April 2007

Report of the

Computational Subsurface Sciences Workshop

January 9 - 12, 2007

Sponsored by the DOE Office of Science
in Cooperation with the Offices of

Environmental Management
Fossil Energy
Civilian Radioactive Waste Management

Hosted by the Office of Advanced Scientific Computing Research

Office of Science
U.S. DEPARTMENT OF ENERGY

Environmental Management

Fossil Energy

Civilian Radioactive Waste Management

Front Cover Collage Acknowledgements

STOMP-WO simulation image showing DNAPL migration through the heterogeneous subsurface at the Brooklawn Site, Louisiana; project funded through NPC Services.

STOMP-CO_2-SC simulation image showing gas saturation of the injection of supercritical CO_2 into the Rose Run formation at the American Electric Power Site, West Virginia; project funded through Battelle, Columbus.

WORKSHOP ORGANIZERS

ORGANIZING COMMITTEE

- Todd Arbogast – University of Texas at Austin
- Ken Eggert – Los Alamos National Laboratory
- Mike Heroux – Sandia National Laboratories
- Tissa Illangasekare – Colorado School of Mines
- Moe Khaleel – Pacific Northwest National Laboratory
- Paul Meakin – Idaho National Laboratory
- Jack Parker – Oak Ridge National Laboratory
- Barry Smith – Argonne National Laboratory
- Andy Tompson – Lawrence Livermore National Laboratory
- Dave Zachmann – Emeritus Colorado State University (Chair)
- George Zyvoloski – Los Alamos National Laboratory

APPLICATION DISCIPLINE SCIENTISTS PANEL LEADERS

- Site Characterization and Model Calibration
  - Tissa Illangasekare – Colorado School of Mines
  - Jack Parker – Oak Ridge National Laboratory
- Validation, Verification, Uncertainty Analysis, and Decision Optimization
  - Dongxiao Zhang – Oklahoma University
  - Jon Helton – Sandia National Laboratories
- Coupled Phenomena
  - Dick Ewing – Texas A&M University
  - Steve Yabusaki – Pacific Northwest National Laboratory
- Carbon Sequestration
  - Diana Bacon – Pacific Northwest National Laboratory
  - Rajesh Pawar – Los Alamos National Laboratory
- Research at Fundamental Scales
  - David Dixon – University of Alabama
  - Tim Scheibe – Pacific Northwest National Laboratory

COMPUTATIONAL SCIENTISTS PANEL LEADERS

- Site Characterization and Model Calibration
  - Omar Ghattas – University of Texas at Austin
  - Roelof Versteeg – Idaho National Laboratory
- Validation, Verification, Uncertainty Analysis, and Decision Optimization
  - Peter Kitanidis – Stanford University
  - Don Estep – Colorado State University
- Coupled Phenomena
  - Clint Dawson – University of Texas at Austin
  - Mark White – Pacific Northwest National Laboratory
- Carbon Sequestration
  - Mary Wheeler – University of Texas at Austin
- Research at Fundamental Scales
  - Brent Lindquist – Stony Brook University
  - Bjorn Engquist – University of Texas at Austin

EXECUTIVE SUMMARY

To determine the basic research needs for computational subsurface sciences, the U.S. Department of Energy (DOE) Office of Science (SC) organized a Computational Subsurface Sciences Workshop. The workshop was held in Bethesda, Maryland, during January 9-12, 2007. Collaborating DOE offices were SC, the Office of Environmental Management (EM), the Office of Fossil Energy (FE), and the Office of Civilian Radioactive Waste Management (RW). The purpose of the workshop was to obtain community input on computational science research needs and opportunities in the subsurface sciences and related areas, with a focus on developing a next generation of numerical models of subsurface flow and process simulation. Highlighted areas included potential terascale (and future petascale) computational algorithms to enable high-fidelity subsurface simulation models that fully couple key physical, chemical, geological, and biological processes with new capabilities to quantify and reduce model uncertainty.

CROSSCUTTING FINDINGS AND CHALLENGES

During the course of the workshop, nearly all of the breakout groups discussed the following crosscutting findings and challenges.

Uncertainty Analysis

Significant uncertainties are inherent in the analysis of subsurface systems. The challenge is to develop effective tools for analyzing the uncertainty of complex, large, multi-physics, multi-scale systems. We need to develop effective computational methods to 1) assess uncertainty in inputs and models, 2) propagate uncertainty and variability through models, and 3) mitigate the effects of uncertainty.

New Multi-Scale Computational Tools

Current up-scaling techniques to couple across scales generally are not adequate for coupled nonlinear processes in heterogeneous media. New computational tools are needed for linking scales and representing processes to obtain high-fidelity predictions beyond the range of conditions and scales at which the models and parameterizations were developed.

High-Fidelity Simulations

This is an opportune time for the subsurface sciences to benefit from advances in high-performance computing, data management, and data visualization. Advances in high-performance computing will enable a new generation of high-fidelity subsurface science simulation models with enhanced physical realism and fewer simplifying approximations. These new simulation capabilities will boost public confidence and regulatory acceptance of scientific predictions generated through numerical simulations.

Increased Collaboration

Increased collaboration between computational scientists and discipline scientists is essential as the effort goes forward. Especially important is the continued collaborative development of new high-performance computing enabling technologies designed to meet the specific needs of subsurface science research. For broad advances in simulation capabilities, we need new parallel linear and nonlinear solvers, high-performance computing data management and data fusion tools and methods for visualizing distributed datasets.

PANEL FINDINGS AND PROPOSED RESEARCH DIRECTIONS

The Priority Research Directions identified by the five breakout panels during their two days of workshop discussions are summarized below.

Integrated Site Characterization

Tools for conceptual model formulation, calibration, engineering design and management, and decision making are available, but are commonly run serially with little iteration and feedback to guide decisions regarding site characterization and monitoring. Integration of these processes could greatly improve decision-making and reduce costs; however, current methods are not readily scalable for large-scale, multi-process models. We need new protocols and computational tools for integrating 1) site and process characterizations, 2) site monitoring, 3) conceptual model formulation, 4) model calibration, 5) uncertainty analysis, and 6) decision optimization. The goal is to develop new computational tools and to link existing tools to provide an integrated framework for modeling coupled subsurface processes to solve specific practical engineering problems (e.g., groundwater remediation, CO_2 sequestration, etc.) to enable optimization of the design and operation of engineered systems and associated characterization and monitoring efforts.

Protocols for "Soft" Data

There are different degrees of uncertainty in the subsurface sciences site and process characterization data (hard to soft). There is a need to assimilate these different data in model calibration, taking into consideration scale and uncertainty issues. We recommend developing protocols and computational tools for using data at disparate time and space scales, and for assimilating "soft" data in model calibration and for assessing associated prediction error. New, practical algorithms are needed for integration of multiple data types in site and process characterization.

Optimized Decision Making

It is essential that we develop improved methods to optimize decision-making capabilities about large-scale subsurface sciences management and develop new methods to update models quickly with incoming data. These methods must be computationally efficient and should incorporate

uncertainty related to errors in conceptual models, empirical data, numerical models, and process characterization. Such methods will provide dramatic improvements over existing computational methods for complex model analysis and predictive science. These algorithms are applicable to a wide range of models arising in many disciplines, as they use high-performance computing to provide a bridge between experimentalists and modelers. Optimizing decisions under uncertainty will lead to more cost-effective solutions for highly expensive DOE subsurface undertakings. Data assimilation and calibration feedback will enable rapid utilization of data as collected.

Cyber Infrastructure

We need to develop 1) strategies for verification and validation of complex models and codes, 2) convergence metrics and measures for stochastic processes, 3) standards for data storage and programming interfaces, 4) infrastructure and tools for handling large simulations and large datasets, and 5) a collection of benchmark problems. A database of benchmark problems will provide a basis for performing model validation and code verification. This effort will provide guidelines and strategies for model validation and code verification under uncertainty and better communication of results of validation and verification to decision makers and stakeholders.

New Algorithms for Coupled Phenomena

Coupled subsurface phenomena (e.g., hydraulic, thermal, mechanical, biological and chemical) are dominated by complex heterogeneities, multi-phase, multi-component flows and interacting, scale-dependent processes. Existing codes are unable to simulate these phenomena with sufficient fidelity for optimal decision-making. There is a strong need to develop next-generation multi-physics subsurface simulation capabilities based on novel coupling and decoupling strategies combined with emerging discretization and high-performance computing solver technologies.

High-Performance Computing Data Management

Challenges such as contaminant transport, CO_2 sequestration, and nontraditional energy production demand accurate coupled process simulations. The integration of dispersed data sources and the management of petascale datasets will enable coupled process simulation and the produce knowledge required by decision makers. Simulating coupled processes at large temporal and spatial scales is becoming possible today because of significant advances in computational infrastructure. To develop large-scale, strongly coupled process simulations requires high-performance computing data storage, data access, and manipulation and visualization of diverse, dispersed, massive datasets. Petascale data management needs arising from coupled processes will demand new data intensive computing capabilities. This is an indispensable paradigm shift for subsurface sciences, with broader application to other scientific disciplines.

Improved Multi-Phase Flow Simulators

Current multi-phase flow simulators can handle only a very limited number of fluid components. Improved modeling of fluid mixtures is needed to assess the impact of CO_2 sequestration achieved by injecting flue gas consisting of five or more gases without expensive separation. Also needed are new compositional models for fluid mixtures and equations of state for fluid mixtures over wide ranges of temperatures and pressures. This will enable assessment of the impact of sequestration gas mixtures and injection pressures on reservoir and cap rock geochemistry. New computational tools for modeling complex fluid mixtures may enable more economically viable carbon sequestration procedures.

New Algorithms for Parallel Computations

Modeling carbon sequestration requires multiple, coupled sub-models, including multi-phase flow, reactive transport, geo-mechanical deformation and heat flow, that may operate at different space and time scales. Coupling of geo-mechanical and geo-chemical modeling is critical to accurate prediction of cap rock and reservoir integrity in CO_2 sequestration. We need improved computational parallelization of coupled process models and must move beyond domain decomposition to improve interoperability between process models and improve methods of up-scaling heterogeneous parameters.

Risk Analysis

To satisfy regulators and to gain public acceptance, new computational methods are needed to calculate risk associated with subsurface sciences prediction in areas such as carbon sequestration and site remediation. We need to develop efficient, robust risk analysis methods to capture process complexity and to improve coupling between risk analysis methods, process models, and monitoring observations.

Community Building

Currently, the subsurface science community is not sufficiently unified. It has been described as tribal (but friendly). For example, in the area of CO_2 sequestration, the petroleum industry, which is a significant player in the subsurface sciences, is under-represented in many collaborations of interest. To expedite community building, a virtual or distributed subsurface sciences institute may be useful. We also endorse the concept of one or more web-based, distributed repositories for storing codes, datasets, and computational tools to facilitate communication among code authors and application-orientated scientists and engineers. These repositories should use a framework that provides global access and preserves local control. Related activities will facilitate collaboration and outreach among the large numbers of disciplines and research groups necessary to attack the challenging multi-scale, multi-physics subsurface science problems, such as CO_2 sequestration, and will accommodate new and emerging high-performance computing environments.

High-Fidelity CO_2 Sequestration Model

Current simulation models are not capable of prediction and optimization of long-term isolation performance for representative (e.g., FutureGen-scale) CO_2 storage sites at space and time resolutions that effectively capture process complexity. Emerging high-performance computational capabilities have the potential to enable the development of site-specific coupled process models at previously unattainable time and space resolutions and to deploy protocols for integrating modeling, site characterization, and monitoring activities. They also will serve to help determine the level of process and model coupling and grid resolution required for simulation of large-scale CO_2 storage applications.

Physically Based Process Representation

High-fidelity computational models are needed for critical chemical, biological, and physical processes at the scales at which they are observed (e.g., from the molecular scale to the nano scale, to the micron scale, to the pore scale, to larger scales). Accurate process representations are needed at the next lower scale to understand the current scale. The computational challenge is to find ways to deal with sampling statistics, electronic behavior, complexity of pores, pore networks, and system size and boundaries. The computational demands are enormous and access to advanced high-performance computing resources is essential. Research in this area will lead to the development of new multi-scale mathematics and new computational science, based on application requirements. Potential benefits are improved conceptual models for advanced oil recovery, remediation technologies, CO_2 sequestration, and reliable predictions of contaminant transport.

New Algorithms for Parallel in Time Computations

At fundamental scales, multiple phenomena evolve on multiple, widely separated temporal scales. Numerical simulations of such phenomena are inherently difficult to parallelize. High-performance subsurface science simulations typically achieve parallel processing performance by relying on algorithms that use a spatial decomposition. For effective representation of phenomena at fundamental scales, we need to develop parallel in time algorithms with support for non-homogeneous time stepping.

Heterogeneity at Multiple Scales

Geometrical and biogeochemical heterogeneity have important impacts on subsurface processes at all scales. In many geological systems, the effects of heterogeneity dominate fluid flow and chemical transport. For example, focused flow is important in the transport of contaminants through aquifers, and heterogeneity is essential for the trapping of oil in reservoirs. At the pore scale, both geometric and chemical heterogeneity have an important impact on multi-phase fluid flow. It is important to include realistic heterogeneities in computer models used for practical applications. Current subsurface simulation models are not capable of realistic representation of complex heterogeneous media, especially if fractures occur or evolve on multiple scales. We

need accurate phenomenological models that provide comprehensive but concise descriptions of subsurface structures and processes, and include quantification of uncertainty. The grand challenge in this area is a comprehensive representation of mechanical discontinuity distributions from the pore scale to the repository scale. This work will require substantial access to leadership-class computational resources to solve the stochastic optimization, uncertainty quantification, parameterization selection, and design of experimental problems that arise in resolving the problem of media parameterization and reconstruction for computational subsurface science. An essential component of this research direction is the experimental, data-based validation of the models and algorithms that will result from this research. Exciting opportunities result from the fact that at the core level (~10 cm) a wealth of experimental data can be generated which can be used for the validation of the relationship between parameterizations, microstructure and subsurface physical phenomena.

Test Bed of Up-Scaling Benchmarks

All subsurface science models involve multiple, non-well separated length and time scales and complex, coupled physical phenomena. Direct numerical simulation from fundamental scales to the field scale is impossible. Analytic understanding of scaling is incomplete, although it is recognized that down-scaling leads to an ill-posed inverse problem. Research in needed to identify the minimum information that must be transferred across scales to maintain model validity. Work is needed on developing hybrid models based on local scale refinement and hierarchical models that maintain accuracy across scales by means of scientifically and mathematically defensible up-scaling techniques. There is currently little or no basis on which to 1) determine appropriate sequences of scales for models and 2) evaluate the impacts of information loss resulting from up-scaling. We recommend developing highly resolved, computationally demanding models of specific processes to serve as a test bed, or set of benchmarks, for evaluating up-scaling methods and up-scaled models. This test bed will require terascale and eventually petascale computational resources for solution of benchmark problems. The benchmark problems need to be validated using laboratory or field data. Advanced algorithms, mathematics, and computational methods will be developed for solving large problems and will provide a validation basis for application-level codes and models.

Pore Scale Particulate Transport

Most subsurface simulation codes are based on mathematical models of single phase and/or multi-phase flow and mass transport formulated at one or more continuum scales. In these models, geo-materials with micro scale and pore scale physical and chemical heterogeneities and complex pore and/or fracture geometries are treated as homogeneous “effective media.” Both fluid flow and mass transport, including mass transport mediated by colloids, nano particles and polymers, through these complex media are represented by averaged flow and transport equations based on concepts such as Darcy flow, relative permeability, and averaged particle transport “trapping and release rates.” An important advance for current codes would be the capability of modeling particulate transport at the pore scale, and coupling this to continuum

scale models. Such a capability would provide a firm scientific basis for developing technologies for controlling the behavior of micron and submicron-sized particles of concern (e.g., radioactive and/or chemically toxic colloids, bacteria, or viruses) in the subsurface. This would enable a more accurate and reliable assessment of the impact of small particles and polymers in problems related to contaminant fate and transport and CO_2 sequestration and enhanced oil recovery, for example.

Pore Scale Reactive Flow

Pore scale modeling has been used for several decades to simulate single-phase and multi-phase fluid flow in fractured and porous media. Recently, these models have been coupled with simple models for precipitation, dissolution, and the transport of dissolved substances. However, improved codes on more-capable computing systems are needed to simulate the behavior of reactive fluids with a wide range of properties under the full range of conditions relevant to subsurface applications. Important advances could be made by efficiently implementing existing methods on petascale computing systems. Many important applications will also require the development of better single-physics algorithms and new codes that couple important physical and chemical processes on a wide range of length and time scales. Eventually, pore-scale models will be coupled with atomistic and/or continuum scale models on computing systems with capabilities that substantially exceed those of current systems. As an interim solution, pore-scale simulations could be used to provide model parameters, and to develop better constitutive equations, for field scale models.

WORKSHOP LOGISTICS

More than 150 experts registered for this invitation-only workshop and the participants included a diverse set of professional disciplines and computational scientists working in the subsurface sciences. DOE program managers, subsurface sciences stakeholders, and decision makers were well represented. The workshop participants came from 11 DOE laboratories, 23 U.S. universities, and 2 European universities.

Technical panel discussions focused on five major areas for which simulation and modeling are clearly relevant to subsurface science research.

1. Site Characterization and Model Calibration
2. Validation, Verification, Uncertainty Analysis and Decision Optimization
3. Coupled Phenomena
4. Carbon Sequestration
5. Research at Fundamental Scales

A common goal of these five panel discussions was to define research needs and opportunities in the subsurface sciences that can benefit from new high-performance, computing-based collaborations involving scientists from various disciplines, applied mathematicians and computational scientists.

The workshop began on Wednesday morning with welcoming remarks by workshop organizers and representatives of the collaborating DOE offices. The workshop agenda was pursued through five breakout groups, each consisting of 20 to 30 panel members. Panel assignments were based on the first and second panel choices indicated by the attendees when they registered for the workshop. The workshop schedule also included four excellent plenary talks given by leading subsurface and computational scientists. Details of the workshop can be accessed from the conference website at http://subsurface2007.labworks.org/.

Discipline scientists led the panel discussions on Wednesday morning with the goal of articulating the research needs in terms of what is needed to better understand the physical processes involved in subsurface phenomena and to develop better decision-making tools. During the Wednesday afternoon panel sessions, computational scientists directed the panel discussions with the goal of identifying ways in which high-performance computing can lead to a new generation of high-fidelity subsurface simulation models. The Wednesday program concluded with a plenary session during which each panel reported on their preliminary Priority Research Directions (PRDs). On Thursday, the second full day of the workshop, discipline and computational scientist panel leaders directed the panel discussions to 1) revisit the PRDs discussed the previous day, 2) suggest PRDs overlooked in the first day of panel discussions, and 3) identify panel members who could serve as lead authors in producing PRD documents that would provide comprehensive, easily understood descriptions of each PRD. The Thursday program concluded with a plenary session in which each panel presented the latest versions of their PRDs.

CONCLUSION

Within the SC and the other DOE programmatic offices that collaborated in this workshop (EM, FE, and RW) are simulation capabilities, scientific software tools, computing platforms, and interdisciplinary researchers without equal in the world. These DOE assets, together with university- and industry-based researchers, represent an extraordinary resource for creating a next generation of computational subsurface science simulation tools and using them to gain understanding of many important physical problems. Examples of subsurface science problems of national importance include nuclear waste repository design and monitoring, transport and fate of contaminants, alternative energy production, bio-remediation, carbon sequestration, and reservoir modeling. The computational subsurface sciences research agenda challenges the limits of current simulation capabilities in every direction and presents many opportunities for significant advances in simulation capabilities and mutually beneficial collaborations, making it a worthy effort to be undertaken.

Favorable trends in computing capability and cost of large-scale simulation are transforming all fields of science and engineering into substantially simulation-based activities. In less than 20 years, we have seen an increase in sustained processing rates of five orders of magnitude accompanied by a cost in sustained computing capability that has dropped by four orders of magnitude over the same period. Despite these favorable trends in the computational landscape,

many challenges exist. Most of the current code base for the computational subsurface sciences was created for computing platforms of much lower capabilities than are now available. A substantial effort will be to rewrite much of the existing code base so that it can benefit from modern software practices and high-performance parallel architectures. Virtually all large-scale data structures in existing codes will need to be replaced with distributed versions. As the software infrastructure is reconstructed, attention can be given to extensibility, reusability, object orientation, modularity, portability, and performance tuning. The computational subsurface science research agenda will require new, specially designed enabling technologies to address 1) parallel programming models for MPI and multithreading, 2) dynamic load balancing, language interoperability, 3) high-performance data management, 4) performance monitoring metrics, 5) debugging tools, 6) distributed data archiving, 7) visualization, 8) data mining, 9) model validation, and code 10) verification.

For high-fidelity computational process simulation in subsurface sciences to achieve its potential, close collaboration will be required between discipline scientists in the DOE applications offices and computational scientists in the Office of Advanced Scientific Computing Research. These collaborations will need to address daunting challenges related to the multi-scale, multi-physics nature of subsurface sciences problems and need to develop new ways for dealing with uncertainty in site and process characterization and associated risk assessment. The successful SciDAC program can serve as a starting point for the further development of a collaboration model for pursuing a computational subsurface science research agenda.

At the workshop, many lively and substantive discussions clearly illustrated the need for new computational simulation capabilities in the subsurface science applications of interest to EM, FE, and RW.

CONTENTS

REPORTS OF THE PANELS

SITE CHARACTERIZATION AND CALIBRATION PANEL REPORT

The focus of this panel was the development of computational methods to improve model accuracy and precision by making optimal use of available data and by integrating decisions regarding data collection with model calibration and error analysis to focus limited resources on efforts that yield the greatest benefit. The problems of interest involve movement of fluids, chemicals and/or heat through geologic media that exhibit complex heterogeneity at multiple scales in time and space.

Conceptual model formulation is a crucial step in the modeling process. Ideally, conceptual model complexity and data collection efforts will be advanced in an iterative manner. At the first iteration, a simplified conceptual model is utilized with initially available data to jointly assess the need for additional data and/or model refinement. As additional data is obtained, the benefit of further data collection and/or model refinements can be evaluated. The process of conceptual model formulation is assumed to be a manual process performed by the project team. That is, at a given stage in the project, a limited number of potential model refinements would be specified for analysis. However, the process of converting candidate conceptual models to numerical code input can be laborious, and the development of improved tools to automate the process would be very useful.

As model complexity increases (i.e., less restrictive assumptions, higher resolution), intrinsic model uncertainty is expected to decrease. However, more parameters will generally required for model calibration. As data availability increases, the level of model complexity that minimizes prediction uncertainty is expected to shift towards greater optimal complexity and conversely with less data the optimal model will shift towards lower complexity (Figure 1). This implies that for a given set of data available for calibration, increasing model complexity will eventually be counterproductive.

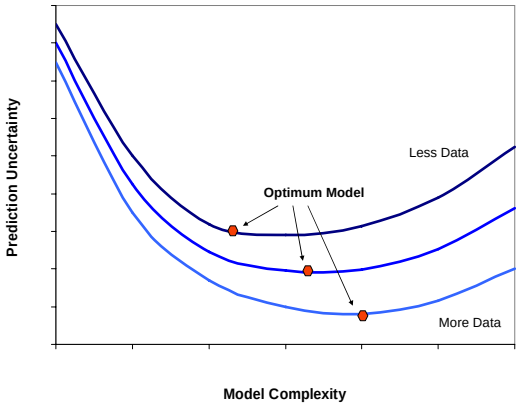


Figure 1. Conceptual relationship between model complexity (time/space scale resolution), data availability, and prediction uncertainty.

Direct determination of all input parameters required in a complex multi-process subsurface model is generally not feasible. The estimation of unknown model parameters must therefore be carried out by an inverse solution approach in which parameter values are determined that minimize deviations between observed and predicted system behavior subject to constraints or plausibility criteria, weighted to consider uncertainty in observations. Since many direct measurements will not be at the same time/space scales as the model, upscaling or downscaling may be necessary to avoid biases due to scale disparities. Uncertainties associated with the scaling process will need to be considered in addition to those due to measurement uncertainty. Furthermore, "soft" data may often be available that exhibit indirect relationships with model input or output that can improve the inverse solution. Model predictions will be uncertain to a degree that depends upon parameter covariances, sensitivity of predictions to uncertain parameters, intrinsic model limitations, and uncertainty in data used for calibration. Prediction probability distributions computed assuming multi-normal or multi-lognormal parameter errors will frequently lie outside feasibility space for calibration data.

In addition to measurements of variables required as model input and of variables that are directly computed as model output (and which can thus be used for inverse modeling), various measurements may also be available that exhibit a significant correlation with input or output variables. The latter are often referred to as "soft" data and the former as "hard" data, reflecting the relative difference in uncertainty in interpretation. The reality is more of a continuum in data uncertainty, due to measurement limitations and scale discrepancies between the model and observations. The panel recognized that a need exists for the development of methods to account for disparity between scales at which parameters are measured and model inputs are defined and to use the full spectrum of data "hardness" (or "softness") in the calibration process, with due consideration given to effects of data uncertainty vis a vis the conceptual model formulation.

Novel multiphysics and multiscale codes are needed to model complex coupled processes in the subsurface. While the exact design, components and implementation of these codes is unknown, in order for these codes to meet their objectives they will need to efficiently couple and integrate diverse code elements developed by different groups, integrate massive amounts of diverse data from diverse sources (including dynamic sensor networks), and take advantage of future developments and enhancements in computational infrastructure. In addition, such codes will need to pass rigorous, auditable verification and validation tests both in the development and implementation phase. A priority research direction is the development and implementation of a rigorous and powerful cyber infrastructure for subsurface modeling. The development of this cyber infrastructure will utilize and build on existing US and international efforts in the development of standards for e.g. sensor web services, self-validating models, service discovery, distributed computing, code adaptivity while at the same time accommodating the unique aspects and attributes associated with the subsurface modeling problem. This cyber infrastructure will provide a presently unavailable but critically needed framework for the efficient development, testing and field deployment of novel multiphysics and multiscale codes. This ability will permit the effective use of novel codes for long-term monitoring, management and optimization of CO_2 sequestration sites, and the cost effective mitigation of subsurface contaminants.

A major challenge in subsurface modeling is the development of scalable parallel algorithms for estimating uncertain parameters from sparse measurements or various types at multiple time and space scales. The need to solve such inverse problems to calibrate model parameters and estimate current system state has been well articulated. However, substantial challenges and opportunities remain for the development of efficient inverse algorithms that can scale to the next generation of petascale computing systems. There is an urgent need to extend to the inverse setting the successes engendered over the past decade by scalable parallel algorithms for *forward* (also known as *direct*) simulations. Since the structure of inverse operators is usually very different from those of forward operators, entirely new classes of scalable numerical algorithms are needed.

While inverse problems in subsurface modeling are characterized by a wide range of inversion parameters (e.g., flow and transport constitutive parameters, initial conditions, boundary conditions, reaction and other model coefficients, source terms, geometry and topography) and simulation equations (e.g., describing multiphase reacting flow, transport, geomechanics, and geophysics), they share a common mathematical structure: an optimization problem with an objective function that represents the data misfit between observations and predictions, constraints in the form of large-scale simulation equations (typically discretized PDEs and ODEs), and inequality constraints on the parameter bounds.

Large-scale simulation-based inverse problems are significantly more difficult to solve than the corresponding forward problem, for the following reasons:

- Inverse solution typically requires numerous forward solutions, and can thus be intractable when the forward solve takes weeks on multi-thousand processor systems.
- The inverse problem is usually ill posed, even when the forward problem is well posed.
- When the forward problem is an evolution equation in space, the inverse problem is a boundary-value problem in space-time. Thus, inverse operators are usually non-causal and non-local, despite the local and causal nature of forward operators.
- New infinite-dimensional operators appear in the inverse problem that are not present in the forward problem—adjoints, Hessians, and KKT operators—and these require operator-specific regularization, iterative solvers, preconditioning, globalization, inexactness, and parallel implementation, which are new frontiers for research in numerical algorithms.
- Non-uniqueness in inverse problems is manifested via families of possible parameter values that are consistent with the observations. There is often a need to not only estimate the most likely parameters from among the families (via regularization or prior model criteria), but to also characterize the uncertainties in their estimates, which stem from uncertainties in the observations, the models, and the priors.

Generic optimization or nonlinear least squares software packages are incapable of exploiting the structure of the embedded forward or inverse operators or their underlying infinite-dimensional nature, and are thus inadequate for problems with large numbers of inversion parameters. As the

underlying forward simulations embrace the multi-teraflops era and move toward the petascale, the chasm between the capabilities of available inversion algorithms and software, and the needs of scientific codes, grows even wider.

The next generation of inverse algorithms must respond to the following difficulties that characterize subsurface inverse problems:

- The forward simulations embedded within inversion are typically characterized by complex, nonlinear, multiscale, multirate, multiphysics, and biogeochemical processes that require terascale and ultimately petascale computing resources;
- The inversion variables often stem from infinite-dimensional spaces (i.e. they represent initial conditions, boundary conditions, heterogeneous material properties, distributed sources, or geometry); and
- Infinite-dimensional inequality bound constraints are required in the problem formulation (such as positivity constraints on heterogeneous parameter fields).

To address these issues, next-generation inverse algorithms are expected to:

- Respect the infinite-dimensional character of the underlying operators, because petascale inverse problems in subsurface modeling involve PDEs, and because algorithms that respect the infinite-dimensional structure of the PDEs can often exploit mesh independence to find solutions in a small number of iterations, even when the dimension of the parameter space is of the order of millions—problems that are impossible for off-the-shelf, algebraic-oriented optimization packages.
- Make use of Hessian information, because derivative-free methods are unable to scale beyond a small number of inversion parameters, and because gradient-only methods (e.g., steepest descent, nonlinear conjugate gradients, quasi-Newton) also do not usually scale well with increasing problem size.
- Be adjoint-based, because such methods offer the only hope for computing derivatives (and Hessians actions) for high-dimensional parameter spaces.
- Be multilevel in nature, because of the great successes of multilevel and multigrid methods for forward simulations and their promise for inverse problems. However, multilevel methods must take on a very different form for inverse problems due to the differing character of forward and inverse operators, and require careful analysis and development to be effective for inversion.
- Parallelize in a fine-grained manner, because of the high-dimensional state and parameter spaces that result after discretization. Parallelism must be exploited at the “grid point” level, following the lead of large-scale forward simulation codes. Additional opportunities exist for coarse-grained parallelism, such as for direct sensitivities (across inversion parameters), adjoint computations associated with active set methods (across active constraints), and

multiple source inverse problems (across each instance of the forward problem), but these must build on a finer-grain parallelism to permit scalability to large numbers of processors.

Subsurface modeling problems are ideal proving grounds for inverse algorithms. The medium and state are not directly observable, thus resulting in uncertainty in parameters, initial/boundary conditions, and source terms. Forward simulation therefore relies on inverse analysis to estimate these parameters (and possibly their uncertainty) from the observables. The forward simulation problems alone are extremely challenging; inverse solution for *high-dimensional* parameter spaces (as results from discretization of heterogeneous property fields) is intractable using current methods. The development of scalable parallel numerical algorithms for inverse solution that can match the scalability and efficiency of the associated forward solvers is thus of paramount importance.

Conceptual model formulation, site characterization, model calibration, uncertainty analysis and decision/design optimization are commonly undertaken serially with minimal *ad hoc* iteration. Without a fully integrated approach to these tasks, global optimization of model formulation, characterization and monitoring activities, and design and operation of engineered systems cannot be achieved. As a result, excessive costs and/or risk may be incurred because overly complex or overly simple models may be adopted than are warranted for the intended purpose; data may be collected that does not substantially reduce decision uncertainty; truly useful data may not be identified and collected; and/or engineered systems may be over- or under-designed. The ability to efficiently and practically integrate conceptual model formulation, site characterization, model calibration, uncertainty analysis and system monitoring for complex coupled subsurface systems is currently impeded by a lack of scalable, powerful tools for integrated analyses. Because such integrated analyses will require a large number of direct simulations, high-performance computing resources and advances in algorithms for direct and inverse solutions, error propagation, and multi-objective design optimization schemes are needed.

A need exists for the development of an integrated framework and computational tools for modeling subsurface systems that enables global optimization of conceptual model formulation, site characterization and monitoring activities, and design and operation of engineered systems with full consideration of prediction uncertainty. A stochastic optimization framework is proposed to simultaneously optimize engineered system design and operation variables, conceptual model formulation, and/or site characterization/monitoring plans to minimize a multi-objective function that may involve cost expectation (including penalty costs associated with failure to meet design requirements), regulatory criteria, and/or other externalities. The integrated framework would couple model calibration (inverse problem solution) using data that may involve “direct” parameter measurements (e.g., hydraulic conductivity), indirect or “soft” data of various types (e.g., borehole descriptions, penetration resistance, geophysical data, etc.) with a generalized error analysis methodology (including measurement and upscaling errors), with a multi-objective stochastic optimization procedure. Implementation will involve adaptation and refinement of existing codes, development and implementation of new methods, and

integration in a manner that provides maximum computational efficiency, robustness, maintainability, and adaptability.

Models are generally developed to resolve specific engineering or planning issues – for example, to determine configurations and flow rates for wells to meet specific production or regulatory goals with a given probability of success and at minimum expected total cost. Various approaches may be used for the formulation of stochastic design optimization problems. Global optimization of an objective function with many parameters is a particularly difficult and challenging problem.

Integration of the foregoing modeling tasks is needed to simultaneously optimize conceptual model formulation, characterization/monitoring plans, and engineering design. In practice, the entire process may be iterated as incremental refinements in the model formulation and data collection are performed until further efforts are not warranted on a cost-benefit basis. Effects of undertaking additional site characterization studies, pilot tests, monitoring procedures, etc. on model predictions and their uncertainty may be evaluated by adding synthetic data at the proposed locations and/or times with appropriate measurements and/or upscaling uncertainty and rerunning the inverse problem, error analysis, and management optimization problem to determine the impact on total cost (including the additional data collection costs) and hence to quantify the net cost or benefit of alternative plans. Integration should be performed in a manner that provides flexibility to the user to execute selected components and bypass others to suit specific needs. Input/output structures and program control information should be implemented in a manner that is code independent, so that the infrastructure can be readily implemented with any direct problem solution.

VALIDATION, VERIFICATION, UNCERTAINTY ANALYSIS, AND DECISION OPTIMIZATION

CURRENT STATUS

A blend of physical experimentation, modeling, simulation, and high performance computing has become the standard toolset for scientific investigation and engineering design in complex multiphysics, multiscale geologic subsurface systems. The practical limitations in experimentation and simulation mean that substantial uncertainty accompanies any description or prediction of the behavior of a subsurface system. Uncertainty is inherent in the inaccessibility of the subsurface to measurement and experiment, data gathered from subsurface systems, the complexity and heterogeneity of the subsurface, mathematical descriptions of the physical processes in the subsurface, and the analysis and computational simulation of complex multiscale, multiphysics models. The need for a scientifically defensible representation of uncertainty in the analysis of complex systems arising in the subsurface sciences is now widely recognized.

For example, we need scientifically defensible decision-making tools for subsurface systems involving contaminant removal (storage). This includes such decisions as the location and pumping rates at wells that are pumping contaminant out of the ground, location of monitoring devices which can include sampling wells, and also soil removal or treatment. For more complex systems, chemicals may be injected into the groundwater to accelerate the biodegradation of the contaminant ("bioremediation") or enhance the effectiveness of extraction pumping (e.g., by adding surfactants), which engenders more decisions including the concentrations of injected materials. If we choose to allow these decisions to vary over time, we have a dynamic decision problem, i.e., a control problem. Other important applications include the use of the subsurface for storage of hazardous wastes and carbon sequestration.

In general, optimization methods have many applications, including:

- Determination of the best conceptual model to fit observational data (model calibration or data assimilation)
- Definition of conditional realizations, to estimate performance (results of forward modeling) with uncertainty estimation
- Selection of new data (what values, where, and when) to be collected that will reduce uncertainty estimates (feedback) or optimize decisions
- Selection between different alternatives, including:
  - collection of additional data
  - remediation actions
  - continued monitoring
  - design alterations.

Optimal decision-making poses extreme computational challenges because the underlying simulation models for groundwater flow and contaminant transport are complex and computationally intensive, and optimization typically requires hundreds or thousands of simulations. It is essential that we develop better algorithms and approaches. The existing methods are not computationally efficient enough or mathematically complex enough to address fully the uncertain environment in which decisions must be made or the complexities associated with multimodal optimization problems arising from nonlinearities and heterogeneities. In addition, methods that deal effectively with coupled simulation models (e.g., operator decomposition solution of multiphysics models) and methods that generate feedback policies are necessary. The uncertainty incorporated should include parameter, data, numerical and model error in the analysis, which is not adequately addressed in current modeling efforts.

The computational demands associated with the analysis of the creation and propagation of uncertainty in a complex multiphysics, multiscale subsurface system severely affect all existing approaches to uncertainty analysis.

One current approach for analyzing uncertainty in a complex subsurface system is the use of random sampling techniques, such as Monte-Carlo methods. For large-scale problems, however, the computational requirements of random sampling are enormous. The situation is even worse when estimating probabilities of rare events. The computational challenges include the overhead involved with organizing multiple simulations of a subsurface system and the large number of simulations required to carry out an accurate statistical analysis. Another approach is based on representing uncertain model inputs and the solution fields in terms of a spectral representation, such as a Karhunen-Loéve series or polynomial chaos expansion and then solving for the spectral coefficients. The coefficients are computed intrusively, by deriving deterministic equations for the coefficients or non-intrusively by random sampling. Either approach entails enormous computational demands and again raises the need for fundamental research in of spectral approaches for efficient implementation on high performance computers.

In some circumstances, the quantification of uncertainty can be attacked directly by deriving equations for statistical quantities, probability density functions or moments, that represent information about the quantities of interest. Computing solutions of these equations yields information about uncertainty directly, for example the expected value of oil/gas recovery versus time and the uncertainty of the expected value. However, solving these equations for large-scale models is again a computationally intensive problem.

Variational analysis based on duality, adjoint equations, and generalized Green's functions present a traditional deterministic approach for analyzing the creation and propagation of uncertainty through a complex system. The computational demands and efficient implementation in high performance computing environments, including for example storage of forward solutions for forming an adjoint problem and solution of the adjoint problem, once again presents severe mathematical and computational hurdles.

In addition to computational hurdles in uncertainty analysis, there are also difficult mathematical issues, such as in finding useful representations of uncertainty and treating different kinds of uncertainty in one analysis. For example, the starting point for analyzing the uncertainty in model predictions is characterizing the uncertainty in the input parameters and data used for a model or in the data that is used to calibrate the model. This presents multiple hurdles because harvesting information from experiment and field measurements in subsurface systems is difficult and costly and there are serious physical limitations on the type and scale of data that can be gathered. The data determined in physical measurements must then be processed into a form useful for mathematical modeling, which is another complex issue.

In short, there is a tremendous need to improve the existing approaches and devise new methodologies for analyzing the creation and propagation of uncertainty in complex subsurface systems. There is also a tremendous need to implement uncertainty analysis tools in high performance computing environments in order to meet the computational demands associated with uncertainty analysis.

SUBSURFACE SCIENCE CHALLENGES

The scientific need for an adequate representation of uncertainty in the analysis of complex systems is now widely recognized while the national importance of the issues involving subsurface systems, e.g. radioactive waste disposal at Yucca Mountain, carbon sequestration, oil and gas recovery, and pollution in groundwater, means that this uncertainty has tremendous scientific, societal, and economic implications.

COMPUTATIONAL SCIENCE CHALLENGES

There is a critical need for significant high performance computational methods for treating uncertainty in all phases of modeling, from uncertainty in experimental results, to uncertainty in the evaluation of models and, ultimately, to the quantification and interpretation of uncertainty in the results of computational analyses. This need raises theoretical and computational challenges in all aspects of modeling a complex system, including data acquisition and treatment, e.g., up or downscaling, model uncertainty, numerical simulation error, operator decomposition of physics and scales, analysis and quantification of uncertainties, sensitivity analysis, data assimilation, and the formulation of optimal decisions. The challenge to the computational community is to develop and implement high performance tools and libraries as well as standardized benchmarks for the analysis of uncertainty as a fundamental part of high performance computational modeling of complex subsurface systems.

Another key challenge is the development and implementation of mathematical frameworks for analyzing the propagation of uncertainties through complex multiphysics, multiscale subsurface systems. Mathematical frameworks must include the effects of data, parameter, and model uncertainty as well as errors produced during simulations, and therefore must deal with uncertainties described with a variety of statistical and deterministic representations. The mathematical approaches must allow for propagation, cancellation, and accumulation of effects in order to produce an accurate quantification of the uncertainty in analysis results. Mathematical techniques for uncertainty analysis must accommodate abrupt and even discontinuous behavior in space and time, e.g., sharp or discontinuous material changes at specified locations in space and discontinuous model changes in time arising from specified conditions, as well as multiple bifurcations present in specified data and parameter ranges. The techniques must also deal with the effects of operator decomposition in a multiphysics model. An associated challenge is to develop implementations of the mathematical frameworks for ultra scale computing platforms because of the intensive computational demands of the analysis. The high performance software for mathematical analysis of uncertainty must be designed “from the ground floor”, as opposed to ad hoc implementations on top of existing high performance simulation frameworks.

Distributing available resources in order to analyze and predict the behavior of a complex geological subsurface system both accurately and efficiently is a related critical computational challenge. There are enormous costs associated with experiment and field measurements as well as severe constraints on the available resolution in computational simulations even on petascale

computers. A challenge is to develop and implement frameworks that can adaptively guide the distribution of resources in order to reduce the uncertainty in a model analysis. The development of truly efficient adaptive frameworks would greatly increase the accuracy in model analysis and system prediction that can be achieved with a given amount of resources.

OPPORTUNITIES AND RESEARCH NEEDS

This panel identified three significant Priority Research Directions.

Uncertainty Representation, Uncertainty Propagation and Sensitivity Analysis for Subsurface Systems

Significant uncertainty accompanies any description or prediction of the behavior of a complex subsurface system and this uncertainty has important scientific, societal, and economic implications. This raises a critical need for high performance computational methods for treating uncertainty in all phases of modeling. The challenge is to develop tools and libraries for uncertainty analysis as a fundamental component of high performance computational modeling of subsurface systems.

Decision Optimization, Data Assimilation, Control under Uncertainty

Decision optimization is the computational process that seeks to determine the most advantageous, in terms of cost or other criteria, decisions that meet specified physical and performance constraints. The challenge is to improve computational methods to optimize decisions in the management of large-scale subsurface systems in the presence of uncertainty, particularly for evolving systems that required dynamic updating a recalibration in response to incoming data. These methods must account for uncertainty, be computationally efficient, and take advantage of ultra scale hardware and software.

Metrics and Benchmarks for Verification and Validation

Application codes establish credibility for accuracy and reliability by performing domain specific verification and validation testing. Developing the problem descriptions and methods of assessment are difficult and time consuming. Therefore, a common database of metrics and benchmarks could greatly improve computational subsurface application verification and validation and the assessment of credibility. The challenge is to produce techniques and benchmark problems for verification and validation specific to subsurface science applications.

CONCLUSION

Meeting the theoretical and computational challenges associated with treating uncertainty in subsurface systems will provide a greatly enhanced ability to fully utilize high performance computational platforms in addressing fundamental scientific and societal questions related to subsurface processes. This will result in increased confidence in subsurface system analyses, due

to enhanced resolution in the analyses and more reliable quantification of the uncertainty in the analysis results. Ultimately, this will allow scientifically defensible policy decisions regarding issues of national importance such as contamination remediation, carbon sequestration and radioactive waste storage.

CARBON SEQUESTRATION

CURRENT STATUS

For at least several decades to come, the United States will rely on fossil fuels for a majority of the required and expanding energy production (coal, oil, methane, and possibly oil shale). This heavy reliance on fossil fuels will result in increased production of carbon dioxide (CO_2) unless the nation makes significant progress towards the development of near-zero emission power plants. The capture, separation and storage or reuse of carbon, carbon sequestration, should be employed to stabilize and eventually reduce concentrations of this greenhouse gas in the atmosphere.

The Department of Energy's Carbon Sequestration R&D program is tasked to provide science-based assessments of costs, long-term stabilities and environmental impacts. Clearly modeling and simulation will serve as the principal tool in evaluating the technology options, but a number of advances in the state-of-the-art are required in a relatively short time frame.

The panel on Carbon Sequestration identified five priority research directions:

1. Improved computation of coupled multi-scale process models for carbon sequestration
2. Efficient risk assessment techniques for carbon sequestration
3. Development of a subsurface science collaboration infrastructure
4. Modeling the sequestration of supercritical fluid mixtures
5. Modeling FutureGen-scale geologic CO_2 sequestration.

SUBSURFACE SCIENCE CHALLENGES

Improved Computation of Coupled Multi-Scale Process Models For Carbon Sequestration

Given the significant increase in carbon dioxide in the atmosphere associated with burning of fossil fuels and the important role that carbon sequestration could play in remediation, significant investment is required to predict the long-term fate of CO_2 pumped into subsurface geologic formations. Accurate and efficient numerical models are needed to predict CO_2 storage capabilities, control injection, and assess risk of leakage to the surface. Fundamental improvements over existing simulation models are needed in several key areas. The subsurface is extremely heterogeneous and so requires new methods of characterization, including 4D seismic and satellite imaging methods, to detect subsurface deformation and potential leakage. New simulators must be able to incorporate this information and provide feedback for gathering

it. Individual process models must be discretized numerically to preserve pertinent physical and chemical principles. Near well-bore environments are critical to model accurately as these zones will likely change dramatically upon injection of CO_2, and abandoned wells create a risk of leakage. Accurate simulation in the presence of heterogeneities will require multi-scale discretizations and computational grids that adapt dynamically to zones of rapid changes. Sequestration of CO_2 in the subsurface is inherently is a complex problem, and accurate simulation will require new algorithms for simultaneous coupling of multiple process models. Coupled and adaptive, multi-physics, multi-scale, multi-resolutions processes present unique computational infrastructure demands due to their scale and heterogeneity in space and time. Significant impacts of this research will extend beyond the carbon sequestration problem. Improved parallelization methods, numerical algorithms, data assimilation techniques, and scaling methods for heterogeneous systems will have dramatic impact on the computational sciences as a whole, and computational subsurface science will benefit more broadly, due to the similarity of processes that affect sequestration and the fate of near-surface contaminant flows.

Efficient Risk Assessment Techniques for Carbon Sequestration

Large scale deployment of geologic CO_2 sequestration will require demonstration of comprehensive analysis of risks, including financial, health, safety, and environment, associated with the technology. Understanding the impact of storing large amount of CO_2 over a long time frame is a complex problem due to a number of coupled physical and chemical processes. Traditional risk analysis approaches cannot be directly applied to geologic systems due to the differences in the fundamental behavior of geological systems and engineered systems such as a nuclear power plant, including dynamic evolution of component behavior, large uncertainties and limited opportunity to perform extensive characterization experiments. This has led to the development of approaches that combines extensive PA calculations supplemented with a risk analysis based on PA results. The PA calculations can become computationally intensive depending on the number of uncertain parameters, typically making it impractical to explore all of the combinations to develop statistically meaningful results required for risk analysis. In addition, these analyses can take significant time, which could be a limitation for development of multiple sequestration sites.

Subsurface Science Collaboration

The subsurface science community has been described as “tribal” (but friendly). Codes, datasets and computational tools are not openly shared until they are finished products, and even then they are sometimes, de facto or de jure, proprietary. An extreme example of this challenge is that the petroleum industry, which houses tremendous expertise in subsurface science, is underrepresented in many collaborations of interest in CO_2 sequestration.

Modeling the Sequestration of Supercritical Fluid Mixtures

Geological carbon sequestration is a promising technology for reducing CO_2 emissions from coal-fired power plants and other CO_2 emitters. However, capturing and separating CO_2 from the

other products of coal combustion can be an expensive process. This has led to interest in the possibility of injection unpurified, or less purified, flue gas from the power plant directly into the subsurface. The flue gas would first be compressed to a supercritical state before injection. While the simulation of the injection of supercritical CO_2 into the subsurface has become fairly routine, an equation of state for a complex mixture of supercritical fluids has not been defined at the temperatures and pressures encountered during geologic carbon sequestration. The challenge is to provide thermodynamic properties of the supercritical fluid mixture, along with the individual aqueous solubility of each of the flue gas components. Molecular dynamics modeling may be used to simulate these properties, as a supplement to the limited experimental data that is available for these complex gas mixtures. Both the molecular dynamics modeling used to determine the equation of state for these supercritical mixtures, and the modeling of the subsurface injection of these mixtures are computationally intensive calculations that may be greatly accelerated by code parallelization and execution on massively parallel supercomputers.

Modeling FutureGen-Scale Geologic CO_2 Sequestration

FutureGen power plants will require the efficient capture and secure geologic isolation of CO_2 emissions. Defensible forecasting of long-term isolation performance demands advanced reactive transport simulators that explicitly couple the interdependent multiphase flow, geochemical mass transfer, and geomechanical deformation processes that control subsurface CO_2 migration and trapping. In addition, field application of these simulators requires site characterization and geostatistical methods to define initial and boundary conditions, myriad site-independent data that underpin the coupled process models to facilitate dynamic system evolution, and time-lapse monitoring data to permit iterative improvement of both site characterization and dependent modeling efforts. Further, because the relevant basin-scale systems are typically characterized by extreme permeability and compositional heterogeneity (often requiring highly resolved domains); because CO_2 injection introduces sharp density and chemical gradients across dynamic plume boundaries (severely limiting time-step magnitude); and because isolation performance must be predicted for 100s-1000s of years, such applications also demand advanced parallel architectures, numerical methods, and space-time discretization schemes. Finally, the uncertainty bounds that characterize these highly complex simulations must be quantified for risk assessment through sensitivity analyses that span the extensive matrix of site-characterization and site-independent parameters that influence predictive results

COMPUTATIONAL SCIENCE CHALLENGES

Improved Computation of Coupled Multi-Scale Process Models For Carbon Sequestration

Currently available numerical simulation packages do not couple all the needed compositional flow, reactive transport, and geomechanics models. A few simulators have some of these capabilities, such as the UT-Austin IPARS parallel multiphase geomechanical geochemistry simulator and the SIMWULF compositional, geomechanical, thermal simulator. In 2004,

Lawrence Berkeley National Laboratory coordinated a study to test and evaluate codes to model geological sequestration, including various versions of the Lawrence Berkeley National Laboratory TOUGH, the Pacific Northwest National Laboratory STOMP, the Los Alamos National Laboratory FLOTRAN and FEHM, the Institut Français du Petrole (IFP) SIMUSCOPP, the University of Stuttgart MUFTE_UG, and the Computer Modeling Group (CMG) of Canada GEM (Pruess et al. 2001; 2004). The study demonstrated substantial agreement between results predicted from different simulations, but also areas with only fair agreement and some significant discrepancies. Most disagreements could be traced to differences in fluid properties. However, all but one of the test cases were for 1D homogeneous media (one heterogeneous 2D test problem was included). While this study was extremely useful, it did not compare 3D reservoirs with complex heterogeneity on different scales and an equation of state compositional flow model along with geochemical and geomechanical effects and non-isothermal phenomena. In summary, there appears to be no single simulator that incorporates and adequately models all of the major physical and chemical processes induced by injection of CO_2 into potential disposal reservoirs that is also coupled to a geophysical model. Moreover, most of the above simulators do not include high fidelity algorithms that include error estimators for adaptivity.

Efficient Risk Assessment Techniques for Carbon Sequestration

Currently, PAs are performed by developing an overall system model that links multiple sub-systems or components. In most cases the sub-systems are defined based on differences in the fundamental physics that governs the performance of the sub-system. It is impossible to develop a single process level numerical model that can couple physics which governs multiple, distinct sub-systems. The approach typically used is to capture the behavior of the sub-systems through simple analytical formulations or computationally efficient numerical models and link the analytical sub-models into the overall system model. The behavior of the sub-systems is characterized through laboratory experiments, field experiments/observations or detailed numerical simulations. The scientific challenge is to develop approaches that effectively capture the behavior at multiple spatial and temporal scales. For example, in geologic CO_2 sequestration application there are multiple coupled processes taking place in the primary sequestration reservoir. The coupled processes include hydrologic, geo-chemical and geo-mechanical processes resulting from interaction of CO_2 with the reservoir fluids and rock. These processes have different time scales, whereas hydrologic processes are relatively fast, geochemical interactions are significantly slow. In addition, these processes are intrinsically coupled. In defining long-term fate of CO_2 in a sequestration reservoir, it is necessary to take into account these coupled processes (which can be done through detailed numerical simulations). The challenge is to capture this coupled behavior effectively in a systems level model for the reservoir, so that it can be used for performance assessment. The computational challenge is to develop numerical algorithms that can describe the physics of the sub-system that care computationally efficient so that PA modeling can be performed in a timely manner.

Subsurface Science Collaboration

Most scientists conduct analysis and run models in several different software and hardware environments, mentally coordinating the export and import of data from one environment to another. Scientific and engineering workflows attempt to formalize this ad-hoc process so that scientists can design, execute, and communicate procedures repeatedly and with minimal effort. These workflows are a blend of scientific problem-solving, engineering design, and traditional business workflow techniques. The workflow research for a collaboration infrastructure should concentrate on components to enable Dynamic Data-Driven Application Systems (DDDAS) and 3D visualization.

In DDDAS, simulations and measurements form a symbiotic feedback control system. In carbon sequestration the monitoring of the CO_2 plume forms a feedback loop with site characterization and the modeling/simulation activities. Integrating enabling components into the scientific and engineering workflow framework to support DDDAS is integral to the success of the carbon sequestration program.

On the other hand, 3D visualization is on the critical path to scientific and engineering understanding and discovery. Providing scientists and engineers with matching computational, analysis and visualization power should be an essential element of a subsurface sciences collaboration infrastructure and visualization is currently the weakest link. This work should provide a significant step forward in integrating hardware-accelerated visualization into the collaboration infrastructure.

Modeling the Sequestration of Supercritical Fluid Mixtures

Simulation of CO_2 injection into subsurface geologic formations is now done routinely (e.g., Xu et al. 2005; Bacon et al. 2006), and the equation of state for the pure CO_2 system is fairly well developed (Span and Wagner 1996). The multiphase flow simulators used to simulate supercritical CO_2 injection (e.g., White and Oostrom 2006; Xu et al. 2006) into the subsurface have typically been used to simulate a limited number of fluid components, with either a single gas phase or a binary gas mixture (e.g., Oldenburg and Unger 2003). A multiphase flow model is needed that incorporates an equation of state for gas mixtures, along with a means for tracking changes in the composition of the gas mixture with time and space. An equation of state for fluid mixtures is needed over a wide range of temperatures and pressures.

The principle species for this system, CO_2, N_2, O_2 and Ar, are closed electronic shell non-reactive molecules. This means that high accuracy can be obtained from a two-body potential description of the system without polarization. There is probably enough data for the pure systems to develop highly-accurate two-body like-like potentials. A site-site interaction producing very high accuracy of many properties, including near critical phase coexistence has already been reported for the most difficult species CO_2 (Zhang and Duan 2005). Preliminary progress has been made for the pure N_2 and O_2 end members (Duan et al.). The unlike interactions (e.g., CO_2-N_2 two-body interactions) in these systems are more difficult. However,

because of the lack of dipole in these systems, two body interaction calculated from high level quantum chemistry methods (e.g., CCSD(T); Bartlett 2005) should provide sufficient accuracy. Again, this is a good system to test the concept of developing a fully first principles thermodynamic model of a useful system.

Given the potential models, the simulations for the flue gas system can proceed, once an efficient implementation of the two-body potential for an MD code has been developed. However, given that free energy and partial molar quantities are desired (requiring very large particle number simulations) it would be a good idea to develop this code for a massively parallel environment. Considerable progress in this direction has recently been reported (Bowers et al. 2006; Fitch et al. 2006). For the prediction of phase equilibria, a Gibbs Ensemble (Panagiotopoulos 1987) or similar method should be used. This means that problems with particle insertion and averaging over a mixture composition will require many floating point operations, again suggesting that a highly scalable code be developed (Duan et al. 2004). In principle this is manageable and progress has been made for similar mixtures. These calculations suggest that even in the critical region simulations (MC) can yield results with near experimental accuracy.

Modeling FutureGen-Scale Geologic CO_2 Sequestration

In terms of model development, a key challenge is to achieve computationally efficient and robust explicit coupling of models that represent interdependent physical and chemical processes operating over a wide range of space-time scales. A fundamental scientific challenge is to determine the spatial dimensions—which may range from plume to basin scale depending on reservoir permeability and lateral continuity—that are required to accurately represent injection-triggered pressure perturbations, dependent evolution of the reservoir/cap-rock/well-bore stress regime and strain response, and the ultimate fate of displaced formation waters. Within this appropriately dimensioned domain, a second key challenge is to determine the spatial granularity—which may be uniformly coarse or fine, telescoped from coarse to fine with increasing plume proximity, heterogeneous fine and coarse to mimic known hydraulic and compositional variations, or even dynamic (adaptive gridding techniques)—that is required to provide sufficiently accurate simulation of CO_2 plume migration, sequestration partitioning among distinct trapping mechanisms, and long-term isolation performance, which depends most importantly on the evolution of cap-rock/well-bore seal integrity per concomitant geomechanical deformation and geochemical alteration processes (Johnson et al., 2005). Further, spatial granularity of the overall domain must be optimized in concert with temporal granularity of (i.e., time stepping through) overall simulation time; e.g., appropriate dovetailing of space/time granularities is particularly crucial in the context of accurately and efficiently representing the migration of steep gradients in fluid-phase saturations and aqueous concentrations, which characterize dynamic plume boundaries.

