saturated subsurface flow (including ice phase) • 3-D reactive transport processes • Land surface processes including surface energy balance and vegetation dynamics	Understanding fate and transport of below-ground carbon in Arctic landscapes under climate change.	DOE laboratories: LBNL, ORNL	Gautam Bisht
AMDF (Long time-scale atomistic simulations)	LANL	Atomistic Dynamics of Materials • Thermodynamics (phase diagrams, free energy barriers) • Kinetics (rates of defect nucleation, diffusion, clustering)	• Nuclear materials • Mechanical deformation of solids • Nanotechnology (film growth, nanoscale friction)	Mostly LANL, a few universities	Danny Perez
Unnamed Code	UT-Austin	Pore-scale and continuum-scale single-phase fluid flow, transport, and reactions • Newtonian and non-Newtonian flow Transport with diffusion and advection (modules for computing dispersion coefficents) • Homogeneous and heterogeneous reactions • Coupling pore-scale regions to continua for hybrid modeling	• Contaminant flow and transport with adsorption/reaction effects • Upscaling flow, transport, and reaction properties • High performance pore-scale domain decomposition and hybrid modeling	Balhoff research group	Matthew Balhoff
Boltzmann (Agent-based Cell Metabolism)	PNNL	General code for simulating sets of coupled reactions using statistical thermodynamics/fluctuation theory	Cell metabolism, nanotechnology/materials	In development, not released	Bill Cannon

Code	Organization	Process Simulated	Primary Applications	User Base	Contacts
Mosart (Model for Scale Adaptive River Transport)	PNNL	Reach-scale transport and transformation of water, heat and biogeochemistry fluxes	• Reach to regional scale channel water stages and velocity variations with explicit accounting for climate and human-induced changes • Reach to regional scale stream temperature variation with explicit account for climate and human-induced changes • Reach to regional scale sediment, C and N cycles with explicit account for climate and human-induced changes	PNNL, WSU	Hong-Yi Li
NWChem (High-performance Computational Chemistry)	PNNL	Chemical reactions; electronic structure of molecular assemblies, biomolecules, nanostructures, and solid-state; ground-and excited-state chemical processes; molecular properties, and relativistic effects	Molecular spectroscopy, catalysis, condensed phase systems in complex environments, thermodynamics of chemical processes, interfacial chemistry	Open source (downloaded ~10,000 times)	Karol Kowalski
Reactive Transport Lattice Boltzmann Method	LANL	Pore-scale multiphase reactive transport • Liquid/gas flow, with phase transition Transport (advection/diffusion) of multi-component solute species • Heat Transfer (convection/diffusion) • Multi-component homogeneous reactions, either treated kinetically or as in instantaneous equilibrium Precipitation/dissolution of multiple minerals, nucleation, sorption, ion exchange • Changes in solid/pore geometry	• Contaminant transport and biomass growth • Multiphase flow and reactive transport involved in CO_2 sequestration • Coupled mass transport and (electro) chemical reaction in fuel cells and micro-reactors	LANL	Qinjun Kang
Taxila LBM	LANL	Pore-scale multiphase fluid flow, limited transport and reaction capability • Lattice-Boltzmann simulations for multiphase flow, using modified Shan-Chen model • SRT, MRT, higher order derivative estimates for stable interfaces	• Comparisons to analogue fluid microfluidic device experiments • Carbon sequestration applications	LANL	Qinjun Kang

Code	Organization	Process Simulated	Primary Applications	User Base	Contacts
AMBER (MM & QM/MM)	UCSD	Classical Particle MM using either MM or QM/MM potentials • Solution Phase biomolecules • Lipid membranes and membrane transport • Membrane bound biomolecules • Protein, small molecule surface interactions	• Protein-Ligand binding free energies • Enzyme reaction mechanisms • Lead discovery and optimization • Biocatalysis design • Lipid transport mechanisms	~1,100 site licenses/version 15,000 users 50 developers	Ross Walker
CrunchFlow (EcoTrait) (Microbial Trait-Based Reactive Transport)	Berkeley Lab	Multicomponent Reactive Transport • Variable density flow, advective-diffusive-dispersive transport • Multicomponent aqueous equilibrium, gas equilibrium, mineral precipitation and dissolution, surface equilibrium, ion exchange, isotopes • Microbially mediated reactions with thermodynamic considerations	• Water-rock interactions, chemical weathering, hydrothermal • Radionuclide transport, exchange and sorption • Contaminant transport and bioremediation • Soil, sediment, water biogeochemical (C, N, Pc, S, Fe etc) cycling with adaptive microbiology informed by metagenomic reconstruction	• CrunchFlow-widely used • EcoTrait-under development at LBNL	Eoin Brodie
eSTOMP (eXtreme Scale Subsurface Transport Over Multiple Phases)	PNNL	Nonisothermal multiphase flow and multicomponent reactive transport, geomechanical deformation	• Field scale simulation of flow and contaminant transport • Field scale simulation of biostimulation Coupled with geomechanics simulator to model ground deformation due to CO_2 injection	• Scalable version used primarily by PNNL staff and one DOE contractor • Serial version used by national laboratories, government agencies, private companies, and universities	Yilin Fang Steve Yabusaki Bruce Palmer Tim Scheibe
SPH and DPD modules in LAMMPS	PNNL	Multiscale, multiphase, and multiphysics modeling • Complex fluids (non-Newtonian fluids, plastic materials, suspensions) • Soft matters (polymers, colloids, biological cells) • Moving interfaces and boundaries	• Ice sheet dynamics (non-Newtonian fluid, moving interface) • Friction stir welding process (plastic large deformation, moving boundary) • Colloid transport in porous media (fully coupling of colloid dynamics and fluid dynamics) • Electrokinetic flow (coupling of hydrodynamics and electrostatics)	PNNL, Brown University, SNL	Wenxiao Pan, Alex Tartakovsky