OPPORTUNITIES AND RESEARCH NEEDS

Improved Computation of Coupled Multi-Scale Process Models For Carbon Sequestration

We identify three main classes of challenges for accurate and efficient computational simulation of underground carbon sequestration: 1) algorithms for multi-scale modeling of specific, individual processes, 2) techniques for modeling the coupling of these specific processes, and 3) computational software infrastructure tools. The specific processes requiring multi-scale modeling include data interpretation and scaling from site characterization, physics-preserving model discretization, well models, accurate representation of natural geologic heterogeneity, and adaptive gridding for error reduction. Because of the complexity and shear size of the problem, carbon sequestration codes must run on the world's largest supercomputers. This necessitates research on improving massively parallel computation of coupled process models, including operator splitting techniques involving both space and time, utilizing implicit, semi-implicit, and explicit couplings and adaptive strategies, and error estimation and propagation of errors between sub-domains and across process modules.

Efficient Risk Assessment Techniques for Carbon Sequestration

A priority direction is the development of numerical algorithms and computational techniques that can be used to perform comprehensive, timely risk analysis of geological CO_2 sequestration sites. These approaches will combine the development of novel methods of risk analysis, up-scaling methods, model reduction approaches, efficient parameter sampling methods, efficient algorithms for large-scale Monte-Carlo simulations (MCS) and application of petascale computing. Developments associated with this research would be beneficial in defining the research priorities related to overall geologic sequestration science based on data needs. In addition, developments made during this research will be beneficial to subsurface computational science and will be applicable to a number of practical areas such as risk analysis of contaminant transport, and the optimization of petroleum reservoir performance.

Subsurface Science Collaboration

We propose that the site characterization, monitoring data, and the instruments themselves, be integrated into the collaboration infrastructure forming virtual observatories, one for each proposed site. These observatories will provide the first carbon sequestration measurement and observation systems designed to both to answer significant scientific questions and to have the multidisciplinary participation necessary to achieve credible subsurface forecasting and prediction. As such, each will transform the way we conduct subsurface science by enabling the integration of research, development and demonstration from modeling to observations to measurements.

The repository would utilize a framework that would promote global access and preserve local control. It will provide collaboration tools for tracking and reviewing the development of new

and improved codes. The framework would also be capable of sustaining a wide and ever-evolving variety of codes from the subsurface science community, in modular formats to provide a platform for continual improvements.

Modeling the Sequestration of Supercritical Fluid Mixtures

Although a general equation of state for supercritical fluid mixtures has been developed (Duan et al. 1996), it is not accurate at the lower temperatures generally encountered in geological carbon sequestration. Furthermore, this equation of state must be expanded to include subcritical temperatures and pressures, in order to perform simulations for risk analysis where the flue gas may rise toward the surface. Simulation of the sequestration of multicomponent supercritical fluid mixtures in the subsurface is a greater computational challenge than modeling pure CO_2 sequestration, due to the more complex equation of state and the need to track relative amounts of each fluid component. Improved scalability of compositional multiphase flow and/or coupled reactive transport and multiphase flow models would be needed to meet this challenge.

Modeling FutureGen-Scale Geologic CO_2 Sequestration

The requisite coupling of diverse process models, the characteristic hydrologic/compositional complexity and heterogeneity of basin-scale geologic systems, and the dependent extreme space-time dimensions and granularities required for numerical simulation of long-term CO_2 isolation performance demands advanced parallel architectures, numerical methods, and space-time discretization schemes. Hence, large computing clusters are required and must be exploited efficiently to carry out these applications; storage issues will be addressed by forming sub-models on individual processors and allowing communication between them through boundary conditions (e.g., Dickinson et al., 2006), while inter-compatibility of sub-models will be achieved through iteration. Great challenges exist in designing such coupling algorithms and load balancing as well as data scatter and gather among the myriad process models; all potential couplings of scale and process should be explored. As a result, studies within this research will help establish guidelines for the design, development, and implementation of robust and efficient massively parallel reactive transport simulators.

CONCLUSION

The work recommended here regarding improved computation of coupled multi-scale process models, modeling of supercritical fluid mixtures, and development of efficient risk assessment techniques will enable more accurate predictions of CO_2 sequestration potential of target geologic formations, improved assessment of risks to public health, better management of the injection process, and long-term modeling of the formation and assessment of human intervention when necessary so as to maintain carbon sequestration over time.

The subsurface science collaboration infrastructure will expedite research and development by avoiding duplication of effort and allowing scientists to “stand on each others’ shoulders.” It will facilitate collaboration among the large numbers of disciplines and research groups necessary to

attack the multiscale, multiphysics nature of CO2 sequestration, and it will accommodate new and emerging high-performance computing environments.

Defensible, accurate forecasting of uncertainty-bound long-term CO_2 isolation performance will be a required component of FutureGen power plants. Sophisticated predictive efforts of this kind absolutely demand advanced reactive transport modeling capabilities. Studies within this PRD will significantly improve and demonstrate such capabilities and, as a result, our ability to identify optimal geologic targets for CO_2 storage, forecast their long-term isolation performance, and optimize injection strategies and monitoring programs to achieve and verify such performance.

REFERENCES

Bacon, D. H., M. D. White, N. Gupta, J. R. Sminchak and M. E. Kelley (2006). Assessment of CO_2 Injection Potential and Monitoring Well Location at the Mountaineer Power Plant Site. 8th International Conference on Greenhouse Gas Control Technologies, Trondheim, Norway, Elsevier, Oxford, UK.

Bartlett, R. J. (2005). How and Why Coupled-Cluster Theory Became the Preeminent Method in Ab-initio Quantum Chemistry. *Theory and Applications of Computational Chemsitry*. C. E. Dykstra, G. Frenking, K. S. Kim and G. E. Scuseria, Elsevier**:** 1191-1221.

Bowers, K. J., E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Greggersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan and D. E. Shaw (2006). "Scalable Algorithms for Molecular Dynamics simulations on Commodity Clusters." *SC2006, November, 2006, Tampa, Florida*.

Duan, Z. H., N. Moller and J. H. Weare "Gibbs Ensemble Monte Carlo Simulation of Vapor/Liquid and Metastable Liquid/Liquid Phase Equilibria in the CO_2-CH_4-N_2 System." *Unpublished research*.

Duan, Z. H., N. Moller and J. H. Weare (2004). "Gibbs ensemble simulations of vapor/liquid equilibrium using the flexible RWK2 water potential." *Journal of Physical Chemistry B* **108**(52): 20303-20309.

Duan, Z. H., N. Moller and J. H. Weare (1996). "A general equation of state for supercritical fluid mixtures and molecular dynamics simulation of mixture PVTX properties." *Geochimica Et Cosmochimica Acta* **60**(7): 1209-1216.

Dickinson, J. E., James, S. C., Mehl S., Hill, M. C., Leake, S. A., Zyvoloski, G. A., and A. Eddebbarh , 2006, "A New Ghost-Node Method for Linking Different Models and Initial Investigations of Heterogeneity and Nonmatching Grids", accepted in *Advances in Water Resources.*

Fitch, B. G., A. Rayshubskiy, M. Eleftheriou, T. J. C. Ward, M. Giampapa, R. S. Pitman and R. S. Germain (2006). "Blue Matter: Approaching the Limits of Concurrency for Classical Molecular Dynamics." *SC2006, November, 2006, Tampa, Florida*.

Johnson, J.W., Nitao, J.J., and Morris, J.P., 2005, Reactive transport modeling of cap rock integrity during natural and engineered CO_2 storage: in *Carbon Dioxide Capture for Storage in Deep Geologic Formations: Vol. 2*, Thomas, D.C., and Benson, S.M., eds., p. 787-813.

Oldenburg, C. M. and J. A. Unger (2003). "On Leakage and Seepage from Geologic Carbon Sequestration Sites: Unsaturated zone Attenuation." *Vadose Zone Journal* **2**: 287-296.

Panagiotopoulos, A. Z. (1987). "Direct Determination of Phase Coexistence Properties of Fluids by Monte-Carlo Simulation in a New Ensemble." *Molecular Physics* **61**(4): 813-826.

Pruess, K., C.-F. Tsang, D. Law, and C. Oldenburg. 2001. Intercomparison of simulation models for CO_2 disposal in underground storage reservoirs. LBNL-47353.

Pruess, K., J. Garcia, T. Kovscek, C. Oldenburg, J. Rutqvist, C. Steefel, and T. Xu. 2004. "Code intercomparison builds confidence in numerical simulation models for geologic disposal of CO_2." *Energy* **29(9-10)**, 1431-1444; LBNL-52211, 2003.

Span, R. and W. Wagner (1996). "A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa." *Journal of Physical Chemistry Reference Data* **25**(6): 1509-1596.

White, M. D. and M. Oostrom (2006). STOMP: Subsurface Transport Over Multiple Phases, Version 4.0, Users Guide, Richland, Washington, Pacific Northwest National Laboratory.

Xu, T. F., J. A. Apps and K. Pruess (2005). "Mineral sequestration of carbon dioxide in a sandstone-shale system." *Chemical Geology* **217**(3-4): 295-318.

Xu, T. F., E. Sonnenthal, N. Spycher and K. Pruess (2006). "TOUGHREACT - A simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media: Applications to geothermal injectivity and CO_2 geological sequestration." *Computers & Geosciences* **32**(2): 145-165.

Zhang, Z. G. and Z. H. Duan (2005). "An optimized molecular potential for carbon dioxide." *Journal of Chemical Physics* **122**(21): 214507.

COUPLED PHENOMENA

CURRENT STATUS

Current mathematical and computational capabilities for modeling coupled phenomena are inadequate for addressing future DOE subsurface science challenges. High-fidelity simulation of subsurface processes must be built on a fundamental understanding of multifluid, multiphase,

multicomponent flow and transport, biochemical and geochemical reactions, thermal and geomechanical effects, all interacting within a highly heterogeneous porous media. Sensitivity to physical processes and model parameters, and uncertainties in parameters, media properties and the current state of the system, must be taken into account. Optimal decision making and risk reduction will require the development of a comprehensive uncertainty/sensitivity analysis framework, within which coupled process models will be executed hundreds to thousands of times under different scenarios. The magnitude of this effort will tax even petascale computational resources; for it to be feasible will require new modeling approaches and the development of efficient and flexible high-fidelity simulators for coupled subsurface processes.

Unlike engineered systems, natural geologic formations are extremely complex. Material properties display high variability and complex spatial correlation structures span a rich hierarchy of length scales. In these subsurface systems, coupled subsurface processes involving multiphase, multicomponent interactions exhibit a wide variety of behaviors over a large range of spatial and temporal scales. It is not possible (or desirable) to include all these length and time scales in mathematical models of coupled flow and reactive transport. The current practice is to develop models based on the continuum scale where Darcy's law applies and the geometrical complexity of the pore space is averaged and replaced by empirically determined effective parameters or constitutive relations. This is adequate for most single phase flow and non-reactive transport. However, for complex, coupled, nonlinear processes (e.g., multiphase flow, biogeochemical reactions, and geomechanics), the lack of rigorous frameworks for upscaling undermines the reliability of predictive simulations used to inform decisions affecting the environment and human health.

Current capabilities for data management (acquisition, transport, storage, calibration, analysis, integration, and visualization) are inadequate for modeling and simulating complex, coupled, multiscale subsurface processes. As the processes requiring understanding and prediction become more complex, coupling of large-scale phenomena becomes more important and data management requirements become more challenging. Since processor speeds are not increasing significantly, coupled subsurface computations will need to use more processes and data management tools that allow for effective use of many processors for coupled bio, geo, and chemical processes. Adaptive decouplings within the code, to take advantage of weak physical couplings, or domain decomposition techniques will require very flexible, modular data management capabilities not currently available.

SUBSURFACE SCIENCE CHALLENGES

Challenges such as contaminant transport, CO_2 sequestration, and nontraditional energy production demand effective coupled simulations. The integration of dispersed data sources and the management of petascale datasets will enable coupled process simulation and the knowledge required by decision makers.

Environmental Management has many challenges in subsurface transport, coupled biogeochemical processes, attenuation, stabilization, modes of release, risk assessment, and regulatory compliance. All of these applications involve complex coupled phenomena, are extremely data intensive, require advanced data management, and ultra scale computational tools. Radioactive Waste challenges are siting, licensing, constructing and operating high-level waste sites. Processes are infiltration of water, corrosion of materials, coupled hydro-thermo-geo-chemical. Chemical transport in diverse media and long-term uncertainty in the data greatly complicate the data management issues. Fossil Energy challenges include CO_2 sequestration, oil shale and oil sands recovery, enhanced oil recovery, improved fossil energy utilization and hydrogen production from fossil fuels.

Since coupled processes are present in each application domain of importance to subsurface sciences, we must overcome the difficulties in treating enormous, disparate data structures that require adaptive access to metadata and complex data retrieval capabilities. Coupling of complex phenomena complicates uncertainty analysis and quantification by introducing additional sources of uncertainty. We need novel methods for dealing with uncertainty in both scale and process coupling. Since the massive coupling of large processes generates enormous, dense matrices in simulation codes, the adaptive utilization of decoupled processes solution schemes for mildly coupled processes must be developed. The development of efficient, effective data management tools for large coupled processes will enable major advances in our modeling and prediction of subsurface processes and will greatly assist in critical decision processes in each aspect of the DOE.

COMPUTATIONAL SCIENCE CHALLENGES

The principal impact of this research on computational science will be the development of scalable multiscale/multiphysics algorithms on petascale computers that use novel approaches to load balancing, data communication across components, and dynamic, adaptive restructuring of computational grids and algorithms in response to time evolution of complex systems. A critical outcome of developing accurate and robust multiscale methods to simulate process coupling is the ability to identify how detail may be sacrificed in exchange for accurately simulating increasingly large systems.

OPPORTUNITIES AND RESEARCH NEEDS

This panel identified three Priority Research Directions.

Mathematical and Numerical Simulations of Multiscale Subsurface Process Coupling

This priority research direction focuses on developing mathematical concepts, computational algorithms and petascale software required to understand and quantitatively predict the nature of coupled subsurface phenomena (hydraulic, thermal, geomechanical and biogeochemical) across

a range of scales. These phenomena are dominated by complex heterogeneities, multiphase, multicomponent flows and interacting, scale-dependent processes. Existing codes are unable to simulate these phenomena with sufficient fidelity for decision-making in part because they do not consistently account for multiscale interactions. The pressing issues facing DOE (e.g., remediation of mixed organic-radionuclide wastes, long-term radioactive waste isolation, impact of underground nuclear tests on water and contaminant movement, geologic CO_2 sequestration, methane hydrate extraction, oil shale development) require models that couple several processes (e.g., flow, biogeochemistry, thermal, geomechanical). For these problems, addressing the complexities of spatial coupling across scales calls for novel multiscale computational frameworks and architectures that can consistently feed information across scales (both upscaling and downscaling) to simulate coupled process dynamics efficiently.

Achieving efficient adaptivity in upscaling is a challenging task for multiphysics, multiscale problems. Identifying coarse regions and corresponding upscaled models, as well as setting up an adaptive coarse grid in the regions without apparent scale separation, is essential for accuracy and robustness of coarse-scale models. Development of mathematical tools, such as error indicators, to guide these coarsening strategies is an important issue that needs to be addressed. Such a multiscale computational framework would require incorporating ideas from computational upscaling/downscaling, mathematical upscaling/downscaling, and data assimilation. It would require taking advantage of the strengths of each area in a consistent manner with uncertainty determined in a quantifiable way. Implementation of such a multiscale framework would require tremendous (petascale) computational resources.

Within a multiscale framework, it will be necessary to propagate uncertainty through process dynamics across a range of scales to accurately quantify uncertainty in predictions of system behavior and risk associated with non-compliance. The mathematical frameworks for uncertainty propagation (e.g., Monte Carlo techniques, approximate covariance propagation equations, or polynomial chaos theories) lead to computational problems that are vastly different from their deterministic counterparts.

The challenging problem is to develop a rigorous methodology that can capture nonlocal effects accurately for heterogeneous systems across multiple scales where a scale separation assumption is inappropriate. For accurate upscaling, development of approaches and concepts that take into account global effects and while remaining applicable to problems without scale separation is essential. Addressing this challenging task for various subsurface problems is important for not only upscaling, but also for efficient preconditioning techniques that rely on accurate interpolation / downscaling techniques.

High Fidelity Petascale Simulation of Coupled Phenomena

There is a consensus within the subsurface community that solving key DOE application problems requires the modeling of a broad range of coupled phenomena in a complex heterogeneous subsurface. Coupling of non-isothermal, multiphase, multicomponent flow with

geomechanics and biogeochemistry is required to adequately capture system behavior. The diagram below connects several DOE subsurface science challenges with the primary, secondary and emerging physical processes of interest to each application. Despite some general understanding of the processes involved, we do not have a complete understanding of how to describe and efficiently simulate these coupled phenomena. As a result, existing code technologies generally rely on one of two strategies. The first strategy is to couple existing legacy codes via external interfaces. The second approach is to couple all processes through fully implicit temporal discretizations combined with extremely robust but low order spatial discretizations to ensure numerical stability. These choices are made at the expense of high fidelity, computational efficiency and the ability to effectively exploit new high-performance parallel architectures. There is a priority need to develop next-generation multiphysics subsurface simulation capabilities/methods that enable a significant increase in the fidelity of subsurface modeling. Achieving this goal will require three basic elements. First, we must develop novel coupling/decoupling strategies based on a detailed analysis of the interaction of processes in subsurface flow. These new coupling/decoupling strategies must then be combined with emerging discretization and gridding approaches to improve the accuracy of subsurface modeling methodology without sacrificing robustness. Finally, these discretizations must be integrated with modern solver technologies so that the computational models can effectively exploit new petascale architectures.

Data Management for Coupled Phenomena

The goal for this research direction is to create integrated, petascale data management tools for disparate types of multiscale data. Advances in computational infrastructure, with petaflop computing and multiple petabyte online storage, are needed to support simulating coupled processes on very large spatial and temporal scales. The architectures and software tools for managing the requisite data and for choreographing the interactions between computation and data, however, are not currently available.

Since the modeling of detailed, multiscale processes in complex subsurface systems leads to data and computationally intense simulations, it is vital to create integrated petascale data management tools that support data access and mining, geostatistical analyis, pre- and post-processing, as well as visualization. Due to advances in computational infrastructure with petascale computing and multiple petabyte online storage, simulating coupled subsurface fluid flow processes on very large spatial and temporal scales is becoming attainable. The architectures and software tools for managing the requisite data and for choreographing the interactions between computation and data, however, are not currently available. The major issues are:

- Input datasets are huge, come from diverse sources in multiple formats, and must be ingested at various different (spatial and temporal) scales. Typically, input data are collected from different collaborators located in various geographic locations and may derive from field measurements, from lab measurements, or from the output of computational models.

Furthermore, it may not be fully known at the outset what datasets will need to be ingested (e.g., dynamically adaptive multiscale simulations).

- Output datasets are similarly large and formatted for use for multiple application objectives. It is also very common that an open-ended number of other collaborators or systems use the output data of one system.
- During the coupled computation, large amounts of data will be transferred among the various computing elements engaged in the computation. Again, due to the need for dynamic multi-scaling, the pattern and volumes of these transfers will not be know at the outset of the computation.
- During the coupled computation, large checkpoint files will need to be written.
- In addition to existing input datasets, real-time data from intelligent sensors will be important inputs to subsurface science computations. These intelligent sensors, when combined with the Dynamic Data-Driven Application Systems (DDDAS) paradigm, make it possible to modify the configuration of the intelligent sensors during the course of the computation.
- It is particularly important for datasets used in subsurface science to be coordinated with accurate GIS information while modeling coupled surface/subsurface systems.

This need for coupled computations to receive inputs from many different sources and to provide outputs to many different applications strengthens the need for data and metadata format standards. One current example, relevant to coupled computations for subsurface science, is the current version of Hierarchical Data Format (HDF5). Metadata for specific kinds of data relevant to subsurface science is needed; cf. the Open Geospatial Consortium (www.opengeospatial.org) for a significant metadata initiative in a related field.

Given these distributed, large-scale, near real-time data sharing/collaboration requirements, a common Data Collaboration Platform that can provide integrative support for data management/sharing and for choreography among data assembly, computation, visualization, and the steering of computations and instrument configuration, is urgently needed.

CONCLUSION

This is an opportune time to develop next-generation multiphysics subsurface simulators that are flexible enough to take advantage of the known coupled physics/chemistry/biology and data at appropriate scales and link them to provide high-fidelity predictions. We envisage the development of robust computational upscaling and downscaling approaches that permit consistent transfer of information across scales for coupled nonlinear processes and feedbacks. Emerging petascale computing platforms and data management systems are essential components for the development of these novel modeling capabilities.

RESEARCH AT FUNDAMENTAL SCALES

CURRENT STATUS AND SUMMARY OF CHALLENGES

Providing safe and reliable energy while protecting the quality of critical groundwater supplies requires significant improvements to uncertainty quantification in simulations of subsurface flow. Meeting CO_2 sequestration MMV (measurement, monitoring and verification) demands as well as needs for the characterization of relevant processes and properties at Department of Energy (DOE) waste legacy sites requires the ability to model the temporal and spatial evolution of fluids and corresponding transport pathways within heterogeneous geological systems. All subsurface science models involve multiple poorly separated length and time scales and complex, coupled physical phenomena. An understanding of the linkages between all of the different temporal and spatial scales is thus critical. No matter how detailed the physics in a reactive-transport model, if the critical underlying physical, chemical and biological data are missing or unreliable, the accurate predictive capability of a model can be compromised. To model processes at one scale requires that we understand the important variables at smaller scales to insure that these variables are accurately represented at the larger scale. The coupling of mathematical models between fundamental scales (atomistic, pore, laboratory, and field) plays a defining role in understanding subsurface phenomena related to aquifer management, waste clean-up and CO_2 sequestration. Laboratory and field scale measurements show that reactive processes at the pore scale may have significant impact on long-range contaminant transport. For example, failed predictions of uranium removal by ambient groundwater flow at the 300 Area of Hanford are increasingly thought to be a consequence of slow leaching a sub-millimeter length scales. Improvement in both the data and the conceptual models will lead to simulations that are more scientifically defensible. Great care must be taken to minimize errors in the data used in sophisticated environmental or chemical process models so that these errors do not accumulate, propagate, and ultimately invalidate the macroscopic-scale models. Currently, direct numerical simulation from fundamental scales to field scale is impossible. Thus, what are the critical phenomena that are important at the individual scales and how does one use data from different scales in coupled models? What are the critical scientific, mathematical and simulation approaches to deal with the complex geological scales important to the DOE's efforts in subsurface modeling. The Panel on Research at Fundamental Scales dealt with these issues.

A summary of the fundamental spatial scales and appropriate computational approaches and critical application scales is given in Table 1.

Table 1. Fundamental Spatial Scales and Computational Methods

Fundamental Scales	**Computational Approach**
Molecular	Schrödinger Equation
Molecular, nano	Molecular Dynamics
Micron (colloidal particles, polymers)	mesoscale modeling
Pore	Discrete grains, continuum or discrete description of

	fluids, Smoothed Particle Hydrodynamics/Discrete Particle Dynamics/ Lattice Boltzmann / Computational Fluid Dynamics
Application Scales	**Critical Physical Phenomena**
Mineral-water interface	Adsorption, mineral growth/dissolution, redox
Pore network	Linked pores and grains
Continuum (porous medium)	Darcy flow, bulk concentration reactions
Aquifer/aquitard	heterogeneous porous media
Geological basin	features that control fluid flow and geochemistry – fractures, stratigraphy

Many fundamental spatial scales are important for contaminant fate and transport modeling in the subsurface ranging from the molecular to the nanoscale to the mesoscale polymer/colloidal), and to the pore scale. There are also many time scales of importance ranging from the femto-second time steps of molecular motions to the millions (and billions) of years of geological time scales. These scales provide the basic parameters which serve as the basis of the physical models important at even larger scales beginning with pore networks and fractures to the continuum porous medium to the aquifer scale and then to the geological basin scale.

The computational approaches to deal with these scales and the coupling between scales led to the development of a set of Proposed Research Directions as outlined in Table 2. These PRDs began at the smallest scale, that of atoms and molecules and went up to the need to develop experimental approaches and datasets to validate models as well as to provide critical data at different scales. Although most of the scales are separated by spatial size, there was a specific PRD on bridging time scales and the individual PRDs on spatial scales also dealt with different time scaling issues.

Table 2. Proposed Research Directions – Fundamental Scales

Computational Model Development for Critical Processes at Fundamental Scales: Molecular and Nano
Pore Scale and Microscale Computational Polymer and Colloidal Science for Subsurface Applications
Modeling Processes at the Pore Scale
Efficient Bridging of Time Scales
Multiscale Analysis
Highly-Resolved Numerical Testbeds for Evaluation of Upscaled Models
Improved Media Parameterization and Reconstruction
Coupled Geophysical Imaging for Subsurface Fluid Identification and Process Characterization

We expect that advances in the application of computational approaches to subsurface science and technology will benefit substantially from the efforts made by computational scientists working on a very wide range of commercially and scientifically important applications. We believe that it will be important for the computational subsurface science community to work closely with this larger community. Advances will require teamwork between computational chemists, biologists applied mathematicians and computer scientists.

COMPUTATIONAL MODEL DEVELOPMENT FOR CRITICAL PROCESSES AT FUNDAMENTAL SCALES: MOLECULAR AND NANO

Models of contaminant fate and transport in the subsurface are built on detailed knowledge of the binding and reaction of contaminants on soil particles as well as transport and reaction in groundwater. There is a critical need for improved data at the molecular and nanoscale for such models in terms of both spatial length and time lengths from the femto-second to seconds and larger. These scales are the most fundamental small scales in both space and time that are relevant to subsurface modeling. These are the scales when coupled to thermodynamics provide the basic equilibrium and kinetic data necessary for all subsurface transport models. Improved data at these fundamental small scales are needed for effective models at larger scales such as colloid and pore scales, which then connect to higher levels at the pore network, fracture and geological basin scales. Substantial advances are needed in electronic structure and molecular dynamics approaches to provide reliable predictions of the thermodynamics and kinetics of complex geochemical and biogeochemical systems. The availability of petaflop computer architectures and the development of computational approaches and algorithms that take full advantage of these architectures will allow us to make great strides in understanding the molecular/nano-scale chemistry of complex subsurface systems and evaluating which data from these scales must be incorporated into larger-scale models for accurate prediction of field-scale processes.

Great progress has occurred over the last two decades in understanding the fundamental chemical processes that govern contaminant fate and transport. This progress has occurred through a combination of experimental advances driven by access to high-end measurement capabilities and of computational chemistry advances made possible by the availability of more advanced computer architectures. However, because of the complexity of the subsurface environment, substantial advances in theory, algorithms, software, and access to high performance computers are still needed for such progress at the small scales to continue. Advances are required in electronic structure methods to deal with complex surfaces, to provide accurate results for the thermodynamics of solutions interacting with mineral surfaces, and to provide reliable predictions of kinetics in such complex systems. In addition, there are serious issues with current dynamical simulation techniques in terms of the length of the possible dynamical simulation, which will require substantial advances in theory and algorithms especially for petaflop computing if we are to reliably simulate the dynamics of large nanoscale systems for times beyond nanoseconds. Finally, there is a need to model biological processes from the protein and cell membrane scale to the cell scale to the community level as all of these can have substantial impacts on subsurface processes. The community scale is at the >mm scale. Improving existing capabilities will have both immediate and long-term impacts.

Pore Scale and Microscale Computational Polymer and Colloidal Science for Subsurface Applications

The prediction of the behavior of small particles and polymers in complex pore spaces and fracture apertures is challenging because multiple physical processes acting on a wide range of length scales must be taken into account. Computer simulations have been used for more than half a century to develop a better understanding of polymers and colloids. During this time, rapid increases in computing speed and memory capacity and speed have enabled steady progress from simple models such as the self-avoiding random walk model for polymer molecules in a good solvent to the realistic simulations of the folding of protein molecules in aqueous environments. In the past, most simulations of polymers and colloids have focused on the discovery and testing of scaling relationships, fractal dimensions, and other "universal" or generic characteristics and behaviors. The models used for this purpose have usually been based on simple particle shapes and interaction potentials – often selected to offset the effects of limited computing capability. The simulation of protein folding is a notable exception, but a detailed molecular dynamics simulation of protein folding, using the fastest computers available today, is possible only for small, rapidly folding, proteins. In addition, most simulations have been carried out using only a single computational method (Monte Carlo simulation, molecular dynamics, dissipative particle dynamics or Brownian dynamics, for example). Such simulations take into account only a limited subset of the physical processes that play an important role in most physical polymer and colloidal systems. In addition, the application of particle-level and molecular simulations to the behavior of polymers, colloids and nanoparticles in the subsurface has been quite limited.

A new generation of multiscale, multiphysics (and multi-algorithm) codes that perform well on petascale computing systems are needed to raise computational polymer and colloid science for subsurface applications to a new level. This new capability will allow system specific simulations that include all relevant physical process to be applied to subsurface problems and applications and will lead to a reduction in the number of experiments. The same approach will also provide the information and insights needed to better predict the behavior of polymers and particles in the subsurface, and eventually pore scale models will be coupled with continuum scale models for a variety of applications including prediction of the fate and transport of toxic particles formed in radioactive waste repositories, and the impact of synthetic or biologically generated particles and polymers on oil recovery. As computing capability grows and improved understanding and experience leads to better algorithms, the accuracy of system specific simulations will increase and they will play an increasingly important role in the development of polymers, colloids and nanoparticles for subsurface applications and in predicting the behavior of anthropogenic and natural polymers and particles in the subsurface. After ten years, we expect that fully coupled multiscale multiphysics methods will allow particle level simulations to inform larger scale subsurface flow and transport models in a fully integrated manner.

Modeling Processes at the Pore Scale

Pore scale modeling has been used for several decades to simulate single phase and multiphase fluid flow in fractured and porous media, and recently these models have been coupled with simple models for precipitation, dissolution, and the transport of dissolved substances. Methods that have been developed to simulate multiphase fluid flow by the astrophysics, statistical physics and computational fluid dynamics communities have been modified and applied to the behavior of fluids in the subsurface. In addition, recent advances in lattice Boltzmann methods have substantially expanded the range of potential applications to subsurface processes. Despite these advances, we currently do not have algorithms that can be applied to the full range of fluids and conditions that are important in subsurface systems. In particular, large viscosity, density and compressibility ratios are a challenge for most methods. In addition, some promising methods such a smoothed particle hydrodynamics and dissipative particle dynamics have been applied only on quite small scales (a small number of pores or fractures). However, improved codes on more capable computing systems are needed to simulate the behavior of reactive fluids with a wide range of properties under the full range of conditions relevant to subsurface applications. Petascale computing is needed to enable the application of these methods (and other methods based on “first principles” approaches) to systems on the scale of those currently simulated using more empirical and approximate pore network models. For the first time, this will enable the gap between pore scale and continuum scale models to be closed using accurate methods based on first principles (at least for those systems for which the concept of a representative elementary volume (REV) is valid and the application of traditional continuum modeling approaches can be justified). However many important applications will also require the development of better single physics algorithms and new codes that couple all-important physical and chemical processes on a wide range of length and time scales. In some cases, information from or coupling with atomistic and mesoscale (between the atomistic scales and the scales on which fluids can be represented by continuum fields) simulations will be required, while in other cases, pore scale models will be based primarily on experimental information such as fluid viscosities and surface tensions. For example, the coupling between fluid flow and fluid-fluid-solid contact line dynamics depends on atomistic and mesoscale physics, and complex evolving sub pore scale morphologies (and the associated changes in surface areas and reactivities) which can change rapidly as geosystems are forced far from equilibrium must be taken into account in many applications. Eventually, pore-scale models will be coupled with atomistic and/or continuum scale models on computing systems with capabilities that substantially exceed those of current systems. Although we expect that, with the right resources, critical advances can be made during the next five-ten years, challenges requiring new ideas, new methods and new, much more capable, computing systems will remain.

Efficient Bridging of Time Scales

The wide range of temporal scales in subsurface science, from the tenths of picosecond motions of atoms in formation fluids and minerals to geological scales in million of years poses an even larger computational challenge then does the large range of spatial scales described above. This

is a major obstacle for accurate predictability from simulations in subsurface science is the wide range of temporal scales. The development of innovative computational techniques for incorporating phenomena on fundamental time and spatial scales into models that are relevant on field scales is needed. Improvement of fast solvers on fundamental scales is necessary. The fundamental algorithms of atomistic modeling need to be improved in order to provide accurate solutions of the electronic Schrödinger equation based on new algorithms that will take full advantage of modern computer architectures in order to use *ab initio* molecular dynamics on realistic systems. New techniques are required for the prediction of the rare events that represent abrupt qualitative changes in the system (e.g. chemical reactions). At atomistic scales, the goal is to extend the time that can be accurately simulated. Rare event methodologies would also apply to other relevant time scales. Coupled multi-scale methods are also important. Recent developments have introduced new paradigms for on-the-fly coupling of, for example, atomistic and continuum models in the same simulation. Subsurface simulation is an ideal environment to develop these paradigms into concrete algorithms. One example is mapping of *ab initio* (calculated from the electronic Schrödinger equation) interactions and dynamics to adaptive effective potentials to be used in a larger time scale molecular dynamics simulation. Another example is the use of pore scale network models to dynamically generate parameters for a Darcy type simulation. The impact of massively parallel computing on simulation fidelity has been fundamental. Parallelization by spatial domain decomposition is standard. Parallelization in the time dimension is much harder. A few new technologies have been introduced. It will be advantageous to adapt these parallel-in time algorithms to the modeling subsurface processes and apply them, for example, to pore-scale flow simulations where friction and dissipation are present. Success on these projects would make it possible to fully take advantage of future generation computer architectures for the reliable simulation of essential subsurface processes.

Multiscale Analysis

Experimental and field scale measurements have demonstrated that reactive processes at the pore scale may strongly affect long-range, even field-scale, contaminant transport. While modeling efforts exist at specific fundamental scales, multiscale techniques that accurately capture the influence of geochemical reactions over a range of strongly coupled, non-well separated scales is in its infancy. Moreover, the required dynamic range of scales in both time and space is sufficiently large that fully resolved reaction simulations are unattainable. Combining the potential resolution and fidelity of direct numerical simulations made possible by petascale computing with current experimental and modeling efforts will stimulate new computational and theoretical advances in multiscale analysis. These advances will play a critical role in conducting simulations with well-characterized uncertainty which, in turn, will drive responsible policy decision and program development. Existing theories and computational tools have limited capability to treat the extreme range of closely coupled time and length scales typically present in reactive transport. New developments in multiscale analysis are required to significantly reduce uncertainty in subsurface flow simulations. A priority research area is the development of rigorous approaches for upscaling in the presence of many spatial and temporal scales. Application of such approaches to models of reactive transport with high reaction rates is

of critical importance. Progress in this direction will drive development in related areas: (a) upscaling flow and non-reactive transport in disordered or random media; (b) downscaling of coarse data used for site characterization; and (c) upscaling over time scales. The computational component of this research is critical and depends on Petascale class computing to advance numerical methods and mathematical models at specific scales, to facilitate bridging of neighboring scales through direct high-resolution simulations, and to conduct large ensemble studies. The synergy of these studies will provide the critically needed, high fidelity predictive simulations, with well characterized uncertainty, that are necessary for predicting accurate flow dynamics at the field scale.

Highly-Resolved Numerical Testbeds for Evaluation of Upscaled Models

All subsurface science models involve multiple poorly separated length and time scales and complex, coupled physical phenomena. Direct numerical simulation from fundamental scales to field scale is currently impossible. Analytic understanding of scaling is incomplete; upscaling commonly requires simplifying assumptions that are often invalid, and downscaling commonly leads to an ill-posed inverse problem. Research is needed to identify the minimum information that must be transferred across scales to maintain model validity. Work is needed on developing hybrid models based on local scale refinement and hierarchical models that maintain accuracy across scales by means of scientifically and mathematically defensible upscaling techniques. There is currently little or no basis on which to (1) determine appropriate sequences of scales for models and 2) evaluate the impacts of information loss due to upscaling. We recommend developing highly resolved, computationally demanding models of specific processes to serve as testbeds, or set of benchmarks, for evaluating upscaling methods and upscaled models. These numerical testbeds should represent the best current understanding of selected subsurface flow, transport, and reaction processes at fundamental scales (e.g., pore-scale or smaller), combined with state-of-the-art non-invasive characterization datasets. This testbed will require tera/petascale computational resources because 1) the simulation of processes at the relevant scales will be highly computationally intensive, and 2) the specific nature of each testbed will allow customized code with minimal overhead to be designed for maximum scalability. Advanced visualization, data mining, and data management tools should be utilized to maximize the availability and usefulness of the numerical testbeds once the simulation process is completed. The benchmark problems need to be validated using laboratory or field data. Advanced algorithms, mathematics, and computational methods will be developed for solving large problems and will provide a validation basis for application-level codes and models.

Improved Media Parameterization and Reconstruction

Critical geologic features, such as mechanical discontinuities, for any subsurface reservoir/repository exhibit heterogeneity that spans multiple scales. Mechanical discontinuities range in scale from a few microns (micro-cracks) to centimeters-meters (fractures, joints) to kilometers (faults). They can occur singly or in sets producing heterogeneity from the pore scale to the reservoir scale. This heterogeneity will vary temporally because of participation in the

hydrogeologic, geochemical, biogeochemical and tectonic cycles, in addition to any human activities/interference. The prediction of the effect of subsurface media heterogeneity on physical phenomena at the decision level scale depends crucially on the parameterization of heterogeneity. In addition, the long time interval simulation of configurations with varying media structures needs the capability of creating samples of media with prescribed heterogeneity parameters (the inverse problem). A priority research direction is the development and validation of efficient media parameterizations that correctly capture the interaction between the various subsurface physical processes at all the relevant scales as well as the development of algorithms to solve the inverse problem. Such parameterizations must be amenable to uncertainty quantification to address the pervasive risk assessment questions that appear in mission-directed subsurface investigations. Statistical uncertainty quantification that has found its way into subsurface modeling usually treats the computational model as a black box. New approaches must bring uncertainty quantification into the model specification and result in computational models that fully incorporate uncertainty propagation rather than compute it through replication or approximately as an afterthought.

An essential component of this research direction is the experimental data-based validation of the models and algorithms that will result from this research. Exciting opportunities result from the fact at the core level (~10 cm) a wealth of experimental data can be generated which can be used for the validation of the relationship between parameterizations, microstructure and subsurface physical phenomena. This work will require substantial access to leadership class computational resources to solve the stochastic optimization, uncertainty quantification, parameterization selection, and design of experiments problems that arise in resolving the problem of media parameterization and reconstruction for computational subsurface science.

Coupled Geophysical Imaging for Subsurface Fluid Identification and Process Characterization

In order to obtain the necessary data for subsurface modeling, one must have the ability to obtain cost effective, volumetric and near time information on the temporal and spatial evolution of fluids and corresponding transport pathways within heterogeneous geological systems. Traditional approaches to monitor and characterize such processes through point sampling sensors in boreholes are expensive and fail to provide sufficient information. Traditional methods to monitor the lateral and volumetric extent of sequestered CO_2 and subsurface contamination and/or determine changes in the lithology, which control transport, involve the use of instrumented wells. Data from such wells provides an insufficient characterization of relevant subsurface processes and properties within heterogeneous geological media. While the information from boreholes is of very high resolution at a very fine scale in the direction along the borehole, it represents a very limited sampling of the subsurface in other spatial directions. In addition, dense temporal sampling of boreholes for all but the simplest parameters is financially infeasible. As the subsurface typically will have significant lateral and vertical variations in geology and physical properties, major features that affect the storage of CO_2 and contaminant migration and transport will often be completely missed by borehole sampling. In addition,

processes will only be detected if they affect the value at the sampling point at the moment of sampling. The only feasible approach to this problem is a comprehensive subsurface imaging effort that effectively exploits both multi modal geophysical data (including seismic and electromagnetic wavefield data and electrical field data) and confirmatory point measurement data. Such a comprehensive subsurface imaging effort must provide for quantitative prediction of subsurface processes, including fluid identification and transport pathways. It will require novel computational and numerical tools in the development and enhancement of effective joint imaging methodologies which can utilize temporally and spatially dense 3D and 4D (time lapse) geophysical data (electrical, electromagnetic and seismic).

To resolve the data sampling problem, joint geophysical imaging technologies need to be advanced and developed which can take advantage of the increasingly more common time lapse geophysical datasets. Development of a coupled geophysical imaging capability will need to be based on novel inverse methodologies as well as on detailed, physics based modeling which links the micro and meso scale properties of the subsurface system and the processes occurring in such systems to the macroscopically observed (change in) geophysical response. The novel inverse methodologies will require faster methods for evaluating gradient, Jacobian or Hessian operators within the chosen inversion/imaging framework as well as the effective use of anticipated advancements in solvers, preconditioners and multi-grid and multi-level methods that utilize Petaflop computing distributed architectures. In addition, coupled inversion will require innovative measures to link different physical models and data (both geophysical and point data) together in a way which honors method and geometry dependent differences in resolutions and uncertainties as well as theoretical and field derived relationships between different physical parameters. The imaging technologies and analysis should be able to provide high resolution information on fluid processes and properties in heterogeneous geological media with sufficient resolution and confidence for both regulatory and operational needs.

Such developments would lead to a significant breakthrough in our ability to image properties and processes in the subsurface, make quantitative predictions on flow fields and saturation levels, and obtain information on the properties that affect subsurface flow and transport at unprecedented resolution. This development is conditional on 1) the availability of extensive and high quality geophysical datasets over relevant field sites and well described test beds and 2) the establishment of extensive interactions and collaborations between geophysicists and computational scientists. This development will result in tools that can be coupled to advanced pore and continuum scale models in other domain areas (e.g., reactive flow and transport models, mechanical models, and models of biogeochemical processes). It will also lead to tools that would give near real time, actionable information on subsurface processes in days instead of months to years, and thus be useable for not only characterization and monitoring, but even for feedback driven control.

PRIORITY RESEARCH DIRECTIONS

SITE CHARACTERIZATION AND MODEL CALIBRATION

INTEGRATED MODELING, SITE CHARACTERIZATION, ERROR ANALYSIS, AND MANAGEMENT OPTIMIZATION FRAMEWORK

Jack Parker, Oak Ridge National Laboratory

ABSTRACT

Conceptual model formulation, site characterization, model calibration, uncertainty analysis, and system monitoring are commonly undertaken serially with minimal *ad hoc* iteration. Without a fully integrated approach to these tasks, global optimization of model formulation, characterization and monitoring activities, and design and operation of engineered systems cannot be achieved. As a result, excessive costs and/or risk may be incurred because models that are more complex or more simple models than warranted may be adopted for the intended purpose; data may be collected that does not substantially reduce decision uncertainty; truly useful data may not be identified and collected; and/or engineered systems may be over- or under-designed. The ability to efficiently and practically integrate conceptual model formulation, site characterization, model calibration, uncertainty analysis, and system monitoring for complex coupled subsurface systems is currently impeded by a lack of scalable, powerful tools for integrated analyses. Because such integrated analyses will require a large number of direct simulations, high-performance computing resources and advances in algorithms for direct and inverse solutions, error propagation, and multi-objective design optimization schemes are needed. This PRD proposes to undertake the development and implementation of protocols and code-independent computational tools for global optimization of conceptual model formulation, site characterization and monitoring activities, and engineering system design and operation under conditions of uncertainty.

EXECUTIVE SUMMARY

The goal of this PRD is to develop an integrated framework for modeling subsurface systems that enables global optimization of conceptual model formulation, site characterization and monitoring activities, and design and operation of engineered systems in consideration of prediction uncertainty. Applications of interest include soil and groundwater remediation, nuclear and other waste disposal systems, CO_2 sequestration, and conventional and unconventional hydrocarbon recovery. A stochastic optimization framework is proposed that simultaneously optimizes engineered system design and operation variables, conceptual model formulation, and/or site characterization/monitoring plans to minimize a multi-objective function that may involve cost expectation (including penalty costs associated with failure to meet design requirements), regulatory criteria, and/or other externalities. The integrated framework would couple model calibration (inverse problem solution) using data that may involve “direct” parameter measurements (e.g., hydraulic conductivity), indirect or “soft” data of various types

(e.g., borehole data, penetration resistance, geophysical data, etc.) with a generalized error analysis methodology (including measurement and up-scaling errors), with a multi-objective stochastic optimization procedure. Implementation will involve adaptation and refinement of existing codes, development and implementation of new methods, and integration in a manner that provides maximum computational efficiency, robustness, maintainability, and adaptability. The proposed effort would result in a major advancement in the ability to make reliable and cost-effective decisions using subsurface models.

SUMMARY OF RESEARCH DIRECTION

This priority research direction is focused on development of an integrated framework, computational tools and infrastructure for global optimization of conceptual model formulation, site characterization and monitoring plans, and design and operation of engineered subsurface systems with due consideration to prediction uncertainty. Specific applications may include groundwater remediation problems, large-scale CO_2 sequestration projects, production of unconventional hydrocarbon resources, etc. The approach addresses obstacles that currently limit the utility of computer models to manage large-scale engineering projects cost-effectively, by considering prediction uncertainty associated with field and laboratory data at various time and space scales, model formulation, and calibration. A computational framework that couples model calibration with a generalized error analysis methodology for performance predictions, in conjunction with a stochastic optimization methodology should be implemented and tested to identify system design parameters and conceptual model formulation to minimize a multi-objective function that may involve cost expectation (including penalty costs associated with failure to meet design requirements), regulatory criteria, and/or other externalities. The stochastic utility function would consider prediction uncertainty resulting from uncertainty in conceptual model formulation, measured and estimated model input, and the calibration process. The inverse solution would use both "direct" parameter measurements (e.g., hydraulic conductivity, reaction coefficients, etc.) and indirect or "soft" data (e.g., borehole measurements, penetration resistance measurements, geophysical test results, etc.) to reduce estimation uncertainty. Achieving this level of model integration will be a major advance in subsurface science.

Problem Overview

Uncertainty in model predictions may be attributed to two sources. First, models of complex systems invoke certain implicit and/or explicit assumptions (e.g., homogeneous, isotropic, valid parametric formulations, etc.). We refer to prediction error caused by deviations from these assumptions as "intrinsic" model error. Intrinsic prediction errors will vary depending on the magnitude of deviations from assumptions and on model sensitivity to these deviations.

Second, even in the absence of intrinsic error, "extrinsic" prediction errors arise from imperfect knowledge of model input associated with 1) uncertainty in measurements used as direct model input, 2) errors in estimates of model inputs derived from direct measurements at time and/or space scales different from the model after up-scaling (or down-scaling) to the model scale, and

3) uncertainty associated with propagation of uncertainty in estimated model parameters. Calibration of larger scale models implicitly requires measurements of primary model variables (e.g., concentrations, fluxes, etc.) averaged over larger spatial/temporal scales. However, averaging scales for direct measurements are often at smaller (or occasionally larger) scales than the model scale, resulting in uncertainty in model values at the assumed scale.

As model complexity increases (i.e., less restrictive assumptions, higher resolution), intrinsic model uncertainty is expected to decrease. However, more parameters will generally be required for model calibration. As data availability decreases (e.g., as financial resources become more scarce), parameter uncertainty will increase, thus leading to greater extrinsic model uncertainty. At some point of data scarcity, ill-posedness of the inverse problem will become so severe that prediction uncertainty grows rapidly. Under such conditions, multiple parameter sets may yield nearly indistinguishable calibration accuracy. However, predictions under conditions different from those during calibration could produce sharply different results. It is likely that as model complexity increases for a given level of data availability, prediction uncertainty will diminish up to a point and, thereafter, will increase as increases in extrinsic error outstrip the rate of decrease in intrinsic error. With increasing model complexity, prediction uncertainty is expected to become more sensitive to data availability. As data availability increases, the level of model complexity that minimizes prediction uncertainty is expected to shift towards greater optimal complexity, and conversely, with less data, the optimal model will shift towards lower complexity (Figure 1). This implies that for a given set of data available for calibration, increasing model complexity will eventually be counterproductive.

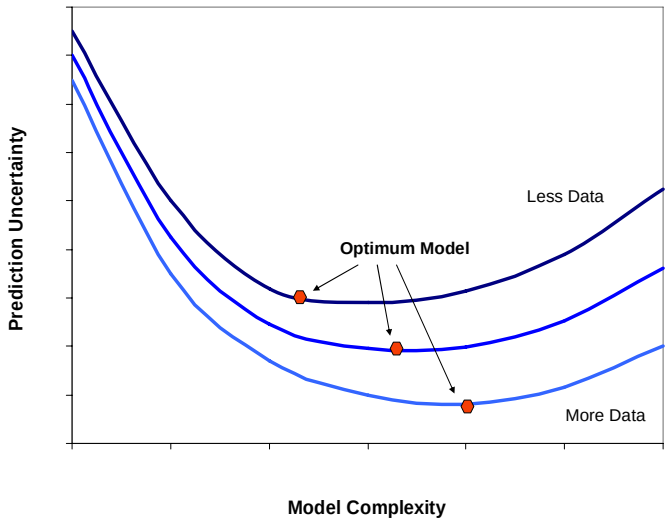


Figure 1. Conceptual relationship between model complexity (time/space scale resolution), data availability, and prediction uncertainty.

To maximize the utility of models for a given engineering application, it is critical to consider interactions between conceptual model formulation, data availability, calibration technology, and prediction uncertainty. Unfortunately, these relationships have been studied very little.

Furthermore, progress will remain impeded until scalable computational tools are available that enable coupling of multi-process subsurface models, inverse solutions, error propagation methods, and multi-objective management optimization schemes. Because such analyses will require a large number of direct model simulations, high-performance computing resources must be used. Global optimization of model formulation, characterization and monitoring activities, and design and operation of engineered systems cannot be achieved without the proposed fully integrated modeling approach. Global optimization avoids overly complex or overly simple models and over- or under-design of engineered systems, while ensuring optimal data collection to meet project objectives.

A high-level view of a possible integrated modeling framework is illustrated in Figure 2. For various specific applications, various components may be bypassed or iteration may be undertaken globally or on or subsets of components. This research would focus on relevant specific protocols, infrastructure, and component tools capable of addressing complex applied subsurface problems in a practical, cost-effective manner.

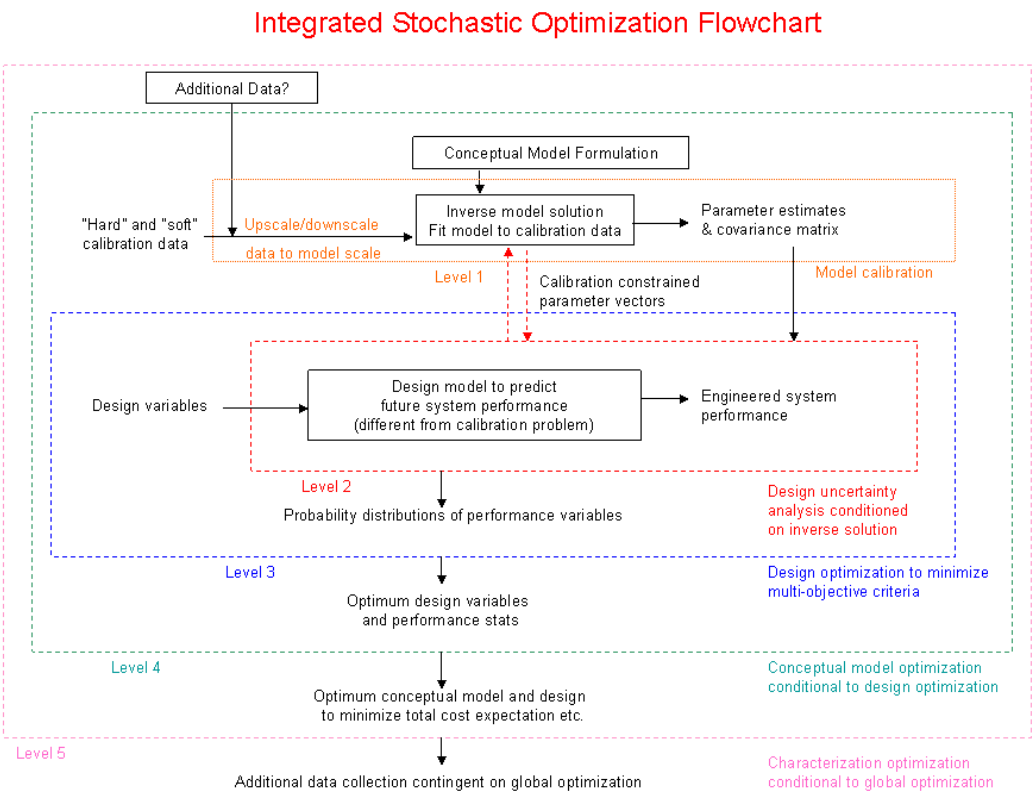


Figure 2. High-level flow chart for an integrated multi-tier stochastic optimization problem.

Conceptual Model Formulation

Conceptual model formulation is a crucial step in the modeling process. Ideally, conceptual model complexity and data-collection efforts will be advanced in an iterative manner. At the first iteration, a simplified conceptual model is used with initially available data to jointly assess

the need for additional data and/or model refinement. As additional data is obtained, the benefit of further data collection and/or model refinements can be evaluated. The process of conceptual model formulation is assumed to be a manual process performed by the project team. That is, at a given stage in the project, a limited number of potential model refinements would be specified for analysis. However, the process of converting candidate conceptual models to numerical code input can be laborious, and the development of improved tools to automate the process would be very useful.

Inverse Problem Solution

Direct determination of all input parameters required in a complex, multi-process subsurface model is generally not feasible. The estimation of unknown model parameters must therefore be carried out by an inverse solution approach in which parameter values are determined that minimize deviations between observed and predicted system behavior subject to constraints or plausibility criteria, weighted to consider uncertainty in observations. Because many direct measurements will not be at the same time/space scales as the model, up-scaling or down-scaling may be necessary to avoid biases caused by scale disparities. Uncertainties associated with the scaling process will need to be considered in addition to those resulting from measurement uncertainty (see also PRDs by the Fundamental Scales panel on up-scaling/down-scaling). Furthermore, "soft" data may often be available that exhibit indirect relationships with model input or output that can improve the inverse solution. In addition, scale conversion issues, "soft" data will also exhibit uncertainty associated with the constitutive relationship to model parameters or output for which it serves as a surrogate. Inverse problem solutions must accommodate the fact that objective function surfaces are commonly pitted with local minima. Furthermore, a large number of direct simulations can be required to obtain a solution to the inverse problem, especially for gradient-based methods with many unknown parameters. A variety of gradient and non-gradient methods have been developed that exhibit various advantages and drawbacks.

Estimation co-variances of inverted parameters may be computed from the Hessian matrix of the inverse solution objective function. Effects of alternative plans for additional site characterization or monitoring data collection on parameter co-variances may be estimated by running the inverse solution with synthetic "data" generated from the direct model solution without the added data.

Separate PRDs have been written that deal specifically with the development of robust, scalable inverse solution algorithms and implementations that would meet the specific needs for incorporation into the proposed integrated modeling protocol.

Error Analysis Problem

Following calibration, the model may be used to simulate future system performance for a specific problem (e.g., to compute concentration at a compliance location over a defined period for an assumed groundwater remediation plan). Model predictions will be uncertain to a degree

that depends upon the magnitude of parameter uncertainty, the magnitude and sign of parameter correlations, and the sensitivity of predictions to each uncertain parameter. Prediction probability distributions computed assuming multi-normal or multi-lognormal parameter errors will frequently lie outside the feasibility space for calibration data. This problem can be overcome by calibration-constrained optimization or Markov Chain Monte Carlo methods.

PRDs have been written that deal with the development and implementation of error analysis methods that would meet of the proposed integrated modeling protocol [see VV/Opt. Panel].

Multi-Objective Optimization

The purpose of the final component is to determine values of specified design and operations parameters that optimize the performance of an engineered system; for example, configurations and flow rates for wells to meet regulatory criteria with a given probability of success and/or at minimum expected total cost. Various approaches may be used for the formulation of stochastic design optimization problems. One approach involves the adoption of chance-constrained design criteria (e.g., finding the lowest cost design that will maintain contaminant concentrations below a certain value at a defined location with a specified level of confidence). The probability of design criteria exceeding critical limits for a given set of design variables (e.g., well locations, pumping rates, etc.) may be computed using various methods as described in the previous section. An alternative approach is to seek the design that minimizes the expectation of *total* cost, which includes penalty costs for not meeting design criteria (i.e., anticipated costs at a future date to rectify the consequences of failure). The expected total cost may be viewed as a probability-weighted average cost (for a given set of design variables) considering model uncertainty plus the failure cost times the probability of failure.

Global optimization of an objective function with many parameters is a particularly difficult and challenging problem. A PRD that focuses on the development and implementation of multi-objective optimization methods applicable to complex subsurface problems is described by the VV/Opt panel.

Component Integration

Integration of the foregoing components would provide a means of simultaneously optimizing conceptual model formulation, characterization/monitoring plans, and engineering design (Figure 2). In practice, the entire process may be iterated as incremental refinements in the model formulation and data collection are performed until further efforts are not warranted on a cost-benefit basis. Effects of undertaking additional site characterization studies, pilot tests, monitoring procedures, etc., on model predictions and their uncertainty may be evaluated by adding synthetic data at the proposed locations and/or times with appropriate measurements and/or up-scaling uncertainty and rerunning the inverse problem, error analysis, and management optimization problem to determine the impact on total cost (including the additional data collection costs) and hence to quantify the net cost or benefit of alternative plans.

Integration should be performed in a manner that provides flexibility to the user to execute selected components and bypass others to suit specific needs. Input/output structures and program control information should be implemented in a manner that is code independent, so the infrastructure can be readily implemented with any direct problem solution.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Computational methods for model calibration, uncertainty analysis, and management/design design optimization are available. However, interactions among conceptual model formulation, site characterization/monitoring plans, model calibration, and engineering design and management are rarely given more than superficial consideration in subsurface science, and global optimization spanning these activities has not been undertaken.

The integrated modeling protocol envisioned here may involve 10^6 or more direct solutions of calibration or forward simulations. This would be a challenging computational effort, even with efficient and highly parallel solution methods. Key challenges are identified below:

- Development of scalable calibration, error analysis and multi-objective optimization methods that enable robust and efficient solution of a complex nested optimization problem
- Development of efficient, flexible and code-independent I/O and program control structures to integrate code components
- Development of methods for up-scaling and down-scaling data for use in model calibration and quantification of resulting calibration uncertainty
- Development of methods for evaluating effects of future data collection activities on model calibration and prediction uncertainty.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

This project would develop new scalable algorithms that would be applicable to a wide variety of engineering problems.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

This PRD would lead to a quantum advance in the ability to use multi-process subsurface models to improve decision-making and minimize costs by simultaneously optimizing site characterization plans, long-term monitoring, engineering design and operations, and model formulation for complex subsurface problems.

TIME FRAME

Significant progress could be achieved in 3 to 5 years and full attainment of the objectives is anticipated within 10 years.

DATA UP-SCALING AND SOFT DATA ASSIMILATION

Tissa Illangasekare, Colorado School of Mines and Srikanta Mishra, INTERA

ABSTRACT

Models that simulate complex subsurface processes are used in decision-making and in regulatory programs. Model calibration for reliable predictions for large systems over long time horizons remains a major challenge. Key input parameters that are adjusted during calibration are spatially variable and depend on the measurement scale. Various site characterization techniques providing initial parameter estimates produce data types with different degrees of reliability and uncertainty. Thus, a need exists for the development of practical methods to account for disparity between scales at which parameters are measured and model inputs are defined. New computational algorithms for parameter up-scaling and assimilation of different data types are expected to produce significant advances in the calibration of subsurface flow and transport models.

EXECUTIVE SUMMARY

Models used for decision-making, site management, remediation, and performance assessment are based on computational algorithms selected to solve mathematical formulations of the complex physical, geo-chemical and biological processes that govern the fate and transport of hazardous chemicals and waste mixtures. To work, these algorithms require that the three-dimensional physical domain be discretized into computational sub-domains. During simulation, the algorithms use parameters defined for each of these computational sub-domains. These parameters, which may vary spatially and/or temporally, characterize the fundamental processes occurring in the soil and/or rock pores. The prediction accuracy and reliability of the simulation model is primarily determined by how accurately these parameters are represented in the model. Different characterization techniques are currently used to estimate these parameters at different measurement scales. The specific data collection methods have their inherent degrees of accuracy and uncertainty and, accordingly, are classified as “hard” or “soft” data. Model calibration involves the determination of these parameters in the model’s computational sub-domain scale. The primary goal of this PRD is to contribute towards the development and validation of two types of algorithms for improving the calibration and application of large field-scale models of subsurface systems. New algorithms are needed for 1) up-scaling of parameters from measurement scale to the computational sub-domain scale and 2) systematic assimilation of ‘hard” and “soft” data into the calibration process. The research needed to develop these proposed algorithms first involves basic research into the development and validation of methods for parameter up-scaling from measurement scale to the computational scale and assimilation of the two data types. The demands placed on the data for proper validation of these algorithms precludes the use of sparse spatially distributed data that is generally available at field sites. Hence, prior to the implementation of the algorithms at field

sites, the data need to be validated using experimental results generated in test systems that allow for the collection of more accurate spatially distributed and temporally varying data.

SUMMARY OF RESEARCH DIRECTION

The priority research direction presented in this PRD focuses on development, validation, and application of new algorithms for up-scaling parameters determined at different measurement scales to the discretized sub-domain scale of the model and use of different data types with varying degrees of reliability and uncertainty. The goal is to improve the calibration process for models that are used in decision-making and regulatory programs. In general, these models should have the capability to simulate complex subsurface processes that occur at DOE-managed sites contaminated with hazardous chemicals and waste mixtures. Data needed for proper validation of these algorithms are not generally available at field sites. Hence, prior to the implementation of the algorithms at field sites, validation using experimental data generated in pilot-scale test systems is needed.

Methods and Algorithms for Parameter Up-Scaling

The subsurface models used to address problems at DOE sites need to have the capability to capture a variety of physical, chemical, and bio-chemical processes. These processes include 1) single phase flow that defines how water flows in saturated zones of aquifers, 2) multi-phase flow that describes water flow in the unsaturated zone and behavior of chemicals that are in the form of non-aqueous phase liquids (NAPLs) in both unsaturated and saturated media, 3) transport of dissolved contaminants, 4) reactive processes (e.g., sorption, etc.), 5) decay processes (e.g., biological and radioactive), and 6) colloidal transport. The physical systems that are of interest include subsurface formations with varying types and geologic and hydro-geologic complexity (e.g., sedimentary, fractured media, aquifers in interaction with surface systems, single continuum, dual porosity, dual permeability, preferential flow paths, large-scale discontinuities such as faults, etc.). The basic processes that control fate and transport occur at the pore scale. However, it is not practical to observe and characterize these processes at these scales. The observations are generally made at much large scales either in the field or in the laboratory using soil samples extracted from the field. These observations are used then to estimate effective parameters that characterize the process at the scale of measurement. The geologic heterogeneity that is manifested at all scales makes these parameters variable in space. Measurements taken at a limited number of spatial locations are not adequate to determine the input parameters of the model that need to be defined for the discretized computational sub-domain of the model. This step of the PRD addresses the development of up-scaling algorithms to account for disparity between the scale at which a physical variable is measured and the scale at which a model is parameterized.

Algorithms for Soft Data Assimilation

Two primary approaches are used in model calibration. The commonly used "trial-and-error" approach is an exercise in the adjustment of system properties until a reasonable "match" is obtained between model predictions and field observations. A sensitivity analysis is performed to gage uncertainty once the "best" estimate of model parameters is determined through calibration. However, the uncertainty associated with such estimates is rarely quantified. Furthermore, trial and error procedures become impractical or infeasible in problems involving the application of complex codes (e.g., unstructured and adaptive meshes) for multi-physics problems (e.g., flow, transport, thermal, reactions, etc.). The second approach to calibration, referred to as "inverse modeling," is a formal attempt to automate the model calibration process using optimization and regression methods. Data collected during site characterization are classified into "hard" and "soft" data types in the groundwater modeling literature. Poeter and McKenna (1995) refer to "hard data" as data with negligible uncertainty and "soft data" are information with non-negligible uncertainty, such as indirect measurements gathered in geophysical surveys and expert opinion regarding geologic fabric or structure. The *ad hoc* nature of trial-and-error calibration allows for the use of both "hard" and "soft" data during model calibration. On the other hand, "inverse modeling" algorithms such PEST (Doherty 2000), UCODE (Poeter and Hill 1998) and the PEST package in MODFLOW-2000 (Hill et al. 2000), use "hard" data. The goal of this task is to develop algorithms to use both "hard" and "soft" data in a systematic way during model calibration.

Validation of Methods and Algorithms

The development of computational algorithms discussed earlier needs to be followed by a validation phase before application to field problems. These algorithms address issues related to accuracy and uncertainty of input data and the corresponding uncertainty and reliability of model predictions. Hence, proper validation requires a demonstration of the versatility of the algorithms to handle a variety of problems involving subsurface processes in geologic and hydro-geologic settings and demonstrating that the methods help in reducing the prediction uncertainty and error. Primary consideration to be given to achieve the goals of task is the availability of accurate data on both system properties and observations. The uncertainty and precision of these data have to be accurately known. Data obtained from typical field sites or even at highly characterized sites are not adequate to gain a good understanding of the state of data uncertainty and accuracy. Adequate control does not exist to conduct field tests to make observations at all necessary points in the aquifer. Hence, the only viable way to obtain data for validation is through controlled pilot-scale experiments in intermediate-scale test tanks or pilot-scale field experiments at well characterized research sites (e.g., Borden, MADE) or at sites that need to be developed and managed by DOE.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Numerical models that simulate the fate and transport of chemicals in the subsurface are widely accepted as tools for decision-making and in regulatory programs. The modeling and simulation community has made great strides in the last 10 years in simplifying and automating the construction and execution of numerical models to support decision makers. The remaining bottlenecks in model application are model calibration and the assignment of uncertainty to model results. As stated by Hill (1998), "…in model calibration, various parts of the model, including the value of model input values, are changed so that the measured values (often called observations) are matched by equivalent simulated values, and, hopefully, the resulting model accurately represents important aspects of the actual system." In modeling aquifers, parameters such as hydraulic conductivity, effective porosity and dispersivity are spatially distributed. In some cases, parameters also may exhibit temporal variability (e.g., mass transfer coefficients, decay coefficients, etc.). In these cases, the number of parameter values can be infinite; however, numerical models that discretize the problem domain require parameter values at a finite number of calculation points (i.e., cells, nodes, or elements) as inputs. In problems involving large physical systems, the number of observations generally are limited and, therefore, are able to support the estimation of relatively few model input values. Before a simulation model may be applied at a site, the model parameters must be tailored to that site through a calibration process. Regulatory agencies now require that the model results be accompanied by detailed risk assessments that require the assignment of certainty to the answers.

Before the implementation of an up-scaling methodology in computational algorithms, fundamental questions about whether and how these parameters are related at different scales must be answered. Any additional tests needed to gather information at larger scales to up-scale the parameters would have to be determined. If the existing characterization methods are found to be inadequate, new methods have to be developed. As measurements are made from limited observations, uncertainty is inherent in the data. In addition to this measurement uncertainty, the-up-scaling process itself will introduce uncertainty to model parameters, resulting in model-prediction uncertainty. Research is needed to quantify these uncertainties and strategies for minimizing them through optimization of data collection (e.g., locations, amounts, and frequency). Another challenging up-scaling issue that needs attention involves determining when variables, such as observations of head made in wells and concentration data collected at monitoring wells (which are point observations), have to be up-scaled to the model sub-domain scale for calibration. Existing methods used to up-scale these point observations are empirical and lack scientific rigor.

Hard data for aquifer characterization is obtained through traditional pump tests and tracer tests. The dependence of accurate flow and transport simulation on the inclusion of heterogeneity in soil properties is well established through both theoretical analysis and numerical modeling. However, field sampling is always too sparse to support detailed characterization of subsurface materials. On the other hand, indirect measurements based on geophysical surveys and expert opinion regarding geologic fabric or structure (i.e., soft data) help constrain and refine model

parameters at the scale where simulations are conducted (e.g., the plume or regional scales). The proposed algorithms should be designed to conduct this assimilation of the two data types in a more systematic way. In more complex modeling problems involving complex processes such as multi-phase flow, constitutive models for relative permeability, capillary pressure, etc., cannot be determined for the modeling scale using hard data. This suggests the need for improved constitutive models between typical soft data sources and corresponding model parameters. As part of the development of algorithms for data assimilation, methods also need to be developed to quantify uncertainty in model parameters resulting from the incorporation of "soft" data.

The implementation of algorithms for field applications poses many computational challenges because of the dimensionality of the problem (three-dimensional space and time), the complex processes that have be simulated, and the large amount of data that have to be used. All algorithms will require computationally intensive numerical strategies. Incorporation of stochastically based approaches (e.g., geo-statistical methods to define heterogeneities) to the algorithms and site characterization will require new strategies. With the advances made in data collection and monitoring based on wireless sensor networking through which system behavior could be monitored dynamically, additional computational challenges will arise to use joint inversion methods for dynamic model calibration.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

When the proposed algorithms are fully developed and functional, they are expected to make major advances in the way subsurface fate and transport models are calibrated. The computational science community has been quite successful in developing innovative numerical algorithms to solve the governing partial differential equations some of which are highly non-linear because of complex constitutive relations. With the phenomenal advances made in computer technologies that allow fast computing, higher precision in computational accuracy has been achieved with the use of finer grid resolution and smaller computational time steps. However, these improvements mean less when uncertainties in the data used to develop the model lead to questions about the accuracy of the models to predict long-term behavior. It is our expectation that a new paradigm in model calibration will result from this research. This new paradigm will allow computational scientists to focus on development of not only computationally accurate models, but also models that are reliable in their predictions. Prediction algorithms with seamless transition from calibration to prediction may be developed. This will require innovation as algorithms for data assimilation, up-scaling, inversion, and forward simulation individually may be computing intensive when dealing with large physical systems that require long-term predictions (e.g., performance assessment at sites with radioactive waste). Effective coupling for practical problem solutions requires innovation.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

This research will lead to better approaches for site characterization and model calibration at all scales. Many advances have been made in understanding the basic science issues associated

with how chemicals interact in the soil-water environment. From a modeling point of view, the practical goal of this research is to characterize these processes and parameterize them to be used as model inputs. With all the advances made in model development, the time has come for researchers working at these smaller scales (i.e., the core and laboratory scales) to appreciate and understand the value of information generated at their test scales and how they are used in models. The up-scaling methods that are researched will allow the researchers to develop better characterization methods that will add value in a significant way to make better model predictions.

Another field where advances can be made in subsurface science as a result of this research is in the area of shallow surface geophysics. The algorithms for data assimilation will use characterization data generated using geophysical techniques. Potential new research areas will develop where geophysical characterization methods could be developed specifically for data assimilation with the goal of model calibration.

TIME FRAME

The research tasks on development of algorithms and validation need to be conducted concurrently. As the foundations for the research are very well established in relevant basic sciences, computational science, and engineering, the goals as stated could be accomplished in 3 to 4 years. The development of the experimental test systems, building on existing systems, will take 2 to 3 years.

REFERENCES

Doherty, J. 2000. *PEST: Model-Independent Parameter Estimation*. Watermark Numerical Computing.

Hill, M.C. 1998. *Methods and Guidelines for Effective Model Calibration.* USGS Water-Resources Investigations Report 98-4005, 90pp.

Poeter, E.P and S.A. McKenna. 1995. "Reducing Uncertainty Associated with Groundwater Flow and Transport Predictions." *Groundwater* 33(6):899-904.

Poeter, E.P. and M.C. Hill. 1998. *Documentation of UCODE, A Computer Code for Universal Inverse Modeling*. U.S. Geological Survey Water-Resources Investigations Report 98-4080, 116p.

CYBER-INFRASTRUCTURE FOR SUBSURFACE MODELING

Roelof Versteeg, Idaho National Laboratory

ABSTRACT

The currently existing cyber-infrastructure for subsurface modeling is inadequate to support and allow for effective development, validation, and use of novel multi-physics and multi-scale codes. To be successful, such codes will need to 1) efficiently integrate solvers, codes, and code components developed by diverse groups; 2) take advantage of current and future computational infrastructures; and 3) allow for rigorous, auditable testing and validation against diverse datasets. An effective cyber-infrastructure for subsurface modeling will need to both use and integrate insights developed in cyber-infrastructure design by other agencies and scientific discipline, and allow for the effective addressing of challenges unique to subsurface science.

EXECUTIVE SUMMARY

Novel multi-physics and multi-scale codes are needed to adequately model complex coupled processes in the subsurface. While the exact design, components, and implementation of these codes is unknown, in order for these codes to meet their objectives they will need to:

- Couple and integrate diverse code elements developed by different groups
- Integrate massive amounts of diverse data from diverse sources (including dynamic sensor networks)
- Take advantage of future developments and enhancements in computational infrastructure.

In addition, such codes will need to pass rigorous, auditable verification and validation tests both in the development and implementation phases.

A priority research direction is the development and implementation of a rigorous and powerful cyber-infrastructure for subsurface modeling. The development of this cyber-infrastructure will build on existing U.S. (NSF 2006) and international efforts focused on development of standards and tools for scientific workflows, sensor web-services (OGC 2006), self validating models, service discovery, distributed computing, and code adaptivity, while at the same time accommodating the unique aspects and attributes associated with the subsurface modeling problem.

This cyber-infrastructure will provide a presently unavailable but critically needed framework for the efficient development, testing, and field deployment of novel multi-physics and multi-scale codes. This ability will permit the effective use of novel codes for long term monitoring, management and optimization of CO_2 sequestration sites, and the cost-effective mitigation of subsurface contaminants.

SUMMARY OF RESEARCH DIRECTION

This PRD is focused on designing and implementing the cyber-infrastructure required to support and allow for effective development, validation, and use of novel multi-physics and multi-scale codes for subsurface modeling. Specifically, such a cyber-infrastructure should provide for and integrate the following elements

- Common code libraries, tools, and models
- Well-described, publicly accessible synthetic, laboratory, and field datasets that can be used for code testing and validation
- Standards for subsurface data and model description, similar to these developed by the Open Geospatial Consortium (OGC)
- A structure allowing for transparent, reproducible, and auditable code execution
- Enabling technologies such as web-based visualization and well-defined authorization and authentication systems that will allow for security in a distributed environment.

Common code libraries, tools and models

The ability to accurately and efficiently model subsurface phenomena will increasingly require the coupling of diverse numerical codes such as geophysical imaging codes, reactive transport codes, and codes describing physical phenomena at scales ranging from pore to macroscopic. Because of the range of expertise required to develop such codes, it is likely that these codes will be developed by geographically separated organizations. At the same time, many of these codes will require common components (such as function optimizers; automatic adjoint calculators; fast, sparse matrix inverters). Thus, a cyber-infrastructure that provides both for repositories for parts of such codes, as well as tools allowing the efficient coupling of such codes through a range of different structures, will be a necessity. The cyber-infrastructure PRD should provide for a framework for code coupling and for the development of necessary middleware (such as high-level standards for code functionality description)

Well described, publicly accessible synthetic, laboratory and field datasets which can be used for code testing and validation

As codes attempting to accurately model subsurface phenomena become increasingly complex, it will be necessary to thoroughly test and validate such codes at many levels. For simple models, one can do this by comparing model output to analytical solutions; however, for models that attempt to model the effect of chemical reactions at the pore scale level on macroscopic geophysical signatures, there are no analytical solutions. Thus, an essential part of a cyber-infrastructure is a set of extremely well-described, publicly accessible synthetic, laboratory, and field datasets which can be used for code testing and validation. An analog to this can be found in the seismic oil industry for which several of these datasets have been created and extensively

used over the years. Such a series of datasets (the specifics of which should be designed in a discussion between modelers and experimentalists) is an essential part of a cyber-infrastructure.

Standards for subsurface data and model description

One of the complicating factors in subsurface modeling is the complete lack of standards for model and data description. This effectively makes comparison of code performance between similar codes (e.g., reactive transport models) extremely complex. While it can be done (at large costs) for a single case, it is not feasible to routinely compare model performance on diverse cases. While some codes have intrinsically different abilities, this can to a large part be alleviated through the definition and acceptance of a suite of standards on subsurface data and model description. This could include the acceptance or participation in the development of XML schemas as developed by the OGC. It could also include the implementation of modeling codes or components of such codes as web services.

A structure allowing for transparent, reproducible and auditable code execution

The validation of the performance of reactive transport models has recently become a major issue at several sites for which DOE scientists have performed modeling efforts. To a large extent, the problems associated with this validation result not from the actual results of the code (or the underlying science), but from the ability to provide an audit trail associated with the code or, alternatively, the challenge of performing reproducible research. Current approaches are antiquated and inefficient. What is needed is a structure that allows for transparent, reproducible, and auditable code execution, and eventually (at a larger scale) reproducible research (Schwab et al. 2000). This poses many interesting computational challenges such as the development of temporal databases and achieving reproducibility, including reproducibility with distributed code components (which could be potentially done through scientific workflow languages).

Enabling technologies

In addition to the cyber-infrastructure described above, this PRD should consider the need to integrate a host of enabling technologies, including web-based visualization technologies and well-defined authorization and authentication systems allowing for security in a distributed environment. Some of these technologies are being developed by either other U.S. government agencies (e.g., the NASA WorldWinds product, which is developed with DOE support) or private industry. Other technologies specific to DOE subsurface modeling needs may need to be actively developed by DOE. Such technologies might include the development of subsurface sensors that can be accessed by numerical models through web-service calls in now-casting or forecasting applications.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The cyber-infrastructure developed under this PRD will allow the computational science community to successfully develop and test their new models.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

Meeting this PRD is a prerequisite to successful subsurface modeling efforts. If this PRD is met with usable codes, then the subsurface sciences community will be able to successfully and confidently use the advanced and novel models developed over the next several years.

TIME FRAME

The challenges described here will need several years to meet. It will take code developers at least 4 years to bring their code into compliance with standards and to effectively use public data for code validation. However, many aspects of this PRD can be achieved in 1 to 2 years.

REFERENCES

NSF (2006). NSF'S Cyberinfrastructure vision for 21st Century Discovery, Version 7.1, National Science Foundation.

OGC (2006). OGC White Paper: OGC Sensor Web Enablement: Overview and High Level Architecture, Open Geospatial Consortium.

Schwab, M., N. Karrenbach and J. Claerbout (2000). "Making scientific computations reproducible." Computing in Science & Engineering **2**(6): 61-67.

PREDICTION-BASED INVERSE MODEL CALIBRATION

Characterization and Calibration Panel, Lead author: Al Aziz Eddebbarh, John Doherty; Scott James, Los Alamos National Laboratory;

ABSTRACT

Models that simulate complex subsurface processes are used in decision-making and in regulatory programs. Calibration of large systems models developed to yield reliable predictions over long time remains a major challenge. Traditional model calibration is conducted ensuring that all important processes are incorporated while parameters important to the model and its predictions may be ignored through lumping. As a result, pertinent data important to prediction may not be identified and collected. In addition, uncertainty analysis can help assess the uncertainty in numerical model outputs that arises from the modeling approximation of the real world and lack of precision in model structures, parameters, boundary conditions, and data manipulation. Thus, a need exists for the development and implementation of computational tools for global optimization during calibration, and uncertainty assessment. Because such optimization and uncertainty analyses require a large number of direct simulations, high-performance computing resources and advances in algorithms for direct and inverse solutions, error propagation, and multi-objective design optimization schemes are needed. The process will provide a model with a set of parameters that allows it to make predictions with minimized potential error and the ability to quantify the extent of this potential error.

EXECUTIVE SUMMARY

The goal of this PRD is to develop a model optimization methodology that focuses not only on important processes but also on the parameters most important to model predictions with minimized potential error and to quantify the magnitude of this potential error. Applications of interest include soil and groundwater remediation, nuclear and other waste disposal systems, CO_2 sequestration, and unconventional hydrocarbon recovery.

SUMMARY OF RESEARCH DIRECTION

This PRD is focused on developing optimization tools for model calibration that provide optimal parameter values yielding predictions with minimized potential error and the ability to quantify the extent of this potential error. The optimization tools, coupled with generalized error prediction analysis methodologies for performance predictions, will help extract maximum information from existing data. Specific relevant applications may include but are not limited to groundwater development and remediation problems; the proposed geologic repository at Yucca Mountain, Nevada; CO_2 sequestration projects; production of unconventional hydrocarbon resources; etc.

PROBLEM OVERVIEW

Models are calibrated so that they make better predictions than if they were not calibrated. Unfortunately, calibrated model predictions can still be wrong. Furthermore, it is now being fully understood that a calibrated model can make even worse predictions than it did before calibration. With traditional approaches to model calibration, there is no way to determine 1) if a calibrated model's predictions are better than those before calibration; 2) if the predictions are determined to be better, how much better; and 3) if their predictions are determined to be wrong, then how wrong. Traditional approaches to calibration are not able to ensure that calibrated models minimize "potential predictive wrongness" while quantifying the remaining uncertainty in the potential predictive wrongness.

The traditional approach to model calibration follows the tenet of the "principle of parsimony" espoused in many modeling texts and guidelines. First, the dimensionality of the calibration problem is reduced to facilitate a tractable model (i.e., fewer parameters are used to ensure their unique estimability) given the dataset available for calibration. The parameter values are then estimated through implicitly or explicitly maximizing some goodness-of-fit criterion. When the fit is judged "sufficient" (usually through minimization of an objective function), the model is deemed "calibrated" and, therefore, suitable for making predictions—predictions that may lay the groundwork for performance assessment calculations. If automatic parameter estimation software is used in the calibration process, some estimates of parameter uncertainty are available. Estimates of the uncertainty of key model predictions can then be made based on the predictions' dependence on the estimated parameters and their uncertainties.

This PRD is proposed to resolve situations in which the traditional approach to model calibration and predictive error analysis based on the parsimony principle is not effective or accurate. Instead, better ways to calibrate a model and to explore predictions errors based on the theory of mathematical regularization are proposed. Some of these theories have being used by many other branches of science in which the analysis of costly and important data demands that maximum information be extracted (e.g., medical imaging). For example, a kidney is not defined prior to processing the data contained within a medical image; instead, the location of the kidney "emerges" as a natural part of the data interpretation process. The same process should be used in subsurface model interpretation. This PRD asserts that there is no alternative but for the groundwater industry to cross the same threshold that has been crossed in other industries, and adopt application of regularized inversion as a methodology for model calibration and uncertainty analysis.

Traditional Model Calibration. A model's complexity is commensurate with the predictions it makes. No processes salient to those predictions will be omitted if the integrity of those predictions would be eroded by their omission. The same thing could be said about parameters. No parameter salient to model prediction should be dropped or lumped into other parameters.

Clearly, ignoring system heterogeneity is a forgone conclusion even if during model calibration it may be justified on the basis that parameters pertaining to this heterogeneity cannot be estimated. Unfortunately, it is not equally justifiable that heterogeneity be ignored when the model is used to make a prediction. If a prediction is sensitive to actual system heterogeneity—heterogeneity that, of necessity, "falls between the cracks" of the calibration—then that prediction may be seriously in error, despite the fact that the model may be "well calibrated" (i.e., a good fit between model outputs and field observations was obtained on the basis of a parameter simple enough to allow unique estimation of parameters, but which possesses just enough complexity to obtain this good fit).

If a model is to be "calibrated," then its parameterization must be simplified so that the "inverse problem" (calibration) of assigning values to model parameters on the basis of an (often limited) calibration dataset has a unique solution. Traditionally, such simplification is undertaken prior to model calibration through adoption of a simplified parameter scheme (e.g., an effective permeability applied to an entire geologic unit). However, other approaches to calibration are needed so that simplifications undertaken during the calibration process itself are far less subjective. Nevertheless, the fact remains that pursuit of the "calibrated model" requires sufficient parameterization simplification to yield a (simplification-dependent) unique set of estimated parameters. Because it is impossible to infer parameterization detail to the same level that hydraulic property detail really exists, the cost of obtaining a calibrated model is therefore a parameter field that must be locally in error, even if it is roughly correct in an average sense.

Parameter error leads to predictive error. Furthermore, to the extent that a prediction depends on hydraulic property heterogeneity that "falls between the cracks" of the calibration process, the potential magnitude of its error grows. In general, predictions different from those comprising the calibration dataset (e.g., hydraulic head calibration, specific discharge data and predictions) are more likely to be in error because they may depend on hydraulic properties that were fixed or grossly averaged over the model domain because of a paucity of information on these properties within the calibration dataset. Unfortunately, this introduces a contradiction because the reason for employing a complex physically based model in the first place is because predictions of just these types (different from the dataset) need to be made (otherwise the prediction would simply be directly measured at the site).

Regularization. "Regularization" is a word that mathematicians use to describe the parameter simplification process necessary to achieve a unique solution to an inverse problem (such as model calibration). In general, with fewer available data, more regularization must be undertaken (and hence a greater degree of parameter simplification). Regularization can be implemented using manual parsimonizing methods such as zonal definitions. Parameter parsimony can also be implemented mathematically such that it is optimized to the calibration dataset and, hence, extracts maximum information as possible from that dataset.

The difference between the regularized inversion method and the traditional approaches is that the former are designed for the estimation of many parameters (possibly numbering in the

hundreds or even thousands) rather than just a few. Thus, we introduce to the model domain a parameterization density that is commensurate with whatever hydro-geological or process complexity that it is necessary for model prediction accuracy. It should be noted that this does not eliminate parameter error and hence model predictive error because parameter simplification in one form or another is an unavoidable precursor to model calibration. However, it does provide the ability to quantify potential parameter and predictive error. Parameter complexity is not sustainable in a model because of inherent limitations in the calibration dataset; therefore, this complexity can be readily reintroduced where predictive "wiggle room" is tested as part of a predictive error analysis procedure.

Parameters whose values are estimated through regularized inversion can be defined on the basis of a large number of small zones of piecewise constancy. This can be accomplished through devices such as pilot points, through local or global basis functions, through combinations of these, or using other methodologies. The point is that if potential variability of hydraulic properties over a certain area is relevant to a prediction, such variability should be recognized in the model's parameterization (and thus included in the calibration and subsequent predictive error analysis). Inclusion in the calibration process ensures that maximum information is extracted from a calibration dataset; inclusion in the predictive error analysis ensures that the level of potential error associated with important model predictions is quantified. To the extent that simplification is required to achieve a unique solution to the inverse problem, mathematical regularization ensures that the calibration dataset is used optimally. Essentially, this produces "smoothed" or "blurred" parameter fields that are no "smoother" and no more "blurred" than necessary. To the extent that a prediction depends on hydraulic property detail that cannot be represented in these smoothed fields, the effect of smoothing on potential predictive error is quantified. Meanwhile, agonizing decisions such as how to supplement, reduce, or adjust an often artificial rectilinear zonation scheme do not need to be made, thus making calibration a far less subjective process.

Some may complain that the use of so many parameters may lead to "over-fitting" to a calibration dataset, pointing out that a close fit between model outputs and historical measurements can indeed be obtained when many parameters are estimated, but that predictive error may be consequently increased. This is easily avoided because regularized inversion allows the modeler to vary the extent to which improved model-to-measurement fit is traded against the potential for model predictive error. Because the potential for such error can now be quantified, it can also be minimized once the level of measurement noise and the level of geological heterogeneity are estimated.

Two broad approaches to regularized inversion have been applied to subsurface model calibration: 1) the "Tikhonov" methodology and 2) the "subspace" methodology. Each has its advantages and disadvantages; however, certain hybrid schemes are able to combine the strengths of both of these without compromising computational efficiency. Complex models with long run times can be assigned thousands of parameters while their calibration can be achieved within a number of model runs less than twice the number of parameters actually used

in the model. Linear and nonlinear predictive error analysis can then be undertaken with similar computational efficiency.

Tikhnov Regularization. The Tikhonov regularization is implemented by re-formulating the inverse problem of model calibration as a constrained minimization problem. First, a "preferred condition" is defined for all parameters used in the model. This can comprise preferred values for these parameters or preferred relationships between them (e.g., estimated or measured hydraulic property homogeneity). A set of parameter values is sought that achieves a certain (user-specified) level of model-to-measurement fit; this level of fit is set in accord with expected levels of measurement noise. Uniqueness is achieved by finding values for parameters that achieve this fit with minimal departure from the preferred parameter condition. If preferred, parameter conditions are sensibly defined on the basis of site characterization studies, a realistic set of parameters is thereby achieved.

Subspace Method. The use of subspace methods recognizes the fact that most calibration datasets are best equipped to provide unique estimates of combinations of parameters and not individual parameters. Mathematical tools (e.g., singular value decomposition) determine what these combinations are and how many such combinations are estimable while inestimable parameter combinations retain their original values. By working with parameter combinations rather than individual parameters (e.g., combinations that are orthogonal in parameter space), the dimensionality of the "calibration solution space" (i.e., the number of parameter combinations that are actually estimated) can be optimized in accord with the level of measurement noise. That is, these combinations are assembled to provide optimal "receptacles" for the information content of the calibration dataset. If initial parameter estimates provided to the inversion process are based upon site characterization studies, then the fact that parameter combinations comprising the inestimable "calibration null space" (which is orthogonal to the calibration solution space) remain unchanged during calibration ensures reasonable parameter values in the calibrated model. The optimal dimensionality of the calibration solution and null spaces depends on the level of model-to-measurement fit desired, which should be set in accordance with measurement noise.

Combined Method. This scheme combines the strengths of both of these methods and reduces the number of effective parameters through pre-definition of estimable parameter combinations (using singular value decomposition or some related methodology), and through maintenance of these combinations through the calibration process as "super parameters." The number of effective parameters is reduced to one comprising the optimal dimensionality of the calibration solution space. The number of model runs required per calibration iteration is, thereby, reduced to the number of super parameters employed in the inversion process, as derivatives are computed with respect to these super parameters rather than with respect to individual model parameters. The model run efficiency of the calibration process may be increased enormously.

Regardless of the regularization method used, the advantages of developing many parameters to characterize hydraulic property complexity and heterogeneity over a model domain, rather than

just a few parameters, are considerable. The parameter estimation process is free to be maximally responsive to the calibration dataset, introducing heterogeneity to estimated spatial parameter fields where the data suggest that such heterogeneity exists, or producing smooth or uniform parameter fields where there are no data to suggest otherwise. Thus, heterogeneity exists within the calibrated model "where it has to exist" because regularized inversion will introduce heterogeneity only where it is stipulated by the data. However, because model representation of heterogeneity may be considerably smoothed compared to what actually exists, parameter and predictive error may still abound.

Predictive error quantification. Because hydraulic properties in any real-world system are much more complex and heterogeneous than the calibrated model parameter fields that represent them, model parameters cannot help but be locally in error. So, too, will be many model predictions, particularly those that depend on hydraulic property detail. Thus, an unavoidable consequence of building and calibrating a model is the introduction of parameter and predictive error with most of this error arising from differences between model and real-world property fields. These differences represent the hydraulic property detail that "slips between the cracks" of the calibration process.

Where the number of parameters used in a model is commensurate with potential hydraulic property complexity, predictive error can be quantified. Through site characterization studies, or simply through geological insight, information will always be available on the range of hydraulic properties that may exist within a study site or region, and on the degree of spatial correlation that these properties may show. Sometimes this information may be encapsulated in geostatistical descriptors such as a variogram. Regardless, reasonable estimates of hydraulic property variability can always be made; after all, a geologist will quickly identify aspects of model parameterization that seem unbelievable. These ideas can be approximately encapsulated in a spatial covariance matrix of hydraulic properties, which provides both a brief statistical summary of the innate variably of hydraulic properties in a study area and the likely continuity of these properties. Often such a matrix can be built easily for most sites. Its approximate nature does not matter because approximation infers uncertainty. High uncertainty infers potentially high hydraulic property variance, which is justifiably translated into potentially high levels of model predictive uncertainty if hydraulic property details are inestimable through the calibration process and predictions of interest are sensitive to them. Using basic matrix manipulation methods, such a probabilistic description of the "heterogeneity that may exist" within the subsurface, when compared with the "heterogeneity which must exist" as represented by the calibrated model parameter field, allows a probabilistic description of model predictive error to be developed, based on the difference between the two.

Subspace regularization methods provide particularly useful insights into the sources of model predictive error. Certain combinations of parameters are estimable through the calibration process; however, these estimates are contaminated by measurement noise in the calibration dataset. This is one source of potential predictive error. The other source of error arises from inestimable parameter combinations comprising the calibration null space. Thus, to the extent

that a prediction depends on the null space (orthogonal) combinations of parameters, its potential error is in no way decreased during the calibration process. The potential wrongness of model predictions that depends on these parameter combinations is thus a function of the innate variability of system hydraulic properties described by the user-supplied covariance matrix. Total predictive error could be computed by combining this term (null space error) with the measurement noise term.

Monte Carlo Analysis. "Calibration-constrained Monte Carlo" analysis could be implemented as another adjunct to regularized inversion. Information forthcoming from the regularized inversion process facilitates generation of stochastic parameter fields that minimally affect the calibrated status of a model. By adding these parameter variations to the calibrated parameter field (using pre-calculated sensitivities to eliminate the need for extra model runs and by correcting for model-to-measurement misfit incurred by model nonlinearity through minor adjustment of parameter combinations comprising the calibration solution space), a suite of parameter fields that calibrate the model while encompassing the innate complexity of hydraulic property reality could be generated. Model predictions made with all fields span the variance of that prediction.

Advantage of this Research. This PRD outlines advantages that could be achieved by using regularized inversion for calibration based on using a small number of parameters in accordance with the "principle of parsimony." In fact, all calibration requires parameter simplification and parsimony. However, where regularization is undertaken by mathematical means, such simplification is optimally tuned to the calibration dataset, thus extracting maximum information from that dataset. This results in a calibrated parameter fields that are indeed a simplified or smoothed version of reality, but are no simpler and no smoother than necessary. Furthermore, the difference between the heterogeneity that must exist to explain the data, and that which may exist in accordance with geological considerations, is explicitly accommodated in a predictive error analysis process during the regularized calibration process.

Model parameterization could be a far simpler matter, based on the tenet that "…if it may affect the prediction, then include it as a parameter" (the same rule that applies to processes simulated by the model). Good fits between model outputs and field data can be achieved on the basis of aesthetically pleasing and geologically reasonable parameter fields, unencumbered by artificialities (such as geologically unsupported rectilinear zones arbitrarily emplaced at locations) contrived to lead to a better fit between model outputs and field measurements. Overall, a modeler will be satisfied that data have been treated with respect, and endowed with a worth equal to its cost because maximum information is extracted from it to make predictions whose potential wrongness is minimized.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

By incorporating all pertinent processes and parameters into model calibration, the proposed research will take advantage of the phenomenal advances made in the computer technologies with fast computing, higher precision in computational accuracy, and the use of finer grid resolution and smaller computational time steps. However, all these improvements are useless if model predictions are wrong and prediction accuracy and uncertainty cannot be assessed based on uncertainties in the data used in developing the model. It is our expectation that a new standard in model calibration will result from this research, allowing the development of not only computationally accurate models but also models that are reliable in their predictions. This will complement research in algorithms for data assimilation and up-scaling of parameters and integration of characterization plans, long term monitoring.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

This research will help lead to better approaches for model calibration and model prediction error assessment. This approach will build on the many advances in subsurface and computational sciences, and the parallel efforts to characterize and parameterize processes to be used as model input. This will advance subsurface models ability to improve decision-making and minimize costs by extracting maximum information from site characterization data.

TIME FRAME

Significant progress could be achieved in 1 to 2 years and full attainment of the objectives is anticipated within 5 years.

SCALABLE PARALLEL ALGORITHMS FOR INVERSE PROBLEMS IN SUBSURFACE MODELING

Characterization and Calibration Panel

Lead author: Omar Ghattas, University of Texas at Austin

With input from Roscoe Bartlett, Sandia National Laboratories ; Lorenz Biegler, Carnegie Mellon University, George Biros, University of Pennsylvania; Eldad Haber, Emory University Matthias Heinkenschloss, Rice University; David Keyes, Columbia University; Bart van Bloemen Waanders, Sandia National Laboratories

ABSTRACT

Estimating uncertain parameters in subsurface simulation models from sparse surface and subsurface measurements is an extremely challenging inverse problem. This PRD describes challenges and identifies opportunities for the development of efficient inverse algorithms that can scale to the next generation of petascale computing systems. There is an urgent need to create inverse algorithms that are as efficient and scalable as the corresponding algorithms for forward simulations. Since the structure of inverse operators is usually very different from those of forward operators, entirely new classes of scalable numerical algorithms are needed.

EXECUTIVE SUMMARY

This PRD addresses the challenges of developing scalable parallel algorithms for estimating uncertain parameters in subsurface simulation models from sparse surface and subsurface measurements. There is an urgent need to create inverse algorithms that are as efficient and scalable as the corresponding algorithms for forward simulations. Because the structure of inverse operators is usually very different from those of forward operators, entirely new classes of scalable numerical algorithms are needed.

Large-scale, simulation-based inverse problems are significantly more difficult to solve than the corresponding forward problem for the following reasons:

- The inverse problem is usually ill-posed, even when the forward problem is well-posed.
- When the forward problem is an evolution equation in space, the inverse problem is a boundary-value problem in space-time.
- New infinite-dimensional operators that are not present in the forward problem appear in the inverse problem, and these require operator-specific regularization, iterative solvers, preconditioning, globalization, inexactness, and parallel implementation.

- There is often a need to not only estimate the most likely parameters from among families of parameters consistent with the observations, but to also characterize the uncertainties in their estimates.

This PRD identifies new directions for research in scalable parallel numerical algorithms for inverse problems that will elevate inverse algorithms to a level of scalability and efficiency on par with associated forward solvers. These inverse algorithms must respect the infinite-dimensional character of the underlying inverse operators, make use of Hessian information, capitalize on adjoint information, exploit multi-level ideas, and parallelize in a fine-grained manner.

SUMMARY OF RESEARCH DIRECTION

This priority research direction addresses the challenges of developing scalable parallel algorithms for estimating uncertain parameters in subsurface simulation models from sparse surface and subsurface measurements. The need to solve such *inverse problems* to calibrate model parameters and estimate current system state has been identified in other PRDs, whereas this PRD describes challenges and identifies opportunities for the development of efficient inverse algorithms that can scale to the next generation of petascale computing systems. This PRD is motivated by the urgent need to extend to the inverse setting the successes engendered over the past decade by scalable parallel algorithms for *forward* (also known as *direct*) simulations. Because the structure of inverse operators is usually very different from those of forward operators (e.g., forward operators are often differential, while inverse operators are compact), entirely new classes of scalable numerical algorithms are needed.

While inverse problems in subsurface modeling are characterized by a wide range of inversion parameters (e.g., flow and transport constitutive parameters, initial conditions, boundary conditions, reaction and other model coefficients, source terms, geometry and topography, etc.) and simulation equations (e.g., describing multiphase reacting flow, transport, geo-mechanics, and geophysics), they share a common mathematical structure, which is an optimization problem with an objective function that represents the data misfit between observations and predictions, constraints in the form of large-scale simulation equations (typically discretized partial differential equations [PDEs] and partial differential differential equations [ODEs]), and inequality constraints on the parameter bounds.

For the following reasons, large-scale, simulation-based inverse problems are significantly more difficult to solve than the corresponding forward problem:

- An inverse solution typically requires numerous forward solutions and, thus, can be intractable when the solution takes weeks on multi-thousand processor systems.
- The inverse problem is usually ill-posed, even when the forward problem is well-posed.

- When the forward problem is an evolution equation in space, the inverse problem is a boundary-value problem in space-time. Thus, inverse operators are usually non-causal and non-local, despite the local and causal nature of forward operators.
- New infinite-dimensional operators appear in the inverse problem that are not present in the forward problem—adjoints, Hessians, and Karush-Kuhn-Tucker (KKT) operators—and these require operator-specific regularization, iterative solvers, preconditioning, globalization, inexactness, and parallel implementation, which are new frontiers for research in numerical algorithms.
- Non-uniqueness in inverse problems is manifested via families of possible parameter values that are consistent with the observations. There is often a need to not only estimate the most likely parameters from among the families (via regularization or prior model criteria) but also to characterize the uncertainties in their estimates, which stem from uncertainties in the observations, the models, and the priors.

Generic optimization or nonlinear least squares software packages are incapable of exploiting the structure of the embedded forward or inverse operators or their underlying infinite-dimensional nature and, thus, are inadequate for problems with large numbers of inversion parameters. As the underlying forward simulations embrace the multi-teraflops era and move toward the petascale, the chasm between the capabilities of available inversion algorithms and software, and the needs of scientific codes, grows even wider.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

In this section, we describe challenges and opportunities for the development of new scalable inverse algorithms. These new algorithms must respond to the following difficulties that characterize subsurface inverse problems:

- The forward simulations embedded within inversion are typically characterized by complex, nonlinear, multi-scale, multi-rate, multi-physics, and biogeochemical processes that require terascale and ultimately petascale computing resources.
- The inversion variables often stem from infinite-dimensional spaces (i.e., they represent initial conditions, boundary conditions, heterogeneous material properties, distributed sources, or geometry).
- Infinite-dimensional inequality bound constraints are required in the problem formulation (such as positivity constraints on heterogeneous parameter fields).

To address these issues, next-generation inverse algorithms are expected to:

- *Respect the infinite-dimensional character of the underlying operators,* because petascale inverse problems in subsurface modeling involve PDEs, and because algorithms that respect the infinite-dimensional structure of the PDEs can often exploit mesh independence to find solutions in a small number of iterations, even when the dimension of the parameter space is

of the order of millions—problems that are impossible for off-the-shelf, algebraic-oriented optimization packages.

- *Make use of Hessian information,* because derivative-free methods are unable to scale beyond a small number of inversion parameters, and because gradient-only methods (e.g., steepest descent, nonlinear conjugate gradients, quasi-Newton) also do not usually scale well with increasing problem size.
- *Be adjoint-based,* because adjoint methods (either hand coded or automatically differentiated) offer the only hope for computing derivatives (and Hessians actions) for high-dimensional parameter spaces.
- *Be multilevel in nature,* because of the great successes of multi-level and multi-grid methods for forward simulations and their promise for inverse problems. However, multi-level methods must take on a very different form for inverse problems because of the differing character of forward and inverse operators, and require careful analysis and development to be effective for inversion.
- *Parallelize in a fine-grained manner,* because of the high-dimensional state and parameter spaces that result after discretization. Parallelism must be exploited at the "grid point" level, following the lead of large-scale forward simulation codes. Additional opportunities exist for coarse-grained parallelism, such as for direct sensitivities (across inversion parameters), adjoint computations associated with active set methods (across active constraints), and multiple source inverse problems (across each instance of the forward problem), but these must build on a finer-grain parallelism to permit scalability to large numbers of processors.

Below, we identify several areas in which excellent opportunities exist to develop scalable inverse algorithms with the above properties, and outline some of the inherent challenges.

Multilevel Solvers for Inverse Problems

The principal difficulty in designing fast solvers for inverse problems is related to the linearized operators associated with the first-order optimality conditions (also known as the KKT conditions): the KKT matrix and the reduced Hessian (the Schur-complement of the inversion variables in the KKT matrix). These two operators are typically available only through their action on a vector (i.e., as matrix-vector multiplications). The discretized KKT matrix inherits the sparsity pattern of the forward problem and is of dimension of the (space-time) state, adjoint, and inversion variables. For petascale applications, this dimension can number in the billion to trillion range for stationary and time-dependent problems, respectively; thus, forming this matrix is nearly impossible. The situation is worse if the reduced Hessian is formed. It is a dense matrix of dimension of "only" the inversion variables, but it is dense and requires as many forward simulations just to construct it—which is intractable even for moderately-sized problems. Therefore, solving KKT or reduced Hessian systems requires Krylov methods, with matrix-free implementations, and possible linearized forward and adjoint solves at each iteration. Clearly, reducing the number of iterations by effective preconditioning is paramount. Multigrid

and domain-decomposition (DD) methods have proven to be extremely powerful in solving forward problems. Given the fact that the optimality conditions for subsurface inverse problems form sets of PDEs (perhaps coupled to ODEs), it is only natural to pursue solution algorithms inspired by successful PDE solvers. Such multilevel methods, however, are difficult to design properly. To achieve optimal algorithmic complexity and parallel scalability, these methods will require extensive analysis and careful implementation. Several new directions must be pursued in parallel:

- *Domain decomposition methods for the KKT system* are inspired by their success for forward operators (e.g., they undergird several of the most powerful preconditioners within PETSc). They facilitate parallelization of the overall inversion algorithm, not only at the forward solver level but also at the optimization level. Moreover, it has been observed numerically that some KKT DD preconditioners are insensitive to the choice of regularization parameter, which makes them attractive for inverse problems. The extension of DD techniques used for the forward solver to the adjoint PDEs arising in the particular inverse context requires careful analysis of all components, such as transmission conditions for the adjoint DD method. Another way in which DD methods can be exploited is through parallelism across the time dimension. As mentioned above, even when the forward problems are evolutionary, the KKT systems are boundary value problems because of coupling with a backward-in-time, terminal-value adjoint problem and the regularization operator. Therefore, unlike the forward problem, the structure of the first order KKT conditions suggests that creating concurrency in the time dimension will be beneficial, if not essential. Even if such methods are not the most flop-efficient, this approach might still be preferable because of the overall wall-clock gains when additional processors are available.
- *Multigrid for reduced space methods.* A Krylov solve with the reduced Hessian requires a forward/adjoint pair of solves for each matrix-vector multiplication, and thus scalable preconditioners to speed up the convergence are absolutely essential. For many problems that are infinite-dimensional in both the state and inversion variables, the Hessian has a "compact plus differential" form; it can be scaled by a preconditioner for the differential part to yield mesh independent convergence. However, the constant can be very large, particularly for highly nonlinear inverse problems. Multigrid offers the hope of greatly reducing the constant, thereby breaking the intractability bottleneck. However, standard multigrid schemes, which have been very successful for solution of PDEs, are not directly applicable to reduced Hessian operators. The main difficulty is that the continuous reduced Hessian is a strongly smoothing, compact, and nonlocal operator (hence its discrete version is a dense matrix). Its eigenvector–eigenvalue correspondence is reversed from the usual case of an elliptic differential operator, with large eigenvalues associated with smooth eigenvectors, and small eigenvalues associated with oscillatory eigenvectors (hence the presence of noise in the unregularized inverse solution). Therefore, typical stationary or Krylov multigrid smoothers are completely inappropriate for the reduced Hessian; in addition

to being expensive to apply, they act as roughers becasus the large eigenvalue components, which correspond to smooth eigenvectors, are typically resolved first, leaving oscillatory components in the error. The challenge then is to construct proper smoothers for the reduced Hessian.

- *Multigrid for full-space methods.* Designing scalable multi-grid solvers for KKT systems is a challenge. Indefiniteness of the linearized KKT operator is one reason; finding efficient smoothers is another. Indeed, designing pre-conditioners and linear solvers for saddle point systems is an active research topic. For time-dependent problems, the structure of the KKT system becomes even more complicated. Algebraic multi-grid is often not an option because access to an assembled matrix if often not possible, especially for evolutionary PDE constraints. The difficulty, as mentioned above, is that the KKT operator corresponds to a boundary-value problem with the forward and adjoint fields evolving in different directions in time. Specially designed multiplicative block and pointwise smoothers have been tested numerically, and have shown promising results for optimization problems with elliptic, parabolic, and parabolic-reactive PDE constraints. Despite the absence of theory, spectral analysis for select linear problems indicates that in certain cases, multi-grid on the full-space KKT system results in significant speedups compared to the reduced space multi-grid, mostly because of avoidance of the adjoint/forward solves at each reduced Hessian iteration. For nonlinear multi-grid on KKT systems, little theory exists. Yet, initial attempts based on the full approximation scheme (FAS) concept have been notably effective with respect to speedups, globalization by detecting negative curvature directions, and iterative regularization for inverse problems. In addition, a nice feature of FAS is that pointwise or block-pointwise nonlinear smoothing schemes are cache-friendly (as they deliver more flops per cache hit); thus, FAS for KKT can be used for latency tolerance. For example, by recursively adjusting the smoother's block size, one can design FAS schemes that adjust to the hardware's latency and bandwidth characteristics.

INTERIOR POINT ALGORITHMS FOR INEQUALITY CONSTRAINTS

Inequality constraints are handled in two ways in most (generic) large-scale optimization codes. First, active set strategies identify potential tight inequality constraints as the algorithm proceeds and, based on multiplier estimates, decide on which constraints should be added to and dropped from the active set. For large-scale problems, however, an expensive combinatorial element still characterizes active set solvers. On the other hand, interior point or barrier approaches have considerably reduced complexity that is independent of the number of active constraints. Inequality constraints are substituted by a penalty formulation with logarithmic barrier terms and an associated penalty parameter. They can be applied to a wide set of problem classes and can exploit problem structure of KKT matrices. However, there are a number of issues must be addressed before interior point methods can be applied successfully to inverse problems involving subsurface modeling. One needs to deal first with ill-conditioning as the barrier term is driven to zero and second with selecting the penalty parameter. Recent work shows that the KKT system can be modified to a doubly augmented positive definite system, which leads to a

well-conditioned preconditioner. However, it can be difficult to determine a good strategy to reduce the barrier parameter at each nonlinear iteration. These difficulties may be circumvented via multigrid techniques. In particular, it may be possible to connect the penalty parameter to the mesh-size, using grid sequencing to slowly refine the mesh and decrease the penalty parameter.

Low Rank Hessian Approximations for Regularization Parameter and Covariance Estimation

Upon solution of many inverse problems, there are several post-optimality analyses that are typically of interest. First, we wish to select the "best" value of the regularization parameter based on current information, and second, we would like to describe the uncertainty in the solution. These tasks formally require operations on the reduced Hessian operator or its inverse. As has been discussed above, for large-dimensional inversion spaces, the reduced Hessian cannot be formed, let alone manipulated. Fortunately, the "compact plus differential" structure of the reduced Hessian implies that the compact part (which is the component whose formation is intractable) can be well-represented by a low-rank estimate, and often on a coarser grid. There are opportunities to exploit this structure using several ideas: 1) DD and multigrid methods used as preconditioners, 2) truncated spectral approximations obtained via partial Lanczos eigenanalysis of the reduced Hessian, 3) information generated by conjugate gradient solution of the reduced Hessian system, 4) limited memory quasi-Newton approximations, and 5) simple coarse grid approximations. These low-rank, reduced Hessian approximations can be used in the following two ways:

- *Regularization parameter estimation.* The solution of inverse problems usually depends critically on the type of regularization and choice of regularization parameter. For PDE inverse problems, two main types of regularization operators are used: 1) Tikhonov regularization (typically L2 or H1) or 2) "robust regularization" such as total variation and the Huber M-estimator. When the noise level is known, the discrepancy principle can be used to estimate the regularization parameter. More difficult is the estimation of the regularization parameter when the noise level is unknown. In this case, more sophisticated methods such as generalized cross validation (GCV) are required. GCV uses the reduced Hessian to assess the quality of a solution for a given regularization parameter and then updates it. Approximation of the reduced Hessian by low-rank estimates will enable this kind of analysis for infinite dimensional inversion spaces.
- *Covariance estimation.* Under the Gaussian hypothesis for uncertainty in the observations, model, and prior (and for a linear inverse problem), solution of the inverse problem yields the most likely model given the observations. It is possible to describe the uncertainty in this solution by invoking the equivalence (or approximate equivalence, for a nonlinear inverse problem) of the inverse of the reduced Hessian with the posterior parameter covariance matrix. For high-dimensional inversion spaces, this is often intractable, for reasons mentioned above. The ability to generate low-rank estimates of the reduced Hessian that capture the significant modes of this operator will permit covariance-based estimation of uncertainty in the inverse solution.

Uncertainty Estimation via Solution of Statistical Inverse Problems

The regularization approach to inverse problems as described above will yield an estimate of the "best" parameter values that simultaneously fit the data and minimize the regularization (or prior) term, and this can be connected to the mean of the posterior parameter probability density when the uncertainties in data, model, and prior are Gaussian (and, strictly speaking, when the inverse problem is linear). When this is not the case, the posterior probability density function (PDF) is no longer Gaussian, and one is led to the Bayesian statistical setting to infer a complete statistical description of the uncertainty in the inverse solution. Thus, we are able to quantify the resulting uncertainty in the parameters, taking into account uncertainties in the data, model, and prior information, without making any assumptions on the form of the uncertainties. However, the computational challenge for the Bayesian approach is to develop methods for exploring the posterior PDF of the model parameters, a task that can be intractable for high-dimensional probability spaces or for expensive forward simulations (the need for which is embedded within the expressions for the posterior density). The simplest posterior quantity to compute is the *maximum a posteriori* (MAP) estimate, which is a point estimate of the parameter values that maximize the posterior density. The structure of the resulting optimization problem resembles those from the regularization approach, and the advanced algorithms described above apply directly.

More generally, however, we are interested in propagating the uncertainties in the model parameters through the forward simulation to obtain uncertainties in the predictions. In this case, we require a tractable representation of the posterior probability density that can be repeatedly queried as needed during forward propagation of uncertainties. Typically, this requires the mean, the covariance, and higher moments of the distribution. Because computing these quantities involves integration over parameter space, straightforward numerical quadrature is out of the question for anything other than a few parameters. Special sampling techniques, such as Markov chain Monte Carlo (MCMC) methods, have been developed to generate sample ensembles, and the sample points then provide approximations of the mean, covariance, etc. Even so, this approach can become intractable as the complexity of the forward simulations and the dimension of the parameter space increase. A major research effort is needed to develop Bayesian inverse methods that can scale to the high-dimensional parameter spaces implied by uncertain heterogeneous parameter and state fields, and the complex nonlinear forward simulations inherent in subsurface simulation.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Subsurface modeling problems are ideal proving grounds for inverse algorithms. The medium and state are not directly observable, thus resulting in uncertainty in parameters, initial/boundary conditions, and source terms. Forward simulation therefore relies on inverse analysis to estimate these parameters (and possibly their uncertainty) from the observables. The forward simulation problems alone are extremely challenging; inverse solution for *high-dimensional* parameter spaces (as results from discretization of heterogeneous property fields) is intractable using

current methods. The development of scalable parallel numerical algorithms for inverse solution that can match the scalability and efficiency of the associated forward solvers, as described in this PRD, is thus of paramount importance. The subsurface modeling setting will serve to motivate and maximally stress these algorithms. Because these algorithms will have widespread potential use in many inverse problems of relevance to the SC mission (and beyond), the potential impacts of this PRD are manifold. Examples of impacted inverse problems include those occurring in atmospheric and ocean modeling (for state estimation), combustion (for reaction parameter estimation), accelerator design (to identify as-built cavity shapes from frequency measurements), and geophysics (earth properties from seismic and electromagnetic measurements).