Code	Organization	Process Simulated	Primary Applications	User Base	Contacts
PFLOTRAN	SNL, ORM Research, LBN,L LANL, Intel	Subsurface reactive multiphase flow, heat transfer, biogeochemical transport, geomechanics • Immisible/miscible two-phase (liquid-gas) flow • Heat transfer • Multicomponent solute transport (advection/hydrodynamic dispersion) • Chemical reactions: aqueous speciation, ion activity models, mineral precipitation-dissolution, ion exchange, isotherm-based sorption, surface complexation (equilibrium, kinetic, multirate), radioactive decay and ingrowth, colloids, microbiological • Multiple interacting continua	• Carbon sequestration • Climate change • Enhanced geothermal energy • Fate and transport of contaminants • Hydraulic fracturing • Nuclear waste disposal • Radioisotope tracers	~10 developers 20+ users (unknown) Mailing lists: pflotran-dev@googlegroups.com pflotran-users@googlegroups.com	Glenn Hammond, Peter Lichtner
COBRA (Metabolic Network Modeling)	UW-Madison	Intracellular metabolic reaction rates, cellular growth, cellular consumption/production rates • Predicts rates of intracellular reactions in genome-scale networks • Predicts cellular growth rates for varying environmental conditions and genetic perturbations. • Predicts cellular consumption/production rates of environmental nutrients	• Metabolic engineering (strain design, process design) • Systems biology: data integration and analysis • Experimental design (network discovery)	>500 users	Jennifer Reed
TETHYS (Transient Energy Transport Hydrodynamics Simulator)	PNNL	• Laminar and Turbulent Incompressible Fluid Flow (Navier-Stokes) • Energy Transport – conduction, convection, buoyancy, conjugate heat transfer • Solute Transport (Eulerian) – diffusion, advection, conjugate mass transfer, homogeneous reactions, surface reactions • Solute Transport (Lagrangian) - diffusion, advection, homogeneous reactions • Estimation of dispersion in complex geometries using particle tracking and volume averaging • Linkage to geochemical code underway	• Environmental hydrodynamics and water quality • Pore-scale flow and transport	PNNL staff	Marshall Richmond, Bill Perkins

Code	Organization	Process Simulated	Primary Applications	User Base	Contacts
CrunchFlow	LBNL, PSU	Flow, transport, and (bio)geochemical reactions at multiple scales (pore (microns) to field scales (10s meters)) • single-phase flow (water) • Dispersion and diffusion • Reaction processes: mineral dissolution/precipitation, surface complexation, ion exchange, microbe-mediated redox reactions	• geological carbon sequestration • environmental bioremediation • chemical weathering, soil formation	LBNL, various universities (PSU, Stanford, Berkeley)	Carl Steefel
DHSVM (Distributed Hydrology Soil Vegetation Model)	PNNL	A distributed, physics-based approach to solve coupled energy and water balance equations describing watershed processes at high spatial (10 to 90-m) and temporal (hourly) resolution • Topographic controls on energy transfer and water movement • Spatial/temporal distribution of canopy/ground snow conditions, evapotranspiration, soil moisture, ground saturation, runoff and streamflow • Distributed erosion and sediment transport • Stream temperature • Nutrient cycling and export	• Land surface energy and moisture fluxes • Climate-hydrology interactions • Impacts of climate and land use change on water resources	Broad and active international user base >100	Mark Wigmosta
PLAS (Biological Systems Analysis)	Georgia Tech	Biological processes; typically metabolic pathway systems	Simulation and analysis of ODE models from the fields of biology and medicine; student training	System biologists	Antonio Ferreira
protoMD (Process Simulation)	IU-Bloomington	Self-assembly, structural transition, and chemical reactions in nanoscale supramolecular assemblies	Development of multiscale algorithms. Computer-aided nanomaterial design with applications in nanoreactors, vaccines, strong materials, and nanoelectronics	Small group of selected users	Peter Ortoleva
GROMACS NAMD Amber CHARMM	ORNL	Molecular Scale, 0.1-10nm; 1femtosec-10microsecs	Bioenergy, Biology, Biogeochemistry, Chemistry, Materials	• GROMACS:16,000 citations; • CHARMM 31,000 etc	Roland Schulz
RAPTOR (Rapid Approach for Proton Transfer and Other	University of Chicago ANL	Reactive MM simulations • Multistate simulation algorithms • Atom-transfer, synthesis, decomposition, mutation reactions • Large-scale condensed phase	• Proton and Hydroxide transport in clusters, liquids, proteins, polymer electrolyte membranes • Fundamental processes in bulk water and at water-air interface	Voth group and collaborators (~20)	Gregory Voth

Code	Organization	Process Simulated	Primary Applications	User Base	Contacts
Reactions) (add-on for LAMMPS)		simulations • Multi-configurational coarse-grained models	• Biomolecular: Proton transport in Influenza A M2, CcO, carbon nanotubes • Electrochemical: Proton and hydroxide exchange membranes		
UCG-MD (Ultra-Coarse-Grained Molecular Dynamics)	University of Chicago, ANL	Large-scale, Coarse-Grained MD • Efficient load balancing of large CG systems via Hilbert space-filling curves (implicit solvent etc.) • Internal states/properties of CG "beads" dynamic at runtime ("Ultra-CG")	• Self-assembly of large, biologically relevant structures (e.g., viral capsids) • Dynamics of large-scale assemblies (membranes, cytoskeleton)	Voth group and collaborators (~5, increasing as code develops))	Gregory Voth