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

As mentioned above, solution of subsurface parameter estimation problems for high-dimensional parameter spaces is currently intractable using available inverse methods. Successful development of inverse algorithms that can scale to large parameter spaces and large processor/core counts, as contemplated by this PRD, will usher in a new era of computational simulation of subsurface geophysical and biogeochemical processes in which uncertainty in inputs are managed via systematic inversion and uncertainty estimation methods.

TIME FRAME

Significant progress could be achieved in 3 to 5 years and full attainment of the objectives is anticipated within 10 years.

BIBLIOGRAPHY

V. Akcelik, G. Biros, O. Ghattas, J. Hill, D. Keyes, and B. van Bloemen Waanders, "Parallel algorithms for PDE-constrained optimization," in *Frontiers of Parallel Computing*, M. Heroux, P. Raghaven, and H. Simon, eds, SIAM, 2006.

L. Biegler, O. Ghattas, M. Heinkenschloss, and B. van Bloemen Waanders, eds., *Large-Scale PDE-Constrained Optimization*, Springer-Verlag, 2003.

L. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. Van Bloemen Waanders, eds., *Real-time PDE-Constrained Optimization*, SIAM, 2007.

VALIDATION, VERIFICATION, UNCERTAINTY ANALYSIS AND DECISION OPTIMIZATION

DECISION OPTIMIZATION, DATA ASSIMILATION, CONTROL UNDER UNCERTAINTY

Coordinators: Dave Engel, Pacific Northwest National Laboratory; Christine Shoemaker, Cornell University

Contributors: Peter Kitanidis, Stanford University; Jon Helton, Sandia National Laboratory; Dongxiao Zhang, University of Oklahoma; Donald Estep, Colorado State University
Reviewer: Thomas Fogwell, Energy Solutions (Fluor Hanford PHMC)

ABSTRACT

Decision optimization is the computational process that seeks to determine the most advantageous, in terms of cost or other criteria, decisions that meet specified physical and performance constraints. The challenge is to improve computational methods to optimize decisions in the management of large-scale subsurface systems, particularly for evolving systems that need to be updated with incoming data. These methods should account for uncertainty, be computationally efficient, and take advantage of current and future computer power (specifically high-performance computing).

EXECUTIVE SUMMARY

An example of an important decision for subsurface systems involves contaminant removal (i.e., storage). This includes such decisions as the location and pumping rates at wells that are pumping contaminants out of the ground, the location of monitoring devices that can include sampling wells [1], and also soil removal or treatment. For more complex systems, chemicals may be injected into the groundwater to accelerate the biodegradation of the contaminant (i.e., bioremediation) or enhance the effectiveness of extraction pumping (e.g., by adding surfactants), which engenders more decisions including the concentrations of injected materials. If we choose to allow these decisions to vary over time, we have a dynamic decision problem (i.e., a control problem. Other pressing examples include the use of the subsurface for storage of hazardous wastes and carbon sequestration.

Common approaches to solving optimization problems employed for subsurface systems assume that the model, boundary conditions, parameter, and data are known with certainty. However, in practical terms, it is impossible to specify subsurface processes and parameters without error. Research is needed to devise optimization methods that allow for uncertainties in the optimization problem and subsequently estimate uncertainties in computed optimal values. As a simple example, deterministic optimization methods can be used for stochastic models by employing random sampling (e.g., Monte-Carlo methods), which involves computing a large number of realizations of the aquifer to characterize uncertainty. There also are optimization methods that are especially designed to deal with uncertainty. All of these approaches have attendant significant challenges in terms of their efficient implementation on high-performance computers.

Data assimilation is the problem of matching information computed from the solutions of a model to actual observations of the physical state being modeled. The goal is to predict system behavior based on the current available information about the system. Data assimilation is often posed as a computationally intensive inverse or optimization problem. The melding of state observations and physically informed mathematical models promises to provide a way of computing accurate and realistic predictions that might not otherwise be feasible. Data assimilation also raises issues regarding the processing of field measurements and data into a form that is useful for comparison to information computed from models described in the *Characterization of data, model, and parametric uncertainty* section of the PRD: "Uncertainty Representation, Uncertainty Propagation, And Sensitivity Analysis For Subsurface Systems."

The practical significance of optimization cannot be overemphasized, since it is through optimization that one can take full advantage of the sophisticated models under development in order to make the best possible decisions. Optimization for decisions under uncertainty will lead to more cost effective solutions for highly expensive DOE subsurface undertakings. Data assimilation will provide the capability of predicting the behavior of complex subsurface systems. Data assimilation and feedback optimization will enable rapid utilization of data as collected at the sites where remediation and cleanup is essential.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

To date, most optimization methods employed for the management of subsurface systems are deterministic [2, 3, 4]. Deterministic approaches for solving decision problems are computationally challenging because the underlying simulation models for groundwater transport are detailed and computationally expensive, and optimization typically requires hundreds or thousands of simulations. For example, the U.S. Environmental Protection Agency has sponsored a demonstration project on the use of optimization methods for groundwater remediation [5] for two field sites involving two competing optimization groups. Both groups required thousands of simulations. The models used in this demonstration project were much simpler and required less computer time than the typical, much-more-detailed models developed by DOE. There is the potential to dramatically improve these methods including developing better numerical methods [6, 7] for complex problems and to develop effective parallel algorithms [8, 9]. In addition, more research is needed to quantify the uncertainty and error in computed optimal solutions because resolution error in subsurface simulations is usually sufficiently large as to affect optimal values.

With improved algorithmic development coupled with high-performance computing resources, dramatic reductions can be made in the analysis approach. This improved approach will produce excellent information that can help make cost-effective decisions for management and remediation design alternatives for DOE subsurface problems. For instance, selection from among different remediation actions is a key element of the cleanup process at many DOE waste sites. As another example, a major issue in the development of the proposed repository for high-level radioactive waste at Yucca Mountain, Nevada, [10] has been whether 1) to design the waste

disposal drifts so that the wall temperatures remain above the boiling point of water for a substantial period of time or 2) to design the waste disposal drifts so that the wall temperatures always remain below the boiling point of water. An adequate resolution of this issue, which is inherently an optimization problem, involves a computationally intensive analysis of a highly complex system involving extensive uncertainties.

In general, optimization methods have many applications, including:

- Determination of the best conceptual model to fit observational data (model calibration)
- Definition of conditional realizations, to estimate performance (results of forward modeling) with uncertainty estimation
- Selection of new data (what values, where, and when) to be collected that will reduce uncertainty estimates (feedback) or optimize decisions
- Selection between different alternatives, including:
  - collection of additional data
  - remediation actions
  - continued monitoring
  - design alterations.

It is essential that better algorithms and approaches be developed to address these challenges. There are some existing methods, but in general they are not computationally efficient enough or mathematically complex enough to fully address the uncertain environment in which decisions must be made or the complexities associated with multimodal optimization problems arising from nonlinearities and heterogeneities. In addition, methods that deal effectively with coupled simulation models (e.g., multi-physics models) and methods that generate feedback policies are necessary. The uncertainty incorporated should include parameter, data, numerical, and model error in the analysis, which is not adequately addressed in current modeling efforts.

The issue of assimilating data into models arises in all scientific areas [11]. In its broadest sense, it is a challenge that arises at the meeting point of data and models. Technology has driven advances on both sides of the equation: new techniques of measurement have led to a significant increase in the amount and types of available data and advances in high-performance computing provide the capability of new levels of computational modeling. The development of effective data assimilation methods has become one of the fundamental challenges in scientific prediction.

The problem of assimilating data into a model of a subsurface system (or any system) is both challenging and fundamental in that it aims at the estimation and prediction of an unknown, true state. It includes both observations and computational models. Measured observations provide direct information of the true state. Such observations are heterogeneous in space, irregular in time, and subject to differing accuracies. Computational models use knowledge of underlying physics and dynamics to provide a description of state evolution in time. Models also are far from perfect because of conceptual model error, uncertainty in the initial conditions and

computational limitations, and the fact that model evolution cannot precisely represent the true state.

The overall problem in the subsurface modeling area (and many other disciplines) is that forward modeling is very computationally intensive, and is becoming even more so. The current implementation of modeling techniques (e.g., Monte Carlo and derivative-based methods) is not adequate to handle the next generation of subsurface models (either computation or implementation).

SUMMARY OF RESEARCH DIRECTION

Currently, algorithms exist for optimization and data assimilation, some being used in subsurface modeling and other disciplines [11, 12, 13, 14, 15, 30, 31, 32, 33]. Most of these methods could be extended with new research to take full advantage of current computational power to deal with problems of realistic complexity.

Decision Optimization

Optimization methods can provide an invaluable tool to identify the best design or management decisions. Optimization is better than a "trial-and-error" approach, which simulates the model for each of a pre-selected set of alternatives. The problem is that the number of alternatives can be very large and the resulting number of simulation runs (if an ensemble method is being used) is not computationally feasible if the models themselves are computationally expensive. Instead optimization methods can be used that will search through the space of parameters without doing a simulation for each possible alternative and evaluate only the alternatives that are promising.

There are different types of optimization algorithms. Their advantages and disadvantages depend upon the type of management problem under consideration. Derivative-based optimization algorithms are very efficient if the derivatives of the model are available and accurate at low computational cost, and if the problem is known to be convex and have a single optimum. These methods have been used in a number of codes for parameter estimation and design decisions [16]. There are heuristic global optimization methods in use [17], but they can typically take thousands of simulations and are too inefficient for computationally expensive simulation models. Obtaining derivatives from a complex nonlinear simulation model (especially with reactive transport with nonlinear relationships) can also 1) be computationally expensive and perhaps inaccurate if done with finite differences, 2) be demanding on human time if done analytically or with automatic differentiation [18,28,29], or 3) be impossible if the source code is not available.

Recent work in the area of derivative-free optimization has shown that these methods are well suited to subsurface simulation models, which require numerical solutions to complex partial differential equations [19, 20, 21]. However, further improvements are needed in this area with more research, especially for parallel algorithms.

The development of software tools that use advanced techniques for subsurface model optimization in the areas of derivative-based (local) optimization and global-optimization methods are needed, including:

- Implementation of current techniques in existing models (both intrusive and non-intrusive methods)
- Development and enhancement of computational models that will use current and new techniques in both local optimization (e.g., derivative-based) and global optimization (adaptive sampling/searching)
- Development of frameworks and toolkits to implement optimization techniques within existing and future computational models.

Data Assimilation

Standard methods for assimilating pressure, rate, and saturation observations into subsurface flow models requires development of complex codes for the flow and transport and simultaneous assimilation of all data. Because of the existence of various simulators in the DOE complex, the effort required to develop codes for each simulator, and to keep each up to date will be large. Automatic adjoint and differentiation codes exist [22, 28, 29] but have not been adapted to recent advances in control theory or to modern scientific environments.

Most subsurface problems involving multi-phase flow and transport are highly nonlinear with the possibility of many modes in the probability density for the subsurface variables. Several major problems arise in the assimilation of data for highly nonlinear problems. Therefore, fundamental theoretical development is needed for data assimilation.

Current data assimilation methods that are used in other disciplines need to be incorporated into exiting subsurface modeling efforts [11,30,31,32]. Future directions/developments in data assimilation techniques need to be addressed. The coordination/cooperation of mathematicians, statistician, geophysicists, and computer scientists is needed to tackle this problem.

In addition to the features for optimization and data assimilation indicated in the previous sections, computational efficiency has also become a vital part for subsurface modeling. In the subsurface modeling area, the focus is on computationally expensive computer simulation models that take many hours to run. For an optimization problem, the optimization part of the calculation is relatively quick compared to the model run times. Hence, the problem is to find algorithms that give an acceptably accurate answer with relatively few simulations (e.g., hundreds, not tens of thousands of model simulations/runs). Such algorithms would make effective use of petascale resources to do a very thorough analysis of subsurface management decisions under uncertainty.

For data assimilation, algorithms are needed to take advantage of high-performance computing. Current methods [23, 24, 25] are computationally intensive and need to be implemented to take

advantage of the current and future computer power. New developments (extensions) are also needed to take advantage of future computer power.

Optimization of Systems Evolving under Uncertainty

The mathematics of optimal control of uncertain systems (or stochastic optimal control) have had little impact so far on improving the management of subsurface systems, primarily because of the very heavy computational requirements [33]. However, the problem is important enough and computational power has increased so much over the last few years that it is desirable to re-visit this research area.

It is well understood that the essential elements of optimal control of uncertain systems are feedback, caution (or hedging), and probing [26]. Some systems behave so that just allowing for feedback (i.e., adjustment of estimates and decision variables through data assimilation as time passes) is sufficient for obtaining the (practically) best solution. Feedback improves performance by varying actions to keep the system close to a desirable trajectory [27]. Other systems, however, are dominated by caution and probing effects. For such systems, one must explicitly incorporate in the optimization the anticipation that the system will be steered to alleviate large losses that may correspond to low-probability events (the "caution" effect) or to excite the system (the "probing" effect) so that through measurements one can gain information about system parameters that results in a substantial improvement of performance. Caution- and probing-dominated optimal control-problems are harder to solve [see 14]. Development and implementation of these algorithms to perform under high-performance computing environments will be critical for decision-making.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The development of new methods will provide potentially very large improvements over existing computational methods for complex, multi-physics, and distributed model analysis. The methods will need to be highly computationally efficient and, hence. will use state-of-the-art methods in numerical analysis and distributed computing. This will then stimulate research by computer scientists in these areas. These algorithms are applicable to a wide range of models arising in many areas of different scientific domains including physical sciences and engineering, where there are many computationally intensive models in need of optimization to help with design and management.

POTENTIAL SUBSURFACE SCIENCE IMPACT

Optimization for decisions will lead to more cost effective and environmentally appropriate solutions under uncertainty for highly expensive DOE subsurface projects. Optimization has been shown in many fields to reduce the cost of management and provide better project design [5]. The amounts of money required for DOE environmental management and remediation are so great that even a small percentage decrease in cost can result in huge cost savings. We would

expect that the decreases would be greater than a small percentage, with further increases in the economic benefits of implementing optimization methods.

Data assimilation and feedback optimization will enable rapid utilization of data as collected.

The use of high-performance computing to provide rapid and accurate connection between incoming data and model predictions will provide a bridge between data gathers (experimental and field data) and modelers.

TIME FRAME

Positive effects on computational sciences and the subsurface sciences from the research described above could be obtained in 2 to 6 years.

REFERENCES

[1] P. Mugunthan and C.A. Shoemaker. 2004. Time Varying Optimization for Monitoring Multiple contaminants under Uncertain Hydrogeology. Bioremediation Journal, 8(3-4) 129-146.

[2] L.-C. Chang, C.A. Shoemaker, and P.L.-F. Liu. 1992. Optimal time-varying pumping rates for groundwater remediation: Application of a constrained optimal control algorithm. WaterResour. Res., 28, 3157– 3173.

[3] T. Culver and C.A. Shoemaker. 1996. Impact of Treatment Facility Capital Costs on Dynamic Optimal Groundwater Reclamation. ASCE Jn. of Water Res. Plan. and Manag. Division 123, pp. 23-29.

[4] Zheng, C. and P.P. Wang, 2002, A field demonstration of the simulation-optimization approach for remediation system design, Ground Water 40(3): 258-265.

[5] D.B. Becker, B. Minser, R. Greeenwald, Y. Zhang, K. Harre, K. Yager, C. Zheng, and R. Peralta. 2006. Reducing Long-Term Remedial Costs by Transport Modeling Optimziation. Ground Water, pp. 864-875.

[6] C.M. Mansfield and C.A. Shoemaker. 1998. Use of Numerical Sparsity in an Aquifer Quality Optimization Algorithm. ASCE Jn. of Water Res. Plan. Manag. 124, pp. 15-21.

[7] P. Mugunthan, C.A. Shoemaker, R. G. Regis. 2005. Comparison of Function Approximation, Heuristic and Derivative-based Methods for Automatic Calibration of Computationally Expensive Groundwater Bioremediation Models. Water Resources Research. Vol. 41, W11427,doi:10.1029/2005WR004134.

[8] H. Caffey, L-Z Liao, and C.A. Shoemaker. 1993. Parallel Processing of Large Scale, Discrete-Time, Unconstrained Differential Dynamic Programming. Parallel Computing. 19,1003-1018.

[9] R. Regis, C.A. Shoemaker. 2006. Parallel Radial Basis Function Methods for the Global Optimization of Computationally Expensive Functions. European Journal of Operations Research in press.

[10] U.S. DOE (U.S. Department of Energy). 2002. Yucca Mountain Science and Engineering Report. U.S. Department of Energy, Las Vegas, NV DOE/RW-0539-1, Rev. 1.

[11] D. McLaughlin, Y. Zhou and D. Entekhabi. 2006. Computational issues for large-scale land surface data assimilation problems. Jour. Hydrometeorology, 7 (3), 494-510.

[12] R. Andricevic and P.K. Kitanidis. 1990. Optimization of the Pumping Schedule in Aquifer Remediation under Uncertainty. Water Resour. Res., 26(5), 875-885.

[13] S.-I. Lee and P.K. Kitanidis. 1991. Optimal Estimation and Scheduling in Aquifer Remediation with Incomplete Information. Water Resour. Res., 27(9), 2203-2217.

[14] S.-I. Lee and P.K. Kitanidis. 1996. Optimization of monitoring well installation time and location during aquifer decontamination. Water Resour. Management, 10, 439-462.

[15] R.C. Peralta. 2001. Remediation simulation/optimization demonstrations. In Proceedings of MODFLOW and Other Modeling Odysseys, ed. S. Seo, E. Poeter, C. Zheng and O. Poeter, 651-657. Golden, Colorado: International Groundwater Modeling Center.

[16] K.D. Jarman and T.F. Russell. 2001. Analysis of 1-D moment equations for immiscible flow. In AMS-IMS-SIAM Joint Summer Research Conference on Fluid Flow and Transport in Porous Media. American Mathematical Society.

[17] J-H Yoon and C.A. Shoemaker. 2001. Improved Real-Coded Genetic Algorithm for Groundwater Remediation. ASCE Jn. of Computing. 15(3), 224-231.

[18] C. Bishcof, G. Whiffen, C. Shoemaker, A. Carle, and A. Ross. 1994. Application of Automatic Differentiation to Groundwater Transport Models. pp. 173-182 in Computational Methods in Water Resources X, A.Peters et al. (editors), Kluwer Academic Publishers, Dordrecht.

[19] R.G. Regis and C.A. Shoemaker. 2007. A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions. INFORMS Journal of Computing, in press.

[20] B. Debusschere, H. Najm, P. Pébay, O. Knio, R. Ghanem, O. Le Maître. 2004. Numerical challenges in the use of polynomial chaos representations for stochastic processes. SIAM J. of Scienctific Computing 26(2) 698-719.

[21] D.W. Engel, A.M. Liebetrau, K.D. Jarman, T.A. Ferryman, T.D. Scheibe, and B.T. Didier. 2004. An Iterative Uncertainty Assessment Technique for Environmental Modeling. Meeting of TIES and ACCURACY 2004: 6th International Symposium on Spatial Accuracy Assessment. PNNL-SA-41199

[22] C.H. Bischof, P. Khademi, A. Mauer, and A. Carle. 1996. ADIFOR 2.0 – automatic differentiation of Fortran 77 programs. IEEE Computational Science & Engineering, 3(3):18-32

[23] C.K. Wikle and L.M. Berliner. 2005. Combining Information Across Spatial Scales. Technometrics, 47:80-91.

[24] L.M. Berliner, R.F. Milliff, and C.K. Wikle. 2003. Bayesian Hierarchical Modeling of Air-Sea Interaction. J. Geophys. Res., 108(C4).

[25] C. Andrieu and Y.F. Atchadé. 2006. On the efficiency of adaptive MCMC algorithms. In Proceedings of the 1st international Conference on Performance Evaluation Methodolgies and Tools (Pisa, Italy, October 11 - 13, 2006).

[26] A.A. Feldbaum. 1965. Optimal Control Systems, Academic Press, NY.

[27] G.J. Whiffen and C.A. Shoemaker. 1993. Nonlinear Weighted Feedback for Groundwater Remediation under Uncertainty. Water Resources Research. 29, 3277-3290.

[28] M. Berz, C. Bischof, G. Corliss, and A. Griewank, 1996. Computational Differentiation: Techniques, Applications, and Tools. Philadelphia, PA: Society for Industrial and Applied Mathematics.

[29] A. Griewank, 2000. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. Philadelphia, PA: Society for Industrial and Applied Mathematics.

[30] McLaughlin, D., O'Neill, A., Derber, J., Kamachi, M., 2005. “Opportunities for enhanced collaboration within the data assimilation community,” Quarterly Journal of the Royal Meteorological Society 613, 3683-3693.

[31] McLaughlin, D., 2002. “An Integrated Approach to Hydrologic Data Assimilation: Interpolation, Smoothing, and Forecasting,” Advances in Water Resources 25, 1275-1286.

[32] Zhou, YH, McLaughlin, D, Entekhabi, D., 2006. “Assessing the performance of the ensemble Kalman filter for land surface data assimilation,” Mon Weather Review 134, 2128-2142.

[33] B.S. Minsker and C.A. Shoemaker. 1998. “Quantifiying the Effects of Uncertainty on Optimal Groundwater Bioremediation Policies,” *Water Resources Research* 34, pp.3615-3625.

UNCERTAINTY REPRESENTATION, UNCERTAINTY PROPAGATION AND SENSITIVITY ANALYSIS FOR SUBSURFACE SYSTEMS

Coordinator: Donald Estep, Colorado State University

Contributors: Bert Debusschere, Sandia National Laboratory; Donald Estep, Colorado State University; Jon Helton, Sandia National Laboratory; Peter Kitanidis, Stanford University; Daniel Tartakovsky, University of California San Diego; Dongxiao Zhang, University of Oklahoma

ABSTRACT

Significant uncertainty accompanies any description or prediction of the behavior of a complex subsurface system, and this uncertainty has important scientific, societal, and economic implications. This raises a critical need for high-performance computational methods that can treat uncertainty in all phases of modeling. The challenge is to develop tools and libraries for uncertainty analysis that can become a fundamental component of high-performance computational modeling of subsurface systems.

EXECUTIVE SUMMARY

A fusion of experiment, modeling, and high-performance computation has become the main tool for scientific investigation and engineering application in complex multi-physics, multi-scale subsurface systems. The practical limitations in experiment and computation mean that substantial uncertainty accompanies any description or prediction of the behavior of a subsurface system. Uncertainty is inherent in data gathered from subsurface systems, the complexity and heterogeneity of subsurface systems, the mathematical descriptions of physical processes in subsurface systems, and the analysis and computational simulation of complex multi-scale, multi-physics models. Because of the complexity and heterogeneity of subsurface systems, and the inaccessibility of the subsurface to measurement and experiment, these uncertainties are often large and difficult to quantify. The scientific need for an adequate representation of uncertainty in the analysis of complex systems is now widely recognized while the national importance of the issues involving subsurface systems—for example, radioactive waste disposal at the proposed repository at Yucca Mountain, Nevada (DOE 2002), carbon sequestration, oil and gas recovery, and pollution in groundwater—means that this uncertainty has tremendous scientific, societal, and economic implications

There is a critical need for significant development of high-performance computational methods for treating uncertainty in all phases of modeling and simulation, from experimental error to uncertainty in the evaluation of subsurface models to the quantification and interpretation of uncertainty in the results of model analysis. The sheer variety of sources and forms of

uncertainty in a subsurface system increases the challenges for development and implementation of new computational methods in a broad range of statistical and mathematical specializations. The different approaches used to treat various uncertainties must be combined to produce a complete and conceptually consistent analysis tool for predictive science and engineering. The challenge to the computational community is to develop and implement high-performance tools and libraries for uncertainty analysis that can become a fundamental part of high-performance computational modeling of complex subsurface systems.

Meeting the theoretical and computational challenges associated with treating uncertainty in subsurface systems will provide a greatly enhanced ability to fully use high-performance computational platforms to address fundamental scientific questions related to subsurface processes. This will result in increased confidence in subsurface system analysis, due to better accuracy in the analysis with more reliable quantification of the remaining uncertainty in the results. Ultimately, this will allow better informed policy decisions regarding issues of national importance, such as contamination remediation, carbon sequestration, and radioactive waste storage.

SUMMARY OF RESEARCH DIRECTION

The complexity of the coupled flow, transport, biogeochemical, and mechanical processes that occur in a heterogeneous subsurface geologic medium and the importance of the decisions that must be made involving these processes have forced a revolution in predictive science and engineering for subsurface systems. A fusion of physical experiment and measurement approaches, mathematical modeling, and high-performance computational simulation has become the primary tool for scientific investigations and engineering applications. This fusion of disciplines derives from the need to combine physical experiments and computational modeling in probing, describing, modeling, and simulating the behavior of subsurface systems.

These limitations mean that substantial uncertainty accompanies any description or prediction of the behavior of a subsurface system. Uncertainty is inherent in (i) data gathered from subsurface systems, (ii) mathematical descriptions of the physical processes in the subsurface, and (iii) the analysis and computational simulation of complex multiscale, multiphysics models. The importance of uncertainty and uncertainty quantification is increased by the inaccessibility of the subsurface to measurement and experiment, and the complexity and heterogeneity of the subsurface. The scientific need for an adequate representation of uncertainty in the analysis of such complex systems is now widely recognized (Christie et al. 2005; Wagner 2003; Ayyub 1997; Nikolaidis et al. 2004; Oberkampf et al. 2002(1); Oberkampf et al. 2002(2); Oberkampf et al. 2004; Roache 1998; Trucano et al. 2006; Babuska et al. 2001; Babuska et al. 2002; Zhang 2002). The national importance of the issues involving subsurface systems—radioactive waste disposal at the proposed repository at Yucca Mountain, Nevada (DOE 2002), carbon sequestration, oil and gas recovery, and pollution in groundwater—means that this uncertainty has tremendous scientific, societal, and economic implications.

Not only does uncertainty arise from many sources in subsurface analysis, but it takes many forms. These sources can be parsed into four conceptual types: 1) the inherent, or aleatory (i.e., the randomness [predictability] of an event), uncertainty in possible future events (e.g., earthquakes) in the system; 2) the heterogeneity and scale differences of properties of the subsurface; 3) error and uncertainty in computationally demanding simulations of mathematical models; and 4) the epistemic uncertainty arising from lack of knowledge about the physical processes and their interactions, the state of a given system, and the representation of the aleatory uncertainty. Dealing with the sheer variety of uncertainties in a subsurface system itself becomes a major hurdle.

The treatment of these uncertainty components must be combined to produce a complete and conceptually consistent analysis that can be used for predictive science and engineering. This raises theoretical and computational hurdles in all aspects of modeling a complex system.

1. Acquisition of data and determination of experimental error
2. Conversion of data to model structures and model inputs
3. Characterization of the epistemic uncertainty associated with model structures and model inputs
4. Linking a system of models for different processes
5. Accurate numerical evaluation of both individual models and systems of linked models and the errors associated with numerical solution
6. Statistical analysis of aleatory uncertainty
7. Statistical analysis of epistemic uncertainty
8. Presentation and interpretation of uncertainty analysis results at the level of individual models and also the entire assembled analysis
9. Development, presentation, and interpretation of sensitivity analysis results at the level of individual models and also the entire assembled analysis
10. Formulation of appropriate decisions on the basis of the analysis (e.g., where to invest resources to reduce uncertainty or a regulatory decision on the licensing of a deep geologic repository for radioactive waste).

There is a substantial need for significant development of new computational, statistical, and deterministic tools for the analysis and control of uncertainty in each of these aspects of modeling as well as the need for new approaches to merge these techniques to deal with uncertainty in the large.

Building and implementing the capability of analyzing, quantifying, and controlling uncertainty in high-performance software for modeling subsurface systems is an enormous computational challenge. ,To put this in perspective, consider that the accurate and efficient solution of differential equations, which has traditionally been a main focus of high-performance

computation, is only the starting point for the analysis and quantification of uncertainty in a complex system. There is a great need for substantial progress in the conceptual design, computational design, and computational implementation of uncertainty assessments for complex systems as a fundamental part of high-performance computational modeling of complex subsurface systems.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The scientific and computational challenges in uncertainty analysis can be parsed into five major areas: 1) characterization of the uncertainty in data, model, and parameters that form the basis for simulating specific physical systems; 2) characterization and propagation of uncertainty in model simulation; 3) control of uncertainty where required and possible; 4) analysis of complex subsurface systems in which uncertainty and variation play a significant role; and 5) development of high-performance tools and libraries for uncertainty analysis. We stress, however, that uncertainty in subsurface model analysis and prediction results from a complex interaction of uncertainties occurring at all stages of mathematical modeling. Indeed, the uncertainty associated with a model of a subsurface geologic system is as complex as the physical processes that collectively constitute the system being modeled. Thus, there are significant connections among the different challenges described below.

Characterization of data, model, and parametric uncertainty

To analyze the uncertainty in model predictions, it is necessary to characterize the uncertainty in the input parameters and data used for a model or in the data used to calibrate the model. This presents two major challenges: 1) harvesting information from experiment and field measurements in subsurface systems is difficult and costly and 2) there are serious physical limitations on the type and scale of data that can be gathered. The data determined in physical measurements must be processed into a form useful for mathematical modeling.

Processing and representing data and its uncertainty. The data that can be gathered from subsurface systems is dictated by physical limitations on experiment and measurement. Typically, the inputs required for data and parameter values in mathematical models of subsurface systems (e.g., detailed pore structure in porous media) cannot be determined by direct measurement. Instead, the necessary information must be obtained by converting available data into a useful form. This introduces additional sources of uncertainty in the quantities actually used as inputs to a model. In addition, the uncertainty of any physical data is transformed along with the data, and the resulting uncertainty is typically complex (e.g., associated with multi-modal, heavy-tailed, or otherwise non-standard distributions). Further research is needed to develop computationally efficient methods (e.g., up-scaling procedures) to construct distributions underlying the available samples of data and parameters and to represent those distributions in a compact and accurate manner amenable for use in high-performance simulations.

In practice, probabilistic descriptions of uncertainty may have to be supplemented with expert review processes (Ayyub 2001; Budnitz et al. 1998; Cooke et al. 2004; Garthwaite et al. 2005; McKay et al. 2000) or possibly replaced by alternative uncertainty representations such as interval analysis, fuzzy set theory, possibility theory, and evidence theory (Halpern 2003; Helton et al. 2004; Klir 2004, 2006; Ross et al. 2002, 2004). Alternative uncertainty representations raise new computational challenges in terms of efficient implementation for large-scale subsurface models in high-performance computing environments.

Translation of uncertainty representations between physics at different scales. Data and parameters are often measured or inferred in one physical system and at one scale or range of scales, but are then used in other model components that represent different physics on different scales. An example is the measurement of corrosion properties on a small metal coupon, where these properties are subsequently used in a performance analysis of a 1.5 m diameter nuclear waste container. Another example is continuum flow in porous media, where parameters governing macro-scale behavior are extracted from localized micro-scale simulations, which are subject to intrinsic noise on top of the other uncertainties that may be present in micro-scale simulations. This raises mathematical and computational challenges in terms of determining accurate and efficient mechanisms for translating data between different physics and scales and assessing the uncertainty associated with these translations.

A challenge is to develop new methods to characterize the full distribution of variables obtained on one scale along with techniques for translating these distributions to other scales so that models on other scales can properly account for any variability and uncertainty in these variables. This will allow a consistent representation of information across all scales in the system. These new methods must be developed in a manner that permits different representations of uncertainty information to be used in high-performance simulation.

Another challenge arises from the fact that translation between scales sometimes involves the solution of computationally intense inverse problems. For example, the problem of down-scaling consists of finding fine-scale features of a process or media with given coarse-scale features and prescribed prior information, which may be deterministic or statistical in nature.

Dealing with model uncertainty. The uncertainty in model predictions often has significant contributions arising from the modeling itself (Kitanidis 1988; Christakos et al. 1990; Mosleh et al. 1994; Neuman 2003). In subsurface models, this typically occurs when not all relevant structure and physics are known or properly described in a model. Research is needed to develop new methods to represent and quantify this type of model uncertainty and to develop high-performance implementations of existing approaches (e.g., the representation of model uncertainty in terms of parameters and data accompanied by uncertainty distributions) sampling various forms of a model from a family of models, or the use of random matrix theory (Soize 2000). Significantly better methods are also needed to infer a model structure along with its uncertainty or to choose between various possibilities to model a given system and the available data for that system. Model structure affects the global stability properties of a model (e.g., the

predictability of its behavior), and matching the stability properties of a mathematical model of a complex system to the stability of the actual system is a difficult mathematical and computational challenge. Additionally, research is needed to advance statistical cross-validation methods that attempt to effectively evaluate the reliability of models by evaluating their predictive ability against actual data (Golub et al. 1979; Box 1980).

The challenge of developing high-performance computational statistics tools and software. The last three decades have seen a tremendous increase in the capability of solving computationally demanding models. This progress is the result of both the evolutionary developments in high-performance computers and advances in the development of computational algorithms that fully use the capabilities of high-performance computers. Given the importance of uncertainty analysis to predictive science and engineering applications, there is a strong need for an analogous development of high-performance computational statistical algorithms and software. For example, Bayesian inference methods have the potential to solve the challenging parameter estimation, uncertainty propagation, and model validation problems encountered in subsurface sciences under significantly fewer restrictions than some traditional statistics approaches (Sivia et al. 2006; Malinverno 2002; Hoeting et al. 1999; Kass et al. 1995; Kennedy et al. 2006; Oakley et al. 2004; O'Hagan 2006). However, significant algorithmic developments are needed to make the Bayesian approach computationally feasible and efficient on high-performance computers.

Analyzing the creation and propagation of uncertainty

A key challenge is the development and implementation of mathematical frameworks for analyzing the propagation of uncertainties through complex multi-physics, multi-scale subsurface systems. Mathematical frameworks must include the effects of data, parameter, and model uncertainty as well as errors produced during simulations and, therefore, must deal with uncertainties described with a variety of statistical and deterministic representations. The mathematical approaches must allow for propagation, cancellation, and accumulation of effects to produce an accurate quantification of the uncertainty in analysis results. Mathematical techniques for uncertainty analysis must accommodate abrupt and even discontinuous behavior in space and time (e.g., sharp or discontinuous material changes at specified locations in space and discontinuous model changes in time arising from specified conditions) as well as multiple bifurcations present in specified data and parameter ranges. An associated challenge is to develop implementations of the mathematical frameworks that are efficient for high-performance computing platforms because of the intensive computational demands of the analysis.

Dealing with the computational challenges of analyzing the propagation of uncertainty. The propagation of uncertainty through a complex system presents severe computational challenges that must be overcome for uncertainty quantification to fulfill its necessary role in predictive science and engineering. Indeed, computational overhead is sufficiently severe that a detailed assessment of the uncertainty associated with flow predictions is rarely performed in current practice. To illustrate, we consider computational issues for four standard approaches.

1. One approach for quantifying uncertainty is the use of random sampling techniques (e.g., Monte-Carlo methods (Zhang 2002; Rubin 2003; Helton et al. 2003). For large-scale problems, however, the computational requirements of random sampling are enormous (see Vaughn et al. 2000; Helton et al. 2000a, b). The situation is even worse when estimating probabilities of rare events. The computational challenges include organizing the efficient solution of multiple simulations of a subsurface system and developing fast random sampling techniques that reduce the number of simulations required to carry out an accurate statistical analysis.
2. Another approach is based on representing uncertain model inputs and the solution fields in terms of an infinite spectral representation (e.g., a Karhunen-Loéve series or polynomial chaos expansion (Ghanem et al. 1991) and then solving for the spectral coefficients. The coefficients may be computed intrusively, by deriving deterministic equations for the coefficients (Ghanem et al. 1991; Ghanem 1998; Zhang et al. 2004), or non-intrusively by random sampling (Tatang et al. 1998; Le Maître et al. 2002). Either approach entails enormous computational demands and again introduces the need for fundamental research in the organization of spectral approaches for efficient implementation on high-performance computers.
3. In appropriate circumstances, the quantification of uncertainty can be attacked directly by deriving equations for statistical quantities (e.g., probability density functions or moments) that represent information about the quantities of interest (Kitanidis 1988; Graham et al. 1989; Zhang 1998, 2002). Computing solutions of these equations yields information about uncertainty directly (e.g., the expected value of oil/gas recovery versus time and the uncertainty about the expected value). However, solving these equations for large-scale models is a computationally intensive problem.
4. Variational analysis based on duality, adjoint equations, and generalized Green's functions present a traditional deterministic approach for propagating uncertainty through a complex system (Cacuci et al. 1981a, b; Marchuk 1995; Eriksson et al. 1995; Marchuk 1996, Lanczos 1996, Estep et al. 2000; Giles et al. 2002; Bangerth et al. 2003; Estep et al. 2006). The computational demands and efficient implementation in high-performance computing environments, including storage of forward solutions for forming the adjoint problems and solution of the adjoint problems, present severe challenges that require further investigation.

The computational demands associated with the analysis of the propagation and analysis of uncertainty through a complex subsurface system elevates the challenge of developing high-performance software for such analyses "from the ground floor," as opposed to *ad hoc* implementations on top of existing high-performance simulation frameworks.

The development of fast random sampling methods. The analysis of uncertainty propagating through a subsurface system is distinguished from purely statistical analysis of data resulting from physical experiments by the assumption that mathematical models capture information about the physical processes. It is important to develop new random sampling methods that make use of available information about the physical processes to improve the accuracy and the

rate of convergence of the statistical analysis and that can be implemented efficiently on high-performance computers. A classic example is to use derivative properties of a mathematical model (e.g., obtained via solution of an adjoint problem) to construct variance reduction and importance sampling techniques for Monte-Carlo methods (Estep et al. 2006a). Further research is needed for the development of physics-based surrogate models for the purpose of accelerating the convergence of random sampling methods (e.g., preconditioning of Markov chain Monte Carlo methods or use in importance sampling). Both approaches raise serious issues for implementation on parallel computers.

The creation and propagation of uncertainty in operator decomposition methods for subsurface simulations. Operator decomposition (Marchuk 1990; Hundsdorfer et al. 2003) is a very widely used technique for computing numerical solutions of multi-physics, multi-scale subsurface models (Arbogast et al. 1996). While operator decomposition yields computational and modeling benefits, it also introduces new forms of uncertainty because it discretizes the instantaneous interaction between different physical processes acting on different scales (Ropp et al. 2005; Estep et al. 2007). It also elevates difficult issues involved with the characterization of data, model, and parametric uncertainty because uncertainty is passed between components of the system model. The challenge is to develop and implement mathematical frameworks for analyzing the production and propagation of uncertainties through an operator decomposition solution of a subsurface system that provides accurate quantification of uncertainty but that can still be implemented efficiently in high-performance computing environments.

Control of uncertainty

Distributing available resources in order to analyze and predict the behavior of a complex geological subsurface system both accurately and efficiently is a critical computational challenge. There are enormous costs associated with experiment and field measurements as well as severe constraints on the available resolution in computational simulations even on petascale computers. A challenge is to develop and implement frameworks that can adaptively guide the distribution of resources to efficiently reduce the uncertainty in a model analysis. The development of truly efficient adaptive frameworks would greatly increase the accuracy in model analysis and system prediction that can be achieved with a given amount of resources.

Adaptive resolution in model simulation. Adaptive numerical solution of partial differential equations takes on new importance in the context of uncertainty analysis and quantification for models of subsurface systems and poses new challenges. In addition to the traditional challenge of efficient solution of computational problems that stress even petascale computers, uncertainty analysis requires the determination of families of solutions for which numerical errors may vary significantly. Another challenge is to develop underlying estimates for guiding adaptivity that allow control of the error in specific quantities of interest computed from solutions of multi-scale, multi-physics subsurface models and that can also be implemented efficiently in high-performance computing environments. For example, approaches based on adjoint equations

(Eriksson et al. 1995; Estep et al. 2000; Becker et al. 2001; Giles et al. 2003; Bangerth et al. 2003) involve significant memory and parallel efficiency issues that have yet to be overcome.

Adaptive representations, sampling and experimental feedback. A critical part of the goal of controlling uncertainty is providing quantitative feedback in order to guide further experiment and field measurement. The challenge is to develop efficient computational techniques for distinguishing those parameters and data that are most important in terms of affecting the uncertainty of computed information and to give indication of the range and nature of allowable error in those inputs. This information can then be used to guide further experiment and measurement or to compute better representations from available data. For example, in the case of spectral representation of uncertain parameters, the order, the number of degrees of freedom, and the basis functions can be adapted depending on the distribution of a parameter. This will require advances in adaptive statistical sampling methods as well as the capability to propagate uncertainties with mixed representations through the model. A related aspect is the fact that accuracy and resolution requirements can vary throughout a simulation and, as a consequence, efficiency dictates allowing methods to change the sampling and resolution requirements as needed depending on what is required at the time and location in the simulation (Le Maître et al. 2004; Wan et al. 2005). Alternatively, if the accumulation of uncertainty renders some observables meaningless, it may be advantageous to switch to different observables that are more meaningful.

Adaptive model selection. Another challenge is to develop mathematical and computational methodologies for allowing the model representation itself to change adaptively. For example in situations in which computational resources are the primary barrier to the fidelity of the physical description, a more comprehensive model could be selected when a simulation progresses into a phase where the currently used models are outside its limits of applicability. However, combining physical descriptions in this way poses severe challenges for efficient implementation on high-performance computers.

Sensitivity analysis of complex subsurface systems

Sensitivity analysis, which is the study of the dependence of the uncertainty in analysis results on the contributions from uncertainties in the individual data and parameter inputs, is a fundamental part of uncertainty analysis (Cacuci et al. 2004; Frey et al. 2002; Helton et al. 2006; Ionescu-Bujor et al. 2004; Saltelli et al. 2000; Saltelli et al. 2004). However, there are significant challenges regarding the efficient implementation of sensitivity analysis tools on high-performance computers. For example, one common class, so-called local methods, are based on estimating partial derivatives of analysis results with respect to uncertain analysis inputs. The generation of the required partial derivatives can be very demanding computationally (e.g. adjoint methods [Lanczos 1996; Marchuk 1995; Marchuk et al. 1996; Cacuci et al. 1980; Cacuci 1981a, b], derivative adding compilers [Griewank 2000]) and pose difficulties for efficient implementation on parallel computers. Owing to the importance of local methods in both

sensitivity analysis and optimization procedures, it is important that efficient methods for their implementation in analysis of computationally demanding models be developed.

Another class of methods, so-called global methods, involves an exploration of the entire input space. For example, one approach is to generate a map between analysis inputs and analysis results (e.g. with Latin Hypercube Sampling [Helton et al. 2003; McKay et al. 1979]) and then to explore this map with a variety of procedures (e.g., correlation analysis, regression analysis and non-parametric regression analysis [Cacuci 2004; Frey et al. 2002; Helton et al. 2006]). Again, the challenge is to develop efficient computational frameworks on high-performance platforms. For example, Sobol' variance decomposition (Saltelli et al. 1999; Sobol' 1993) provides information-rich sensitivity analysis results but computationally efficient methods are needed for its implementation. The development of non-parametric regression analysis tools (Bowman et al. 1997; Hastie et al. 1990; Hastie et al. 2001; Ruppert et al. 2003; Schimek, 2000; Simonoff 1996; Storlie 2006) would be an important contribution to both sensitivity analysis and the use of surrogate models in the analysis of complex systems.

Development of high-performance tools and libraries for uncertainty and sensitivity analysis

While the previous sections focused on advances needed to make uncertainty and sensitivity analysis feasible for the most complex of problems, it is important not to lose sight of the tremendous amount of algorithmic development that has already taken place in this field. It is fair to say that many problems are already amenable to uncertainty and sensitivity analysis with methods that are available today. Unfortunately, adding uncertainty and sensitivity analysis capabilities to existing legacy codes in a fashion that is efficient on high-performance computers is a very daunting challenge. There is a critical need to develop tools and libraries that facilitate both 1) the development of high-performance model simulation software that includes tools for uncertainty and sensitivity analysis as a fundamental components and 2) the (automated) retrofitting of legacy codes to make them amenable to uncertainty and sensitivity analysis. Such software must provide a variety of tools for uncertainty and sensitivity analysis and the capability for a user to combine various tools. These tools and libraries need to be widely promoted and made available to the broad scientific community.

Direct support is needed to foster interdisciplinary collaborations between the application scientists and engineers, the mathematicians and statisticians developing new algorithms, and the computer scientists implementing algorithms on high-performance platforms.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Research addressing the computational challenges described above will provide scientists and engineers with a great increase in the ability to fully use high-performance computational platforms to address fundamental scientific questions and issues of national importance involving computationally intensive mathematical models.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

Research in the indicated areas will have a positive impact on the subsurface sciences by enhancing 1) the resolution and numerical efficiency of analyses involving subsurface systems, 2) enhancing the resolution and quality of uncertainty and sensitivity assessments performed as necessary and integral parts of such analyses, and 3) increasing the quality, appropriateness and usefulness of decisions made on the basis of such analyses.

TIME FRAME

Positive effects on computational sciences and the subsurface sciences from the indicated research could be obtained in 2 to 6 years.

REFERENCES

Arbogast, T., S. Bryant, C. Dawson, F. Saaf, C. Wang, and M.F. Wheeler, 1996. "Computational methods for multi-phase flow and reactive transport problems arising in subsurface contaminant remediation," *J. Comp. Appl. Math.* 74, 19-32.

Ayyub, B. M. (ed.), 1997. *Uncertainty Modeling and Analysis in Civil Engineering*, Boca Raton, FL: CRC Press.

Ayyub, B. M., 2001. *Elicitation of Expert Opinions for Uncertainty and Risks*, Boca Raton, FL:CRC Press.

Babuska, I., and T. Strouboulis, 2001. *The Finite Element Method and its Reliability*, New York, NY: Oxford University Press.

Babuska, I., and T. Stroubulis, 2002. "Can We Trust the Computational Analysis of Engineering Problems, " *Math. Model. Numer. Simul. Contin. Mech.*, edited by I. Babuska, T. Miyoshi, P. G. Ciarlet, 169-183.

Bangerth, W. and R. Rannacher, 2003. *Adaptive Finite Element Methods for Differential Equations*, Boston: Birkhauser.

Becker, R. and R. Rannacher, 2001. "An Optimal Control Approach to a Posteriori Error Estimation in Finite Element Methods," *Acta Numerica,* 1–102.

Bowman, A. W. and A. Azzalini, 1997. *Applied Smoothing Techniques for Data Analysis*, Clarendon: Oxford.

Box, G. E. P., 1980. "Sampling and Bayes' inference in scientific modelling and robustness," J. Roy. Stat. Soc. A 143, 383-430.

Budnitz, R., J., G. Apostolakis, D. M. Boore, L. S. Cluff, K. J. Coppersmith, C. A. Cornell, and P. A. Morris, 1998. "Use of Technical Expert Panels: Applications to Probabilistic Seismic Hazard Analysis," *Risk Analysis* 18, 463-469.

Cacuci, D. G ,C. F. Weber, E. M. Oblow, and J. H. Marable, 1980. "Sensitivity Theory for General Systems of Nonlinear Equations," *Nuc. Sci. Eng.* 75, 88-110.

Cacuci, D. G., 1981. "Sensitivity Theory for Nonlinear Systems. I. Nonlinear Functional Analysis Approach," *J. Math. Phy*. 22, 2794-2802.

Cacuci, D. G., 1981. "Sensitivity Theory for Nonlinear Systems. II. Extensions to Additional Classes of Responses," *J. Math. Phys*. 22, 2803-2812.

Cacuci, D. G. and M. Ionescu-Bujor, 2004. "A Comparative Review of Sensitivity and Uncertainty Analysis of Large-Scale Systems--II: Statistical Methods," *Nucl. Sci. Eng*. 147, 204-217.

Christakos, G., 1990. "A Bayesian Maximum Entropy View to the Spatial Estimation Problem," *Math. Geol*. 22,763-777.

Christie, M. A., J. Glimm, J. W. Grove, D. M. Higdon, D. H. Sharp, and M. M. Wood-Schultz, 2005. "Error Analysis and Simulations of Complex Phenomena," *Los Alamos Science* 29, 6-25.

Cooke, R. M. and L. H. J. Goossens, 2004."Expert Judgement Elicitation for Risk Assessment of Critical Infrastructures," *J. Risk Research* 7, 643-656.

Eriksson, K., D. Estep, P. Hansbo, and C. Johnson, 1995. "Introduction to Adaptive Methods for Differential Equations," *Acta Numerica*, 105–158.

Estep, D.J., M.G. Larson, and R.D. Williams, 2000. "Estimating the Error of Numerical Solutions of Systems of Reaction-Diffusion Equations," *Mem. Amer. Math. Soc*, 146,1–109.

Estep, D. and D. Neckels, 2006. "Fast and Reliable Methods for Determining the Evolution of Uncertain Parameters in Differential Equations," *J. Comput. Phy*. 213, 530-556.

Estep, D.,V. Ginting, D. Ropp, J. Shadid, and S. Tavener, 2007. "An A Posteriori Analysis of Operator Splitting", submitted to SIAM J. Numer. Anal.

Frey, H.C. and S. R. Patil, 2002. "Identification and Review of Sensitivity Analysis Methods," *Risk Analysis* 22, 553-578.

Garthwaite, P.H., J. B. Kadane, and A. O'Hagan, 2005. "Statistical Methods for Eliciting Probability Distributions," *J. Amer. Stat. Assoc*. 100, 680-700.

Ghanem, R. and P. Spanos, 1991. *Stochastic Finite Elements: A Spectral Approach*, New York: Springer Verlag.

Ghanem, R., 1998. "Probabilistic Characterization of Transport in Heterogeneous Media," *Comput. Methods in Applied Mech. Engrg*. 158, 199 – 220.

Giles, M. and E. Suli, 2002. "Adjoint Methods for Pdes: A Posteriori Error Analysis and Postprocessing by Duality," *Acta Numerica* 145–236.

Golub, G. H., M. Heath, and G. Wahba, 1979. "Generalized cross-validation as a method of choosing a good ridge parameter, " *Technometrics 21*, 215-223.

Graham, W.D. and D. McLaughlin, 1989. "Stochastic analysis of nonstationary subsurface solute transport: 1. Unconditional moments, " *Water Resour. Res.* 25, 215 – 232.

Griewank, A., 2000, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, Philadelphia, PA: SIAM.

Halpern, J. Y., 2003. *Reasoning about Uncertainty*, Cambridge, MA: MIT Press.

Hastie, T.J. and R. J. Tibshirani, 1990. *Generalized Additive Models*, London: Chapman & Hall.

Hastie, T., R. Tibshirani, and J. Friedman, 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York, NY: Springer.

Helton, J. C., J. E. Bean, K. Economy, J. W. Garner, R. J. MacKinnon, J. Miller, J. D. Schreiber, and P. Vaughn, 2000. "Uncertainty and Sensitivity Analysis for Two-Phase Flow in the Vicinity of the Repository in the 1996 Performance Assessment for the Waste Isolation Pilot Plant: Undisturbed Conditions," *Reliability Eng. System Safety* 69, 227-261.

Helton, J. C., J. E. Bean, K. Economy, J. W. Garner, R. J. MacKinnon, J. Miller, J. D. Schreiber, and P. Vaughn, 2000. "Uncertainty and Sensitivity Analysis for Two-Phase Flow in the Vicinity of the Repository in the 1996 Performance Assessment for the Waste Isolation Pilot Plant: Disturbed Conditions," Reliability Eng. System Safety 69, 263-304.

Helton, J. C. and F. J. Davis, 2003. "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Eng. System Safety* 81, 23-69.

Helton, J. C., J. D. Johnson, and W. L. Oberkampf, 2004. "An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions," *Reliability Eng. System Safety* 85, 39-71.

Helton, J. C., J. D. Johnson, C. J. Sallaberry, and C. B. Storlie, 2006. "Survey of Sampling-Based Methods for Uncertainty and Sensitivity Analysis," *Reliability Eng. System Safety* 91, 1175-1209.

Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky, 1999. “Bayesian Model Averaging: A Tutorial,” *Stat.l Science* 14, 382 – 417.

Hundsdorfer, W. and J. G. Verwer, 2003. *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, New York:Springer.

Ionescu-Bujor, M. and D. G. Cacuci, 2004. "A Comparative Review of Sensitivity and Uncertainty Analysis of Large-Scale Systems--I: Deterministic Methods," *Nuc. Sci. Eng.* 147, 189-2003.

Kass, R.E. and A.E. Raftery, 1995. “Bayes Factors”, *J. Amer. Stat. Assoc.*, 90(430),773 – 795.

Kitanidis, P. K., 1988. "The concept of predictive probability and a simple test for geostatistical model validation," in *Consequences of Spatial Variability in Aquifer Properties and Data Limitations for Groundwater Modelling Practice*, Peck, A. , Gorelick, S., DeMarsily, G. Foster, S., and Kovalevsky, V. [edt.], Int. Assoc. of Hydrol. Sci., IAHS Publication No 175, 178-190, URL http://www.stanford.edu/group/peterk/pdf/IAHS175.pdf.

Kitanidis, P.K., 1988. "Prediction by the method of moments of transport in a heterogeneous formation," *J. Hydrol.* 102, 453 – 473.

Kennedy, M.C., C.W. Anderson, S. Conti, and A. O'Hagan, 2006. "Case Studies in Gaussian Process Modelling of Computer Codes," *Reliability Engin. Sys. Safety* 91, 1301-1309.

Klir, G. J., 2004. "Generalized Information Theory: Aims, Results, and Open Problems," *Reliability Eng. System Safety* 85, 21-38.

Klir, G. J., 2006. *Uncertainty and Information: Foundations of Generalized Information Theory*, Hoboken, NJ: Wiley.

Lanczos, C., 1996. *Linear Differential Operators*, Philadelphia: SIAM.

Le Maître O.P., Reagan, M.T., Najm, H.N., Ghanem. R.G. and Knio, O.M., 2002. "A Stochastic Projection Method for Fluid Flow. II. Random Process", *J. Comput. Phys.* 181, 9 – 44.

Le Maître, O.P., H.N. Najm, R.G. Ghanem, and O.M. Knio, 2004. “Multi-resolution Analysis of Wiener-type Uncertainty Propagation Schemes”, *J. Comput. Phy.*, 197,502 – 531.

Malinverno, A., 2002. “Parsimonious Bayesian Markov Chain Monte Carlo Inversion in a Nonlinear Geophysical Problem”, *Geophys. J. Int.* 151,675 – 688.

Marchuk, G. I., 1990. "Splitting and Alternating Direction Methods," *Handbook of Numerical Analysis*, North Holland, Amsterdam, 197–464.

Marchuk, G.I., 1995. *Adjoint Equations and Analysis of Complex Systems*, Boston: Kluwer.

Marchuk, G.I., V. I. Agoshkov, and V. P. Shutyaev, 1996. *Adjoint Equations and Perturbation Algorithms in Nonlinear Problems*, New York: CRC Press.

McKay, M.D., R. J. Beckman, and W. J. Conover, 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21, 239-245.

McKay, M. and M. Meyer, 2000."Critique of and Limitations on the Use of Expert Judgements in Accident Consequence Uncertainty Analysis*," Rad. Prot. Dosimetry* 90, 325-330.

Mosleh, A.,N. Siu, C. Smidts, and C. Liu, 1994. "Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification," *NUREG/CP-0138*. Washington, D.C.

Neuman, S.P., 2003. "Maximum Likelihood Bayesian Averaging of Uncertain Model Predictions," *Stoch. Envir. Res. & Risk Assess.* 17, 291-305.

Nikolaidis, E., D. M. Ghiocel, and S. Singhal (eds), 2004.*Engineering Design Reliability Handbook*, Boca Raton, FL: CRC Press.

A. O'Hagan, 2006. "Bayesian Analysis of Computer Code Outputs: A Tutorial," *Reliability Engin. Sys. Safety* 91, 1290-1300.

Oakley, J.E. and A. O'Hagan, 2004. "Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach," J. Royal Stat. Soc. B66, 751-769.

Oberkampf, W.L., S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin, 2002. "Error and Uncertainty in Modeling and Simulation," *Reliability Eng. System Safety 75*, 333-357.

Oberkampf, W.L. and T. G. Trucano, 2002. "Verification and Validation in Computational Fluid Dynamics," *Prog. Aerospace Sci*. 38, 209-272.

Oberkampf, W.L., T. G. Trucano, and C. Hirsch, 2004. "Verification, Validation, and Predictive Capability in Computational Engineering and Physics," *Appl. Mech. Rev. 57*, 345-384.

Roache, P.J., 1998. *Verification and Validation in Computational Science and Engineering*, Albuquerque, NM: Hermosa Publishers.

Ropp, D. L. and Shadid, J. N., 2005. "Stability of operator splitting methods for systems with indefinite operators: reaction-diffusion systems," *J. Comput. Phys.* 203, 449-466.

Ross, T.J., J. M. Booker, and W.J. Parkinson (eds.), 2002. *Fuzzy Logic and Probability Applications: Bridging the Gap*, Philadelphia, PA:SIAM.

Ross, T.J., 2004. *Fuzzy Logic with Engineering Applications*, 2nd ed. New York, NY: Wiley.

Rubin, Y., 2003. *Applied Stochastic Hydrogeology*, New York: Oxford University Press.

Ruppert, D., M. P. Ward, and R. J. Carroll, 2003. *Semiparametric Regression*, Cambridge: Cambridge University Press.

Saltelli, A., S. Tarantola, and K. P.-S. Chan, 1999. "A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output," *Technometrics* 41, 39-56.

Saltelli, A., K. Chan, and E.M. Scott (eds), 2000. *Sensitivity Analysis*, New York, NY: Wiley.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto, 2004. *Sensitivity Analysis in Practice*, New York, NY: Wiley.

Schimek, M.G., 2000. *Smoothing and Regression: Approaches, Computation, and Application*, New York, NY: Wiley.

Simonoff, J.S., 1996. *Smoothing Methods in Statistics*, New York, NY: Springer-Verlag.

Sivia, D. and J. Skilling, 2006. *Data Analysis: A Bayesian Tutorial (2nd edition)*, New York: Oxford University Press.

Sobol, I.M., 1993. "Sensitivity Analysis for Nonlinear Mathematical Models," *Math. Mod. Comput. Exper*. 1, 407-414.

Soize, C., 2000. "A Non-parametric Model of Random Uncertainties for Reduced Matrix Models in Structural Dynamics," *Probabil. Eng. Mech*. 15(3),277–94.

Storlie, C.B. and J. C. Helton, 2006. "Multiple Predictor Smoothing Methods for Sensitivity Analysis," Sandia National Laboratories, Albuquerque, NM SAND2006-4693.

Tatang, M.A., W. Pan, R.G. Prinn, and G. J. McRae, 1997. "An efficient method for parametric uncertainty analysis of numerical geophysical models," *J. Geophy. Res., 102* (D18), 21925-21931.

Trucano, T.G., L. P. Swiler, T. Igusa, W. L. Oberkampf, and M. Pilch, 2006. "Calibration, Validation, and Sensitivity Analysis: What's What,*" Reliability Eng. System Safety* 91, 1331-1357.

U.S. DOE (U.S. Department of Energy), 2002. "Yucca Mountain Science and Engineering Report," U.S. Department of Energy, Las Vegas, NV DOE/RW-0539-1, Rev. 1.

Vaughn, P., J. E. Bean, J. C. Helton, M. E. Lord, R. J. MacKinnon, and J. D. Schreiber, 2000."Representation of Two-Phase Flow in the Vicinity of the Repository in the 1996

Performance Assessment for the Waste Isolation Pilot Plant," *Reliability Eng. System Safety* 69, 205-226.

Wagner, R.L., 2003. "Science, Uncertainty and Risk: The Problem of Complex Phenomena," *APS News* 12, 8.

Wan, X. and G.E. Karniadakis, 2005. "An Adaptive Multi-element Generalized Polynomial Chaos Method for Stochastic Differential Equations," *J. Comput. Phy*., 209, 617 – 642.

Zhang, D., 2002. *Stochastic Methods for Flow in Porous Media: Coping with Uncertainties*, San Diego, CA: Academic Press.

Zhang, D. and Z. Lu, 2004. "An Efficient, High-Order Perturbation Approach for Flow in Random Porous Media via Karhunen-Loeve and Polynomial Expansions, " *J. of Comp. Phy.*, 194, 773-794.

METRICS AND BENCHMARKS FOR VERIFICATION AND VALIDATION

Coordinator: Richard Drake, Sandia National Laboratory

Contributors: Donald Estep, Colorado State University, Jon Helton, Sandia National Laboratory

ABSTRACT

Domain specific verification and validation testing establish the credibility of application codes for accuracy and reliability by performing. Developing the problem descriptions and methods of assessment are difficult and time consuming; therefore, a common database of metrics and benchmarks could greatly improve computational subsurface application verification and validation and the assessment of credibility. The challenge is to produce techniques and benchmark problems for verification and validation specific to subsurface science applications.

EXECUTIVE SUMMARY

Applying verification and validation techniques to application codes is the primary way to assess the credibility for accuracy and reliability of simulation results. Credibility plays a key role in many high-consequence, decision-making processes, such as waste storage site assessment and contaminant remediation planning. However, the development, documentation, and communication of metrics and benchmark problems for verification and validation in the subsurface sciences have not been addressed in a systematic format containing common techniques and a database of problems.

The challenge is the development of metrics and benchmarks for verification and validation using problems and metrics relevant to subsurface science. Furthermore, the metrics should apply to quantities important for decision-making and those which have the most direct link to practical laboratory and field measurements.

SUMMARY OF RESEARCH DIRECTION

Model verification and model validation are two related, but different, concepts. Two widely used definitions are [Ref. [6], p. 3]:

Verification is the process of determining that a model implementation accurately represents the developers' conceptual description of the model and the solution to the model.

Validation is the process of determining the degree which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

In both cases, a critical component is the method used to determine accuracy, which is achieved by applying a metric (or measure) to the results. Equally important is the choice of benchmark problem, which must be applicable to and capture the essential aspects of the target physics problem and parameter regime as well as the available data from the physical problem. Though vital to future decision-making, the development, documentation, and communication of metrics and benchmark problems for verification and validation in the subsurface sciences has not been addressed in a systematic format containing common techniques and a database of problems. There are examples of efforts to develop benchmark databases for verification and validation in the world of computational fluid dynamics, such as the National Agency for Finite Element Methods and Standards [1]; the National Project for Applications-Oriented Research in Computation Fluid Dynamics (NPARC) Alliance validation database [2]; and the European Research Community on Flow, Turbulence, and Combustion validation collection [3]. There have also been efforts to provide definitions, processes and case studies for verification and validation, such as in [4] and in [5]. Additional discussion is available in several recent surveys [7-13].

The challenge is the development of metrics and benchmarks for verification and validation using problems and metrics relevant to subsurface science. Furthermore, the metrics should apply to quantities important for decision-making and those that have the most direct link to practical laboratory and field measurements.

A successful benchmark database would cover a range of problems important to a specific subsurface analysis, be fully documented with references, provide the metrics with which to determine accuracy, fully disclose the characteristics of the data (for validation problems), and use common methods of describing the aspects of the problem and for storing any associated data.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Development of high-quality benchmark problems demands significant time and expertise. Especially challenging for problems in subsurface sciences is that analysis is required on spatial and temporal scales, which precludes the traditional validation approach of direct comparison with experimental results. Collection and communication of a database of benchmarks requires the cooperation and participation by many members of the scientific community, both from a technical and a policy perspective. The peer review process must be employed to maintain quality and relevance but requires contributions from many. Because of all this, a big challenge will be to maintain the commitment and resources to develop and deploy such a database.

Once an application adopts a fair number of benchmark problems, the computational resources needed to execute them can quickly become overwhelming. Verification can involve mesh convergence studies, order of accuracy calculations, and sensitivity and uncertainty analyses, all of which require repeated execution. Similarly, validation problems can require fine-grained simulations and repeated calculations for uncertainty analysis.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The computational science community will benefit from a thrust towards improved verification and validation in the form of standards and examples for developing benchmark problems and the importance and application of metrics. The ability to apply the benchmarks to enhance credibility of application codes will be recognized and spur efforts in other computational science areas. More technical aspects will be methods for dealing with uncertainty in the models and experimental data.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

This work will provide guidelines and strategies for developing verification and validation benchmark problems in subsurface sciences, including specific issues and challenges related to these problems (e.g., multi-scale heterogeneity). Use of benchmark problems that have been developed will provide a standard and basis for comparison of application codes and the development of credible applications. Finally, a standard benchmark set will provide an additional device for communication with decision makers.

TIME FRAME

Individual benchmarks can be applied to an application code immediately after the problem is developed. This will impact the development of the application and/or help improve its credibility. Once a benchmark database becomes large enough and is recognized in the community, it will impact the development and qualification of application codes. This may take several years. Finally, as the database becomes widely used and understood, it will impact the ability of decision-makers to evaluate the credibility of differing application codes. This would most likely take 5 to 10 years.

REFERENCES

[1] NAFEMS, "NAFEMS Website," National Agency for Finite Element Methods and Standards, 2006, www.NAFEMS.org.

[2] NPARC, "CFD Verification &Validation: NPARC Alliance," NPARC Alliance, 2000, http://www.grc.nasa.gov/WWW/wind/valid/homepage.html.

[3] ERCOFTAC, "Portal to Fluid Dynamics Database Resources," European Research Community on Flow, *Turbulence and Combustion*, 2000, http://ercoftac.mech.surrey.ac.uk.

[4] P. Knupp and K. Salari, *Verification of Computer Codes in Computational Science and Engineering*. Boca Raton, FL: Chapman and Hall/CRC, 2002.

[5] W. L. Oberkampf, T. G. Trucano, and C. Hirsch, "Verification, Validation, and Predictive Capability in Computational Engineering and Physics," *Applied Mechanics Review*, vol. 57, pp. 345-384, 2004.

[6] American Institute of Aeronautics and Astronautics, "AIAA Guide for the Verification and Validation of Computational Fluid Dynamics Simulations," American Institute of Aeronautics and Astronautics, Reston, VA AIAA G-077-1998, 1998.

[7] I. Babuska and J. T. Oden, "Verification and Validation in Computational Engineering and Science: Basic Concepts," *Computer Methods in Applied Mechanics and Engineering*, vol. 193, pp. 4057-4066, 2004.

[8] W. L. Oberkampf, S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin, "Error and Uncertainty in Modeling and Simulation," *Reliability Engineering and System Safety*, vol. 75, pp. 333-357, 2002.

[9] W. L. Oberkampf and T. G. Trucano, "Verification and Validation in Computational Fluid Dynamics," *Progress in Aerospace Sciences*, vol. 38, pp. 209-272, 2002.

[10] P. J. Roache, "Quantification of Uncertainty in Computational Fluid Dynamics," *Annual Review of Fluid Mechanics*, vol. 29, pp. 123-160, 1997.

[11] P. J. Roache, *Verification and Validation in Computational Science and Engineering*. Albuquerque, NM: Hermosa Publishers, 1998.

[12] P. J. Roache, "Building PDE Codes to be Verifiable and Validatable," *Computing in Science & Engineering*, vol. 6, pp. 30-38, 2004.

[13] T. G. Trucano, L. P. Swiler, T. Igusa, W. L. Oberkampf, and M. Pilch, "Calibration, Validation, and Sensitivity Analysis: What's What," *Reliability Engineering and System Safety*, vol. 91, pp. 1331-1357, 2006.

CARBON SEQUESTRATION

IMPROVED COMPUTATION OF COUPLED MULTI-SCALE PROCESS MODELS FOR CARBON SEQUESTRATION

Diana Bacon, Pacific Northwest National Laboratory; John Weare, University of Californial San Diego; Glenn Hammond, Pacific Northwest National Laboratory

ABSTRACT

Increasing levels of carbon dioxide (CO_2) in the atmosphere have been associated with global warming. One promising mitigation strategy is injection and storage of this greenhouse gas in deep geological formations (carbon sequestration). A technical challenge is to ensure that this injected gas remains underground for hundreds to thousands of years. Prediction of the long-term movement and risk of escape of vast amounts of CO_2 will require assimilation of huge datasets into a new generation computer simulator, incorporating coupled models that capture the physical and chemical evolution of the system, including rock deformation, multi-phase fluid flow, and reaction in heterogeneous rocks and deep saline aquifers. Current simulators are inadequate, and require new methods of site characterization, process model coupling, scaling laws to account for spatial heterogeneity, and the development of computational infrastructure, numerical algorithms, and methods for parallel computing. Investment in this area would also yield dramatic improvements to the computational sciences as a whole, and would have broader impact in other fields of subsurface science, such as improved simulation of underground contaminant migration and remediation.

EXECUTIVE SUMMARY

Given the significant increase in CO_2 in the atmosphere associated with burning of fossil fuels and the important role that carbon sequestration could play in remediation, significant investment is required to predict the long-term fate of CO_2 pumped into subsurface geologic formations. Accurate and efficient numerical models are needed to predict CO_2 storage capabilities, control injection, and assess risk of leakage to the surface. Fundamental improvements over existing simulation models are needed in several key areas. The subsurface is extremely heterogeneous and new methods of characterization, including four-dimensional seismic and satellite imaging methods, are needed to detect subsurface deformation and potential leakage. New simulators must be able to incorporate this information and provide feedback for gathering it. Individual process models must be discretized numerically so as to preserve pertinent physical and chemical principles. Near well-bore environments are critical to model accurately as these zones will likely change dramatically upon injection of CO_2, and abandoned wells create a risk of leakage. Accurate simulation in the presence of heterogeneities will require multi-scale discretizations and computational grids that adapt dynamically to zones of rapid changes. Sequestration of CO_2 in the subsurface is inherently a complex problem, and accurate simulation will require new algorithms for simultaneous coupling of multiple process models. Coupled and adaptive, multi-physics, multi-scale, multi-resolution processes present unique computational infrastructure

demands because of their scale and heterogeneity in space and time. Significant impacts of this research will extend beyond the carbon sequestration problem. Improved parallelization methods, numerical algorithms, data assimilation techniques, and scaling methods for heterogeneous systems will have dramatic impact on the computational sciences as a whole, and computational subsurface science will benefit more broadly because of the similarity of processes that affect sequestration and the fate of near-surface contaminant flows.

SUMMARY OF RESEARCH DIRECTION

Atmospheric levels of greenhouse gases such as CO_2 have increased dramatically since the industrial revolution, largely because of the burning of fossil fuels. These gases enhance the atmosphere's ability to retain radiant heat energy, which has the potential to cause dramatic warming on a global scale. One strategy for remediation involves capturing greenhouse gases and sequestering them in underground storage sites such as petroleum reservoirs and deep, saline aquifers. The sequestration of carbon in geologic formations poses many risks and creates many scientific and engineering challenges, but may provide a key method for mitigating the rise of CO_2 in the atmosphere. This PRD involves the science and mathematics needed to develop accurate and efficient computational models and algorithms for the simulation of the processes governing the underground storage of carbon, especially CO_2. These computational tools could be used by specialists to predict the future behavior of these underground systems, to manage the injection of CO_2, to develop strategies for human intervention if the fidelity of the cap-rock seal is compromised, and to assess risks and impacts on public health.

Developing a predictive simulation capability for carbon sequestration in geologic formations is particularly challenging. It involves a wide range of complex natural chemical, physical, and geological processes. More specifically, the system involves the compositional, multi-phase flow and reactive transport of chemical components, complex phase behavior and mineralization, mechanical and thermal effects under conditions at or near regimes of instability. This occurs in a deep, difficult to characterize natural geologic environment which is extremely heterogeneous over a wide range of scales, the properties of which change under the effects of geo-mechanical deformation and fracturing in response to the movement and pressurization of fluids. The system itself is extremely large, and involves processes that evolve over thousands of years on many interconnected temporal and spatial scales.

Currently available numerical simulation packages do not couple all the needed compositional flow, reactive transport, and geo-mechanics models. A few simulators have some of these capabilities, such as the University of Texas-Austin IPARS parallel multi-phase geomechanical geochemistry simulator and the SIMWULF compositional, geomechanical, thermal simulator. During 2004, Lawrence Berkeley National Laboratory coordinated a study to test and evaluate codes to model geological sequestration, including various versions of the Lawrence Berkeley National Laboratory TOUGH, the Pacific Northwest National Laboratory STOMP, the Los Alamos National Laboratory FLOTRAN and FEHM, the Institut Français du Petrole (IFP) SIMUSCOPP, the University of Stuttgart MUFTE_UG, and the Computer Modeling Group

(CMG) of Canada GEM (Pruess et al. 2001, 2004). The study demonstrated substantial agreement between results predicted from different simulations, but also areas with only fair agreement and some significant discrepancies. Most disagreements could be traced to differences in fluid properties. However, all but one of the test cases were for one-dimensional homogeneous media (one heterogeneous two-dimensional test problem was included). While this study was extremely useful, it did not compare three-dimensional reservoirs with complex heterogeneity on different scales and an equation of state compositional flow model along with geochemical and geo-mechanical effects and non-isothermal phenomena. In summary, there appears to be no single simulator that incorporates and adequately models all of the major physical and chemical processes induced by injection of CO_2 into potential disposal reservoirs that is also coupled to a geophysical model. Moreover, most of the above simulators do not include high-fidelity algorithms that include error estimators for adaptivity.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

We identify three main classes of challenges for accurate and efficient computational simulation of underground carbon sequestration: 1) algorithms for multi-scale modeling of specific, individual processes; 2) techniques for modeling the coupling of these specific processes; and 3) computational software infrastructure tools.

Multi-scale Modeling of Processes

Most of the processes involved in carbon sequestration are not unique to this application; nevertheless, these processes tend to involve more critical behaviors and tend to require more accurate modeling because they are more tightly interconnected

Data interpretation and scaling from site characterization. Deep porous formations are difficult to characterize. They are extremely heterogeneous, with many small features such as highly permeable fractures or highly impermeable thin shale rock strata that dramatically alter the movement of fluids and dramatically affect the integrity of the cap rock to contain them. Not only is there a lack of data, but the traditional sources of data from wells is concentrated vertically along the well bore, rather than areally. Novel sources of data, such as four-dimensional seismic (space plus time) to probe the subsurface, and satellite and surface detection of CO_2 for monitoring leakage from the storage site, offer the potential to provide a reasonable characterization of the subsurface site. These also offer the ability to detect changes over time in the site as the ground compacts or swells in response to pressurization. Research is needed to allow these novel sources of data to be interpreted and used in subsurface simulation.

Physics preserving discretizations. Models of multi-phase reactive flow and transport generally take the form of coupled PDEs subject to constraints governing the behavior of the fluid pressures and velocities, chemical species concentrations, phase densities, temperature, etc. (the *solution*). The PDEs tend to have mathematical singularities and degeneracies as chemical species and solid and fluid phases form or disappear. Computer simulation requires that the

PDEs and constraints be approximated accurately through *discretization* (i.e., by reducing the infinite dimensional character of the solution to finite dimensions). However, it is imperative that this discretization preserve pertinent physical and chemical principles as reflected in the PDEs, such as mass balance (mass can neither be created nor destroyed) and the maximum principle (heat and chemical species, for example, tend to diffuse, so they cannot artificially build up at some point in space).

Current numerical methods are promising, but inadequate to fulfill the needs of carbon sequestration. Additional research is needed on many fronts, including but not limited to the following:

- Physics preserving numerical modeling of degenerate diffusion on unstructured grids, possibly using higher order and discontinuous finite difference, finite element, and finite volume methods.
- Numerical approximation of nonlinear, multi-phase, multi-component, reactive, advective processes, such as by discontinuous Galerkin and characteristic methods. The latter avoids the Courant-Levy-Friedrich constraint so that long time periods may be simulated.
- Computational techniques for modeling of nonequilibrium phase behavior for the complex equations of state expected to govern CO_2 sequestration at critical conditions.

Well models. The modeling of wells presents a serious challenge to carbon sequestration efforts. First, the injection of CO_2 gas at a well must be modeled accurately, because this is a mechanical process that drives the flow. The near well-bore environment will change drastically over time as fluids are injected. Mechanisms that increase injectivity include pressurization leading to increases in porosity or local fracturing of the ambient rock, and dissolution of rock. Second, potential leakage of CO_2 gas from abandoned wells must be modeled appropriately to allow engineering assessment of risk and design of remedial action. Currently, well models are often designed as stand-alone single well models that are unsuitable for full field applications involving the complexity envisioned above, so additional research is needed.

Natural geologic heterogeneity. Geologic formations are extremely heterogeneous on essentially all length scales. It is critical to characterize and develop modeling strategies that can deal with strong spatial variations in mineralogy, rock type and multi-phase fluid compositions, because these greatly influences estimates of sequestration potential. Regions or strata of high permeability, or even rock fractures, faults, or interconnected vugs (i.e., cavities) often exist, which become natural conduits for the movement of injected carbon. This channeling (or *fingering*) potentially causes the injected carbon to bypass large regions of the formation. Heterogeneity also affects the adsorption and chemical environment of the reaction dynamics. In addition, heterogeneity evolves in time as mineralization and matrix deformation occur.

To handle the modeling of fluid flow through geologic heterogeneity, multi-scale approaches have been developed in recent years, including multi-scale finite elements, mortar finite elements, variational multi-scale methods, and heterogeneous multi-scale methods. This work

and new approaches need to be supported, and also extended to the modeling of discrete fractures and "vuggy" systems, with concurrent geochemical reactions. Additional dual-porosity and related hierarchical distributed micro-structure models also need to be developed.

Gridding and adaptivity. The computational simulation must be monitored to determine if errors are building up in the solution. Recent work on error estimation and indication can be used to adapt the grid to reduce errors in the computed solution. This line of work must be developed especially for the approximation of degenerate problems, as arise in CO_2 sequestration, and for the approximation of time dependent problems involving thousands of years.

Multi-Physics Coupling of Processes

Modeling carbon sequestration requires the coupling of multiple submodels (multi-phase flow, reactive transport, mineral reactions, geomechanical deformation, heat flow, etc.). The difficulty is that these submodels are tightly coupled and operate at different space and time scales. Of particular importance to carbon sequestration is the coupling of geo-mechanical and geo-chemical models, because this is critical to accurately predict mineralization and cap-rock integrity.

There is a need to move beyond the simple strategies now in common use. Of particular importance is to investigate hybrid approaches in space and time that allow multiple sub-domains to interact in sophisticated ways. Each sub-domain should simulate only those processes relevant within its extent, and to a level of scale appropriate to accurate simulation. This will require significant research in the transfer of data across sub-domain boundaries. One important need is to research the appropriate form of a common interface model and definitions of consistent state information for improved interoperability between sub-domains and process models. Also involved is research on the appropriate way to transfer data across scales.

Because of the complexity and shear size of the problem, carbon sequestration codes must run on the world's largest supercomputers. This necessitates research on improving massively parallel computation of coupled process models, including operator splitting techniques involving both space and time, utilizing implicit, semi-implicit, and explicit couplings and adaptive strategies, and error estimation and propagation of errors between sub-domains and across process modules.

Computational Infrastructure Tools

The above coupled and adaptive, multi-physics, multi-scale, multi-resolution processes present unique computational infrastructure challenges because of their scale, heterogeneity in space and time, and dynamism. Software needs include, but are not limited to, programming and runtime system support for petascale parallelism, solvers and other generic numerical components, support for adaptivity and model coupling, and mechanisms for efficient data streaming.

Emerging petascale systems and underlying multi-core processor and system architectures present significant programming and runtime management complexity. Efficient use of these

systems require parallelism to be expressed at multiple levels and granularities, and to be effectively managed at runtime to avoid excessive communication and synchronization overheads. Dynamic adaptations require appropriate dynamic data-management strategies and support for dynamic partitioning and load-balancing. Addressing these challenges requires research in programming models and runtime tools that can enable the multiple types and levels of parallelism to be simply expressed and effectively exploited.

Achieving efficient, flexible and scalable coupling of physics models and parallel application codes also presents significant algorithmic, numerical, and computational challenges. A key issue is the dynamic and complex communication and coordination patterns required by these applications, which depend on state of the phenomenon being modeled, and so are known only at runtime. The coupled simulations, each typically running on a distinct parallel system or set of processors with independent (and possibly dynamic) distributions, need to periodically exchange information. This requires that 1) communication schedules between individual processors executing each of the coupled simulations are computed efficiently, locally, and on-the-fly, without requiring synchronization or gathering global information and without incurring significant overheads on the simulations, and 2) data transfers are efficient and happen directly between the individual processors of each simulation. Furthermore, specifying these coupling behaviors between the simulation codes using popular message-passing abstractions requires matching sends and receives to be explicitly defined for each interaction. As the individual simulations become larger, more dynamic and heterogeneous and their couplings more complex, implementations using message passing abstractions can quickly become unmanageable. As a result, realizing coupled simulations requires an efficient, flexible and scalable coupling framework and simple high-level programming abstractions and systems.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Many of the challenges noted above are special to carbon sequestration simulation efforts; however, several of these lines of research would generalize to and apply more broadly in the computational sciences. Improved algorithms and technologies for parallel simulation would result for some of the most challenging problems in the field of computational science. Included are 1) data interpretation and assimilation for large datasets organized spatially; 2) consistent, physics preserving discretizations of degenerate diffusive processes, reactive advective processes, and multi-phase flows; 3) multi-scale understanding and approximation problems involving medium heterogeneity; 4) aspects of grid adaptivity; and 5) the simulation of multiple, coupled processes involving changes of scales and physics. The computational infrastructure tools developed should apply directly to the entire field of computational sciences.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

The work recommended here will enable more accurate predictions of CO_2 sequestration potential of target geologic formations, improved assessment of risks to public health, better

management of the injection process, and long-term modeling of the formation and assessment of human intervention when necessary so as to maintain carbon sequestration over time.

The sequestration of CO_2 is one of the most challenging problems in the subsurface sciences, and an accurate and efficient simulation of its processes would have a far-reaching effect. Many of the processes and computational challenges are not unique to carbon sequestration simulation; rather, research is required in this area because the technical issues tend to be more critical and demanding for this application. The modeling and tools envisioned here should apply directly to other areas of subsurface simulation and allow simulation of particularly difficult Superfund and other sites requiring the modeling of geo-mechanical effects, complex multi-phase flows, extremely heterogeneous formations, or many, tightly interconnected multi-scale processes.

TIME FRAME

Significant progress can be made in as short as a 3 to 5 year time frame on many of the challenges identified above. However, it will take longer to achieve the goal of developing a simulator tool that could be used by specialists to predict the behavior of and manage sequestration efforts, and to assess risks and develop strategies for engineered intervention. A reasonable prototype might be developed within perhaps 5 to 7 years. However, a full resolution of all the challenges mentioned above may take considerably longer—possibly 10 to 20 years.

REFERENCES

Pruess, K., C.-F. Tsang, D. Law, and C. Oldenburg. 2001. Intercomparison of simulation models for CO_2 disposal in underground storage reservoirs. LBNL-47353.

Pruess, K., J. Garcia, T. Kovscek, C. Oldenburg, J. Rutqvist, C. Steefel, and T. Xu. 2004. "Code intercomparison builds confidence in numerical simulation models for geologic disposal of CO_2." *Energy* 29(9-10), 1431-1444; LBNL-52211, 2003.

EFFICIENT RISK ASSESSMENT TECHNIQUES FOR CO_2 SEQUESTRATION

Rajeesh Pawar, Los Alamos National Laboratory

ABSTRACT

Large-scale deployment of geologic CO_2 sequestration will require developing confidence in long-term performance of the technology. Such confidence can be developed through application of comprehensive decision support systems to perform cost/risk analyses of the entire system. Traditional risk analysis techniques have limited applicability to geologic sequestration sites. However, comprehensive performance assessments (PAs) coupled with efficient risk analysis techniques have been used to perform risk analysis of geologic sites. There is a need to develop novel risk analysis approaches; efficient computational techniques that can be applied to large-scale, Monte-Carlo simulations; efficient sampling methods; model abstraction methods; systems modeling approaches; and efficient coupling of process level models to systems models. Success in this PRD will lead to a number of ground-breaking advances in numerical and computational sciences and will be beneficial to a number of subsurface flow and transport applications.

EXECUTIVE SUMMARY

Large-scale deployment of geologic CO_2 sequestration will need demonstration of comprehensive analysis of risks, including financial, health, safety, and environmental risks, associated with the technology. Understanding the impact of storing large amount of CO_2 over a long time frame is a complex problem because of a number of coupled physical and chemical processes. Traditional risk analysis approaches cannot be directly applied to geologic systems because of differences in the fundamental behavior of geological systems and engineered systems such as a nuclear power plant, including dynamic evolution of component behavior, large uncertainties, and limited opportunity to perform extensive characterization experiments. This has led to the development of approaches that combine extensive PA calculations supplemented with a risk analysis based on PA results. The PA calculations can become computationally intensive depending on the number of uncertain parameters, typically making it impractical to explore all of the combinations to develop statistically meaningful results required for risk analysis. In addition, these analyses can take significant time that could be a limitation for development of multiple sequestration sites. A priority direction is the development of numerical algorithms and computational techniques that can be used to perform comprehensive, timely risk analysis of geological CO_2 sequestration sites. These approaches will combine the development of novel methods of risk analysis, up-scaling methods, model reduction approaches, efficient parameter sampling methods, efficient algorithms for large scale Monte-Carlo simulations, and application of petascale computing. Developments associated with this PRD would be beneficial in defining the research priorities related to overall geologic sequestration

science based on data needs. In addition, developments made during this PRD will be beneficial to subsurface computational science and will be applicable to a number of practical areas such as risk analysis of contaminant transport, optimization of petroleum reservoir performance, etc.

SUMMARY OF RESEARCH DIRECTION

This PRD is focused on developing numerical and computational algorithms required to efficiently quantify the overall risks associated with implementation of geologic storage of CO_2.

The overall goal of any risk assessment model is to develop the probability of occurrence of an event and to analyze the consequence of the event. The current popular approach to calculating the probability of occurrence of an event is to perform PAs and to supplement them with consequence analysis to determine overall risk. This approach can be extremely computationally intensive. This PRD focuses on two parallel paths:

1. Development of novel risk analysis methods applicable to the inherently unique characteristics of geologic systems
2. Increasing the efficiency of the current approaches.

Risk analysis

Traditional risk analysis methods use event trees, fault trees, and scenarios that cannot be directly applied to a geologic system because the behavior of its components is complex, inter-dependent, and evolves over time. In addition, geologic systems have a large number of uncertain parameters. There is a need to either modify traditional risk analysis methods or develop novel methods to be applicable to geologic CO_2 sequestration sites. Ghosh and Apostolakis (2006) propose an approach that can be used to capture meaningful risk information from traditional PA results. Efficient techniques that can integrate the PA calculations, and risk analysis can lead to novel a risk analysis approaches.

Efficient PA models combined with risk analysis

Systems Models. PA models are typically developed after a detailed analysis of features/events and processes (FEPs) that affect the overall model performance. These models decompose a system into analytically manageable subsystems—modeling the subsystems in detail and linking the subsystems in an overall systems model. There is a need for comprehensive systems models for analysis of geologic CO_2 sequestration technology that can effectively and efficiently capture performance of all components at multiple spatial and temporal scales (Figure 1). Current PA methods applied to geologic storage repositories (Helton et al. 1997) as well as those proposed for CO_2 sequestration sites (Pawar et al. 2006; Bowden and Rigg 2004; Walton et al. 2004) can be extended to develop PA methods for CO_2 sequestration. The systems model needs to effectively capture behavior of subsystems, information on which is available through laboratory experiments, field experiments or detailed process level numerical simulations. There is a need for numerical approaches that efficiently capture the system/subsystem behavior. Up-scaling and model reduction methods are also critical part of this. Development of novel up-scaling

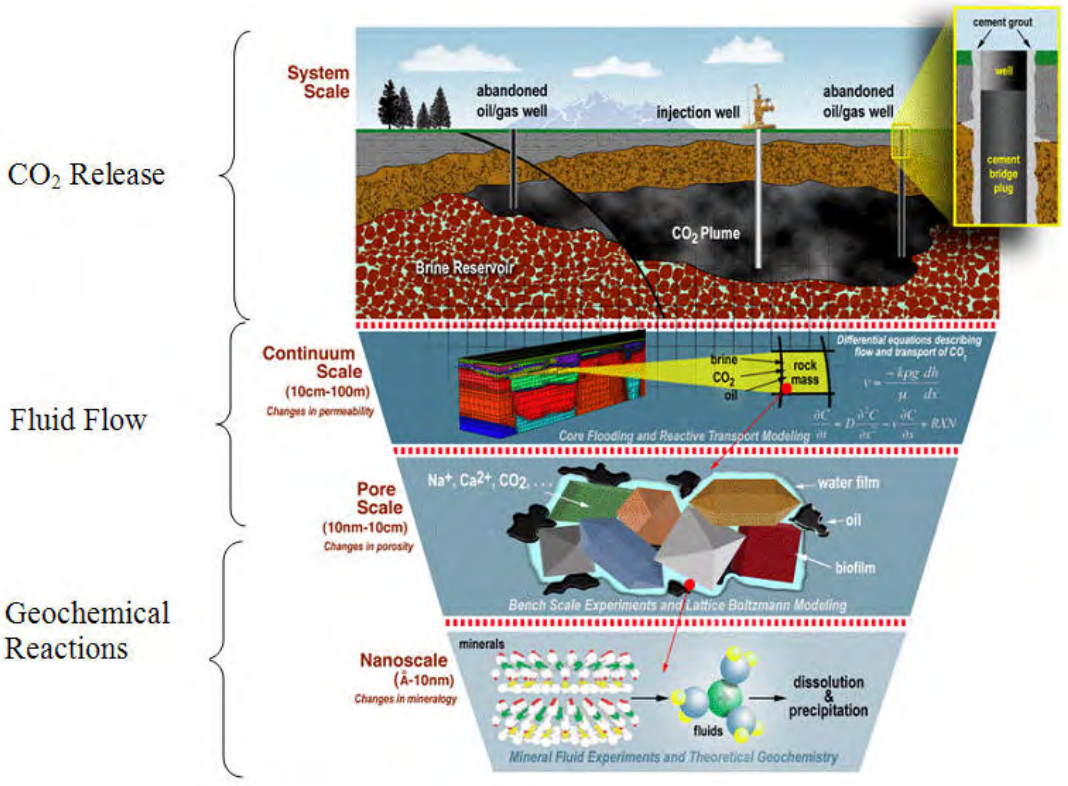


Figure 1. Schematic Illustrating Important Processes that Could Result in a Geologic CO_2 Sequestration Reservoir at Multiple Spatial Scales

approaches is the focus of another PRD. There is a need to develop novel model reduction methods. Methods such as proper orthogonal decomposition (POD) are being increasingly applied to subsurface fluid flow applications (Van Doren et al 2004; Gildin et al. 2006; Vermeulen et al. 2006). Application of these methods needs to be extended to multi-phase, reactive-transport problems; at the same time, novel, efficient model reduction methods need to be developed.

Monte-Carlo Simulations. Because of a large number of uncertain parameters/processes in geologic systems, PAs are performed by applying Monte-Carlo simulation approaches combined with different parameter sampling methods including Latin hypercube sampling (LHS) (McKay et al. 1979; Helton and Davis 2003). Depending on the number of uncertain parameters/processes the number of Monte-Carlo simulations can be extremely large. Such problems would be ideal for petascale computing. There is a need to develop efficient computational algorithms to perform large scale Monte-Carlo simulations as well as efficient and effective parameter sampling methods. In addition to capturing the effect of parameter uncertainty, there is a need for the development of efficient numerical approaches to capture model uncertainty. Applicability of approaches such as those proposed by (Ghosh and

Apostolakis 2002; Neuman and Wierenga 2003) need to be extended to CO_2 sequestration problems, and at the same time, new methods for determining model uncertainty need to be developed.

Risk Analysis. Results of PA calculations have to be analyzed to capture meaningful risk information. Traditionally techniques such as response surface methodology, differential analysis, Kolomogrov-Smirnov tests, etc., have been applied to make risk-informed decisions. Ghosh and Apostolakis (2006) proposed an approach to capture risk information from traditional PA results. There is a need for efficient numerical and computational approaches to synthesize multi-dimensional results generated from multi-dimensional, uncertain input data to quantify risk.

Data integration and optimization. Risk analysis and PA models have to be capable of efficiently integrating data generated during site characterization, site operations, and post-closure to mitigate and minimize risk. Efficient data integration and optimization methods will need to be developed to achieve this capability.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Acceptance of geologic sequestration technology for long-term storage of large amounts ($\sim 10^9$ tons/year) of CO_2 will require consideration of hundreds of storage sites, which in turn requires demonstration of comprehensive, effective, and efficient risk analysis. Understanding long-term fate of CO_2 in geological reservoirs is a complex problem because of a number of complex, coupled physical and chemical processes resulting from interactions between CO_2 and geologic environments. Comprehensive risk analysis can be used to understand the overall, long-term impact of geologic sequestration, but there are significant scientific and computational challenges.

Development of risk analysis approaches that can be applied to geologic systems

Traditional risk analysis approaches use fault trees, event trees, or scenario analyses to define overall risks. There are inherent differences in systems to which traditional approaches are applied (e.g., nuclear reactors) and geologic CO_2 sequestration sites. The major differences are dynamic evolution of components and substantial uncertainties in parameters and processes. It is not practical to conduct experiments on geologic systems at large spatial and temporal scales. As a result, long-term performance needs to be inferred from short-term, controlled field and laboratory experiments; data from natural analogs; and results of detailed numerical simulations. This leads to extensive uncertainties in modeling risks. Traditional risk analysis methods have been modified to apply to geologic repositories because of inherently distinct nature of geologic environments. Typically, in analyzing risk from geologic repositories, results of PAs are used with supplemental analyses. Capturing risk specific information from such analysis is not straight forward and efficient. The challenge is to understand the limitations of current approaches and develop the next generation of risk analysis approaches that can be used to

efficiently determine overall risks specific to geologic systems. This is a grand challenge, but at the same time, steps can be taken to improve current approaches for risk analysis.

Develop computationally efficient PA techniques

Currently, PAs are performed by developing an overall system model that links multiple subsystems or components. In most cases, the bases for the subsystem definition are the differences in the fundamental physics that governs the performance of the subsystem. It is impossible to develop a single process-level numerical model that can couple the physics that governs multiple, distinct subsystems. The approach typically used is to capture the behavior of the subsystems through simple analytical formulations or computationally efficient numerical models (Scott et al. 2005; Eslinger et al. 2006) and link the analytical sub-models into the overall system model. The behavior of the subsystems is characterized either through laboratory experiments, field experiments/observations, or detailed numerical simulations. The scientific challenge is to develop approaches that effectively capture the behavior at multiple spatial and temporal scales. For example, in geologic CO_2 sequestration application there are multiple coupled processes taking place in the primary sequestration reservoir. The coupled processes include hydrologic, geochemical and geo-mechanical processes resulting from interaction of CO_2 with the reservoir fluids and rock. These processes have different time scales, whereas hydrologic processes are relatively fast, geochemical interactions are significantly slow. In addition, these processes are intrinsically coupled. In defining long-term fate of CO_2 in a sequestration reservoir, it is necessary to take into account these coupled processes (which can be done through detailed numerical simulations). The challenge is to capture this coupled behavior effectively in a systems-level model for the reservoir so it can be used for PA. The computational challenge is to develop numerical algorithms that can describe the physics of the subsystem that are computationally efficient so that PA modeling can be performed in a timely manner.

Because of the uncertainties in geologic systems, typically PAs are run by assuming PDFs for uncertain parameters and performing Monte-Carlo simulations coupled with various parameter sampling methods. Depending on the number of uncertain parameters, this can lead to too many calculations to generate statistically adequate results. For example, if there are 100 uncertain parameters and it is desired to test all combinations of 1 percent interval range of uncertain parameters, one would need to run 100^{100} simulations. This would require significant computational resources. To overcome this, in practice Monte Carlo simulations are coupled with different sampling methods such as LHS. The sampling methods ensure that the entire range of parameter distribution is sampled, but do not necessarily make sure that all combinations of different parameter distribution are sampled. Improving the efficiency of Monte Carlo simulations is a significant scientific and computational challenge. Novel approaches are needed to efficiently sample the parameter space and reduce computational requirements for large-scale Monte Carlo calculations. Another approach for performing large-scale Monte Carlo calculations will be application of petascale computing. Further research is required to develop algorithms and architecture necessary for such large-scale computational effort.

PA calculations typically result in information such as flux of CO_2 at a defined location at different times. Such information is then used in risk analysis calculations. Inferring meaningful statistical information from large-scale PA calculations is a daunting task. There are a number of uncertain parameters and processes that result in a multi-dimensional response surface. This response surface needs to be translated in a risk analysis framework. As there are multiple, inter-connected subsystems, a number of possible combinations of subsystems (scenarios) need to be studied to define which combinations (scenarios) will lead to high-risk scenarios. In addition, it is required to identify which parameters/processes control the overall system behavior and whether the uncertainty in those parameters can be reduced. Such analyses can become complex and resource intensive. Research is needed to develop efficient approaches to synthesize multi-dimensional result and parameter space as well as sensitivity analyses.

In addition to using the results of risk analysis for risk-informed decision making, the results can also be used to develop monitoring approaches to minimize and mitigate risk. There can be different sources (e.g., below-surface, above-surface), types (e.g., geophysical, geochemical) and scales (e.g. core or well-bore). It is a big challenge to efficiently integrate the data in risk analysis calculations and see the impact of additional data on overall risk.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Success in this PRD will lead to development of novel numerical algorithms for model reduction, model abstraction including up-scaling methods, and efficient coupling of process level models to systems level models. It will also lead to efficient parameter sampling algorithms and Monte-Carlo simulation techniques. This PRD will also lead to demonstration of applicability of petascale computing.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

Success in this PRD will lead to development of novel approaches for effective risk assessment for geologic repositories. Model reduction algorithms will have wide applicability in a range of subsurface fluid flow applications, including petroleum reservoir management, aquifer managements, and contaminant fate and transport. Efficient approaches will be developed for incorporation of monitoring/characterization data for risk mitigation and minimization.

TIME FRAME

The overall time frame for this PRD should be 3 to 5 years.

REFERENCES

Bowden, A. R. and A. Rigg, 2004. "Assessing Risk in CO_2 Storage Projects," *APPEA Journal*, 682.

Ghosh, S. T. and G. E. Apostolakis, 2002. "Assessing Structural Model Uncertainty in Performance Assessments," *Proceedings of 6th International Conference on Probabilistic Assessment and Management*, ed. E. J. Bonano, A. L. Camp, M. J. Majors and R. A. Thompson.

Ghosh, S. T. and G. E. Apostolakis, 2006. "Extracting Risk Insights from Performance Assessments for High-Level Radioactive Waste Repositories," *Nuclear Technology* **153**, 70.

Gildin, E., Klie, H., Rodriguez, A., Wheeler, M. F. and R. H. Bishop, 2006. "Development of Low-Order Controllers for High-Order Reservoir Models and Smart Wells," *Proceedings of the 2006 SPE Annual Technical Conference and Exhibition*.

Helton, J. C. and F. J. Davis, 2003. "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Engineering & System Safety*, **81**, 23.

Helton, J. C., Anderson, D. R., Marietta, M. G., and R. P. Rechard, 1997. "Performance Assessment for the Waste Isolation Pilot Plant: From Regulation to Calculation for 10 CFR 191.13," *Operations Research* **45**, 157.

McKay, M. D., Beckman, R. J. and W. J. Conover, 1979. "Comparison of 3 Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, **21**, 239.

Neuman, S. P. and P. J. Wierenga, 2003. "A Comprehensive Strategy for Hydrologic Modeling and Uncertainty Analysis for Nuclear Facilities and Sites," *NUREG/CR-6805*.

Pawar, R. J., Carey, J. W., Chipera, S. J., Fessenden, J., Kaszuba, J. P., Keating, G. N., Lichtner, P. C., Olsen, S. C., Stauffer, P. H., Viswanathan, H., Ziock, H. J., and G. A. Guthrie, 2006. "Development of a framework for long-term performance assessment of geologic CO_2 sequestration sites," *Proceedings of the 8th International Conference on Greenhouse Gas Control Technologies*.

Vermeulen, P. T. M., te Stroet, C. B. M. and A. W. Heemink, 2006. "Model Inversion of Transient Nonlinear Groundwater Flow Models Using Model Reduction," *Water Resources Research*, 42, W09417.

Van Doren, J., Markovinovic, R. and J. D. Jansen, 2004. "Reduced Order Optimal Control of Waterflooding Using POD," *Proceedings of 9th European Conference on the Mathematics of Oil Recovery*.

Walton, F. B., Tait, J. C., LeNeveu, D. and Sheppard, M. I., 2004. "Geological Storage of CO2: A Statistical Approach to Assessing Performance and Risk", *Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies*.

SUBSURFACE SCIENCE COLLABORATION

Patrick O'Leary, Idaho National Laboratory; Myron Allen, University of Wyoming;
Ken Manchester, MSE Technology Applications

ABSTRACT

The panel on carbon sequestration expressed concern that the subsurface sciences community is not sufficiently unified to address this type of multidisciplinary problem. The participants recommended that an infrastructure for improved subsurface sciences collaboration be established with publicly accessible, online repositories, observatories, collaboration tools, and a framework for enabling scientific and engineering workflows. In addition, DOE should promote open participation by stipulating that federal funding be predicated on agreement of Principal Investigators to post deliverables, including numerical codes, in open-source format.

EXECUTIVE SUMMARY

For at least several decades in the future, the United States will rely on fossil fuels (i.e., coal, oil, methane, and possibly oil shale) for the majority of our required and expanding energy needs. This heavy reliance on fossil fuels will result in increased production of CO_2 unless the nation makes significant progress towards the development of near-zero emission power plants. The capture, separation, and storage or reuse of carbon (i.e., carbon sequestration) should be employed to stabilize and eventually reduce concentrations of this greenhouse gas in the atmosphere.

The DOE Carbon Sequestration R&D program is tasked to provide science-based assessments of costs, long-term stabilities, and environmental impacts. Clearly modeling and simulation will serve as the principal tool in evaluating the technology options, but a number of advances in the state-of-the-art of applicable technologies are required in a relatively short time frame.

A prerequisite for building effective collaboration is the desire to engage a scientific or engineering challenge, typically one having the scale of a "grand challenge." Carbon sequestration is exactly the type of problem that will require the collaboration of a large cross section of the subsurface sciences community, including applied mathematicians, geoscientists, chemical and petroleum engineers, chemists, and materials scientists.

Improved collaboration in the subsurface sciences has the potential to benefit DOE, and the scientific and engineering community in general, by expanding the resources available to individual researchers, increasing the efficiency of our research system through open information sharing, and by coupling basic and applied research efforts more tightly to national goals.

SUMMARY OF RESEARCH DIRECTION

The subsurface sciences community has been described as "tribal" (but friendly). Codes, datasets, and computational tools are not openly shared until they are finished products, and even then they are sometimes, *de facto* or *de jure*, proprietary. An extreme example of this challenge is that the petroleum industry, which houses tremendous expertise in subsurface sciences, is underrepresented in many collaborations of interest in CO_2 sequestration.

The participants of the Carbon Sequestration Panel propose that a subsurface sciences collaboration infrastructure be established to enable research, development and demonstration for the "grand challenge" scientific and engineering goal of numerically modeling carbon sequestration. The collaboration infrastructure will include a number of unique features including code and data repositories, observatories, software collaboration tools, and scientific and engineering workflows.

An additional requirement for developing successful collaboration is an incentive to collaborate. The participants recommend that DOE should promote collaboration by making Federal funding dependent on agreements from Principal Investigators' that post-deliverables be made available in open-source format on a commonly accessible electronic platform.

Repositories

The virtual repositories consist of a distributed databases, folders and source control management systems, each created from multiple federated resources, providing integrated global access with local controls. The distributed database will maintain reliable, consistent, and pedigreed datasets for benchmarking, validation, and verification activities. Distributed folders would make available community computational tools and software stacks for interoperability and collaboration. Finally, the singularly presented distributed source control management system for open-source software will enable both sharing and community development of subsurface sciences codes.

Observatories

The FutureGen Alliance states, "FutureGen is a $1 billion (U.S.) public-private partnership to design, build, and operate the world's first coal-fueled, near-zero emissions power plant. The commercial-scale plant will prove the technical and economic feasibility of producing low-cost electricity and hydrogen from coal while nearly eliminating emissions. It will also support testing and commercialization of technologies focused on generating clean power, capturing and permanently storing carbon dioxide, and producing hydrogen. In the process, FutureGen will create unique opportunities for scientific exploration, education, and stakeholder engagement." A separate PRD from the Carbon Sequestration panel suggests that FutureGen be used as a laboratory for research, development, demonstration, validation, and verification of computational techniques for modeling carbon sequestration. The four sites that have been chosen for the next step of the FutureGen selection process are:

- Mattoon, Illinois
- Tuscola, Illinois
- Heart of Brazos near Jewett, Texas
- Odessa, Texas.

We propose that the site characterization, monitoring data, and the instruments themselves be integrated into the collaboration infrastructure by forming virtual observatories, one for each proposed site. These observatories will provide the first carbon sequestration measurement and observation systems designed to both to answer significant scientific questions and to have the multidisciplinary participation necessary to achieve credible subsurface forecasting and prediction. As such, each will transform the way we conduct subsurface science by enabling the integration of research, development, and demonstration from modeling to observations to measurements.

Collaboration Tools

Collaboration tools are software packages designed to boost productivity and responsiveness. They are essential to large multidisciplinary research projects spanning multiple organizations to tackle scientific and engineering 'grand challenge' goals. The tools should block the scientist or engineer from being interrupted in the 'productivity zone', but allow them to respond, attend or collaborate at the appropriate time. A typical subset of collaboration tools includes the following: announcements, calendars, chat rooms, discussions, drop boxes, email, email archives, help, membership, message center, workspaces, news, permissions, roles, post its, preferences, resources, schedules, site info and wikis.

Scientific and Engineering Workflows

Most scientists conduct analysis and run models in several different software and hardware environments, mentally coordinating the export and import of data from one environment to another. Scientific and engineering workflows attempt to formalize this *ad hoc* process so scientists can design, execute, and communicate procedures repeatedly and with minimal effort. These workflows are a blend of scientific problem-solving, engineering design, and traditional business workflow techniques. The workflow research for subsurface sciences collaboration should concentrate on components to enable dynamic data-driven application systems (DDDAS) and three-dimensional visualization.

In DDDAS, simulations and measurements form a symbiotic feedback control system. In carbon sequestration, the monitoring of the CO_2 plume forms a feedback loop with site characterization and the modeling/simulation activities. Integrating enabling components into the scientific and engineering workflow framework to support DDDAS is integral to the success of the carbon sequestration program.

On the other hand, three-dimensional visualization is on the critical path to scientific and engineering understanding and discovery. Providing scientists and engineers with matching computational, analysis, and visualization power should be an essential element of the subsurface sciences collaboration infrastructure, and visualization is currently the weakest link. Subsurface science specific visualization tools, as depicted in Figure 1, with interfaces designed for the domain expert, computational scientist and regulator. This work should provide a significant step forward in integrating hardware-accelerated visualization into subsurface sciences research collaborations.

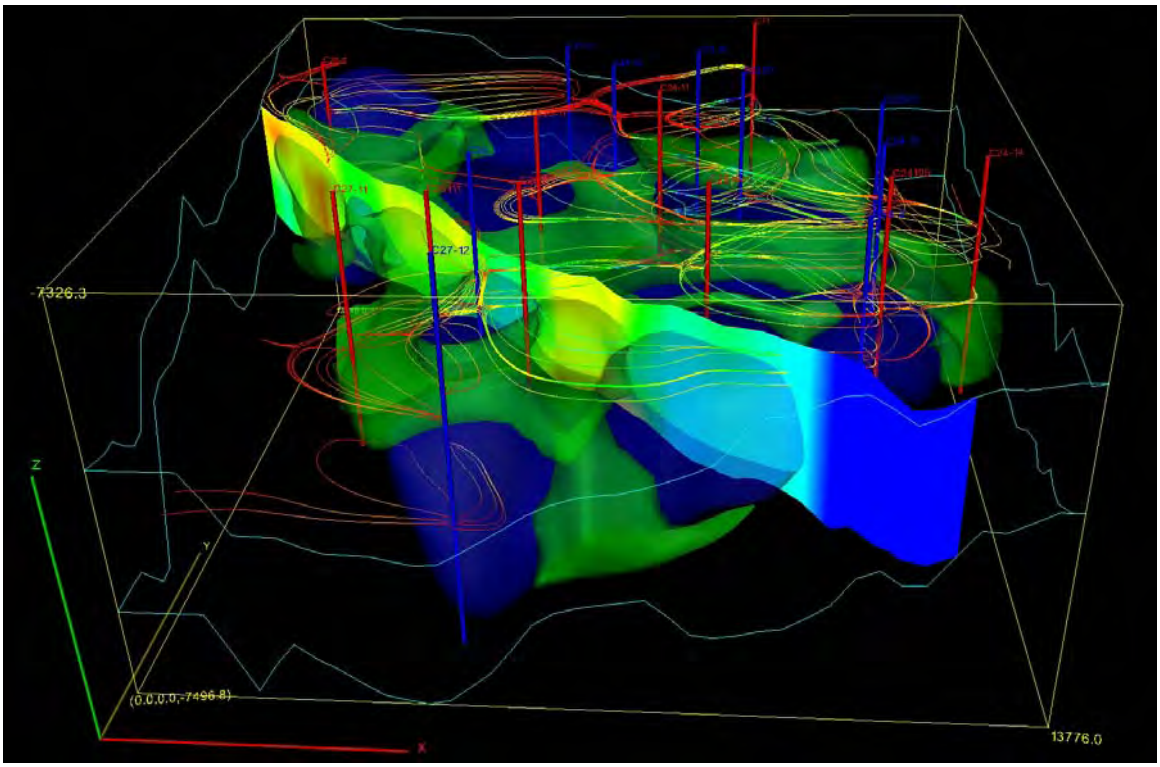


Figure 1. Visualization of a computational oil reservoir demonstrating a fusion of visualization techniques designed specifically for use by subsurface scientists and engineers (image from P. O'Leary).

We propose a web-based, virtual distributed repository for storing codes, datasets and computational tools. This would facilitate communication between code authors and application-orientated scientists and engineers. Federal funding should be predicated on agreement by Principal Investigators' to post deliverables in open-source format in the repository.

The repository would use a framework that would promote global access and preserve local control. It will provide collaboration tools for tracking and reviewing the development of new and improved codes. The framework would also be capable of sustaining a wide and ever-evolving variety of codes from the subsurface sciences community, in modular formats to provide a platform for continual improvements.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Currently the subsurface sciences community is not sufficiently unified. It has been described as tribal (but friendly). Codes, datasets, and computational tools are not openly shared until they are finished, sometimes proprietary, products. This is in contrast, for example, to the collaborative community associated with the Linux operating system. An extreme example of this challenge is that the petroleum industry, a significant player in subsurface science, is underrepresented in many collaborations of interest in CO_2 sequestration.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The repository will serve as an example of a collaborative environment for many other fields requiring the development, maintenance, improvement, and application of sophisticated scientific software.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

The repository will expedite research and development by avoiding duplication of effort and allowing scientists to "stand on each others' shoulders." It will facilitate collaboration among the large numbers of disciplines and research groups necessary to attack the multi-scale, multi-physics nature of CO_2 sequestration, and it will accommodate new and emerging high-performance computing environments.

TIME FRAME

The FutureGen project is currently progressing at a rapid rate. Therefore, the time frame for developing the subsurface sciences collaboration infrastructure needs to be accelerated. Milestones within the accelerated time frame follow.

- FY 2007. Launch Subsurface Sciences Collaboration Portal with a "Best-in-Class" subset of the computational and collaboration software stack.
- FY 2007. Establish a framework for implementing scientific and engineering workflows.
- FY 2008. Deliver initial release of middleware to enable global access to distributed source control management systems with local control.
- FY 2008. Develop software stack for installation of a collaboration appliance for remote community members.
- FY 2008. Integrate FutureGen observatory into the collaboration infrastructure.
- FY 2008. Incorporate web-based, three-dimensional visualization components in scientific and engineering workflow framework.

- FY 2009. Deliver update of middleware for grid-enabled distributed source code.
- FY 2009. Update workflow framework with dynamic data components.
- FY 2009. Develop workflow tools for the carbon sequestration dynamic data-driven application system associated with FutureGen.

REFERENCES

Baker, M. A. & Lakhoo, R., 2006. “Application Reuse through Portal Frameworks.” Fifth UK e-Science All Hands Meeting.

Bly, S., Keith, K. M., & Henline, P. A., 1997. “The work of scientists and the building of collaboratories.” General Atomics Report GA-A22619.

Chin, G., Jr., & Lansing, C. S., 2004. “Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory.” Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, 409-418, New York: ACM Press.

Collins-Sussman, B. Fitzpatrick, B. W. & Pilato, C. M., 2004. Version Control with Subversion. O’Reilly.

Douglas, C. C., Efendiev, Y., Ewing, R., Ginting, V., Lazarov, R., Cole, M. J., Jones, G., & Johnson, C. R., "Dynamic data-driven contaminant simulation," Computational Sciences - ICCS 2005, 2 (2005), pp. 640-647.

Finholt, T. A., 2003. “Collaboratories as a new form of scientific organization.” Economics of Innovation and New Technology, 12(1), 5-25.

FutureGen Alliance, http://www.futuregenalliance.org.

NEON - Extended Description - 2002, National Science Foundation.

Olson, G. M., Atkins, D. E., Clauer, R., Finholt, T. A., Jahanian, F., Killeen, T. I., Prakash, A., & Weymouth, T., 1998. “The Upper Atmospheric Research Collaboratory (UARC).” ACM Interactions, 5(4), 48-55.

Parashar, M., Matossian, V., Bangerth, W., Klie, H., Rutt, B., Kurc, T. M., Catalyurek, U. V., Saltz, J. H. & Wheeler, M. F., 2005. "Towards Dynamic Data-Driven Optimization of Oil Well Placement." *Lecture Notes in Computer Science*, 3515, pp. 656-663.

Qiao W., McLennan M., Kennell R., Ebert D. & Klimeck G., 2006. "Hub-based Simulation and Graphics Hardware Accelerated Visualization for Nanotechnology Applications," IEEE

Transactions on Visualization and Computer Graphics (Proc. IEEE Visualization 2006), 12(5).

Singh, M. P. & Vouk, M. A., 1996. "Scientific Workflows : Scientific Computing Meets Transactional Workflows." NSF Workshop on Workflow and Process Automation in Information Systems: State-of-the-Art and Future Directions, May 8-10.

Sonnenwald, D. H., Bergquist, R. E., Maglaughlin, K. L., Kupstas-Soo, E., & Whitton, M. C., 2001. "Designing to support collaborative scientific research across distances: The nanoManipulator environment." in Churchill, E. F., Snowdon, D. N., & Munro, A. J. (editors). Collaborative virtual environments: Digital places and spaces for interaction. (pp. 202-224). London: Springer-Verlag.

U.S. Department of Energy, Fossil Energy, Carbon Sequestration Program, http://www.fossil.energy.gov/programs/sequestration/.

U.S. Department of Energy National Collaboratories, http://www.doecollaboratory.org.

Wulf, W. A., 1989. “The national collaboratory - a white paper.” In J. Lederberg & K. Uncaphar (Eds.). Towards a National Collaboratory: Report of an invitational workshop at the Rockefeller University, March 17-18, 1989. (pp. Appendix A). Washington, D.C.: National Science Foundation, Directorate for Computer and Information Science Engineering.

Wyatt, E. C. & O'Leary, P., 2004. "Interactive Poster: Grid-Enabled Collaborative Scientific Visualization Environment." IEEE Visualization Proceedings of the conference on Visualization '04, 598:18.

MODELING THE SEQUESTRATION OF SUPERCRITICAL FLUID MIXTURES

Diana Bacon, Pacific Northwest National Laboratory; John Weare, University of Californial San Diego; Glenn Hammond, Pacific Northwest National Laboratory

ABSTRACT

Injecting flue gas directly from a coal-fired power plant into the subsurface without first separating CO_2 from the gas stream would reduce the cost of geological carbon sequestration. Unfortunately, current multi-phase flow and reactive transport simulators are not able to accurately model the injection of complex supercritical fluid mixtures into the subsurface because an equation of state for these mixtures has not been defined. Molecular dynamics simulations offer a solution.

EXECUTIVE SUMMARY

Geological carbon sequestration is a promising technology for reducing CO_2 emissions from coal-fired power plants and other CO_2 emitters. However, capturing and separating CO_2 from the other products of coal combustion can be an expensive process. This has led to interest in the possibility of injection unpurified, or less purified, flue gas from a power plant directly into the subsurface. The flue gas would be compressed to a supercritical state before injection. While the simulation of the injection of supercritical CO_2 into the subsurface has become fairly routine, an equation of state for a complex mixture of supercritical fluids has not been defined at the temperatures and pressures encountered during geologic carbon sequestration. The challenge is to provide thermodynamic properties of the supercritical fluid mixture, along with the individual aqueous solubilities of each of the flue gas components. Molecular dynamics modeling may be used to simulate these properties, as a supplement to the limited experimental data that is available for these complex gas mixtures. Both the molecular dynamics modeling used to determine the equation of state for these supercritical mixtures, and the modeling of the subsurface injection of these mixtures are computationally intensive calculations that may be greatly accelerated by code parallelization and execution on massively parallel supercomputers.

SUMMARY OF RESEARCH DIRECTION

National and international concern about the possible effects of greenhouse gases such as CO_2 on climate is rising. Evidence suggests that the rising level of atmospheric CO_2 primarily results our expanding use of fossil fuels for energy. Although the development of more efficient and alternative energy systems will lead to a reduction in greenhouse gas emissions, it is likely that the burning of fossil fuels will continue to provide a considerable proportion of the nation's energy well into the next century. One strategy for reducing the emission of greenhouse gases to the atmosphere is to capture and separate carbon dioxide from the products of fossil fuel burning

and sequester it below the earth's surface. Mechanisms of CO_2 sequestration in the subsurface include physical trapping by low-permeability formations, solubility trapping by dissolution of CO_2 into the aqueous phase, and mineral trapping by conversion of CO_2 into a mineral form such as calcium carbonate.

Separation or purification of CO_2 from flue gas is an expensive step in carbon sequestration, and a great deal of research is focused on developing less expensive methods (Song 2006). Recently, there has been increasing interest in the possibility of injecting flue gas directly from a coal-fired power plant into the subsurface without first separating CO_2 from the gas stream. As an example, oxyfuel combustion in a pulverized-fuel, coal-fired power station produces flue gas containing contaminants such as water vapor, oxygen, nitrogen, and argon derived from the excess oxygen, impurities in the oxygen used, and any air leakage into the system. There also are acid gases present, such as SO_3, SO_2, HCl, and NO_X produced as products of combustion. The impact of sequestering a complex gas mixture in a deep geologic formation, rather than pure CO_2 (Figure 1), has not been investigated. Co-sequestration of N_2, O_2, and acid gases along with CO_2 in a reducing environment may have unexpected impacts on formation mineralogy and microbiology, as well as the extent of aqueous and mineral trapping of CO_2.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Simulation of CO_2 injection into subsurface geologic formations is now done routinely (e.g., Xu et al. 2005; Bacon et al. 2006), and the equation of state for the pure CO_2 system is fairly well developed (Span and Wagner 1996). The multi-phase flow simulators used to simulate supercritical CO_2 injection (White and Oostrom 2006; Xu et al. 2006) into the subsurface have typically been used to simulate a limited number of fluid components, with either a single gas phase or a binary gas mixture (Oldenburg and Unger 2003). A multi-phase flow model is needed that incorporates an equation of state for gas mixtures, along with a means for tracking changes in the composition of the gas mixture with time and space. An equation of state for fluid mixtures is needed over a wide range of temperatures and pressures.

Prior to injection into the subsurface, flue gas is compressed to a supercritical fluid. A supercritical fluid is any fluid at a temperature and pressure above its thermodynamic critical point. It has the unique ability to diffuse through solids like a gas, and dissolve materials like a liquid. Additionally, supercritical flue gas density is extremely sensitive to temperature or pressure. Although a general equation of state for supercritical fluid mixtures has been developed (Duan et al. 1996), it is not accurate at the lower temperatures generally encountered in geological carbon sequestration. Furthermore, if this equation of state is to be used to perform simulations for risk analysis where the flue gas may rise toward the surface, it must be expanded to include subcritical temperatures and pressures.

Simulation of the sequestration of multi-component supercritical fluid mixtures in the subsurface is a greater computational challenge than modeling pure CO_2 sequestration because it is a more

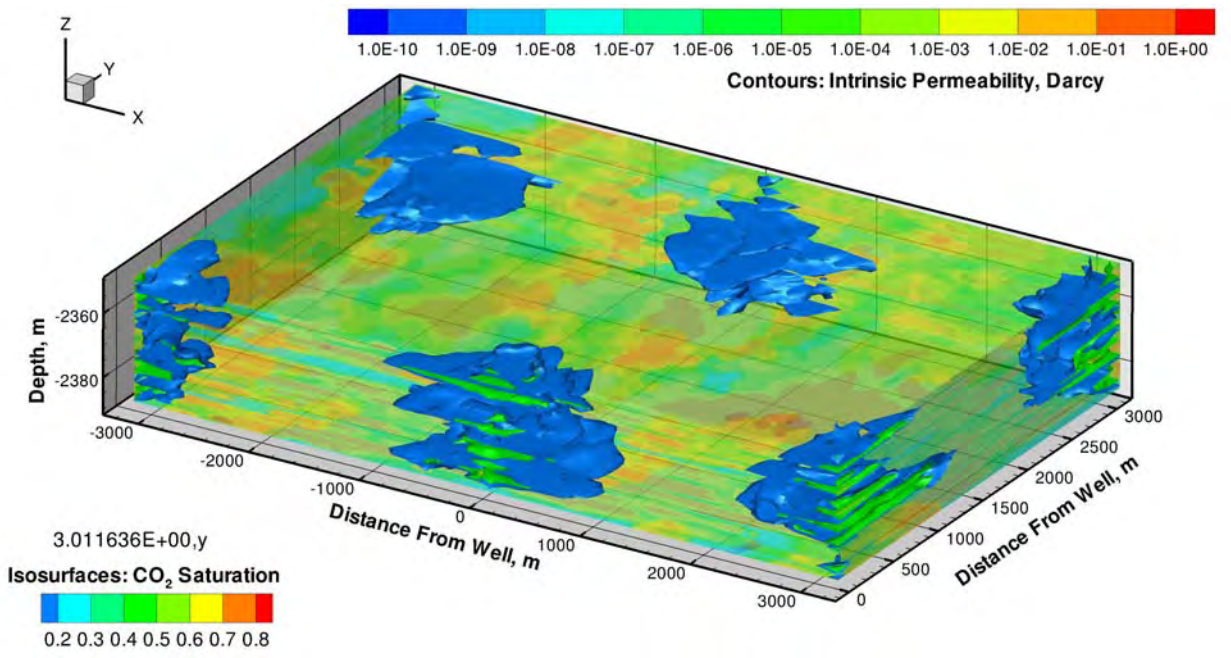


Figure 1. Simulation of Supercritical CO_2 Injection from Multiple Wells into a Brine-Filled Sandstone Formation

complex equation of state and relative amounts of each fluid component must be tracked. Improved scalability of compositional multi-phase flow and/or coupled reactive transport and multi-phase flow models would also be needed to meet this challenge.

Molecular Dynamics Modeling of Flue-Gas Mixtures

The challenge is to provide thermodynamic (e.g., free energy of the mixture, fugacities, enthalpy, subcritical phase coexistence, etc.) and physical properties (e.g., phase density, viscosity and diffusivity, thermal conductivity of the mixtures, etc.) of mixtures in this system. In addition the solubilities of the gases in coexisting aqueous phases are required. Typical flue gas emissions are mixtures of noncondensible gases primarily CO_2 and N_2 (approximately 90 percent), with lesser amounts of O_2 and Ar (approximately 9 percent) and relatively small amounts of SO_2 and NO_x (NO and NO_2). When injected into a mineral formation, the flue gas mixture may encounter and react with formation fluids (brines) and mineral phases. Typical temperatures range in this application from roughly 20°C to roughly 100°C and pressures from 450 bars to near atmospheric. With the exception of CO_2 (T_c=31°C, P_c=73.8 bar), all the pure gases in this system are supercritical in this range. Surprisingly, even for this commonly encountered system there are very few data for mixtures. It is proposed that Molecular Dynamics and Monte Carlo Gibbs Ensemble (for phase coexistence) simulation methods be used to supplement the limited experimental data.

Molecular Dynamics and Monte Carlo methods are increasingly being used to interpret the behavior of complex, strongly interacting, many-body environmentally important materials. While these methods are providing important insights into these systems, there are limitations to their application. These limitations include:

- The accuracy of a many-body simulation is dependent on the ability of the model of the interactions in the system at the atomic or molecular level to capture the real forces in a practical effective potential. These potentials must be established either from empirical data or computed from first principle methods (e.g., from the electronic Schrödinger equation for some few-body subsystem [in favorable cases, two-body]). Depending on the desired accuracy, this may be very difficult. For example, if the molecules have strong dipoles and are polarizable, this must be taken into account, often leading to complex, density-dependent interaction potentials (e.g., as in water-water interactions).
- For problems in which chemical reactions are important, it may not be practically possible to develop an empirical potential. Therefore direct, first principle simulations are necessary (e.g., *ab-initio* molecular dynamics methods [Car and Parrinello 1985]). These are very time-consuming computational problems.
- For many of the properties required, an estimate of the mixture-free energy is necessary. This means that some sort of thermodynamic integration is required, which is computationally expensive.
- For phase coexistence calculations, it is necessary to use Gibbs ensemble methods (Panagiotopoulos 1987). For dense liquid systems, these require very large numbers of potential evaluations and special particle insertion strategies, particularly for the angular dependant potentials that are required for most molecular systems (Duan et al. 2004).
- For dilute species, special methods must be used because of their small particle numbers in a practically-sized simulation.

Given all of the above, there is the issue of summarizing all the experimental and computed data in an appropriate equation of state. This is a non-trivial task, particularly in the critical region where special scaling methods must be introduced to ensure the proper behavior (Kiselev and Rainwater 1998). It usually is more efficient to use several equations of state to model the special properties of different regions of pressure and temperature space (e.g., incompressible fluids vs. gas mixtures).

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Fortunately, for the flue gas system the challenges described in the previous section should be fairly manageable. Given the importance of this system, it is ideal for testing the use of simulation strategies to supplement experimental data that is very difficult to obtain in sufficient quantities. The principal species for this system, CO_2, N_2, O_2, and Ar, are closed electronic shell non-reactive molecules. This means that high accuracy can be obtained from a two-body

potential description of the system without polarization. There is probably enough data for the pure systems to develop highly accurate two-body like-like potentials. A site-site interaction producing very high accuracy of many properties, including near critical-phase coexistence, has already been reported for the most difficult species CO_2 (Zhang and Duan 2005). Preliminary progress has been made for the pure N_2 and CO_2 end members (Duan et al. [unpublished results]). The unlike interactions (e.g., CO_2-N_2 two-body interactions) in these systems are more difficult. However, because of the lack of dipole in these systems, two-body interaction calculated from high-level quantum chemistry methods (e.g., CCSD(T) [Bartlett 2005]) should provide sufficient accuracy. Again, this is a good system to test the concept of developing a fully first principles thermodynamic model of a useful system.

Given the potential models, the simulations for the flue gas system can proceed, once an efficient implementation of the two-body potential for a Molecular Dynamics code has been developed. However, given that free energy and partial molar quantities are desired (requiring very large particle number simulations), it would be a good idea to develop this code for a massively parallel computer environment. Considerable progress in this direction has recently been reported (Bowers et al. 2006; Fitch et al. 2006). For the prediction of phase equilibria, a Gibbs Ensemble (Panagiotopoulos 1987) or similar method should be used. This means that problems with particle insertion and averaging over a mixture composition will require many floating point operations, again suggesting that a highly scalable code be developed (Duan et al. 2004). In principle, this is manageable, and progress has been made for similar mixtures. These calculations suggest that even in the critical region simulations (Monte Carlo) can yield results with near experimental accuracy.

Given the simulated data, the next issue is the development of a succinct representation of the free energy as function of the intensive variables. Progress has been made but this is a non-trivial problem. The ability to generate reliable data on demand with a Molecular Dynamics or Monte Carlo simulation will make an enormous difference in the development of an appropriate phenomenology. As mentioned above, the best choice of the equation of state form will depend on the nature of the phase present in for the temperature, pressure and composition (TPX) conditions of interest. For the critical regions, a scaling equation of state can be developed (Kiselev and Rainwater 1998). Again, the ability to calculate accurate thermodynamic data will be very important. To our knowledge, a comprehensive free energy model in this region has not been reported.

Finally, the issue of the interaction of the flue gas phase with a coexisting aqueous solution needs to be addressed. Water is a very difficult system to simulate. However, there has been recent progress (Paricaud et al. 2005). Methods to calculate the energy of insertion of a solute into an aqueous phase exist (Lisal et al. 2004). However, these methods require accurate potential interactions between the water species and the solute species. Given these insertion free energies, the solubility of the gases in the flue gas system in an aqueous phase can be calculated using Henry's law and the partial fugacities of the flue gas phase (King et al. 1992). While these calculations are not particularly difficult, the required accuracy of the H_2O-flue gas species in a

water environment may be difficult to obtain. Presumably this could be done using very large quantum chemistry type calculations. Improvement of the performance of these methods is a grand challenge problem.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

Injection of pure flue gas, or even less-pure CO_2, may enable carbon sequestration to be economically viable. Multi-phase reactive transport simulators would benefit from a general equation of state for fluid mixtures. A general equation of state for fluid mixtures would be useful in other applications involving geothermal fluid geochemistry.

TIME FRAME

Development of the molecular dynamics models necessary to define the equation of state for fluid gas mixtures, and the implementation of this equation of state in a multi-phase flow and reactive transport simulator would take place within 3 to 5 years.

REFERENCES

Bacon, D. H., M. D. White, N. Gupta, J. R. Sminchak and M. E. Kelley (2006). Assessment of CO_2 Injection Potential and Monitoring Well Location at the Mountaineer Power Plant Site. 8th International Conference on Greenhouse Gas Control Technologies, Trondheim, Norway, Elsevier, Oxford, UK.

Bartlett, R. J. (2005). How and Why Coupled-Cluster Theory Became the Preeminent Method in Ab-initio Quantum Chemistry. Theory and Applications of Computational Chemsitry. C. E. Dykstra, G. Frenking, K. S. Kim and G. E. Scuseria, Elsevier**:** 1191-1221.

Bowers, K. J., E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Greggersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan and D. E. Shaw (2006). "Scalable Algorithms for Molecular Dynamics simulations on Commodity Clusters." SC2006, November, 2006, Tampa, Florida.

Car, R. and M. Parrinello (1985). "Unified Approach for Molecular-Dynamics and Density-Functional Theory." Physical Review Letters **55**(22): 2471-2474.

Duan, Z. H., N. Moller and J. H. Weare "Gibbs Ensemble Monte Carlo Simulation of Vapor/Liquid and Metastable Liquid/Liquid Phase Equilibria in the CO2-CH4-N2 System." Unpublished research.

Duan, Z. H., N. Moller and J. H. Weare (1996). "A general equation of state for supercritical fluid mixtures and molecular dynamics simulation of mixture PVTX properties." Geochimica Et Cosmochimica Acta **60**(7): 1209-1216.

Duan, Z. H., N. Moller and J. H. Weare (2004). "Gibbs ensemble simulations of vapor/liquid equilibrium using the flexible RWK2 water potential." Journal of Physical Chemistry B **108**(52): 20303-20309.

Fitch, B. G., A. Rayshubskiy, M. Eleftheriou, T. J. C. Ward, M. Giampapa, R. S. Pitman and R. S. Germain (2006). "Blue Matter: Approaching the Limits of Concurrency for Classical Molecular Dynamics." SC2006, November, 2006, Tampa, Florida.

King, M. B., A. Mubarak, J. D. Kim and T. R. Bott (1992). "The Mutual Solubilities of Water with Supercritical and Liquid Carbon-Dioxide." Journal of Supercritical Fluids **5**(4): 296-302.

Kiselev, S. B. and J. C. Rainwater (1998). "Enthalpies, excess volumes, and specific heats of critical and supercritical binary mixtures." Journal of Chemical Physics **109**(2): 643-657.

Lisal, M., W. R. Smith and K. Aim (2004). "Analysis of Henry's constant for carbon dioxide in water via Monte Carlo simulation." Fluid Phase Equilibria **226**: 161-172.

Oldenburg, C. M. and J. A. Unger (2003). "On Leakage and Seepage from Geologic Carbon Sequestration Sites: Unsaturated zone Attenuation." Vadose Zone Journal **2**: 287-296.

Panagiotopoulos, A. Z. (1987). "Direct Determination of Phase Coexistence Properties of Fluids by Monte-Carlo Simulation in a New Ensemble." Molecular Physics **61**(4): 813-826.

Paricaud, P., M. Predota, A. A. Chialvo and P. T. Cummings (2005). "From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a Gaussian charge polarizable model." Journal of Chemical Physics **122**(24): 244511.

Song, C. S. (2006). "Global challenges and strategies for control, conversion and utilization of CO2 for sustainable development involving energy, catalysis, adsorption and chemical processing." Catalysis Today **115**(1-4): 2-32.

Span, R. and W. Wagner (1996). "A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa." Journal of Physical Chemistry Reference Data **25**(6): 1509-1596.

White, M. D. and M. Oostrom (2006). STOMP: Subsurface Transport Over Multiple Phases, Version 4.0, Users Guide, Richland, Washington, Pacific Northwest National Laboratory.

Xu, T. F., J. A. Apps and K. Pruess (2005). "Mineral sequestration of carbon dioxide in a sandstone-shale system." Chemical Geology **217**(3-4): 295-318.

Xu, T. F., E. Sonnenthal, N. Spycher and K. Pruess (2006). "TOUGHREACT - A simulation program for non-isothermal multi-phase reactive geochemical transport in variably

saturated geologic media: Applications to geothermal injectivity and CO2 geological sequestration." Computers & Geosciences **32**(2): 145-165.

Zhang, Z. G. and Z. H. Duan (2005). "An optimized molecular potential for carbon dioxide." Journal of Chemical Physics **122**(21): 214507.

MODELING FUTUREGEN-SCALE GEOLOGIC CO_2 SEQUESTRATION

James Johnson, Lawrence Livermore National Laboratory; George Zyvoloski, Los Alamos National Laboratory

ABSTRACT

FutureGen power plants will require the efficient capture and secure geologic isolation of CO_2 emissions. Defensible forecasting of long-term isolation performance demands advanced reactive transport simulators that explicitly couple the interdependent multi-phase flow, geochemical mass transfer, and geo-mechanical deformation processes that control subsurface CO_2 migration and trapping. In addition, field application of these simulators requires site characterization and geo-statistical methods to define initial and boundary conditions, myriad site-independent data that underpin the coupled process models to facilitate dynamic system evolution, and time-lapse monitoring data to permit iterative improvement of both site characterization and dependent modeling efforts. Further, because the relevant basin-scale systems are typically characterized by extreme permeability and compositional heterogeneity (often requiring highly resolved domains), because CO_2 injection introduces sharp density and chemical gradients across dynamic plume boundaries (severely limiting time-step magnitude), and because isolation performance must be predicted for hundreds to thousands of years, such applications also demand advanced parallel architectures, numerical methods, and space-time discretization schemes. Finally, the uncertainty bounds that characterize these highly complex simulations must be quantified for risk assessment through sensitivity analyses that span the extensive matrix of site-characterization and site-independent parameters that influence predictive results

The scientific and computational challenges sketched above lie between modeling applications demonstrated to date and those required to provide uncertainty-bound forecasts of CO_2 isolation performance associated with FutureGen power plants. Hence, there is an urgent need to address these challenges and achieve inaugural simulations of this kind.

EXECUTIVE SUMMARY

Optimal siting of a FutureGen power plant in large part hinges on demonstrating appropriate matching of 1) projected incremental and cumulative CO_2 emissions with predicted injectivity and storage capacity of the target geologic reservoir (source/sink matching) and 2) required and predicted long-term CO_2 containment (regulatory compliance). Appropriateness of both matches can only be assessed quantitatively using advanced reactive transport models that are sufficiently well constrained by state-of-the-art site characterization efforts, geo-statistical methods, and site-independent data. Although a number of simulation tools currently exist that would appear capable of such assessment, appropriate source/sink matching and regulatory compliance has yet

to be demonstrated quantitatively in this manner for the space-time-complexity dimensions and resolutions required to address FutureGen-scale CO_2 sequestration projects. The goal of this PRD is achieve inaugural demonstrations of this kind by addressing the key scientific and computational challenges that have precluded their realization to date. Proposals are sought that will incorporate the required advances within existing reactive transport simulators, then apply these improved models to forecast the long-term isolation performance of a representative FutureGen-scale CO_2 storage site.

SUMMARY OF RESEARCH DIRECTION

Although advanced reactive transport simulators are uniquely well suited to forecast the long-term isolation performance of geologic CO_2 storage sites (Johnson et al. 2004, 2005), such utility has yet to be demonstrated for a FutureGen-scale sequestration project. Thus, there is a pressing need to design, develop, and successfully carry out a simulation effort on this scale by identifying, addressing, and resolving a number of key technical issues that separate current and required capabilities. These include scientific challenges (e.g., to define the space-time dimensions and granularity required to accurately predict plume migration, sequestration partitioning among distinct trapping mechanisms, and long-term isolation performance) as well as computational challenges (e.g., to determine requisite hardware, numerical methods, discretization techniques, time-step algorithms, and model coupling approaches). Ideally, this work will be based on the well-characterized geologic setting of a final-four candidate (perhaps even the selected) FutureGen site, although analogous data from an alternative scale-equivalent isolation project could be used, as could representative synthetic data as a third choice. The envisioned research effort will provide the sequestration modeling community with a unique forum for:

- Gaining critical experience in simulating CO_2 injection, migration, and trapping within representative basin-scale systems characterized by multi-scale physical and compositional heterogeneity
- Interacting with geo-statistically-based site characterization efforts to compare and inform with respect to achievable and required resolution of such heterogeneity
- Interacting with generic modeling efforts to obtain advanced computational tools exploitable in this application
- Comparing predictive results from diverse CO_2 sequestration models in the interest of honing and unifying collective capabilities
- Identifying and prioritizing future model development and site characterization activities.

The fundamental outcome of this effort will be demonstrated significant progress along each of these directions.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The focus of this applied modeling study is to:

- Address technical challenges associated with explicitly coupling many distinct process models that represent a broad range of subsurface physics, chemistry, and mechanics
- Greatly expand, while more finely resolving, the space-time scales of geologic CO_2 storage modeling
- Accurately predict plume migration, sequestration partitioning among trapping mechanisms, and long-term isolation performance
- Design and optimize injection strategies and monitoring programs
- Identify key screening criteria that can be used to select optimal sites
- Quantify the uncertainty of these highly complex predictive models.

In terms of model development, a key challenge is to achieve computationally efficient and robust explicit coupling of models that represent interdependent physical and chemical processes operating over a wide range of space-time scales. A fundamental scientific challenge is to determine the spatial dimensions—possibly ranging from plume-to-basin scale depending on reservoir permeability and lateral continuity—that are required to accurately represent injection-triggered pressure perturbations, dependent evolution of the reservoir/cap-rock/well-bore stress regime and strain response, and the ultimate fate of displaced formation waters. Within this appropriately dimensioned domain, a second key challenge is to determine the spatial granularity—possibly uniformly coarse or fine, telescoped from coarse to fine with increasing plume proximity, heterogeneous fine and coarse to mimic known hydraulic and compositional variations, or even dynamic (adaptive gridding techniques)—that is required to provide sufficiently accurate simulation of CO_2 plume migration, sequestration partitioning among distinct trapping mechanisms, and long-term isolation performance, which depends most importantly on the evolution of cap-rock/well-bore seal integrity per concomitant geo-mechanical deformation and geochemical alteration processes (Johnson et al. 2005). Further, spatial granularity of the overall domain must be optimized in concert with temporal granularity of (i.e., time stepping through) overall simulation time; that is, appropriate dovetailing of space/time granularities is particularly crucial in the context of accurately and efficiently representing the migration of steep gradients in fluid-phase saturations and aqueous concentrations, which characterize dynamic plume boundaries.

Once space-time dimensional and granularity issues have been addressed, challenging applications for the simulation capability—beyond site-specific performance prediction—include the design and optimization of injection strategies (well locations, spacing, and fluxes), such that desired injectivity, plume migration paths, ultimate storage capacity, and accumulation location are achieved, and monitoring programs (sampling/imaging techniques, locations, frequencies), such that predicted performance can be efficiently metered and verified. Another fundamental

application of the envisioned capability is to identify—among myriad site-specific and site-independent parameters—key hydrological, compositional, mechanical, and structural properties (screening criteria) whose values most strongly influence long-term isolation performance, which will help define and prioritize both supporting experimental and site characterization activities. A final challenging and much-needed application is to quantify the uncertainty bounds on forecasting CO_2 isolation performance, which can be achieved through sensitivity analysis over the extensive matrix of interdependent site-specific/independent parameters that influence the overall calculation; at present, this critical field of study remains in infancy.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The requisite coupling of diverse process models, the characteristic hydrologic/compositional complexity and heterogeneity of basin-scale geologic systems, and the dependent extreme space-time dimensions and granularities required for numerical simulation of long-term CO_2 isolation performance demands advanced parallel architectures, numerical methods, and space-time discretization schemes. Hence, large computing clusters are required and must be exploited efficiently to carry out these applications; storage issues will be addressed by forming sub-models on individual processors and allowing communication between them through boundary conditions (Dickinson et al. 2006), while inter-compatibility of submodels will be achieved through iteration. Great challenges exist in designing such coupling algorithms and load balancing as well as data scatter and gather among the myriad process models; all potential couplings of scale and process should be explored. As a result, studies within this PRD will help establish guidelines for the design, development, and implementation of robust and efficient massively parallel reactive transport simulators.

Further potential impacts include 1) improved understanding of how long-term plume migration and cap-rock/well-bore integrity predictions vary with adopted space-time granularities, 2) development of new methodologies for quantifying the uncertainty envelope that surrounds predicted isolation performance as a function of those surrounding site characterization efforts, site-independent data, process models, and process coupling, and 3) benchmark comparison of predictive results from diverse CO_2 sequestration models to hone and unify our collective capabilities.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

Defensible, accurate forecasting of uncertainty-bound long-term CO_2 isolation performance will be a required component of FutureGen power plants. Sophisticated predictive efforts of this kind absolutely demand advanced reactive transport modeling capabilities. Studies within this PRD will significantly improve and demonstrate such capabilities and, as a result, our ability to identify optimal geologic targets for CO_2 storage, forecast their long-term isolation performance, and optimize injection strategies and monitoring programs to achieve and verify such performance.

These improvements will be obtained though successfully addressing a number of key questions that separate current and required modeling capabilities:

- What spatial magnitude is required to accurately depict the injection-triggered pressure perturbation and dependent evolution of the reservoir/cap-rock/well-bore stress regime and strain response?
- What space-time resolutions and style of these granularities is required to provide sufficiently accurate simulation of migration paths, trapping mechanisms, and long-term isolation performance?
- What are the quantitative uncertainty bounds on these complex simulation efforts? What are the optimal injection strategies and monitoring programs for achieving and verifying optimal isolation performance?
- What are the key screening criteria whose values most strongly influence long-term isolation performance; that is, how should we define and prioritize both laboratory experimental and site characterization activities that support CO_2 storage modeling work?

The proposed FutureGen application will provide a unique venue for addressing each of these key issues, which—although explicitly cast here in terms of CO_2 storage—are directly or indirectly relevant to a broad range of basin-scale subsurface modeling activities.

TIME FRAME

Research projects within this PRD could be completed in 3 years.

REFERENCES

Dickinson, J. E., James, S. C., Mehl S., Hill, M. C., Leake, S. A., Zyvoloski, G. A., and A. Eddebbarh , 2006, “A New Ghost-Node Method for Linking Different Models and Initial Investigations of Heterogeneity and Nonmatching Grids", accepted in *Advances in Water Resources*

Johnson, J.W., Nitao, J.J., and Knauss, K.G., 2004, Reactive transport modeling of CO_2 storage in saline aquifers to elucidate fundamental processes, trapping mechanisms, and sequestration partitioning: in Geol. Soc. London Spec. Pub. 233 *Geologic Storage of Carbon Dioxide*, Baines, S.J., and Worden, R.H., eds., p. 107-128.

Johnson, J.W., Nitao, J.J., and Morris, J.P., 2005, Reactive transport modeling of cap rock integrity during natural and engineered CO_2 storage: in *Carbon Dioxide Capture for Storage in Deep Geologic Formations: Vol. 2*, Thomas, D.C., and Benson, S.M., eds., p. 787-813.

COUPLED PHENOMENA

HIGH-FIDELITY PETASCALE SIMULATION OF COUPLED PHENOMENA

Clint Dawson, The University of Texas at Austin; John Bell, Lawrence Berkley National Laboratory; Tom Russell, NSF; Ivan Yotov, University of Pittsburgh; Ivar Aavatsmark, Centre for Integrated Petroleum Research; Malgo Peszynska, Purdue University; Richard Mills, Oak Ridge National Laboratory; Mike Heroux, Sandia National Laboratory

ABSTRACT

Current capabilities in mathematically and computationally modeling coupled phenomena are inadequate for addressing future DOE subsurface science challenges. High-fidelity simulation of subsurface processes must be built on a fundamental understanding of multi-phase/multi-component flow and transport, biochemical and geochemical reactions, thermal and geo-mechanical effects, all interacting within a highly heterogeneous porous media. Sensitivity to physical processes and model parameters, and uncertainties in parameters, media properties, and the current state of the system must be taken into account. Informed decision-making will require the development of a comprehensive uncertainty/sensitivity analysis framework, within which coupled process models will be executed hundreds to thousands of times under different scenarios. The magnitude of this effort will tax even petascale computational resources; for it to be feasible will require new approaches to developing efficient and flexible high-fidelity simulators for coupled subsurface processes.

EXECUTIVE SUMMARY

A number of important DOE applications, such as environmental remediation at the Hanford Site, Savannah River Plant, etc.; storage of depleted nuclear reactor fuels; and carbon sequestration require accurate models of the subsurface that integrate a number of different physical processes. Simulation tools used to study these phenomena must couple hydraulic, thermal, geo-mechanical and biogeochemical effects. Simulations must treat multi-phase, multi-component flow in a subsurface dominated by complex heterogeneities, and interacting, scale-dependent processes. Existing codes are unable to simulate these phenomena with sufficient fidelity and certainty for timely decision-making and risk management. The development of new simulation capabilities that can model the full range of coupled phenomena required to meet these needs is a priority research direction. To realize this goal, three fundamentally important tasks are outlined. These include 1) developing mathematical and numerical analysis tools for intelligently coupling/decoupling physical processes across spatial and/or temporal scales, 2) the development of discretization and gridding techniques that model both individual and coupled processes accurately, and 3) the development of robust nonlinear and linear iterative solution methods for efficient implementation of coupled process models on petascale computational resources.

SUMMARY OF RESEARCH DIRECTION

There is consensus within the subsurface community that solving key DOE application problems requires the modeling of a broad range of coupled phenomena in a complex heterogeneous subsurface. Coupling of non-isothermal, multi-phase, multi-component flow with geo-mechanics and bio-geochemistry is required to adequately capture system behavior. Figure 1 connects several DOE subsurface science challenges with the primary, secondary, and emerging physical processes of interest to each application. Despite some general understanding of the processes involved, we do not have a complete understanding of how to describe and efficiently simulate these coupled phenomena. As a result, existing code technologies generally rely on one of two strategies. The first strategy is to couple existing legacy codes via external interfaces. The second approach is to couple all processes through fully implicit temporal discretizations combined with extremely robust but low-order spatial discretizations to ensure numerical stability. These choices are made at the expense of high fidelity, computational efficiency, and the ability to effectively exploit new high-performance parallel architectures. There is a priority need to develop next-generation, multi-physics subsurface simulation capabilities/methods that enable a significant increase in the fidelity of subsurface modeling. Achieving this goal will require three basic elements. First, we must develop novel coupling/decoupling strategies based on a detailed analysis of the interaction of processes in subsurface flow. These new coupling/ decoupling strategies must then be combined with emerging discretization and gridding approaches to improve the accuracy of subsurface modeling methodology without sacrificing robustness. Finally, these discretizations must be integrated with modern solver technologies so that the computational models can effectively exploit new petascale architectures.

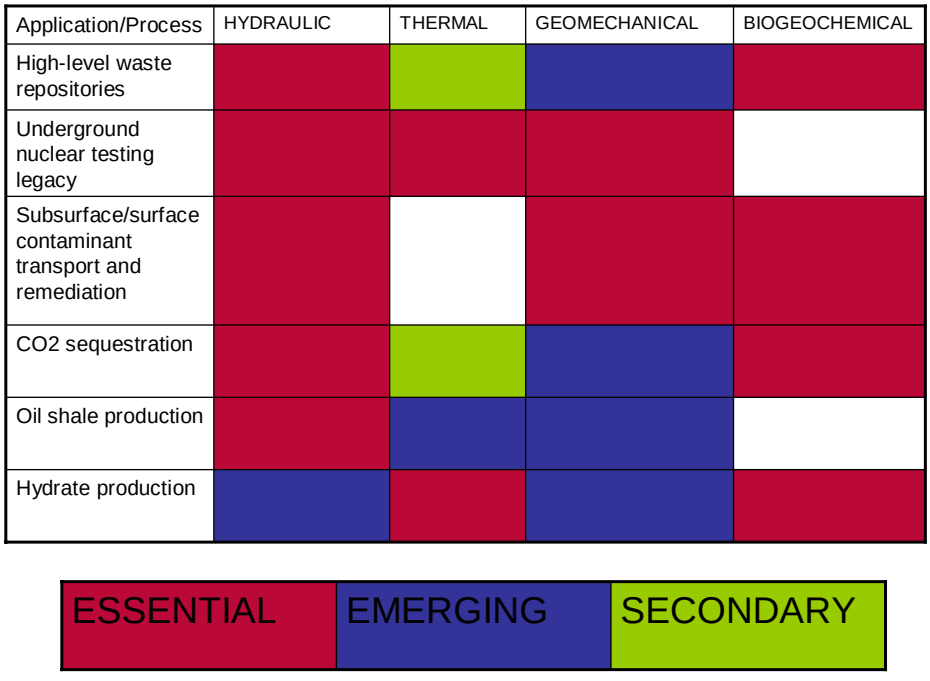


Figure 1. Connections Among Several DOE Subsurface Science Challenges with the Primary, Secondary, and Emerging Physical Processes of Interest to Each Application

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Novel process coupling/decoupling strategies:

Models that couple the broad range of phenomenology needed for next-generation subsurface simulations are incredibly complex. Determining a mathematical model that faithfully represents coupled processes, domains, phases, and components of interest now and potentially in the future is a demanding task in itself. If such a model is accomplished, a simulation algorithm that maintains these couplings at all intermediate steps is unlikely to be practical. For petascale simulation, it is crucial to construct intelligent, robust decoupling algorithms that decouple processes judiciously at intermediate steps to perform efficient computations on parts of the larger systems, where couplings may then be restored upon iterative convergence or accurately approximated through time-splitting techniques. Thus, as a first step toward developing new high-fidelity simulation tools, we need to categorize the different modes of behavior that are embedded within these models. In essence, what is required is a mathematical dissection of the overall model into its principal components, combined with an analysis of how these components interact. This type of analysis will elucidate the mathematical structure of the system and provide the basis for developing robust techniques for mathematically and/or numerically coupling/decoupling different physical processes within a time step or iteration. The objective in developing these types of strategies is to provide an optimal balance between accuracy of the approximation to distinct processes and accuracy of the coupling between processes. This could be viewed as a generic setting for intelligent operator splitting.

Some familiar examples will illustrate the generic idea. An elementary case is decomposition of a self-adjoint operator into orthogonal invariant eigenspaces, which are fully decoupled with respect to the action of the operator. The action of the full operator is decomposed into the actions on the subspaces, and no couplings remain. In subsurface flow of two incompressible liquid phases (e.g., water and either oil or a nonaqueous-phase liquid), it is well-known that the total Darcy velocity of the two phases is a comparatively weak function of the phase saturation, whereas the individual phase Darcy velocities are strong functions; thus, the total velocity is relatively weakly coupled to the saturation. A model expressed in terms of the total velocity and the saturation is therefore more favorable for iterative decoupling than alternative formulations would be. In reactive flows, it is common to split off the reactive processes and phase mass transfer from the rest of the system; this works most efficiently if the interactions between reactions and other processes, such as transport, are relatively weak. This could suggest posing such flows in a Lagrangian frame of reference, where mass transport in phases is better separated from reactions and mass transfer than it is in an Eulerian frame. Another example is the choice of primary variables in a multi-phase multi-component system; certain sets of components could be lumped together into pseudo-components so as to produce weak couplings between pseudo-components. In multi-scale modeling, space-time adaptive mesh refinement can be viewed as a way of choosing patches whose couplings to the larger system are not necessarily weak in the strict sense, but whose exchange terms are nearly static or linear at the level of space-time

resolution, so that the couplings can be accurately preconditioned by linear terms and are mitigated in the nonlinear iteration.

For applications of interest to the DOE, the number of coupled processes, their scale, and relative importance, cannot easily be determined *a priori*. Moreover, new processes may need to be added to account for new technology or previously unknown phenomena. Therefore, there is a tremendous need for an entirely different approach to the modeling of coupled processes. The new paradigm would include all phenomena of interest to a given application and at a given resolution or scale, and guide the user to choose the right level of modeling and computational accuracy reflected in how couplings between the various processes are realized. For success of this strategy, some metrics for measuring decoupling errors would have to be defined. The cross-cutting research objective would be to uncover any unifying principles that could enable a broad range of systems to be analyzed in this way, potentially reaching well beyond subsurface science.

Discretizations of coupled processes

The mathematical structure of the system and the associated coupling/decoupling strategies that decompose the overall model into components provides the starting point for developing new high-fidelity discretizations of the component processes. Discretizations for different physical processes should accurately model each process without introducing errors or instabilities in coupling. The mathematical complexity of some of the basic processes presents a number of challenges to the construction of high-fidelity, robust discretization techniques. For example, the interplay between the nonlinear wave structures associated with Darcy flow and the phase behavior of multi-phase, multi-component mixtures introduces considerable complexity in the construction of accurate discretization schemes for multi-phase flow. The component process discretizations must also provide the flexibility to incorporate interaction with other components and allow for high-accuracy coupling between processes. The broad range of scales appropriate to the different processes introduces considerable complexity into the development of stable and accurate discretization of the coupling.

High-fidelity simulations of coupled subsurface phenomena therefore present new challenges for developing accurate and efficient discretization and gridding methods. Different discretizations and grids (structured or unstructured) may be suitable for different processes, depending on their physical nature, spatial and temporal scales. The accuracy and physical fidelity of each discretization method becomes even more critical in this coupled setting. For example, local mass conservation errors in the flow discretization may lead to spurious mass sources and sinks in the transport simulation. The numerical error for a given process will affect the accuracy in the other processes. Moreover, a loss of stability may occur if the different grids/numerical methods are not chosen appropriately. Interpolations/projections between different grids may be needed, which leads to additional numerical errors. All discretization errors (process specific and coupled) must be controlled through *a priori* and *a posteriori* error analysis. Such analysis is even more challenging when different types of partial differential equations are coupled. In

some cases it may be possible to use existing discretization methods for the different processes. However some couplings may require designing completely new numerical methods to guarantee accuracy and stability.

In addition to addressing the coupling of disparate processes, subsurface simulations must also be able to operate effectively across a wide range of spatial scales and deal with the geometric complexity of subsurface heterogeneity. There are a wide range of modern gridding technologies based on both unstructured and logically-structured grid paradigms that have the potential to accurately represent subsurface complexity. For unstructured grids, the computational mesh can be constructed using tetrahedral, general hexahedra or combinations of both. In the context of structure grids, methodologies based on multi-block curvilinear gridding strategies and embedded boundary representations provide mechanisms for representing complex geometries while retaining some of the computational advantages of structured grids.

These base gridding methodologies need to be augmented with some form of dynamic adaptive mesh refinement to handle the varying spatial resolution requirements. Adaptivity will enable the localization of resolution to track key features such as a geochemical reaction fronts or boundaries of pollutant plumes as they propagate through the subsurface. The heterogeneity of the subsurface introduces a number of novel issues in the construction of adaptive algorithms. In particular, adaptive approaches to subsurface flow need to address the change in the equations that reflect the change of scales at the interface between grids of different resolution.

Iterative solvers and pre-conditioners for coupled processes on petascale computers

High-fidelity simulation of coupled processes leads to very large algebraic systems to be solved at each time step. The solution of these systems often constitutes the vast majority of computation time for subsurface applications, and is usually the part of the calculation least amenable to parallel implementation. The use of smart decoupling strategies can help reduce the demands on the solver and enhance parallelism, but large, sparse, ill-conditioned algebraic systems can still arise from discretizing individual processes. A well-known example is the "pressure equation'' that arises in multi-phase flow models. Furthermore, the development of new discretization/gridding strategies as discussed above requires the development of new solver technologies suitable for these discretizations.

Although substantial progress has been made in the past few decades on solution techniques for large, sparse systems, straightforward applications of these solvers often fare poorly in the face of highly nonlinear and heterogeneous subsurface models. Furthermore, as problem size and complexity grows, solver costs grow disproportionately, making effective solution techniques even more important (Figure 2). This difficulty is further compounded by the fact that many of the most effective solvers exhibit poor parallel scalability on the massively parallel machines required to solve such problems—machines that will grow to the size of tens or hundreds of thousands of processors as we approach the petascale. Effective computational solutions for

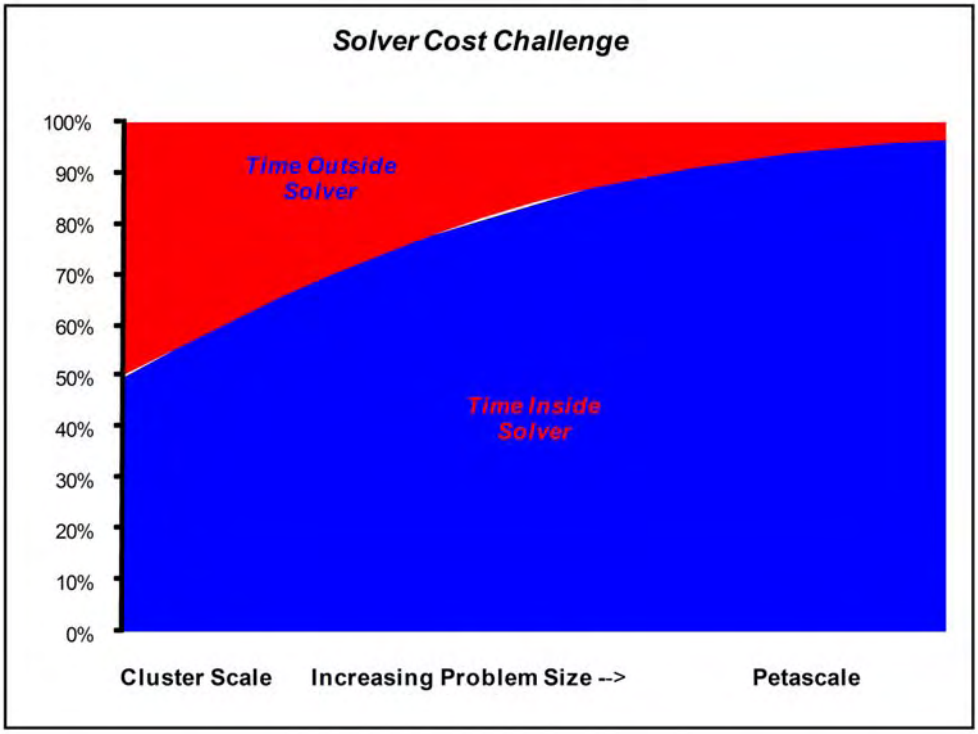


Figure 2. Solver costs increase with problem size and fidelity unless solution algorithms and implementations improve.

complex subsurface problems will demand innovative iterative methods and physics- or process-based pre-conditioners that not only possess numerical properties tailored to the unique challenges of subsurface systems, but are also designed to effectively use emerging petascale architectures. Furthermore, we must invest in the integration of existing software libraries and computational tools that make the development of scalable solutions possible.

In applications where it is advantageous to iteratively converge to a more fully coupled solution, decoupling methods may be used to guide the development of physics-based pre-conditioners. These methods reduce the computational complexity of the large-scale implicit problem by building pre-conditioners based on algebraically decoupling certain phenomena. This opens the door to lower component complexity, the use of scalable multi-grid/multi-level pre-conditioners on subsets of the problem, use of Jacobian-free methods that save memory and computational cost by avoiding explicit calculation of the Jacobian, and use of Schur complements to reduce overall iterative solver complexity. Investigation of the unique challenges in applying multilevel solvers and pre-conditioners (either alone or inside physics-based pre-conditioners) is required. Because many hydrogeologic properties (e.g., permeability) are both position and scale-dependent, coarsening and interpolation methods must be carefully chosen.

Domain decomposition algorithms are another class of powerful parallel solvers. These highly scalable solution algorithms are very suitable for massively parallel petascale computer simulations, because they localize computations and reduce interprocessor communication.

Developing efficient domain decomposition methods for coupled problems is especially challenging, since the convergence of the iterative process is affected by nonlinearities due to complex couplings.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Advances in this framework would lead to payoffs in other aspects of subsurface and computational science. Algorithms based on iterative decoupling of weakly coupled parts would be more amenable to implementation on massively parallel petascale hardware than strongly coupled formulations would be. Couplings complicate uncertainty analysis and quantification by introducing additional sources of uncertainty. If subsystems are found that are known to be weakly coupled, these additional uncertainties are known to be small; hence, the uncertainty analysis can be performed on the subsystems instead of the full system. A similar benefit applies to sensitivity analysis; if subsystems are comparatively insensitive to each other, then the sensitivity of the whole can be analyzed in terms of the sensitivities of the parts.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

This research will lead to advancements in many DOE applications, including contaminant migration, underground nuclear testing, nuclear repositories, oil shale production, and extraction of methane hydrates.

TIME FRAME

The time frame for the proposed research is 3 to 7 years.

MATHEMATICAL AND NUMERICAL SIMULATIONS OF MULTI-SCALE SUBSURFACE PROCESS COUPLING

Lynn Bennethum, University of Colorado Denver; Yalchin Efendiev, Texas A&M University; Scott James, Sandia National Laboratory; Sreekanth Pannala, Oak Ridge National Laboratory; Harihar Rajaram, University of Colorado, Boulder; George Redden, Idaho National Laboratory; Al Valocchi, University of Illinois at Urbana-Champaign; Steve Yabusaki, Pacific Northwest National Laboratory; Marek Zaluski, MSE Technology Applications, Inc.

ABSTRACT

Current multiscale coupling techniques are adequate for nonreactive single phase flows, but not multiphase or reactive flows. In these cases, a framework for robust upscaling and downscaling is required. Next-generation multiphysics subsurface simulators will require novel, adaptive mathematical concepts/computational algorithms that are flexible enough to take advantage of the known coupled physics/chemistry/biology and data at appropriate scales and link them to provide high-fidelity predictions.

EXECUTIVE SUMMARY

Accurate and reliable numerical simulations of coupled subsurface processes are fundamental to critical decisions facing the remediation of contaminated aquifers, siting and design of radioactive waste repositories, geologic sequestration of CO_2, development of alternative energy resources (e.g., methane hydrates, oil shale), enhanced mineral extraction, and water resource management. However, there is considerable uncertainty in the predictions made using current numerical models of coupled subsurface phenomena. Unlike engineered systems, natural geologic formations are highly complex—material properties display high variability and complex spatial correlation structures that span a hierarchy of length scales. In these subsurface systems, coupled subsurface processes involving multi-phase, multi-component interactions exhibit a wide variety of behaviors across a large range of spatial and temporal scales. It is not possible, or is it desirable, to include all these length and time scales in mathematical models of coupled flow and reactive transport. This is due, in part, to the severe computational burden of resolving a wide range of length and time scales to obtain accurate predictions of the subsurface processes. The current practice is to develop models for the continuum scale at which Darcy's law applies and to replace the geometrical complexities of the pore space with empirically determined effective constitutive relations. This is adequate for single phase flow and nonreactive transport. However, for complex, coupled, nonlinear processes (e.g., multiphase flow, biogeochemical reactions, and geomechanics), the lack of rigorous frameworks for upscaling undermines the reliability of predictive simulations used to inform decisions affecting the environment and human health.

Addressing these complexities in next-generation multi-physics subsurface simulators requires *novel, adaptive mathematical concepts/computational algorithms that are flexible enough to take advantage of the known coupled physics/chemistry/biology and data at appropriate scales and link them to provide high-fidelity predictions.* We envisage the development of robust up-scaling and down-scaling, multi-physics approaches that permit the consistent transfer of information across scales through dynamic, adaptive algorithm restructuring on petascale computers. This will facilitate more reliable predictions of subsurface coupled phenomena at the field scale through new knowledge and the accurate portrayal of the interplay between fundamental physical, chemical, and biological processes in complex porous and fractured media.

SUMMARY OF RESEARCH DIRECTION

This priority research direction is focused on developing mathematical concepts, computational algorithms and petascale software required to understand and quantitatively predict the nature of coupled subsurface phenomena (hydraulic, thermal, geomechanical and biogeochemical) across a range of scales. These phenomena are dominated by complex heterogeneities, multiphase, multicomponent flows and interacting, scale-dependent processes. Existing codes are unable to simulate these phenomena with sufficient fidelity for decision-making in part because they do not consistently account for multiscale interactions. The pressing issues facing DOE (e.g., remediation of mixed organic-radionuclide wastes, long-term radioactive waste isolation, impact of underground nuclear tests on water and contaminant movement, geologic CO_2 sequestration, methane hydrate extraction, oil shale development) require models that couple several processes (e.g., flow, biogeochemistry, thermal, geomechanical). For these problems, addressing the complexities of spatial coupling across scales calls for novel multiscale computational frameworks and architectures that can consistently feed information across scales (both upscaling and downscaling) to simulate coupled process dynamics efficiently.

What is needed to develop next-generation multiphysics subsurface simulators are novel, adaptive mathematical concepts/computational algorithms that are flexible enough to take advantage of the known coupled physics/chemistry/biology and data at appropriate scales and link them to provide high-fidelity predictions. We envisage the development of robust computational upscaling and downscaling approaches that permit consistent transfer of information across scales for coupled nonlinear processes and feedbacks.

For example, a multiphase multicomponent reactive transport model might take into account (1) contact angles and variable surface chemistry at the pore scale using the Navier-Stokes equations to describe flow of each phase through complex pore geometry; (2) permeability, suction-saturation, bulk reaction parameter and rate measurements incorporated into conventional multiphase flow equations at the Darcy scale; and (3) upscaled, nonlocal in time and space flow and transport equations at the field scale. An adaptive mathematical/numerical model could incorporate this information in a way that scales either the model or the data in a consistent manner so as to minimize the overall prediction uncertainty. Thus, multiscale modeling may not always require scaling data to fit a particular model scale, but could involve the model adapting to the data provided.

Achieving efficient adaptivity in upscaling is another challenging task for multiphysics/ multiscale problems. Identifying coarse regions and corresponding upscaled models, as well as setting up an adaptive coarse grid in the regions without apparent scale separation, is essential for accuracy and robustness of coarse-scale models. Development of mathematical tools, such as error indicators, to guide these coarsening strategies is an important issue that needs to be addressed. Such a multiscale computational framework would require incorporating ideas from computational upscaling/downscaling, mathematical upscaling/downscaling, and data assimilation. It would require taking advantage of the strengths of each area in a consistent manner with uncertainty determined in a quantifiable way. Implementation of such a multiscale framework would require tremendous (petascale) computational resources.

Within a multiscale framework, it will be necessary to propagate uncertainty through process dynamics across a range of scales to accurately quantify uncertainty in predictions of system behavior and risk associated with non-compliance. The mathematical frameworks for uncertainty propagation (e.g. Monte Carlo techniques, approximate covariance propagation equations, or polynomial chaos theories) lead to computational problems that are vastly different from their deterministic counterparts. The challenge of quantifying uncertainty in nonlinearly interacting multiscale subsurface systems will require novel computational concepts and in turn enhance the computational science arena. Also, there is a need to develop novel flexible frameworks that implement a range of upscaling models (from classical homogenization to quasi-empirical constitutive relations to direct pore-scale simulation) that adaptively balance accuracy and computational burden.

The challenging problem is to develop a rigorous methodology that can capture nonlocal effects accurately for heterogeneous systems across multiple scales where a scale separation assumption is inappropriate. For accurate upscaling, development of approaches and concepts that take into account global effects and while remaining applicable to problems without scale separation is essential. Addressing this challenging task for various subsurface problems is important not only for upscaling, but also for efficient preconditioning techniques that rely on accurate interpolation/ downscaling techniques.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Background

Predictive numerical simulation of dynamic flow and reactive transport processes in natural geologic formations is the primary computational tool for managing and engineering contaminant remediation, radioactive waste disposal, carbon sequestration, extraction of alternative energy resources, and sustainable watershed and aquifer use.

Contaminant Remediation

A critical issue at DOE waste sites is determining whether natural processes are sufficient to contain or remediate contaminants, or if physical, chemical, and/or biological intervention is

necessary to sequester or remove contaminants *in situ*. For example, a 1992 study predicted that 10 years of natural flushing in the Hanford 300 Area aquifer would be sufficient to satisfy the uranium drinking water standard. In fact, elevated uranium concentrations continue to persist. Recent investigations have identified a significantly more complex conceptual model of multi-scale, coupled processes including hourly water level fluctuations that access contaminated lower vadose zone sediments, uranium mobility that is dependent on mixing between river and groundwater, and rate-limited mass transfer at the micro-scale.

Radioactive Waste Repositories

Hydrothermal consequences of heat associated with radioactive wastes must be understood and predicted far into the future to evaluate the effectiveness of geologic system barriers. At Yucca Mountain, fundamental interactions at the interface between engineered and natural materials resulting from heat-driven, multi-phase fluid flow and multi-component reactive transport must be simulated to accurately predict the potential migration of wastes in the short-term, near-field environment as well as the long-term, far-field environments. In this case, the long time scales, lack of experience with analogous systems, and potential consequences demand a mechanistically detailed, multi-scale, scientific basis for the predictive models.

Carbon Sequestration

The success of geologic sequestration of CO_2 will depend on understanding:

- Changes in porosity that affect storage capacity
- Multi-phase flow in the context of additives and catalysts to improve CO_2 stability
- The impact of biological and chemical reactions on reservoir permeability (particularly cap rock permeability and fracture sealing)
- Buoyancy effects and capillary trapping
- Rates of mixing between reacting and non-reacting fluids.

The response of a reservoir to injected CO_2 involves both spatial and temporal scaling challenges. Simulations to date have relied on a number of simplifying assumptions to make the modeling tractable. Improved understanding of the behavior of CO_2 reservoirs will require a comprehensive multi-scale computational framework.

Alternative Energy Resource Development

The DOE is committed to exploring alternative energy resources such as oil shale and submarine methane hydrates. Designing approaches for exploiting subsurface thermal and carbon-based energy resources depends heavily on process coupling across multiple scales. Examples include the installation of freeze walls and the sequence of heat-driven reactions involved in oil shale development, optimization and performance of geothermal reservoirs affected by fluid-rock

interactions, and the impact of procedures and amendments that are used to modify permeability and fluid flow. High-fidelity simulations have been extremely computationally intensive, requiring several months to execute. Novel multi-scale approaches will provide valuable insights on how to improve the efficiency of these simulations.

Mineral Resource Extraction

In situ mobilization of valuable elements that are mediated by biological and chemical reactions have important economic and environmental advantages. How amendments are distributed to locations where they can have impact, and how the subsequent reactions modify system properties, such as permeability, will determine overall performance and economic benefits.

Water Resource Management

Most people get their water from groundwater sources, yet it is a scarce resource with competing industrial and agricultural demands as well as threats to its quality. Houston, for example, faces complex water availability issues resulting from land subsidence from groundwater extraction, salt water intrusion, and decades of industrial groundwater contamination. The unique subsurface environment comprising geo-pressured fluids near the Gulf Coast has resulted in fluid-dependent fault movement that has damaged highways. The ability to use and protect any subsurface resource will depend on understanding and possibly manipulating permeability and flow through physical, chemical and biological processes.

SUBSURFACE SCIENCE CHALLENGE

These subsurface application areas share common attributes including naturally complex, multi-scale heterogeneous porous media (with physical, chemical, biological properties) and coupled interacting scale-dependent processes (e.g., hydrologic, thermal, multi-phase, biogeochemical, geo-mechanical).

Heterogeneity

Natural subsurface systems are physically, chemically, and biologically heterogeneous across a spectrum of scales. Moreover, unlike many engineered systems, the heterogeneity is large, difficult to quantify, and can evolve with time as a result of engineered intervention. Disordered pore structure and variability in physico-chemical surface properties at the micron to millimeter (secondary porosity and grain surface) scale impacts the definition of Darcy-scale porous medium properties such as permeability, dispersivity, and reaction rates. Heterogeneity at the centimeter-meter (core) scale influences equivalent permeability and dispersion tensors and reaction rates at the 10- to 100-meter scale. Geologic heterogeneity is manifest at the 10-meter to kilometer scale, and ultimately impacts site-scale behavior.

It is not possible or practical to deterministically represent property fields (permeability, porosity, stratigraphy, composition, reaction rates, etc.) at all scales at a typical field site or

formation. It is necessary to develop valid methods for correlating the averaged properties of increasingly large volumes to the smaller-scale heterogeneities within the volumes, and to understand how heterogeneous distributions of properties (e.g., porosity, reactive surfaces, biomass) can change in response to various natural or engineered stimuli. See Figure 1. Progress in such scaling approaches will require the use of advanced computing applied to the continuum of scales for different types of systems.

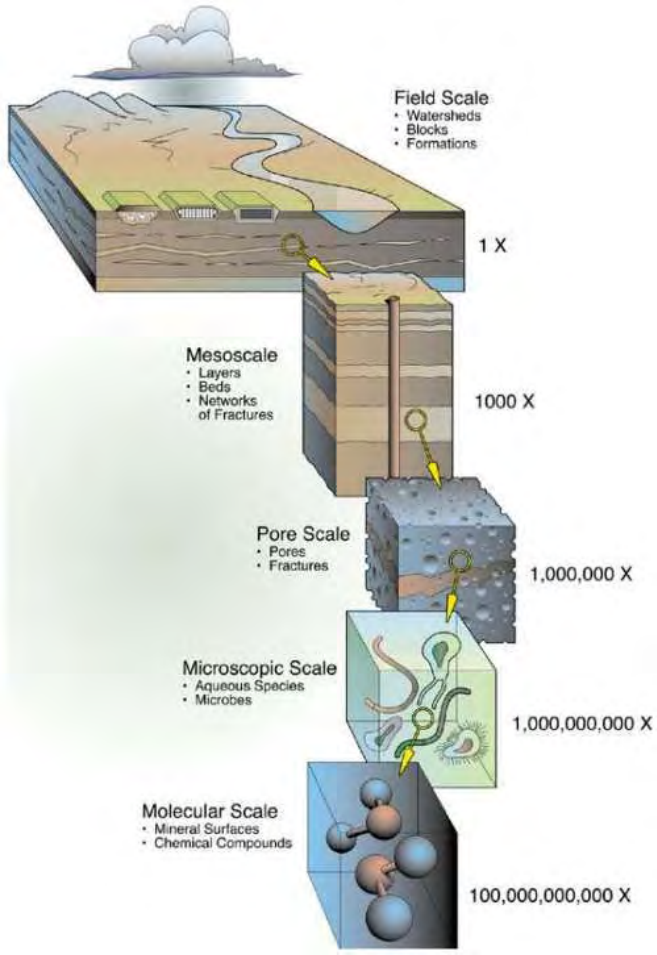


Figure 1. Spanning scales from molecular to the field

Process Coupling

Coupling between physical, chemical, and biological processes involving multi-phase, multi-component interactions and alteration of media properties produces system behavior with enormous complexity across a large range of spatial and temporal scales. Such variability can go unnoticed unless there is an explicit effort to obtain information at multiple scales and under ranges of conditions that can elicit the integrated response of process coupling. For example, geochemical precipitation/dissolution processes and biological growth can alter pore space

geometry resulting in permeability changes, which can in turn affect the flux of reactive fluids, thereby yielding nonlinear feedbacks. Understanding process coupling and scaling becomes even more critical in systems that are characterized by strong, changing and/or episodic gradients. Biogeochemical reactions and physical characteristics of geologic systems along thermal and chemical gradients can often be where the most critical changes take place even though they represent relatively small volumes or even micro-environments. Engineered and natural gradients can change rapidly in degree and location.

Characteristic time scales for interacting processes are often highly disparate. Furthermore, the governing equations and/or constitutive relationships for the same process can potentially take on different forms at different scales. In a given spatial region, temporal evolution as a result of coupled processes can dramatically alter the local dynamics and controlling processes (i.e., associated governing equations).

Understanding macroscopic phenomena (e.g., large-scale transport of fluids, heat, and reactive components in complex, geo-mechanical systems) in terms of fundamental processes at the scales of molecules, cells, solid-solution interfaces, and pores is increasingly being explored to address long-standing issues with the characterization of appropriate field-scale models and parameterizations.

Coupled subsurface processes typically involve phenomena occurring across a wide range of scales such that the existence of a single representative elementary volume (REV), assumed to exist for many of the discrete multi-scale models, may be different for different processes or may not even exist. For example, biologically mediated processes involving micro-environments surrounding biofilms or microbial colonies are typically not explicitly represented in contemporary models that are formulated at the Darcy or REV-scale. Nor are the evolution of these micro-environments and colony structures explicitly related to changes in REV-scale nutrient concentrations resulting from field-scale injection schemes. Significant challenges that are likely to be encountered in developing such a framework to address these issues include potentially different sets of governing equations at different scales, involving different sets of property fields; different types of reaction kinetics at different scales; and vastly disparate time-scales associated with coupled interacting phenomena.

MATHEMATICS/COMPUTATIONAL SCIENCE CHALLENGE

It is often necessary to resolve a wide range of length and time scales in order to obtain accurate predictions of the subsurface processes. There are, however, major computational and mathematical challenges for coupling the sub-continuum to the continuum scale. Most existing mathematical and numerical approaches have been developed for addressing a particular physical process using information transferred from one scale to another scale. For example statistical variability of core-scale measurements can be used in a stochastic up-scaling approach. Darcy-scale properties can be estimated from pore network structure via homogenization or averaging, or under a continuous hierarchical matrix representation, projection operators,

Green's functions, or Fourier/wavelet transforms can be used (Cushman et al. 2002). All of these approaches assume information is given at one scale (or for a continuous hierarchical matrix, an assumed functional form). In reality data is often sparse, incomplete, given at a variety of scales, and subject to error. Thus while the formulation may be accurate, the unavailability of complete data at the appropriate scale limits the usefulness of the standard approaches.

Unfortunately, for the coupled nonlinear processes of interest, the classical up-scaling methods break down and empirical approaches must be adopted. As an example, several investigators (Okubo and Matsumoto 1979; Clement et al. 1996; Thullner et al. 2002) studying the phenomenon of bio-clogging have developed approaches to relate substrate flux and biomass growth to permeability changes. These approaches usually assume idealized biofilm accumulation in simplified porous geometries (e.g., a capillary tube). These empirical up-scaling approaches can in some cases be fit to experimental or field data, but they generally break down when broadly applied to a wide range of conditions. There are no rigorous approaches for risk assessments and engineering designs that work for complex coupled processes like multi-phase flow, biogeochemical reactions, and geo-mechanical processes, which are needed for risk assessments and engineering designs on the target applications above.

Because rigorous mathematical frameworks for up-scaling highly nonlinear processes are lacking, one potential approach is to conduct direct numerical simulations of pore-scale processes at critical spatial locations where high accuracy is required. The severe computational burden of pore-scale simulation has limited the viability of this approach. In practice, however, some type of coarsening (or up-scaling) of the detailed geologic model is usually performed before the model can be used to simulate complex displacement processes.

Thus, the current state-of-the-art multi-scale/multi-physics simulation algorithms/techniques are not sufficient to construct a multi-scale framework that can be used to solve the degree of complexity encountered in coupled subsurface phenomena. A recent National Science Foundation Blue Ribbon panel report (Oden et al. 2006) entitled *Simulation-Based Engineering Science (SBES)* states, "Researchers in the worldwide race toward miniaturization, nanoscience, molecular modeling of drugs and biological systems, advanced materials, and other applications, all of which involve events on atomistic and molecular levels, have run into a formidable roadblock: the tyranny of scales." In addition, the report's finding is that "…formidable obstacles remain in linking highly disparate length and time scales and in bringing together the disciplines involved in researching simulation methods."

The advent of petascale computational resources opens the door to developing robust multi-scale frameworks whose need has long been recognized even in the context of classical subsurface science problems, despite being viewed as computationally infeasible. The emergence of several challenging new application problems listed above, combined with the availability of modern state-of-the-art computational resources, makes PRD timely.

Computational Science Impact

The principal impact of this research on computational science will be the development of scalable multi-scale/multi-physics algorithms on petascale computers that use novel approaches to load balancing, data communication across components, and dynamic/adaptive restructuring of computational grids and algorithms in response to time evolution of complex systems. A critical outcome of developing accurate and robust multi-scale methods to simulate process coupling is the ability to identify how detail may be sacrificed in exchange for accurately simulating increasingly large systems.

Subsurface Science Impact

Advanced multi-scale computational methods on petascale computers will enable progress on two very challenging fronts of natural subsurface environments:

- How is the interdependency between individual processes in complex assemblages expressed as macroscopic, system-level, observable phenomena?
- Can detailed information about coupled processes at small scales be used as the basis for practical (i.e., simpler) and more accurate models of large-scale systems and the macroscopic parameters used to characterize them?

Furthermore, computational simulations that have both resolution commensurate with the scale of study, and sufficient speed to permit testing of competing hypotheses on experimental observations will lead to improved conceptual models of process coupling and quantitative predictions. This is an important step in bridging the gap between laboratory knowledge and field systems where such knowledge should be applied. Building the scientific basis for commonly used parameters, and evaluating model parameters for different conditions are critical not only for advancing the subsurface science, but for providing the basis for policy and regulatory decisions.

In summary, the potential impact of advanced computing on subsurface science can be profound. It is not a question of simulating subsurface environments in explicit detail at all scales, but rather understanding first how fundamental physical, chemical and biological processes are interdependent in complex porous media and how the interdependence is manifested for different rates of system perturbation. Ultimately, we expect to test approaches for translating knowledge of coupled processes from the basic scales of coupling to simplified, but science-based, models applied to the larger scales of interest.

TIME FRAME

A three-year program, with first projects funded during FY08, would permit the development of an integrated, petascale data management toolkit for coupled data, driven by specific subsurface science applications and exhibiting complex workflows, including intelligent sensors, real-time visualization, and steering.

REFERENCES

Clement, T.P., B.S. Hooker, and R.S. Skeen, Macroscopic models for predicting changes in saturated porous medium properties caused by microbial growth, *Ground Water*, 34(5), 934-42, 1996.

Cushman, J.H., L.S. Bennethum and B.X. Hu (2002) A primer on up-scaling methods for porous media. Adv. Water Resour. 25:1043-1067.

Oden T. J. et al., "A Report of the National Science Foundation Blue Ribbon Panel on Simulation-Based Engineering Science: Revolutionizing Engineering Science through Simulation," http://www.ices.utexas.edu/events/SBES_Final_Report.pdf

Okubo, T. and J. Matsumoto, Effect of infiltration rate on biological clogging and water quality changes during artificial recharge, *Water Resources Research*, 15(6) 1536-1542, 1979.

Thullner, M., J. Zeyer, and W. Kinzelbach, Influence of microbial growth on hydraulic properties of pore networks, *Transport in Porous Media*, 49(1), 99-122, 2002.

DATA MANAGEMENT FOR COUPLED PHENOMENA

Richard Ewing; Guan Qin; Guy Almes, Texas A&M University; Wei Zhao, Renesslar Polytechnic Institute; Patrick O'Leary, Idaho National Laboratory; Hong Wang, University of South Carolina; Craig Douglas, University of Kentucky

ABSTRACT

Current capabilities for data management (i.e., acquisition, transport, storage, calibration, analysis, integration, and visualization) are inadequate for modeling and simulating complex, coupled, multi-scale subsurface processes. As the processes requiring understanding and prediction become more complex, coupling of large-scale phenomena becomes more important and data management requirements become more challenging. Because processor speeds are not increasing significantly, coupled subsurface computations will need to use more processes and data management tools that allow for effective use of many processors for coupled bio-, geo-, and chemical processes. Adaptive decouplings within the code, to take advantage of weak physical couplings, or domain decomposition techniques will require very flexible, modular data management capabilities that currently are not available.

EXECUTIVE SUMMARY

For many years, the DOE has invested in subsurface computational research programs to exploit information technology to strengthen subsurface modeling and simulation technologies for solving major environmental problems. With rapid advances in information technologies, and especially telecommunication technologies in the past decade, the new computational infrastructure that can be featured in petascle computing and multi-petabyte online storage is on the near-term horizon. There is the need for the DOE to revisit the grand challenges in subsurface computational sciences to develop next-generation multi-physics subsurface simulation capabilities based on novel coupling/decoupling strategies combined with emerging discretization, gridding, and solver technologies that will fit into today's computing environment.

Subsurface geology exhibits multiple scales because of sequences of geological events that can create layers, fractures, cavities, caves, and faults at various length scales. Coupled subsurface phenomena among hydraulic, thermal, geo-mechanical, and biogeochemical processes are dominated by complex heterogeneities, multi-phases, multi-component flows, and interacting, scale-dependent processes. Consequently, subsurface modeling and simulation exhibit extremely complicated computational challenges that can be characterized in the couplings among multiple physics as well as multiple scales behaviors.

Data is at the center of subsurface computations. Efficient and robust data handling and management is a critical requirement for these computations. Subsurface data are in multiple scales with various formats and must be addressed and used dynamically. Consequently, we

urgently need to develop an integrated petascale data management tool that can fully leverage the state-of-the-art computing environments and accommodate the needs in modeling complicated physical phenomena. There are great challenges involved in developing effective and efficient data management tools for data acquisition, storage, querying/analysis, integration, and visualization.

SUMMARY OF RESEARCH DIRECTION

The goal for this PRD is to create integrated, petascale data management tools for coupled data. Advances in computational infrastructure, with petascale computing and multi-petabyte online storage on the near-term horizon, are needed to support simulating coupled processes on very large spatial and temporal scales. The architectures and software tools for managing the requisite data and for choreographing the interactions between computation and data, however, are currently not available.

The geological modeling process integrates available geological information, such as seismic data, well observation data, core samples, and basin simulation results, to construct a discrete model that describes the geological features of the area of interest. Because of uncertainties in the subsurface, geological modeling is an evolving process that, over time, integrates newly available dynamic data such as production history data through wells or new interpretations of legacy datasets. A team of experts in disciplines and, sometimes, in various geographic locations contribute to this modeling process. The geological modeling process generally requires an integrated data center that has large storage capacity, efficient capabilities to manage and handle large datasets, and the flexibility to create and modify workflows.

Once an adequate geological model has been developed, subsurface fluid flow modeling and simulation are performed on a reservoir model that, because of the limitation of computing power, is often an up-scaled geological model. Generally subsurface fluid flow processes involve the flow of multi-phase, multi-component fluids that may be coupled with chemical, thermal, geo-mechanical, and biochemical phenomena. These flow processes are also often dominated by complicated geological structures involving heterogeneities and scale-dependent phenomena. Modeling and simulating such a complicated flow process generally involve large a system of nonlinear partial differential equations; the discretized form of the system involves many unknowns. Numerical solutions of the resulting system require many iterations, and a large linear system must be solved for each iteration. This modeling and simulation process is very computationally intense, involving constant interactions between computations and large datasets. Huge mathematical challenges arise in solving these simulation problems, in terms of pre-conditioner technologies for linear solvers, domain decomposition technologies for the parallel computing environment, and new discretization technologies.

Because all modeling and simulation processes are data and computationally intense, it is vital to create integrated petascale data management tools for coupled data. Because of advances in computational infrastructure, with petascale computing and multi-petabyte online storage on the

near-term horizon, simulating coupled subsurface fluid flow processes on very large spatial and temporal scales is becoming attainable. The architectures and software tools for managing the requisite data and for choreographing the interactions between computation and data, however, are not currently available. The major issues for developing such a data center are as follows:

- Input datasets are huge, come from diverse sources in multiple formats, and must be ingested at various different (spatial and temporal) scales. Typically, for example, input data are collected from different collaborators located in various geographic locations and may derive from field measurements, from laboratory measurements, or from the output of computational models. In addition, because of the need for dynamic multi-scaling, it may not be fully known at the outset which datasets will need to be ingested.
- Output datasets are similarly large and formatted for use by multiple different applications. It is also very common that an open-ended number of other collaborators or systems use the output data of one system.
- During the coupled computation, large amounts of data will be transferred among the various computing elements engaged in the computation. Again, because of the need for dynamic multi-scaling, the pattern and volumes of these transfers will not be know at the outset of the computation.
- During the coupled computation, large checkpoint files will need to be written.
- In addition to existing input datasets, real-time data from intelligent sensors will be important inputs to subsurface science computations. These intelligent sensors, when combined with the dynamic data-driven application systems paradigm, make it possible to modify the configuration of the intelligent sensors during the course of the computation. (Douglas et al. 2005)
- It is particularly important for datasets used in subsurface science to be coordinated with accurate geographic information systems (GIS) information while modeling coupled surface/subsurface systems.

For example, a geological model is usually based on an integration of all available geological information such as seismic data, production data, well log data, and rock and fluid data. The resulting geologic models have hundreds of millions of elements with billions of unknowns. Another example is that the near real-time satellite imagery data collected/analyzed by the National Oceanic and Atmospheric Administration (NOAA) are widely used by scientific simulations systems in predicting precipitation and temperature, which are in turn used by other simulation systems (e.g., fire risk simulation and disease spreading simulation).

This need for coupled computations to receive inputs from many different sources and to provide outputs to many different applications strengthens the need for data and metadata format standards. One current example, relevant to coupled computations for subsurface science, is the current version of Hierarchical Data Format (HDF5). Metadata for specific kinds of data

relevant to subsurface science is needed; for example, the Open Geospatial Consortium (www.opengeospatial.org) for a significant metadata initiative in a related field.

Further, subsurface scientists need to be able to craft workflows that flexibly receive and store the data, pass it among elements of the workflow computation, and then store, visualize, and deliver the data to other applications. In many cases, as shown in the early work on dynamic data-driven application systems, the workflow may become a "symbiotic feedback control system."

Aspects of subsurface science and of coupling intensify these problems, common in several fields of computational science of interest to SciDAC. Subsurface science is heavily multi-scale and multi-phase (due, for example, to the radically different viscosities of oil, water, and gas, and to the radically different permeabilities of limestone, salt, and sand). Coupling adds to this by requiring data relevant to several different phenomena to be brought together in a single computation, thus amplifying the volume and diversity of data to be ingested.

Given these distributed, large-scale, near real-time data sharing/collaboration requirements, a common data collaboration platform is urgently needed. This collaboration platform should provide integrative support for data management/sharing and for choreography among data assembly, computation, visualization, and the steering of computations and instrument configuration,

In addition to software challenges, coupled subsurface computations are likely to need large primary memories (to deal with the large matrices common in these computations), interconnect fabrics capable of high-volume MPI calls, and very high throughput between/among processors, local disks/storage area networks (SANs), and wide area networks (WANs). Similarly, these applications will challenge our abilities in scientific visualizations, including three-dimensional visualizations.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Figure 1 illustrates the data flow model in a typical scientific computing system. Data are collected from different locations and stored locally. The stored data are then integrated into formats that can be easily queried by local scientific modules which perform the various scientific analyses. The output of the analyses are then put back to the data store and made public to other researchers through the data publishing process. For reasons given below, this simple data flow model is inadequate for coupled subsurface systems.

Data assembly challenges

The objective of data assembly is to receive large amounts of data from distributed autonomous sources. The size of collected data can be daunting. For example, seismic data are generally in the terabytes for each seismic survey and the size of data can be grown very fast, especially in offshore explorations if on-site data processing permits. Typically, seismic data tapes are

transported to a centralize location because they are not very time sensitive and are very large for online transfer due to the limitation of bandwidth. These data sizes result in several very challenging tasks:

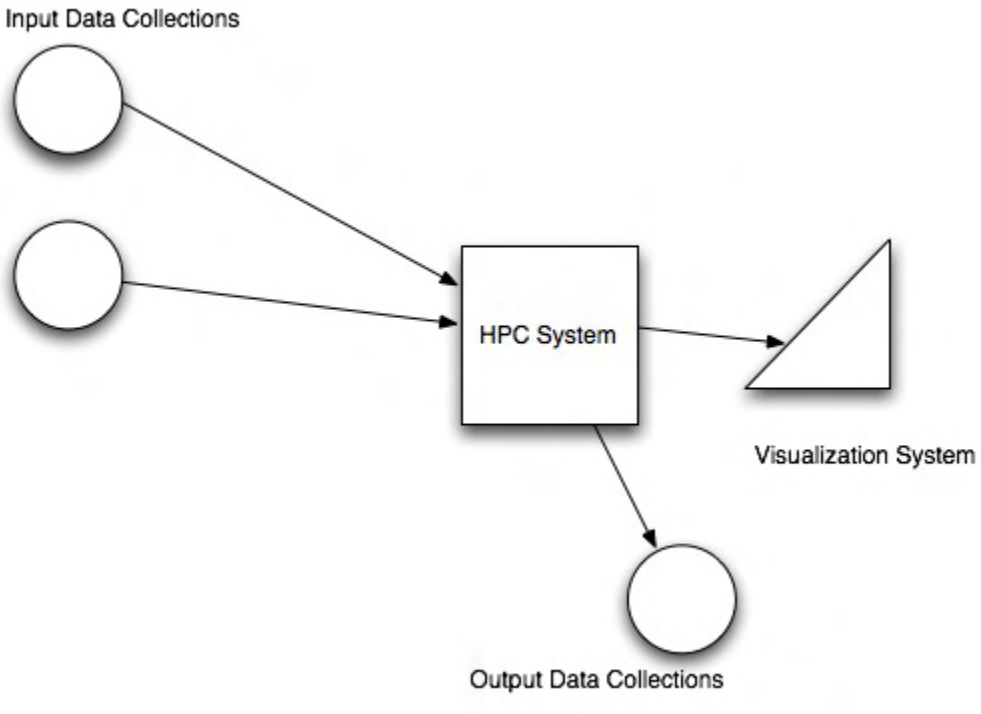


Figure 1. Simple View of the Typical Data Flow in a Scientific Computing System

- How can the relevant systems receive such volumes over constrained WAN resources and with constrained disk and SAN resources?
- How can the relevant systems manage and support the different formats of the data collected from various sources? It is very difficult to automate the transformation between different data formats in an efficient manner, especially when the data size is significantly large.

While in the simple model data are stored centrally, the data in modern subsurface systems come from multiple diverse sources, both as stored data collections and as real-time data from field and laboratory intelligent instruments. In these cases, on-the-fly or just-in-time data assembly are required. There is no centralized data repository, but data are instead assembled (and formats are transformed) in real-time from the distributed sources when needed. This strategy saves the cost for massive data storage and avoids some legal issues associated with the centralized storage of collected data. Nonetheless, it requires common data and metadata to be (either automatically or manually) generated as an integration of the views of data on all distributed sources. It remains a challenge to efficiently map between the common and local schemas of distributed sources, especially when such schemas change dynamically. Data centers that adopt this strategy include the Geosciences Network (GEON), the Community Data Portal (CDP), and Google.

Interacting with intelligent sensors

Several U.S. funding agencies are challenging the sensor community to redesign all sensors that were last overhauled in the 1960s. The response is producing new sensors that contain processing capabilities, program memory, and flash memory in lieu of conventional hard drives. Many new sensors communicate using radio signals and can be reprogrammed and/or reconfigured in the field.

The data center receives large amounts of data, which will come from new, smart sensors in short order. As the data comes in and are filtered and processed, the sensors may need to be reconfigured by steering algorithms to provide different types, resolutions, and accuracies of data. While there might be a human (aided by real-time visualizations) in the loop, automatic procedures need to be developed.

These intelligent sensors are valuable to couple subsurface science because the performance of subsurface science computations depend on the effectiveness of domain decomposition in enabling parallelism. The effectiveness of domain decomposition depends, in turn, on the richness and accuracy of input data. Intelligent sensors and the DDDAS techniques they enable help provide this rich accurate data.

Data storage challenges

The objective of data storage is to organize scientific data with different characteristics in a well-disciplined physical structure, thus facilitating efficient data retrieval. The exponential growth of storage requirement exceeds Moore's law, and is expected to continue in the future. Given such unprecedented size of data storage, traditional file- or relational-database-based methods will not suffice. Advanced scientific databases, such as statistical databases and data warehouses, are needed to support both regular queries on data records and also aggregate queries on the statistical properties of data. Such databases need to support the highly heterogeneous and large datasets common in subsurface science.

Data querying/analysis challenges

The objective of data querying is to help scientists analyze/interpret data and discover useful knowledge. The design of data analysis should provide analytical processing and knowledge discovery tools on large amounts of data. A major challenge to prepare the backend data store structure (e.g., indexing, metadata indexing) is how to design its interface to maximize ease of use and power. This may sound trivial, but considering the significant size of the data, how to find the subset of data of interest (which may still be of significant size) from this huge set itself is a difficult task, let alone the challenge in how to perform advanced analysis (e.g., from basic sorting and averaging to advanced analysis such as statistical distribution).

There are two kinds of design strategies for data analysis:

- *Online analytical processing*. This strategy generates analytical results on the stored data in response to user/application requests. Because the results are generated on the fly, there is no need to store the generated results; thus, this strategy has less demand on storage. However, because of the rigid online response time requirement, this strategy may not support many complicated analytical applications that are required by scientific data analysis.
- *Offline analysis of data*. Here, analyses are done periodically or automatically when the data repository changes (e.g., when new data records are received).

Modules and data flows

Because of the multi-physics, multi-scale nature of subsurface modeling and simulation processes, the proposed data management tools need to move large datasets efficiently and to have the flexibility to create and modify workflows that integrate several computational modules through the process. To accommodate these requirements in subsurface applications, modern forms of modularization are needed. These forms need to move beyond subroutine libraries as found in C++ and Fortran. CASE (computer aided system engineering) tools in the business community, which recognize data flows between objects, support a much more adequate programming model. The modules should accept data from the data center and return a data flow. Figure 2 shows a workflow from SCIRun that, in some respects, illustrates these ideas.

Filtering the data into a format that the module accepts is a necessary overhead that will be feasible using inexpensive, fast processors.

Legacy codes provide one of the most significant places for improvement. While legacy codes provide stability, they also contain long-time imbedded bugs and designs based on obsolete computer designs. They are typically hard to maintain and require employees versed in archaic technologies. Providing modules that can be updated and replaced over time as computer architectures and standards change will provide highly efficient computations for the computing platform and pay for itself over time. To be fast, the modules will be architecture-dependent. Multiple modules that deal efficiently with different architectures should automatically be run by the data center (e.g., one for SMPs would differ from one for distributed memory systems or hybrid systems).

A reliable, time-efficient, computation, not dependent on a historically used language is the goal. Using standards for data flow will allow field users to connect many different modules specific to possibly proprietary methods.

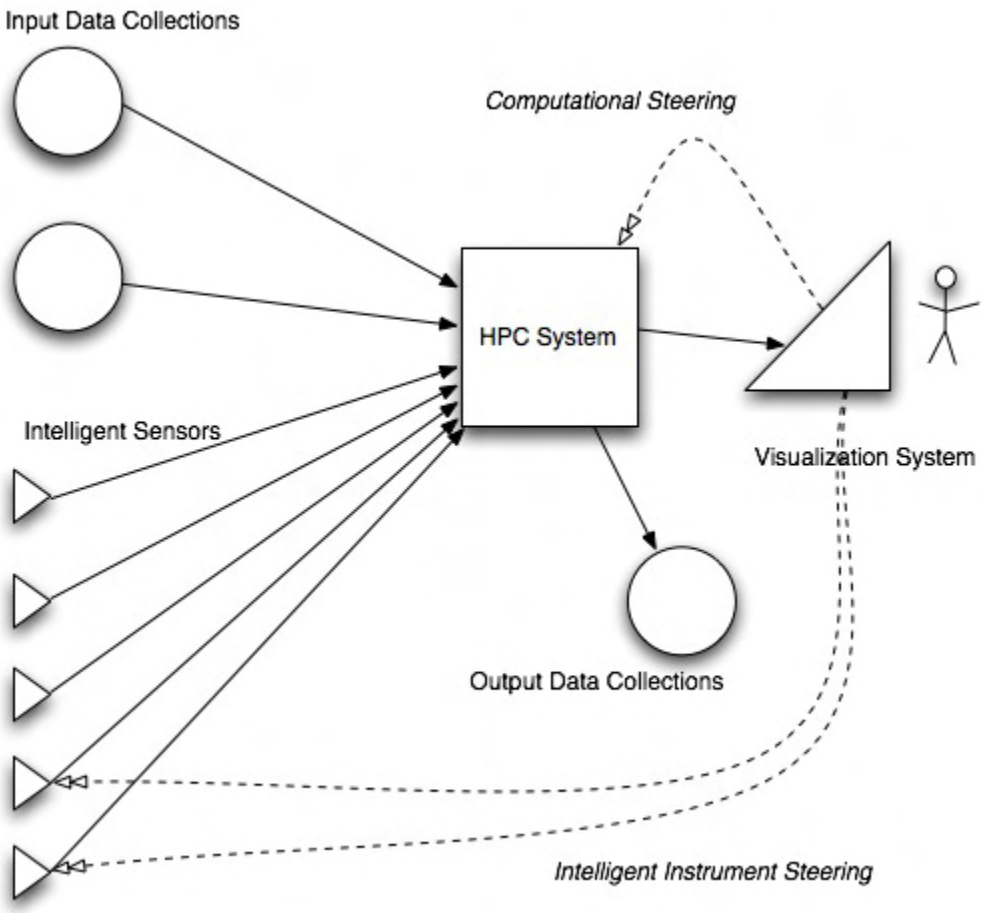


Figure 2. Workflow from SCI Run

Visualization and data analysis

Research efforts in computational subsurface sciences and, in particular, modeling of carbon sequestration processes will require new tools for visualization and data analysis for both development and discovery. Visualization and data analysis will help scientists and engineers predict the effectiveness of proposed algorithms, models, and sequestration sites, while assisting regulators in risk assessment and monitoring. These new tools must support the exploration of large remote complex datasets produced by petascale computing resources. Over the past two decades, both visualization and data analysis have been shown to enhance understanding of large complex data by reducing the dimensionality of the data.

"Visualization researchers should collaborate closely with domain experts who have driving tasks in data-rich fields to produce tools and techniques that solve clear real-world needs." [Chris Johnson, Universityof Utah]

The availability of visualization and data analysis varies greatly across the subsurface science community. Other than a few simple commercial (e.g., EarthVision) and proprietary products,

access is limited to general-purpose visualization and analysis open-source packages such as VTK, SCIRUN, VISIT, and PARAVIEW, which require a high level of expertise to use. These general-purpose tools typically require extensive customization to be useful for this discipline. The subsurface science community requires a unique set of visualization tools tailored to their discipline and design to enhance usability with unique interfaces for domain experts, computational scientists, and regulators. These domain experts demand tools that "automagically" detect, extract, and track the features of interest and present them in a consistent manner. Figure 3 is an example of a visualization that is subsurface-specific (Johnson 2002, 2006; Parker 2006; Weinstein 2005).

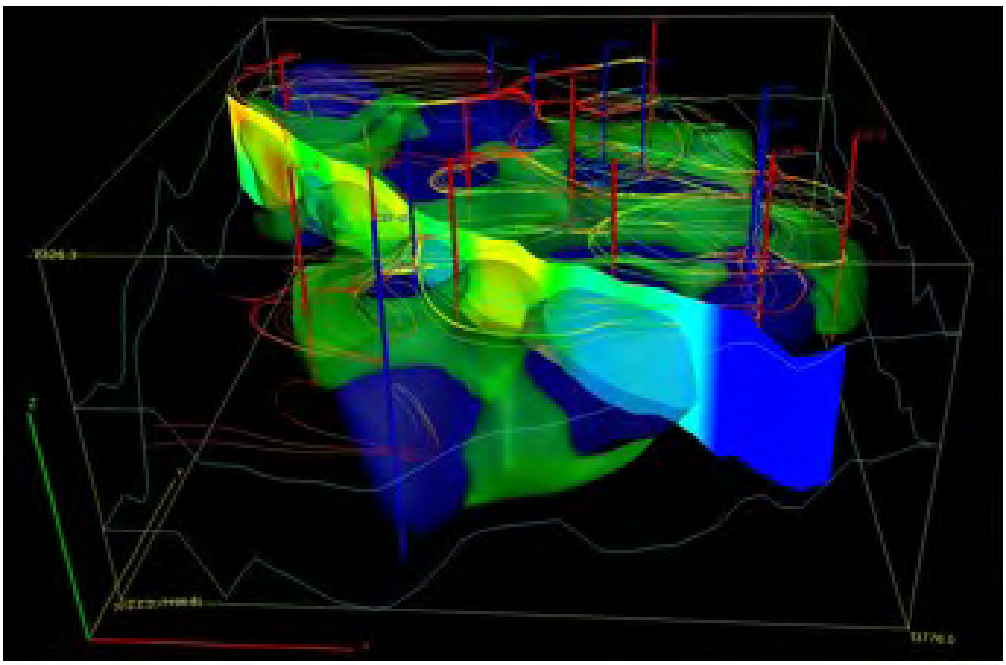


Figure 3. Subsurface-Specific Visualization Mapped onto Three-Dimensional Space

Both model development and risk assessment will result in a number of simulations that require both comparative and uncertainty visualization techniques. In Monte Carlo simulations, comparing the ensemble average to the extreme may provide insight. The amount of uncertainty in the proposed research requires visualization to map uncertainty to simulation results.

As computational subsurface science simulations increase in resolution, they will produce extremely large datasets. Even if current post-processing tools were able to handle these datasets, the network and local storage would be pressed to accommodate them. Therefore, large-scale and remote visualization and analysis tools that scale from the desktop to advanced immersive environments must be developed. These tools should provide a fusion of visualization techniques for multi-field scalar, vector, tensor, and flow data.

Finally, Geographic Information Systems are proving invaluable and must be integrated with visualization and data analysis tools.

Data standardization

Coupled computations rely on data from diverse sources. Data translation into common formats, where possible, will reduce the programming workload for the scientist. Formats such as HDF5, along with relevant metadata, are needed for all inputs and outputs. Furthermore, standardizing computational modules and their interactions will streamline model implementation.

Hardware implications

The hardware used for HPC in the past five years has trended toward multicore processors, faster system buses, and faster cluster interconnect fabrics. This enables very high cluster processing capacity without marked increases in processor clock rates. This trend will likely continue; in effect each cluster node will be a respectable higher-end SMP system. Moore's law will continue to hold, but will result in greater parallelism (e.g., more cores per node and more nodes per cluster totaling several tens of thousands of cores per cluster) rather than in faster processors.

Thus, greater performance will only be attained if greater concurrency can be achieved. The coupling present in subsurface models, however, makes achieving this concurrency more difficult. In part, this concurrency can be achieved through domain decomposition, but only if rich, accurate input data can be provided. Concurrency can also be increased by using rich, accurate data and decoupled techniques in early iterations of model solution; fully coupled techniques can then be used more effectively in final stages of model solution.

Taking these ideas together, Figure 4 shows a complex workflow, including intelligent sensors, real-time visualization, and steering, all critical elements of an integrated, petascale data management toolkit for coupled data.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Petascale data management arising from coupled processes will demand the emerging data-intensive capability. This is an indispensable paradigm shift for coupled subsurface sciences, with broader application to other scientific disciplines. By removing effective data handling as a barrier to the implementation of large coupled simulations, computational science will be strengthened both as a field in itself and in its application to a number of increasingly data-intensive applications. As the number, size, and diversity of data collections ingested, used, and produced during coupled computations increases, the need for effective approaches to data quality assurance will be enhanced. This will drive our need to verify and validate scientific datasets, to consolidate the data formats and metadata standards, and to visualize them effectively. Further, it will improve our ability to apply unstructured grids in a variety of computational science application areas.

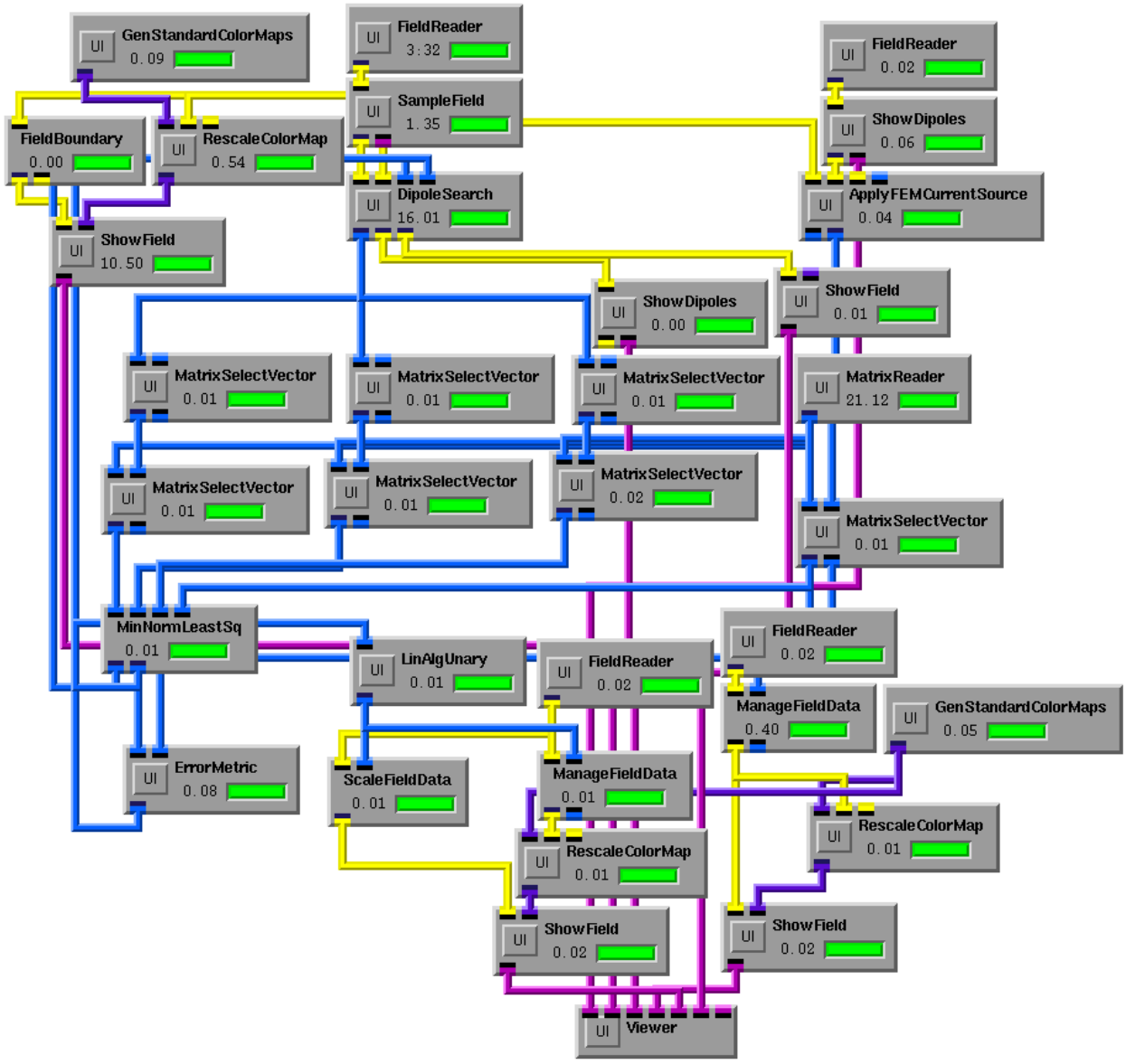


Figure 4. Complex Data Flow with Intelligent Instruments and Steering

POTENTIAL IMPACT ON SUBSURFACE SCIENCES

Challenges such as contaminant transport, CO_2 sequestration and nontraditional energy production demand effective coupled simulation. The integration of dispersed data sources and the management of petascale datasets will enable coupled process simulation and the knowledge required by decision-makers. Following are brief descriptions of the challenges facing DOE programmatic offices:

- Challenges for EM are in subsurface transport, coupled biogeochemical processes, attenuation, stabilization, modes of release, risk assessment, and regulatory compliance. All of these applications are extremely data intensive and require robust data management tools.

- Challenges for RW are siting, licensing, constructing, and operating high-level waste sites. The processes involved include infiltration of water, corrosion of materials, and coupled hydro-, thermo-, and geochemical processes. Chemical transport in diverse media and long-term uncertainty in the data greatly complicate the data management issues.

- Challenges for FE include CO_2 sequestration, oil shale and oil sands recovery, enhanced oil recovery, improved fossil energy utilization, and hydrogen production from fossil fuels.

Because coupled processes are present in each application domain of importance to subsurface sciences, we must overcome the difficulties in treating enormous, disparate data structures that require adaptive access to metadata and powerful data retrieval capabilities. Because the massive coupling of large processes generates enormous, dense matrices in simulation codes, the adaptive utilization of uncoupling processes for mildly coupled processes must be developed. The development of efficient, effective data management tools for large coupled processes will enable major advances in our modeling and prediction of subsurface processes, and will greatly assist in critical decision-making processes for the DOE.

TIMEFRAME

A three-year program, with first projects funded during FY 2008, would permit the development of an integrated, petascale data management toolkit for coupled data, driven by specific subsurface science applications and exhibiting complex workflows, including intelligent sensors, real-time visualization, and steering.

REFERENCES

Douglas, C. C., Efendiev, Y., Ewing, R., Ginting, V., Lazarov, R., Cole, M. J., Jones, G., & Johnson, C. R., 2005"Dynamic data-driven contaminant simulation," Computational Sciences - ICCS 2005, 2, 640-647.

Johnson, C. R., Moorehead, R., Munzner, T., Pfister, H., Rheingans, P., and Yoo, T. S., (Eds.), 2006: NIH-NSF Visualization Research Challenges Report; IEEE Press, ISBN 0-7695-2733-7.

Johnson, C. R., Parker, S., Weinstein, D., and Heffernan, S. 2002. "Component-based problem solving environments for large-scale scientific computing", *Concur. Comput.: Practice and Experience*, 14, 1337–1349.

Parker, S., Zhang, K, Damevski, K, and Johnson, C. R., 2006. "Integrating component-based scientific computing software", in *Frontiers of Parallel Processing For Scientific Computing*, Edited by M.A. Heroux, P. Raghavan, and H.D. Simon, SIAM Press.

Weinstein, D. M., Jones, G. M., Parker, S., Simpson, J,. Zimmerman, K., 2005. "The SCIRun visualization system", in *The Visualization Handbook*. Editors Chris Johnson and Charles Hansen, Elsevier Academic Press, New York, 615-632.

http://www.sci.utah.edu, SCIRun problem solving environment.

http://www.hdfgroup.org, Hierarchical Data Format project.

http://www.dddas.org, Dynamic Data-Driven Application Systems project.

RESEARCH AT FUNDAMENTAL SCALES

MULTI-SCALE ANALYSIS

Leonid Berlyand, Pennsylvania State University; W. Brent Lindquist, Stony Brook University; Konstantin Lipnikov, David Moulton, Los Alamos National Laboratory

ABSTRACT

Providing safe and reliable energy while protecting the quality of critical groundwater supplies requires significant improvements to uncertainty quantification in simulations of subsurface flow. Specifically, experimental and field scale measurements have demonstrated that reactive processes at the pore scale may strongly affect long-range, even field-scale, contaminant transport. While modeling efforts exist at specific fundamental scales, multi-scale techniques that accurately capture the influence of geochemical reactions over a range of strongly coupled, non-well separated scales is in its infancy. Moreover, the required dynamic range of scales in both time and space is sufficiently large that fully resolved reaction simulations are unattainable. Combining the potential resolution and fidelity of direct numerical simulations made possible by Petascale computing with current experimental and modeling efforts will stimulate new computational and theoretical advances in multi-scale analysis. These advances will play a critical role in conducting simulations with well characterized uncertainty which, in turn, will drive responsible policy decision and program development.

EXECUTIVE SUMMARY

The coupling of mathematical models between fundamental scales (i.e., atomistic, pore, laboratory, and field) plays a defining role in understanding subsurface phenomena related to aquifer management, waste clean-up, and CO_2 sequestration. Laboratory and field-scale measurements show that reactive processes at the pore scale may have significant impact on long-range contaminant transport. For example, failed predictions of uranium removal by ambient groundwater flow at the 300 Area of the Hanford Site are increasingly thought to be a consequence of slow leaching at sub-millimeter length scales. Existing theories and computational tools have limited capability to treat the extreme range of closely coupled time and length scales typically present in reactive transport. New developments in multi-scale analysis are required to significantly reduce uncertainty in subsurface flow simulations. Priority research directions must target the development of rigorous approaches for up-scaling in the presence of many spatial and temporal scales. Application of such approaches to models of reactive transport with high reaction rates is of critical importance. Progress in this direction will drive development in related areas: 1) up-scaling flow and non-reactive transport in disordered or random media; 2) down-scaling of coarse data used for site characterization; and 3) up-scaling over time scales. The computational component of this research is critical and depends on petascale class computing to advance numerical methods and mathematical models at specific scales, facilitate bridging of neighboring scales through direct high-resolution simulations, and conduct large ensemble studies. The synergy of these studies will provide the critically needed,

high-fidelity predictive simulations, with well characterized uncertainty, that are necessary for predicting accurate flow dynamics at the field scale.

SUMMARY OF RESEARCH DIRECTION

This PRD is focused on developing the mathematical and computational algorithms, heavily leveraged by petascale computing, that are required to understand and quantitatively predict the nature of up-scaling for geochemical reactions in subsurface flow.

Why now?

Petascale computing will facilitate direct numerical simulation (DNS) treatments over a small range of scales that will allow observation of certain critical multi-scale phenomena (Figure 1). These benchmark computations will provide a basis for developing new computational and theoretical tools to bridge ranges of scales, ultimately up-scaling to the field scale.

Figure 1. A pore level resolving, reactive lattice Boltzmann computation (involving two fluids and four reactant species) could simulate reactive flow and transport in a 1 cm^3 volume of 20 percent porosity media using 10^2 processors, assuming each processor had 1 GByte of independent memory. To simulate 1 m^3 of material would require 10^5 processors (i.e. petascale computing ability). Such volumes are approached by current laboratory-based experiments such as the depicted column experiment that provides detailed measurements of the release of U(VI) under controlled flow conditions. (Figure courtesy N. Qafoku and J. Zachara, PNNL).

The mathematical models are well understood for each fundamental scale; however, the link between the models and their parameters at different scales is missing. Advances in applied analysis coupled to a significant increase in computational power will enable solution over many scales. For example, recent development of regularity properties for partial differential equations (PDEs) with very rough coefficients (Cafarelli et al. 1996) and understanding of the connection between regularity of divergent and non-divergent PDEs allows one to deal with a large number of non-separated scales (Owhadi and Zhang 2007). Other challenges involve the solution of hybrid systems of differential equations and computational approaches to stochastic equations (SCaLeS 2004, pg. 49). Recent developments of the discrete network approximation (Borcea and Papanicolaou 1998; Berlyand and Kolpakov 2001) and the heterogeneous multi-scale methods (HMM) method (E and Engquist 2002) may provide a basis for addressing these challenges.

Efficient progress on reaction scale-up cannot exist as an isolated effort. However a focus on scaling in a research effort combined with other fundamental scale PRDs (identified below) is necessary to provide significant progress.

Why reactions?

Up-scaling analysis for reactions is relatively unexplored. Historically the scaling problem for reactions has been viewed as secondary to the more immediate problem of up-scaling flow. While there are still many unanswered questions in flow up-scaling, we have clearly reached the point where accurate characterization of reaction up-scaling is critical to both short and long time prediction capabilities required for expanding missions of DOE (e.g., CO_2 sequestration, aquifer management) and for improvement in continuing missions (e.g., waste clean-up, surface water contamination).

The reaction challenge "in a nutshell"

Up-scaling of reactions must deal with problem of up-scaling across an extreme range of scales, both spatial and temporal, that may not be well separated. This is complicated by the potentially large number of reactive species that may be involved, and hence a large increase in the number of equations that must be modeled.

The research direction

The proposed research direction encompasses three categories: A) computational, B) mathematical, and C) coordination with other PRDs.

A.1. Different numerical models, tuned for computations at specific fundamental scales, currently exist. Thus, effective scaling over a range of fundamental scales can be approached through coupling such models. An example would be lattice-Boltzmann computations feeding accurate, computed, pore-scale parameter information into a network flow model that, in turn, feeds computed core-scale parameter information into a

continuum-scale model. This direction requires an understanding of accurate "information passing" (Fish 2006) from one model to the next (i.e., from one scale to the next).

A.2. Development of efficient means for solving very large systems of equations on high-performance computers. New adaptive hierarchical and multilevel methods are critical, as these play two important roles: 1) as efficient scalable solvers at a particular scale and 2) as providing discrete up-scaling methodologies and insight into phenomenological continuum models.

A.3. Development of discrete network and hybrid (discrete/continuum) models for continuum problems with many scales.

A.4. Development of "adaptive up-scaling" techniques (local up-scaling, model hybridization) for performing reaction up-scaling variably over the domain of computation (resolve fine scales locally).

B.1. Development of mathematical up-scaling/down-scaling techniques in the presence of large number of spatial and temporal scales which may not be well separated.

B.2. Development of mathematical techniques for analysis of uncertainty propagation and quantification in new multi-scale methods.

C.1. Coordination of effort with, and adoption of methods proposed by, the PRD entitled "Computational Model of Critical Processes at Fundamental Scales." This PRD will focus on the nanometer, colloidal, and pore scales. The first two of these length scales typically need not be considered in scale-up of non-reactive flows.

C.2. Coordination of effort with, and use of the developments to validate scaling between neighboring length scales proposed by, the PRD: "Highly-resolved numerical testbeds for evaluation of upscaled models". Such testbeds are critical for validating the accuracy of scaleup between "neighboring scales".

C.3. Coordination of effort with, and use of methods proposed by, the PRD entitled "Efficient Bridging of Time Scales." This PRD proposes to address fast solvers, techniques for handling rare events, and parallel-in-time algorithms.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

We briefly summarize the current state of the art from both mathematical and computational perspectives. In general, scale-up studies have concentrated on flow, addressing scale-up of absolute permeability (Neuman 1994; Christie 1996; Durlofsky et al. 1997), relative permeability (Christie 1996; Durlofsky 1998; Wallstrom et al. 1999), and transport of conservative tracers (dispersivity) (Furtado et al. 1991; Glimm et al. 1992; 1993; Glimm and Lindquist 1992; Neuman 1990; Neuman and Zhang 1990; Zhang 1995). Only relatively recently has scale-up of reactive transport been considered (Knauss et al. 2005; Li et al. 2006; Kang et al. 2006). Results from such studies indicate the importance of the rate of transport of reacting species. If reaction

rates are sufficiently slow (i.e., no concentration gradients exist in individual pore) than up-scaling formulations based on volume averaging are adequate (Kang et al. 2006). However, if reaction rates are higher, reaction rate up-scaling becomes velocity dependent (Li et al. 2006). This transport rate effect has recently been nicely demonstrated (Knauss et al. 2005) in the interplay between diffusion and transport processes in the modeling of reactive transport in columns.

Mathematically, over 30 years of research (Bensouassan et al. 1978; Jikov et al. 1994; Cioranescu and Donato 1999) has provided a clear understanding of how to perform scale-up for processes in periodic geometry that involve two well-separated length scales. The results are applicable for uncoupled phenomena. Removal of restriction, periodicity, or well-separated length scales invalidates the conclusions of the mathematical approach. Non-periodic geometries have been addressed in recent work (Caffarelli et al. 2005; Torquato 2002; Berlyand and Khruslov 2004). There have been select works that consider scale up of coupled processes (Gibiansky and Torquato 1998). The theoretical work (Glimm and Lindquist 1992; Neuman 1990; Neuman and Zhang 1990; Zhang 1995) tackling scale-up of the transport of a conservative tracer through a permeability field involving a continuum of interacting length scales illustrates the difficulty of dealing with non-well separated scales. This body of work, which uses a perturbation expansion approach, involves an infinite collection of averages of correlation coefficients. Working with the theory involves closure assumptions to truncate the expansion at a usable, low-order. Verification of the closure assumptions is usually experimentally intractable.

Most geologic systems are not periodic (disordered or random) and typically have many, non-well-separated scales (i.e., long spatial correlations). The question of scaling the parameters of the model vs. scaling the form of the model itself must be addressed. The question of parameter vs. form scaling is intimately connected with the question of what minimal information (parameter, phenomenological, solution) at one scale must be carried to the next scale (i.e., appropriate definition of averages); however, this latter question has never been systematically addressed.

Computationally, scale-up studies have largely been performed in idealized model geometries. Many of the studies are “one-off,” involving a single geometry (or one that can be parametrically varied). The length scale range of these studies are typically restricted (e.g., pore to core, core to mesh block, mesh block to field) by computational resources (i.e., effective mesh size). The computational limitations to mesh block size have been the motivation for targeted development of numerical techniques to address scale-up (Aranson and Tsimring 2006; Arbogast 2002; Chen et al. 2003; Durlofsky 2002; Efendiev et al. 2004; Engquist 2005; Fish 2006; Hou and Wu 1997; MacLachlan and Moulton 2006; Tadmor et al. 1996; Wallstrom et al. 2002).

The computational challenge driving up-scaling work is that the DNS approach requires unreasonable computation resources.[1] Up-scaling produces additional model complexity; the fundamental scale DNS model is usually simpler (e.g., the grid for the up-scaled model may require feature alignment and tensor parameterization rather than scalar parameterization). Thus, the up-scaled model typically trades complexity in terms of constitutive equations for fewer degrees of freedom. Thus, questions concerning optimal hardware architecture and conceptual and algorithmic complexity for high-performance computing must be addressed. In addition, robust optimization methods and efficient solvers are needed. Order of accuracy and important physical properties, such as positivity of the solution, are issues that must be addressed in the up-scaled models.

Up-scaled models must be calibrated to ensure that major features of the fine scales have been retained. The accuracy (i.e., error estimates) of the up-scaled models needs to be developed for the technique to have practical use (Wallstrom et al. 1999). To establish error estimates, it will be necessary to run ensembles of statistically equivalent realizations of the model. This additional burden scales linearly with the number of realizations. As errors ideally (large n limit) decrease as the square-root of the size of the ensemble, the computational burden associated with ensemble computation is effectively $4n$ (i.e., halving the error requires $n \rightarrow 4n$).

In this general discussion of challenges, it is important to note that down-scaling (the inverse problem to up-scaling in which information from coarse scales is used to recover fine scale data), which is used for site characterization, is not decoupled from up-scaling, as solution of the down-scaling problem usually involves iterative solution of a very coarse, up-scaled (forward) model. The forward model is usually expensive to compute, and the domain required for characterization is typically physically large. Thus, synergy between down- and up-scaling is critical for reducing uncertainty in site parameter value assignment and in guiding the decisions for additional data acquisition. Down-scaling has the additional, well-recognized challenge of being an ill-posed problem (Cherkaev 2001).[2]

The challenges listed above are common to all up-scaling. However, scaling of reactive transport brings the following new challenges:

- Both time and length up-scaling must be considered.
- The range of length scales must be extended. While flow and non-reactive transport typically involves scaling from pore (10^{-6} m) to field (10^{5} m) scales, reactive transport involves reactions (bulk and surface) that extend length scales down to (10^{-8} m).

---

1 Using a lattice Boltzmann computation that fully resolves at the pore scale for a CO_2 sequestration simulation in a storage basin of volume of $5 \cdot 10^{8}$ m^3 having 20percent porosity, assuming two fluid phases and only a few reacting species, would require 10^{15} processors, assuming each processor had 1 GByte of independent accessible memory. Total CPU time needed to address policy related questions that would cover 100 years (3×10^{9} sec) of storage time is difficult to estimate—estimates vary from an optimistic one-year to more pessimistic CPU times.

2 Here "ill-posed" refers to lack of a unique solution.

- The time scales involved are even longer (10^{-12} s to 10^{12} s for CO_2 sequestration) than the range of length scales.

- Up-scaling over time is inherently a one-dimensional problem and is not amenable to the same distribution-over-processors techniques as are used in spatial decomposition techniques for handling spatial scales. Promising work in this area includes van Lent and Vandewalle (2002) and Griebel et al. (2002).

- The inclusion of reactions can significantly increase the number of equations (corresponding to the number of chemical species that must be reacted/transported) that must be solved. Thus reaction modeling (even without up-scaling) places heavy need on efficient ways to solve very large systems of equations on high-performance computers.

While the challenges for reaction up-scaling are prodigious, as stated in the previous section, we have clearly reached the point where advances in reaction up-scaling are required for the expanding missions of the DOE. To achieve maximum gain from an effort to develop reactive up-scaling techniques, the effort should be coordinated with three of the other Fundamental Scale PRDs: 1) "Computational Models of Critical Processes at Fundamental Scales," 2) "Highly Resolved Numerical Testbeds for Evaluation of Upscaled Models," and "Efficient Bridging of Time Scales".

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

As related in the previous section, past efforts to understand scale-up of flow and transport have resulted in the development of theoretical tools and computational methods to incorporate small scale effects on coarsened grid computations. To date, the effort in scale-up for reactions has been small in comparison. It can clearly be expected that new theories and computational tools for treating non-periodic, multiple, non-separated time and length scales will emerge from a concentrated effort in up-scaling of reactions.

Efforts for up-scaling of both reactive and non-reactive transport will benefit from high-performance computing (reactive transport more so because of the extra range of length scales and the coupling to the extreme range of time scales). Past work (Wallstrom et al. 1999; Furtado et al. 1991; Glimm et al. 1992) has pointed to the clear need for ensemble computation to set realistic error estimates. The need for local refinement of length (Durlofsky et al. 1997) and, undoubtedly, time scales will also benefit from utilization of highly parallel computing platforms.[1]

[1] The question of the number of realizations required for an ensemble computation is not easily answered; the answer depending critically on what variable is to be predicted, the size of the sample space (number of variables to be sampled), and the quality of the data (hard vs. soft data). For two-dimensional computations of oil production rate, 50 to 100 realizations were used to investigate scale up from 100 x 100 to 10x10 grids (Glimm et al. 2001); while for three-dimensional network flow (pore to core scale) computations for residual saturations in two-phase flow, 10^4 realizations were employed (Sok et al. 2002)

Understanding and quantifying the limits in atomistic-to-continuum coupling is a critical component in computational materials science and other applications of interest to SC. Multi-scale methods and techniques of up-scaling in subsurface problems will be used in error, stability, convergence analysis, and uncertainty quantification of coupling atomistic and continuum models.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

DOE's interest in subsurface sciences can be summarized as directed to the improved ability to meet the energy demands of the nation while providing the necessary protection of our environment and water resources. Many of these expanding and continuing DOE missions involve subsurface flow and transport. Improved understanding of up-scaling of reactive transport is essential for providing the ability for quantitative prediction at field scales. As a specific example, consider the effects on CO_2 sequestration modeling. Mineralization reactions are critical to understanding the long-term viability of CO_2 sequestration. These reactions influence 1) pH buffering of formation acidity, 2) changes in porosity and permeability, and 3) sequestration capacity. The extent to which these reactions are expected to occur in the time scales relevant for carbon sequestration is poorly understood because of uncertainties in reaction kinetics in porous media. First, the slow rates of these reactions preclude short-term laboratory study. To study such reactions in the laboratory, geochemists usually resort to idealized mineral systems that eliminate the porous media effects that limit reaction rates. Second, laboratory studies can accurately measure reaction rates only at very small spatial scales. To describe reaction kinetics in potential storage basins, hydrogeologists require reaction rates at scales that are consistent with their models.

Improvements in understanding of reaction scale-up will have beneficial coupling effects to experimental programs. Such improvement will allow better use of experimental data. Conversely, parameters from properly up-scaled models will have greater physical relevance and, thus, will be "easier" to measure experimentally.

TIME FRAME

With a targeted call for proposals, one could expect significant progress to be achieved in two funding periods (6 years).

REFERENCES

Arbogast, T. 2002. "Implementation of a locally conservative numerical subgrid up-scaling scheme for two-phase Darcy flow", *Computational Geosciences* **6**, 453.

Aranson, I. and L. Tsimring. 2006. "Patterns and collective behavior in granular media: Theoretical concepts", *Reviews of Modern Physics* **78**, 641.

Berlyand, L. and A. Kolpakov. 2001. "Network approximation in the limit of small interparticle distance of the effective properties of a high--contrast random dispersed composite", *Archives for Rational Mechanics and Analysis* **36**,179.

Berlyand, L. V., and E. Y. Khruslov. 2004. "Homogenized non-Newtonian viscoelastic rheology of a suspension of interacting particles in a viscous Newtonian fluid", *SIAM Journal of Applied Mathematics*, **64**, 1002.

Borcea, L. and G. Papanicolaou. 1998. "Network approximation for transport properties of high contrast conductivity", *Inverse Problems* **4,** 501.

Bensoussan, A., J.-L. Lions, and G. Papanicolaou. 1978. Asymptotic Analysis for Periodic Structures, North-Holland, Amsterdam.

Caffarelli, L. A., M. G. Crandall, M. Kocan and A. Swipolhkech. 1996. "On viscosity solutions of fully nonlinear equations with measurable ingredients", *Communications on Pure and Applied Mathematics* **49**, 365.

Caffarelli, L. A., P. E. Souganidis and L. Wang,. 2005. "Homogenization of fully nonlinear, uniformly elliptic and parabolic partial differential equations in stationary ergodic media", *Communications on Pure and Applied Mathematics* **58**, 319.

Chen, Y., L. J. Durlofsky, M. Gerritsen, and X. H. Wen. 2003. "A coupled local-global upscaling approach for simulating flow in highly heterogeneous formations", *Advances in Water Resources* **26**, 1041.

Cherkaev, E. 2001. "Inverse homogenization for evaluation of effective properties of a mixture", *Inverse Problems* **17**, 1203.

Christie, M. A. 1996. "Upscaling for reservoir simulation", *Journal of Petroleum Technology* **48**, 1004.

Cioranescu, D. and P. Donato. 1999. Introduction to Homogenization, Oxford University Press, Oxford.

Durlofsky, L. J., R. C. Jones and W. J Milliken. 1997. "A nonuniform coarsening approach for the scale-up of displacement processes in heterogeneous porous media", *Advances in Water Resources* **20**, 335.

Durlofsky, L. J. 1998. "Coarse scale models of two phase flow in heterogeneous reservoirs: volume averaged equations and their relationship to existing upscaling techniques", *Computational Geosciences* **2**, 73.

Durlofsky, L. J. ed. 2002. Upscaling of Geological Models for Reservoir Simulation: Issues and Approaches, *Computational Geosciences* **6.**

E, W, and B. Engquist. 2002. "The heterogeneous multi-scale methods", *Communications in Mathematical Sciences* **1**, 87.

Efendiev, Y., T. Y. Hou and V. Ginting. 2004. "Multiscale finite element methods for nonlinear problems and their applications", *Communications in Mathematical Sciences* **2**, 553.

Engquist, B. 2005. "Multi-scale modeling" in Perspectives in Analysis, *Mathematical Physics Studies* **27**, 51. Springer, Berlin. M. Benedicks, P. W. Jones and S. Smirnov (Eds.).

Fish, J. 2006. "Bridging the scales in nano engineering and science", *Journal of Nanoparticle Research* **8**, 577.

Furtado, F., J. Glimm, W.B. Lindquist, F. Pereira and Q. Zhang. 1991. "Time dependent anomalous diffusion for flow in multi-fractal porous media", *Lecture Notes in Physics* **398**, 79.

Gibiansky, L., and S. Torquato. 1998. "Rigorous connection between physical properties of porous rocks", *Journal of Geophysical Research – Solid Earth* **103**, 23911.

Glimm, J., S. L. Hou, H. J. Kim, Y. Lee, D. H. Sharp, K. Ye and Q. S. Zou. 2001. "Risk management for petroleum reservoir production: A simulation-based study of prediction", *Computational Geosciences* **5**, 183.

Glimm, J., W. B. Lindquist, F. Pereira and Q. Zhang. 1993. "A theory of macrodispersion for the scale up problem", *Transport in Porous Media* **13**, 97.

Glimm, J., and W.B. Lindquist. 1992. "Scaling laws for macrodispersion", *Proceedings of the Ninth International Conference on Computational Methods in Water Resources* **2**, 35-49.

Glimm, J., W.B. Lindquist, F. Pereira and R. Peierls. 1992. "The multi--fractal hypothesis and anomalous diffusion", *Revista Brasileira de Matematica Aplicada e Computacional* **11**, 31.

Griebel, M., D. Oelitz, and P. Vassilevski. 2006. "Space-time approximations with sparse grids", *SIAM Journal on Scientific Computing* **28**, 701.

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

Jikov, V. V., S. M. Kozlov, and O. A. Oleinik. 1994. Homogenization of differential operators and integral functionals, Springer-Verlag, Berlin.

Kang, Q. J., P. C. Lichtner and D. X. Zhang. 2006. "Lattice Boltzmann pore-scale model for multicomponent reactive transport in porous media", *Journal of Geophysical Research-Solid Earth* **111**, B05203.

Knauss, K.G., J. W. Johnson and C. I. Steefel. 2005. "Evaluation of the impact of CO2, co-contaminant gas, aqueous fluid and reservoir rock interactions on the geologic sequestration of CO2", *Chemical Geology* **217**, 339.

Li L., C. A. Peters and M. A. Celia. 2006. "Upscaling geochemical reaction rates using pore-scale network modeling", *Advances in Water Resources* **29**, 1351.

MacLachlan, S. P. and J. D. Moulton. 2006. "Multilevel upscaling through variational coarsening", *Water Resources Research* **42**, W02418.

Neuman, S. P. 1994. "Generalized scaling of permeabilities – Validation and effect of support scale", *Geophysical Research Letters* **21**, 349.

Neuman, S. P. 1990. "Universal scaling of hydraulic conductivities and dispersivities in geologic media", *Water Resources Research* **26**, 1749.

Neuman, S. P. and Y. K. Zhang. 1990. "A quasi-linear theory of non-Fickian and Fickian subsurface dispersion. 1. Theoretical-analysis with application to isotropic media", *Water Resources Research* **26**, 887.

Owhadi, H. and L. Zhang. 2007. "Metric-based upscaling", *Communications on Pure and Applied Mathematics*. DOI: 10.1002/cpa.20163.

SCaLeS. 2004. A Science-Based Case for Large Scale Simulation, Vol. 2. Report of the DOE Office of Science Workshop, Arlington VA, 6/24-6/25, 2003. *http://www.pnl.gov/scales*

Sok, R. M., M. A. Knackstedt, A. P. Shpeppard, W. V. Pinczewski, W. B. Lindquist, A. Venkatarangan, and L. Paterson. 2002. "Direct and stochastic eneration of network models from tomographic images: Effect of topology on residual saturations", *Transport in Porous Media* **46**, 345.

Tadmor, E., R. Phillips and M Ortiz. 1996. "Mixed atomistic and continuum models of deformation in solids", *Langmuir* **12**, 4529.

Torquato, S. 2002. Random Heterogeneous Materials, Springer-Verlag, Berlin.

van Lent, L. and S. Vandewalle. 2002. "Multigrid waveform relaxation for anisotropic partial differential equations", *Numerical Algorithms* **31**, 361.

Wallstrom, T. C, S. L. Hou, M. A. Christie, L. J. Durlofsky and D. H. Sharp. 1999. "Accurate scale up of two phase flow using renormalization and nonuniform coarsening", *Computational Geosciences* **3**, 69.

Wallstrom, T. C., S. Hou , M. A. Christie, L. J. Durlofsky, D. H. Sharp and Q. Zou. 2002. "Application of effective flux boundary conditions to two-phase upscaling in porous media", *Transport in Porous Media* **46**, 155.

Zhang, Q. 1995. "Transient-behavior of mixing induced by a random velocity field", *Water Resources Research* **31**, 577.

EFFICIENT BRIDGING OF TIME SCALES

Bjorn Engquist, UT-Austin; Max Gunzburger, Florida State; John Weare, UCSD; Eric Bylaska, PNNL

ABSTRACT

The wide scope of temporal scales in subsurface science, ranging from tenths of picosecond motions of atoms in formation fluids and minerals to geological scales in millions of years, poses an even larger computational challenge then does the large range of spatial scales. The focus of this PRD is the development of innovative computational techniques for incorporating phenomena on fundamental time and spatial scales into models that are relevant on field scales.

EXECUTIVE SUMMARY

A major obstacle for accurate predictability from simulations in subsurface science is the wide range of temporal scales. The range of scales spans many orders of magnitude from atomistic motion and reactions to geological time scales. Advancement of the state of the art of computational methods in subsurface science, as described by the examples given in this PRD, will require interaction between applied mathematicians, computational scientists and applied scientists.

Improvement of fast solvers on fundamental scales is necessary. The fundamental algorithms of atomistic modeling need to be improved to provide accurate solutions of the electronic Schrödinger equation on the basis of new algorithms that will take full advantage of modern computer architecture. New techniques are required for the prediction of the rare events that represent abrupt qualitative changes in the system (e.g., chemical reactions). At atomistic scales, the goal is to extend the time that can be accurately simulated. Rare event methodologies would also apply to other relevant time scales. Coupled multi-scale methods are also important. Recent developments have introduced new paradigms for on-the-fly coupling of, for example, atomistic and continuum models in the same simulation. Subsurface simulation is an ideal environment to develop these paradigms into concrete algorithms. One example is mapping of first-principle-based (i.e., calculated from the electronic Schrödinger equation) interactions and dynamics to adaptive effective potentials to be used in a larger time scale Molecular Dynamics simulation. Another example is the use of pore-scale network models to dynamically generate parameters for a Darcy type simulation. The impact of massively parallel computing on simulation fidelity has been fundamental. Parallelization by spatial domain decomposition is standard. Parallelization in the time dimension is much harder. A few new technologies have been introduced. It will be advantageous to adapt these parallel-in-time algorithms to the modeling subsurface processes and apply them, for example, to pore-scale flow simulations where friction and dissipation are present. Success on these projects would make it possible to

fully take advantage of future generation computer architectures for the reliable simulation of essential subsurface processes.

SUMMARY OF RESEARCH DIRECTION

Fast solvers on fundamental scales

In heterogeneous subsurface sediments, the reactive transport of metals ions (e.g., Na^{+}, Sr^{2+}, Cs^{2+}), anions (e.g., Cl^{-}, NO_3^{-}, CO_3^{2-}), metal oxyanions (e.g., Cr(VI), U(VI), Tc(VII)), organic solvents (e.g., CCl_4, TCE, TCP), and noncondensible gases (e.g., CO_2, O_2, N_2) are controlled by processes occurring across a range of spatial and time scales. Key microscopic (pore) scale processes include sorption, dissolution-precipitation, heterogeneous and/or homogenous reduction of oxides and aqua ions by Fe(II) and Cr(III), oxidation by Mn(III/IV) oxides, and mass transfer processes. Beyond the fundamental scale, key physical and chemical heterogeneity processes include hydraulic conductivity and pore sizes, the type and distribution of reactive minerals, and the location and mineralogical residence of the species.

Many of the microscopic (pore) scale processes that occur in the subsurface occur on time and length scales related to the atomic structure of the system (e.g., time scales of 0.1 to 100 ps and space scales of 100s of Å), but are observed in the subsurface on much larger time and length scales (e.g., minutes to 10s of years and inches to miles). Whereas there is significant averaging involved in reducing the dimensionality and coarsening time resolution in these problems, the specificity of the atomic level chemistry is essential to predicting the behavior of the system at the larger scales. The computational challenge of the simulation of such systems is to predict their behavior for very large spatial and temporal scales while retaining information from the atomic level.

Additional motivation for the development of methods that can up-scale atomic behavior to the macroscopic level is provided by the recent extensive investment of the DOE in large-scale facilities such as the synchrotron light sources. These new probes such as, x-ray and neutron scattering, time resolved spectroscopy, STM, etc., (see for example Figure 1) yield atomic-level information of unprecedented detail about important subsurface behaviors such as the distribution and reactions of species at mineral surfaces and the structure and stability of toxic and radioactive solutes in solution. However, without appropriate theoretical and computational advances, this new understanding of atomic-level chemistry cannot easily be translated to observations at larger scales. To fully realize the potential of these advanced experiments in the analysis of practical level environmental and energy problems, new theoretical and computational tools are required that can span

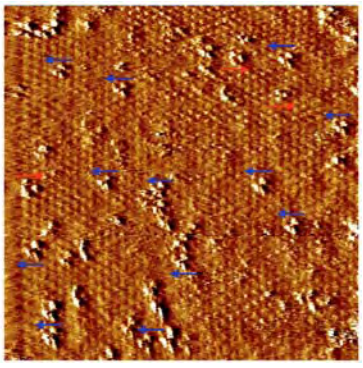


Figure 1: Atomically resolved STM image (20 × 20 nm) of hematite (001) collected in air showing “triplet” Fe clusters on the surface.

the atomic space and time scales and extrapolate these atomic scale measurements to the much larger time and length scales for environmental applications.

At the atomic/molecular scale, the forces between atoms reflect complex many-body effects (e.g., bond breaking and formation and electron transfer leading to redox phenomena). To reliably capture these effects in a predictive theory, it is necessary calculate interactions reliably. If bonds are not being made or broken then classical force fields can provide insight into many types of phenomena given the constraint that they are properly parameterized. To treat bond making and breaking and electron transfer, one must use a first-principles-theory based method that solves the electronic Schrödinger equation as accurately as possible given the available computational resources. The dynamics on the time scale of molecular motion (a few picoseconds) can be solved by using classical molecular dynamics or quantum dynamics. We note that we have separated the problem of nuclear motion from the electronic motion. The dynamics of the system can be simulated either by using quantum mechanical potential evaluations to develop analytical interaction potentials (including potentials adaptively refined on the fly [Csanyi, 2004, Rasamny, 1998]) or by the use of first principles force calculations on the fly in a dynamical algorithm (*ab initio* Molecular Dynamics, AIMD [Car 1985]). The systems of interest are often defected on an atomic scale as is illustrated in Figure 1. Large numbers of atoms are, therefore, required to accurately represent the structural features (isolated defects, grain boundaries, etc.). For problems in which the atomic level chemistry is essential to the interpretation of a macroscopic process, the reliable prediction of the chemistry at the molecular scale will dominate the simulation even if an efficient method to upscale the calculation to the macroscopic level is available.

Currently, there are two approaches to the reliable prediction of electronic structure:
1) molecular orbital (MO) theory and density functional theory (DFT). MO theory has the nice property that one can attain an exact solution by improving the approximations to the solution but at great computational cost. In addition, high-level MO methods require substantial memory and disk storage, which is not available on current and planned high end computer architectures. DFT on the other hand is more efficient and gives high quality results for reasonable computational cost but currently one does not know *a priori* how to improve the exchange-correlation functional to get the exact solution. DFT is the most used approximation to the electronic Schrödinger equation for large systems. While DFT provides a computationally practical approximation for many applications, there are two major bottlenecks to its application to a much wider class of problems at the fundamental scale (Valiev 2002; Marx 2000):

- The DFT level of approximation is not sufficiently accurate to treat many materials problems (e.g., energies in transition and actinide elements and long range van der Waals forces).
- The scaling of existing solution methods is not sufficient to exploit the performance of the next generation parallel computers (10s to 100s of thousands of processors) although preliminary results exhibit good scaling.

There is a real need to improve the exchange-correlation functional in DFT approaches to treat for example metals and metal oxides with highly localized *d* and *f* electrons in their valence shells for predicting the transport of toxic and ore forming metal ions in ground waters. Examples of improved DFT approaches which include electron exchange in a self consistent way (e.g., exact exchange, self interaction correction [SIC], LDA+U) provide at least semi-quantitative accuracy and relatively high efficiency (Kudin 2002; Kulin, 2006). An example of the effect of exchange on electron localization in a mineral system is given in Figure 2. In this problem, DFT calculations (left side of Figure 2) yield a completely delocalized spin density in contrast to the predictions of an exchange localized spin density when self-interactions are taken into account (right side of Figure 2). Scalable algorithms that include exact exchange need to be developed for first principle dynamics.

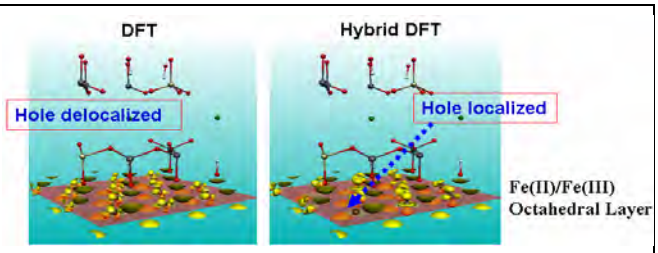


Figure 2: Spin density of the octahedral layer of annite with a hole using DFT and DFT +exact exchange (Eric Bylaska, PNNL).

A more difficult problem with the DFT method is that long range non-bonded (van der Waals) interactions are not included formally in the theory. These interactions dominate many problems of interest to subsurface modeling. The inclusion of exchange is not sufficient to improve the accuracy. This problem can be treated by using a high accuracy MO-based methods (e.g., MP2, or CCSD(T) [Bartlett 2005] or Quantum Monte Carlo methods [Grossman 2002, 2005]), but these methods are currently too costly. However, new techniques may broaden their applications to much more complex systems (Kowalski 2005, 2006). A critical issue is to develop new functionals that can be used to predict long-range interactions in a computational efficient manner.

Significant improvements in algorithm performance, scalability, and implementation are required to treat the important applications using any of these approaches. For example, currently an *ab initio* Molecular Dynamics simulation of 128 water molecules for 0.1 ps requires about 1000 CPU hours and a 30 ps simulation takes about 34 CPU years on a single CPU. The projected size of the next generation supercomputers (10,000 to 100,000 processors) suggests that simulation times and particle size limitations may be overcome by brute force increases in computer size. However, available implementations of present methods do not scale much beyond 1000 processors although recent reports of 10 Tflop scaling using cpmd have been reported. The question of proper approximation both in terms of discretization as well as the (separate) questions of proper choice of iterative methods needs to be addressed. More efficient and/or better scaling methods could be based on the use of completely unstructured simplex finite element (Bank 2003) or wavelet techniques (Harrison 2004), built adaptively using a multilevel solve-estimate-refine iteration. A substantial problem is the scaling of the parallelization of the fast Fourier transform (FFT) method critical to solving the electronic motion problem with plane wave DFT. Currently, this does not scale well to very large number of processors hampering advances in this area.

New techniques for rare events

The time scales for many subsurface processes of interest are on the several hundred picosecond level or much longer so even with major improvements in performance from the development of scaleable AIMD algorithms on very large computers may not address these problems. The time scale issue is extremely important as we do not currently have the techniques available to appropriately sample all of the important regions of phase space necessary to sample long time trajectories or rare events. This is relevant to being able to predict reactivity and to be able to predict the properties of complex couple systems. One effective approach to extend time scales while retaining an atomistic level of description is to introduce rare event methods. In these approaches, it is presumed that the important changes in the system are dominated by the rare occurrence of single concerted events (e.g., chemical reactions). For chemical systems these events usually arise from the abrupt transition form one local minimum on the free energy surface to another separated by a barrier. An example of such a reaction is given in Figure 3, which shows the exchange of a water molecule in the first hydration shell of a Ca^{2+} ion in solution with one in the second shell (note, the reactant, transition state, and product state are illustrated on the left; particle trajectories are illustrated on the right). For this system, this is a fairly fast reaction on the order of 40ps. It is useful to note that no electron pairs have been broken so this is a simpler reaction than many other important ones. For other ions, such as Al^{3+}, the same reaction may take on the order of minutes. Tools are needed to estimate the likelihood of these rare events and their importance in the long time dynamics of the system. An example of a coupling approach is the kinetic Monte Carlo method.

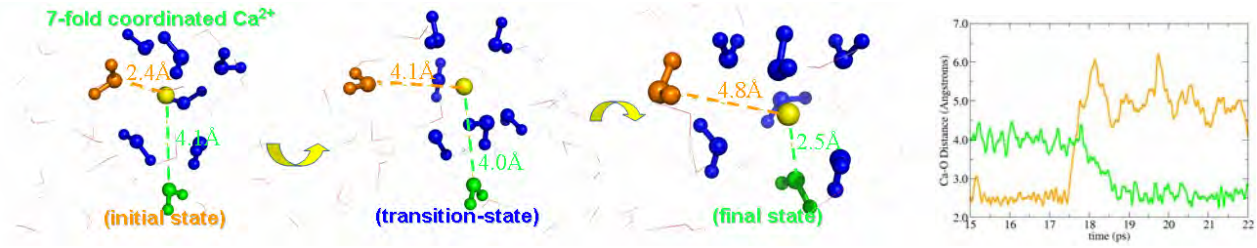


Figure 3. Water Exchange in the Hydration Shell of Ca^{2+}+64H_2O First Principle Simulations (John Weare, University of California, San Diego)

There are a number of ways to search for rare events; however, most of these have been designed for problems with only a few degrees of freedom. To find the effective transition mechanism, one needs to average out the effects of the medium as well as thermal noise. One class of methods accelerates or makes more efficient the exploration of phase space either by using higher temperature dynamics or by modifying the energy landscape so that rare event become more probable (Iannuzzi 2003; Sorenson, 2000). This class of methods has the advantage that the local minimum in the system do not have to be determined before the search starts. In another class of methods, it is presumed that the local minima are known (very often this is the case in chemical reactions with known mechanisms), and the focus of the method is on determining the barrier to transition. There are several methods for doing this; for example, the nudged elastic band (NEB) method (Henkelman, 2000), which works well for systems with a

few degrees of freedom. However, in systems with many degrees a method of averaging out unimportant degrees of freedom is necessary. New methods need to be developed that search for reaction pathways, including minima in noisy, many-dimensional environments. An example of such a method combines minimum free energy path finding and sampling to average out the environmental noise (Weinan 2005).

Coupled multi-scale methods

In this class of methods, models for different temporal and spatial scales as well as different physical processes are coupled in the same simulation. A traditional version is heterogeneous domain decomposition (Quateroni 1999) in which different models are applied in different spatial domains (type A problems in [Engquist 2003]). There are some natural applications of this approach to subsurface science. One example is chemically active surfaces. Adequate modeling of a surface, interface or contact line in porous media flow is typically different from models in the bulk. Another example on a larger scale is the behavior of a geological repository for radioactive waste. Different physical phenomena are prominent in the casing, near the casing and in the far field. In this context, new and robust numerical methods for the interface conditions or the so-called handshake domains between the different models need to be developed. Even more important is the coupling between models for different scales or different physics within the same spatial domain (type B in [Engquist 2003]). For this class of methods, it is possible to incorporate effects of temporal processes of very fast scales at a computational cost, which is related to the one for the modeling of a much slower process. Examples include the quasi-continuum method (Miller 2002, the heterogeneous multi-scale method (Engquist 2003), and the projective integration technique (Kevrekidis 2003). The heterogeneous multi-scale method is a framework for derivation and analysis of multi-scale methods that should apply very well to subsurface sciences.

One example is the coupling of networks on the pore scale with the larger scale a Darcy type flow model. Effective parameters that are needed in the Darcy law simulation can locally be calculated on the fly by network simulations that are local in time and space. New coupling operators for these processes need to be developed. In this way, the higher fidelity of network models should be available at a computational cost that is closer to the cost for the macro-scale simulation. Because the micro-scale modeling is performed locally in time and space, this procedure is very well suited for massively parallel architectures.

Parallel-in time algorithms

There has been much progress in the field of parallel large-scale simulations with respect to spatial domains based on domain decomposition. For temporal domains, significantly less progress has been made. However, there have been some attempts at developments of algorithms that effect parallelization in time. Currently, the parareal algorithm (Lions 2001) is the most commonly used methodology for this purpose. It has been applied to problems posed on very disparate time scales (e.g., the Navier-Stokes equations [Fischer 2003], reservoir

simulation [Garrido 2007], financial modeling [Bal 2002], and molecular dynamics [Baffico 2002]). The algorithm has the flavor of a multiple shooting method. A coarse time step is used to integrate a system in time. The coarse grid solution is used to define initial conditions for fine time-step integration within each coarse grid solution, and then a coarse grid correction is effected. This process is repeated to define an iterative method. All fine-time step calculations are done in parallel. The algorithm terminates in a finite number of steps roughly equal to the number of coarse time intervals, but for some problems (e.g., dissipative differential equations) accurate solutions can be obtained in much less iterations. If the ratio of the coarse and fine time steps is large enough, it is possible that the algorithm results in savings even if it proceeds to termination. There are many opportunities in subsurface science for the application of the parareal concept. Certainly, at the Darcy time scale, it should perform well as it has for the parabolic test problems for which it was designed (Lions 2001). At the pore level, where the Stokes or Navier-Stokes equations must be solved, it should also apply. However, the extension to the much more complex setting encountered in subsurface science will require intensive research as one must deal with highly heterogeneous environments, multiple phases and complex geometries.

The parareal algorithm lends itself very well to multi-scale coupling applications involving different models. There is no reason that in the algorithm, the model used for fine-scale temporal integration has to be the same as that used at the coarse scale. Thus, for example, one could use a model at the pore scale for small-time-step integrations, and then use a Darcy model for the coarse grid corrections. The parareal algorithm may also be useful at other scales below and above the pore and Darcy scales (e.g., coupling Darcy models to diurnal and seasonal models).

Parallelization in time is inherently difficult but could have a very high payoff for subsurface science problems in the environment of massively parallel computer architecture. For problems solved at a single time scale and in settings involving Darcy and Navier-Stokes equations, substantial savings are possible. This may also hold in other settings over very disparate time scales. For coupled multi-scale applications, the payoff may be even greater.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The wide range of temporal scales of all processes in subsurface science poses an even larger computational challenge then does the range of spatial scales. One reason is that the range is larger from picosecond motions of atoms in chemical reactions to geological scales in million of years. The other problem is the inherent difficulty of parallelization resulting from causality. The computing time typically scales linearly with processor speed rather than number of processors. The overarching goal for research in this area is the derivation of models and algorithms that describe macro-scale phenomena and include important atomistic or other micro-scale processes at the larger scale.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Research efforts in this area would replace current different *ad hoc* techniques by methods more based on first principles in subsurface modeling problems. This paradigm will be applicable in many other application areas with multi-scale challenges. These new classes of computational techniques will benefit from new massively parallel architectures. The computational parallel-in-time algorithms that are based on iterative and non-homogeneous time stepping will be useful in subsurface modeling and beyond.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

The development of highly parallel/highly scalable simulation methods is a critical need in essentially all applications of simulation to complex grand challenge problems in areas such as toxic waste containment, CO_2 sequestration, high-performance material synthesis, and energy efficient catalysis. Progress in bridging temporal scales will substantially improve predictability in subsurface simulations. The possibility of applying the higher fidelity models that include fundamental scales can be used for better understanding of basic processes, for benchmarking, or as components in improved hybrid models on field scales.

TIME FRAME

Some parts of the proposed research should yield applicable results within 2 to 3 years. This could certainly be the case for some aspects of the adaptation of the coupled multi-scale methods to subsurface science. Other research areas should generate important computational science progress in 3 to 5 years with the practical impact to follow soon thereafter.

REFERENCES

Car R. and Parrinello M. (1985) Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters* **55**(22), 2471-2474.

Valiev M., Bylaska E. J., Gramados A., and Weare J. H. (2002) First principles molecular dynamics simulations using density-functional theory. In *Reviews In Modern Quantum Chemistry : A Celebration of the Contributions of R. G. Parr,* Vol. II.

Marx D. and Hutter J. (2000) Ab-initio molecular dynamics: theory and implementation. In *Modern Methods and Algorithms of Quantum Chemistry* (ed. G. J), pp. 301-449. NIC, Forschungszentrum.

R.J. Bartlett, How and Why Coupled-Cluster Theory Became the Preeminent Method in Ab initio Quantum Chemistry (R.J. Bartlett). , in *Theory and Applications of Computational Chemistry: The First Forty Years*, edited by C.E. Dykstra, G. Frenking, K.S. Kim, and G.E. Scuseria (Elsevier, 2005), pp. 1191-1221.

Kudin K. N., Scuseria G. E., and Martin R. L. (2002) Hybrid density-functional theory and the insulating gap of UO2. *Physical Review Letters* **89**(26), 266402.

Kulik H. J., Cococcioni M., Scherlis D. A., and Marzari N. (2006) Density functional theory in transition-metal chemistry: A self-consistent Hubbard U approach. *Physical Review Letters* **97**(10), 103001.

Grossman J. C., Rohlfing M., Mitas L., Louie S. G., and Cohen M. L. (2001) High accuracy many-body calculational approaches for excitations in molecules. *Physical Review Letters* **86**(3), 472-475.

Grossman J. C. and Mitas L. (2005) Efficient quantum monte carlo energies for molecular dynamics simulations. *Physical Review Letters* **94**(5), 056403.

Kowalski K. (2005) Implementation of the locally renormalized CCSD(T) approaches for arbitrary reference function. *Journal of Chemical Physics* **123**(1), 014102-.

Kowalski K. and Valiev M. (2006) Asymptotic extrapolation scheme for large-scale calculations with hybrid coupled cluster and molecular dynamics simulations. *Journal of Physical Chemistry A* **110**(48), 13106-13111.

Bank R. and Holst M. (2003) A New Paradigm for Parallel Adaptive Mesh Refinement. *SIREV* **45**(2), 291-323.

Harrison R. J., Fann G. I., Yanai T., Gan Z., and Beylkin G. (2004) Multiresolution quantum chemistry: Basic theory and initial applications. *Journal of Chemical Physics* **121**(23), 11587-11598.

Iannuzzi M., Laio A., and Parrinello M. (2003) Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Physical Review Letters* **90**(23), 238302.

Sorensen M. R. and Voter A. F. (2000) Temperature-accelerated dynamics for simulation of infrequent events. *Journal of Chemical Physics* **112**(21), 9599-9606.

Henkelman G. and Jonsson H. (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, Vol. 113, pp. 9978-9985.

Weinan E., Ren W. Q., and Vanden-Eijnden E. (2005) Finite temperature string method for the study of rare events. *Journal of Physical Chemistry B* **109**(14), 6688-6693.

A Quateroni and A Valli, Domain decomposition methods for partial different, Oxford University Press, 1999

W. E. and B. Engquist, The heterogeneous multi-scale method, Comm. Mat, Science, 1, 87-133, 2003

R. E. Miller and E. B. Tadmor, The Quasicontinuum Method: Overview, Applications and Current Directions", Journal of Computer-Aided Materials Design, 9, 203-239 (2002).

I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg and C. Theodoropoulos. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks. Comm. Math. Sci.., 1, 715-762, 2003.

J.-L. Lions, Y. Maday, and G. Turinici, Resolution d'e par un schema en temps parareel, C.R. cad. Sci. Paris Ser. I Matt., 335, 2001, 335-392.

P. Fischer, Investigation of the parareal algorithm for semi-implicit incompressible Navier-Stokes simulations, Proc. 15th Domain Decomposition Conf., 2003.

I. Garrido, G. Fladmark, and M. Espedal, Parallelization in time for reservoir simulation, to appear.

G. Bal and Y. Maday, A parareal time discretization for nonlinear pde's with applications to the pricing of and American put, in Recent Advances in Domain Decomposition Methods, Springer, 2002, 189-202.

L. Baffico, S. Bernard, Y. Maday, G. Turinici, and G. Zerah, Parallel-in-time molecular-dynamics simulations, Phys. Rev. E, 66, 2002, 057701.1-057701.4.

Rasamny M., Valiev M., and Fernando G. W. (1998) Interatomic potentials via the effective-action formalism. *Physical Review B* **58**(15), 9700-9704.

Csanyi G., Albaret T., Payne M. C., and De Vita A. (2004) "Learn on the fly": A hybrid classical and quantum-mechanical molecular dynamics simulation. *Physical Review Letters* **93**(17), 175503.

HIGHLY-RESOLVED NUMERICAL TESTBEDS FOR EVALUATION OF UPSCALED MODELS

Tim Scheibe
Pacific Northwest National Laboratory

ABSTRACT

All subsurface science models involve multiple, poorly separated length and time scales and complex, coupled physical phenomena. Direct numerical simulation from fundamental scales to field scale is impossible. Analytic understanding of scaling is incomplete; up-scaling commonly requires simplifying assumptions that are often invalid, and down-scaling commonly leads to an ill-posed inverse problem. Research is needed to identify the minimum information that must be transferred across scales to maintain model validity. Work is needed on developing hybrid models based on local scale refinement and hierarchical models that maintain accuracy across scales by means of scientifically and mathematically defensible up-scaling techniques. There is currently little or no basis on which to 1) determine appropriate sequences of scales for models and 2) evaluate the impacts of information loss due to up-scaling. We recommend developing highly resolved, computationally demanding models of specific processes to serve as a test bed, or set of benchmarks, for evaluating up-scaling methods and up-scaled models. This test bed will require tera/petascale computational resources for solution of benchmark problems. The benchmark problems need to be validated using laboratory or field data. Advanced algorithms, mathematics, and computational methods will be developed for solving large problems and will provide a validation basis for application-level codes and models.

EXECUTIVE SUMMARY

We recommend that research be undertaken with the goal of development of a number of rich numerical datasets (numerical test beds) for use in benchmarking studies, in particular of up-scaling and down-scaling methodologies. These numerical test beds should represent the best current understanding of selected subsurface flow, transport, and reaction processes at fundamental scales (e.g., pore-scale or smaller), combined with state-of-the-art non-invasive characterization datasets. Leadership computational resources can be brought to bear effectively on these problems because 1) the simulation of processes at the relevant scales will be highly computationally intensive, and 2) the specific nature of each test bed will allow customized code with minimal overhead to be designed for maximum scalability. Advanced visualization, data mining, and data management tools should also be used to maximize the availability and usefulness of the numerical test beds once the simulation process is completed.

SUMMARY OF RESEARCH DIRECTION

This priority research direction is focused on the application of leadership computational resources to develop very highly-resolved numerical descriptions of selected subsurface problems (test beds) for use in benchmarking and evaluation of up-scaling algorithms and up-scaled models.

Although the application of state-of-the-art, high-performance computing to DOE problems on a site-by-site, problem-specific basis is a worthy goal and is being pursued for some key problems, it is currently not a widespread practice for a number of reasons. Accordingly, it remains the case that most applied models of subsurface processes currently use (and will continue to use in the foreseeable future) relatively coarse grids with effective model parameters that are ill-defined (non-uniquely identified), difficult, or impossible to relate to measurable quantities, and in some cases fundamentally flawed. Many such models are endowed with credibility by applied subsurface scientists not because strong evidence of their fundamental soundness exists, but rather on the basis that they follow accepted standards of practice. This perceived credibility is communicated to policy and decision makers, which provides them with an often inappropriate level of confidence in model predictions. When model predictions and the resulting designs subsequently fail, not only are the negative impacts of incorrect decisions suffered, but public confidence in future policy and decisions is adversely impacted. In contentious issues where opposing parties each have supporting scientific experts with differing opinions and models, public confidence in science-based approaches to decision-making suffers further.

Given that many near-term decisions will be made based on the best available information, which practically speaking will come in many cases from relatively coarsely-resolved simulations, it is critical that the mathematical process descriptions and parameters used in those simulations be consistent with fundamental understanding at smaller scales to the greatest degree possible. This will be made possible by 1) improved theories and algorithms for up-scaling fine-scale information and 2) a means of rigorously testing and validating those theories and algorithms. This PRD addresses the latter.

Numerical test bed concept

The underlying concept of the proposed numerical test bed approach is that a simulation explicitly developed for this purpose can use mathematical process descriptions and parameterizations that are representative of our best scientific understanding of fundamental physical, chemical, and biological processes. In contrast, technical and practical constraints limit applied simulations to process descriptions and parameterizations that include a greater degree of abstraction, phenomenology, and simplifying assumptions. Bringing to bear state-of-the-art computational methods and technology will enable development and solution of numerical testbed simulations with a degree of realism that will far exceed that of typical applied simulations. The results of these simulations (model outputs) will be placed in accessible archives for use by the broad scientific community as benchmark datasets. Assuming that the

numerical test beds represent the closest approximations to "truth" currently possible, they can then be used as comparative standards against which to measure the performance of up-scaled models (coarser resolution) or models with different (more abstracted or phenomenological) process representations. In this manner, strengths and weaknesses of various approaches used in model applications can be assessed and their potential impacts on model predictions (and ultimately policy and management decisions) can be determined. It is also expected that the benchmarking exercises will lead to new ideas regarding how to improve up-scaling methods for use in model applications.

Coordination with Detailed Characterization and Experimentation

Because the numerical test beds will be treated as comparative standards of true system behaviors, it is essential that key aspects of the test bed simulations be validated against experimental observations. While our ability to observe in detail all aspects of model behavior that might be simulated is limited, experiments can be designed in coordination with the numerical test bed efforts that will provide key observations for model validation. Recent advances in non-invasive characterization methods, particularly at local (e.g., pore) scales can provide nearly exhaustive datasets in three dimensions for use in highly-intensive simulations for development of numerical test beds.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The critical scientific challenge is the identification and specification of target processes and/or problems for which this approach can yield significant benefits. Subsurface flow, transport, and reaction processes that can be relatively well defined at a selected scale or for a particular case but less well understood at larger scales, or for general situations, they can comprise the set of preferred candidates for development of numerical test beds. Problems and processes for which field-scale behaviors cannot be defined in a straightforward manner terms of local descriptions are of particular interest. These include processes that are strongly coupled and/or non-linear, and problems in which media properties are highly heterogeneous at multiple scales and/or temporally varying.

The primary computational challenge is the scaling of numerical simulation codes to significantly expand the limits of the size and complexity of simulations that can be executed. Because each numerical test bed to be developed represents a very specific problem with unique properties, it is likely that such an effort will not use general-purpose codes but will instead develop custom codes with only those capabilities explicitly needed. These codes can be designed to take advantage of the properties of the problem under consideration and the most advanced algorithms available to maximize the scalability of the simulations.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The numerical test bed environment will provide a setting for the application of the most advanced computational algorithms and technologies. In most subsurface applications, computational scientists are constrained by a number of factors including budget and time constraints, ties to pre-existing codes, data limitations, and so forth. In a test bed environment, those constraints will be removed, for the most part, and the focus will be on pushing the computational limits as far as possible for a specific problem.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

Many other fields of science have extensively used the numerical test bed approach to facilitate significant scientific advances. Some examples include cloud ice nucleation studies (Cotton et al. 2007), laser simulation (Gross et al. 2002), and astrophysics (Yo et al. 2001). Perhaps the prime example is the field of atmospheric science, in which major resources have been devoted to development of combined numerical/experimental test bed facilities (e.g., see the National Center for Atmospheric Research's "Data Assimilation System Test Bed," website at http://www.mmm.ucar.edu/ modeling/dasystem/index.php). The predominant expected impact is more rigorous understanding of the strengths and limitations of existing models, which will in turn lead to research that will reduce those limitations and enhance confidence in model predictions.

TIME FRAME

- *Proposal development phase.* Identification of target processes, supporting experimental datasets, and definition of numerical test bed problem
- *1 to 3 years*. Development of one or two prototype numerical test beds and solution on terascale computers
- *3 to 5 years*. Based on prototypes, define the broader scope of an integrated numerical test bed system with supporting infrastructure
- *5 to 10 years*. Development of multiple numerical test beds within the integrated structure and solution on petascale computers.

REFERENCES

Cotton, R.J., Benz, S., Field, P.R., Mohler, O. and Schnaiter, M., 2007. A numerical test-bed for detailed ice nucleation studies in the AIDA cloud simulation chamber. Atmos. Chem. Phys., 7: 243-256.

Gross, M.S., Black, I. and Mueller, W.H., 2002. Numerical testbed for laser materials processing Proc. SPIE, 4631: 254-263.

Yo, H.-J., Baumgarte, T.W. and Shapiro, S.L., 2001. Numerical testbed for singularity excision in moving black hole spacetimes. Physical Review D, 64(12): 124011.

COMPUTATIONAL MODEL DEVELOPMENT FOR CRITICAL PROCESSES AT FUNDAMENTAL SCALES: MOLECULAR AND NANO

David A. Dixon, University of Alabama; Andy R. Felmy, Pacific Northwest National Laboratory; Louise J. Criscenti, Sandia National Laboratory

ABSTRACT

Models of contaminant fate and transport in the subsurface are built on detailed knowledge of the binding and reaction of contaminants on soil particles as well as transport and reaction in groundwater. There is a critical need for improved data at the molecular and nanoscale for such models in terms of both spatial length and time lengths from the femto-second to seconds and larger. These scales are the most fundamental small scales in both space and time that are relevant to subsurface modeling. Improved data at these fundamental small scales are needed for effective models at larger scales, such as the colloid and pore scales that then connect to higher levels at the pore network, fracture, and geological basin scales. Substantial advances are needed in electronic structure and Molecular Dynamics approaches to provide reliable predictions of the thermodynamics and kinetics of complex geochemical and biogeochemical systems. The availability of petaflop computer architectures and the development of computational approaches and algorithms that take full advantage of these architectures will allow us to make great strides in understanding the molecular/nano-scale chemistry of complex subsurface systems and evaluating which data from these scales must be incorporated into larger scale models for accurate prediction of field-scale processes.

EXECUTIVE SUMMARY

Many fundamental spatial scales are important for contaminant fate and transport modeling in the subsurface ranging from the molecular to the nanoscale to the mesoscale (polymer/colloidal), and to the pore scale. There are also many time scales of importance ranging from the femtosecond time steps of molecular motions to the millions (and billions) of years of geological time scales. These scales provide the basic parameters which serve as the basis of the physical models important at even larger scales beginning with pore networks and fractures to the continuum porous medium to the aquifer scale and then to the geological basin scale. Of critical importance are the fundamental small scales: those of molecules and nanoparticles. These are the scales that when coupled to thermodynamics provide the basic equilibrium and kinetic data necessary for all subsurface transport models. Great progress has occurred over the last two decades in understanding the fundamental chemical processes that govern contaminant fate and transport. This progress has occurred through a combination of experimental advances driven by access to high-end measurement capabilities and of computational chemistry advances made possible by the availability of more advanced computer architectures. However, because of the

complexity of the subsurface environment, substantial advances in theory, algorithms, software, and access to high-performance computers are still needed for such progress to continue. Advances are required in electronic structure methods to deal with complex surfaces, to provide accurate results for the thermodynamics of solutions interacting with mineral surfaces, and to provide reliable predictions of kinetics in such complex systems. In addition, there are serious issues with current dynamical simulation techniques in terms of the length of the possible dynamical simulation that will require substantial advances in theory and algorithms especially for petaflop computing if we are to reliably simulate the dynamics of large nanoscale systems for times beyond nanoseconds. Finally, there is a need to model biological processes from the protein and cell membrane scale to the cell scale to the community level as all of these can have substantial impacts on subsurface processes. The community scale is at the >mm scale. Improving existing capabilities will have both immediate and long-term impacts. Advances will require teamwork between computational chemists, biologists, applied mathematicians, and computer scientists.

SUMMARY OF RESEARCH DIRECTION

Models of contaminant fate and transport in the subsurface are built on detailed knowledge of contaminant sorption and reactivity with soil particles as well as contaminant complexation and transport in groundwater. Reliable models of the impact of mobile contaminants on humans and of the risk of proposed remediation technologies will be critical for developing the safest and most cost-effective approaches to site cleanup, geological disposal (e.g., CO_2 sequestration or radionuclides), and clean aquifers. Accurate models are needed for critical chemical, biological, and physical processes at the scales at which they occur. Such a research effort needs to span the interfacial regime from bare solid surfaces to complex, solution-phase surface chemistry, covering a range of time and length scales. The spatial scales critical to subsurface science are shown in Table 1. An understanding of the linkages between different temporal and spatial scales is an important need. No matter how detailed the physics in a reactive-transport model, if the critical underlying physical, chemical and biological data are missing or unreliable, the accurate predictive capability of a model can be compromised. To model processes at one scale requires that we understand the important variables at smaller scales to ensure that these variables are accurately represented at the larger scale. Improvement in both the data and the conceptual models will lead to simulations that are more scientifically defensible. Great care must be taken to minimize errors in the data used in sophisticated environmental or chemical process models so that these errors do not accumulate, propagate, and ultimately invalidate the macroscopic-scale models.

Table 1. Fundamental Spatial Scales and Computational Methods

Fundamental Scales	**Computational Approach**
Molecular	Schrodinger Equation
Molecular, nano	Molecular Dynamics
Micron (colloidal particles, polymers)	mesoscale modeling
Pore	Discrete grains, continuum or discrete description of fluids, Smoothed Particle Hydrodynamics/Discrete Particle Dynamics/ Lattice Boltzmann / Computational Fluid Dynamics
Application Scales	**Critical Physical Phenomena**
Mineral-water interface	Adsorption, mineral growth/dissolution, redox
Pore network	Linked pores and grains
Continuum (porous medium)	Darcy flow, bulk concentration reactions
Aquifer/aquitard	heterogeneous porous media
Geological basin	features that control fluid flow and geochemistry – fractures, stratigraphy

Multi-scale modeling is having a broad impact on many areas of research as shown by recent summaries (Yip 2005; McCurdy 2005; Colella et al. 2003; Reed 2003), which show how multi-scale approaches can link molecular-level processes to macroscopic phenomena.

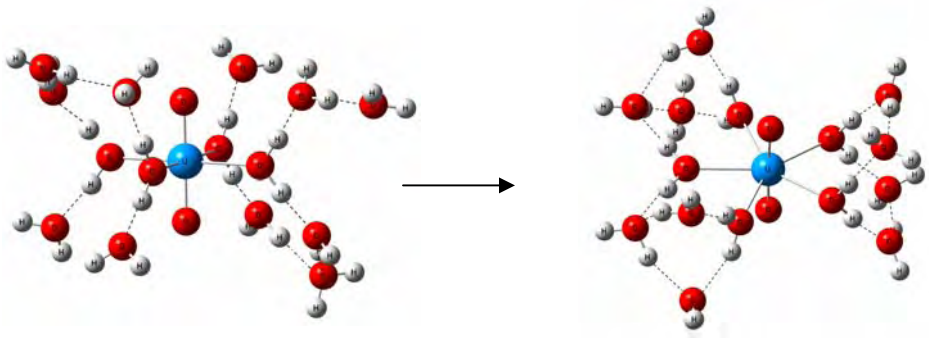
THE MOLECULAR SCALE AND THE NANOSCALE

Advances in computing resources, algorithms, theory, and software have enabled computational methods to make tremendous advances, for example, in computational electronic structure of molecules, molecular dynamics simulations of fluids and proteins, and in pore scale modeling. Coupling these computational science advances with the advanced measurement capabilities at DOE user facilities such as the light sources (e.g., the Advanced Photon Source, the Advanced Light Source, and the National Synchrotron Light Source), neutron sources (e.g., the Spallation Neutron Source), Nanoscience Centers (e.g., the Center for Integrated Nanotechnologies), and the W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL), have enabled unprecedented understanding of complex processes at the fundamental molecular scale and the nanoscale. The results from these computational and experimental advances need to be used to develop improved conceptual models for subsurface science and to provide improved data for higher level models at larger scales.

To support the development of innovative technologies for remediating contaminated soils and groundwater, we need reliable models of contaminant fate and transport in the subsurface built on a detailed understanding of the sorption and reaction of contaminants on soil particles as well as diffusion and reaction in the groundwater. Computational science research needs to span the interfacial regime from bare solid surfaces interacting with a gas to complex, solution phase surface chemistry, and cover a range of time and length scales.

The reactivity of mineral interfaces with aqueous and non-aqueous solutions is a key area of simulation research that is now becoming possible because of recent computational advances. This area of research will require substantial advances to provide detailed information about complex phenomena such as sorption of ions on surfaces, precipitation processes, dissolution reactions, and changes in mineral coatings. Results from this effort will lead to the development of molecular-level information which can be incorporated into thermodynamic, kinetic, and hydrological models of macroscopic subsurface reactive transport models.

Another key need is the ability to predict the behavior of various ions, molecules, and nanoparticles in solution, particularly aqueous solution. Predicting the thermodynamic and kinetic behavior, for example, of contaminants in aqueous solution, is extremely difficult because of the large size of the system that needs to be modeled over substantial simulation times to make a reliable prediction of the entropy. Although, there has been some success in this area, it is extremely computationally intensive to get reliable results just for thermodynamics (Gutowski 2006; Zhan 2002) (See Figure 1). For example, first-principles predictions of the pK_a of inorganic acids are still difficult to accomplish within a couple of pK_a units. To establish the formation constants and distribution of ion pairs and other species in aqueous solution, large-scale classical Molecular Dynamics simulations are required with force field potentials consistent with calculated results based on first principles (Larentzos and Criscenti, in press). In redox chemistry, it is still extremely difficult to predict the kinetics of the electron transfer process in complex systems (Rosso 2002, 2003, 2004). Yet, such electron transfer reactions govern the redox state of ions in solution which for many contaminants governs the solubility and sorption properties of the contaminant, and hence control its fate and transport. We must also be able to computationally treat elements across the Periodic Table, which means that computationally expensive, relativistic effects must be included (Wilson 1988; Liu et al. 1995).



$UO_2(H_2O)_5(H_2O)_{10}^{2+}$ $UO_2(H_2O)_4(H_2O)_{11}^{2+}$

Figure 1. Prediction of the equilibrium reaction $UO_2(H_2O)_4^{2+} + H_2O \leftrightarrow UO_2(H_2O)_5^{2+}$ measured to be -1.2 1.2 ± 0.4 kcal/mol in aqueous solution [L Soderholm, S Skanthakumar, J Neuefeind, *Anal. Bioanal. Chem.* **2005**, *383*, 48] required high level MP2 calculations including relativistic effects of UO_2^{2+} complexes including part of the second solvent shell and a continuum model for the rest of the solution to obtain a value of -2.0 kcal/mol for good agreement with experiment [K. E. Gutowski, and D. A Dixon, *J. Phys. Chem. A*, **2006**, *110*, 8840].

The interactions among microbes, dissolved metal ions, and mineral surfaces must be known to understand the fate and transport of contaminant species in groundwater. Recent advances in simulations enable the modeling of the transport of metal ions from a mineral surface to the outer cell membrane (see Figure 2). However, we are far from being able to model the uptake and redox chemistry of metals by a microbial cell at the molecular to cellular levels.

Molecular-level simulations including a treatment of the electronic structure have developed to the point where studying mineral-water interactions including water structure, solute adsorption, and diffusion along mineral surfaces are now possible. Electronic structure methods include DFT (Parr and Yang 1989; Koch and Holthausen 2001) and molecular orbital theory (Cramer 2002; Yarkony 1995). Computational issues that are still confronted include the ability to study the electronic structure and dynamics of large (>1 nm^3) systems over long time periods (nanoseconds) on a regular basis In classical dynamics simulations, there is a need to go to substantially longer time scales (> 100s of nanoseconds). This will result in a data management/analysis issue of how to deal with the storage of trajectory data for all atoms at many time steps. These data are required to obtain a statistically sound collection of charge-distribution data for the mineral-water interface to define the interfacial electric field. The charge distribution from a mineral-water interface into bulk solution is a fundamental consideration in the development of surface complexation models that describe molecular-scale interfacial reactions in terms of bulk continuum properties and are incorporated into field-scale reactive-transport models.

Molecular-level simulations on aqueous systems and for mineral-water interfaces are currently performed on simple systems that ignore much of the complexity found in natural systems. For example, aqueous solutions are often modeled as solutions of NaCl or $CaCl_2$, but not mixtures involving more than one cation and anion. In addition, these solutions are either modeled as "infinite-dilution" systems or >1 M systems. Simulations, run for over 1 ns often still reflect the initial conditions of the simulation. Therefore, statistical approaches for creating several starting configurations and then integrating the results of several simulations together are necessary. For mineral-water interfaces, scientists have focused on the adsorption of contaminants to neutral cleavage planes or faces of crystals instead of growth steps and edges. Establishing valid adsorption constants for appropriate surface species through molecular simulation requires that we examine the relative Gibbs free energy of adsorption to the mineral surface sites that do in fact dominate adsorption in the field. The influence of defects including vacancies, crystal terminations of different sorts and the effect of impurities on adsorption needs to be investigated.

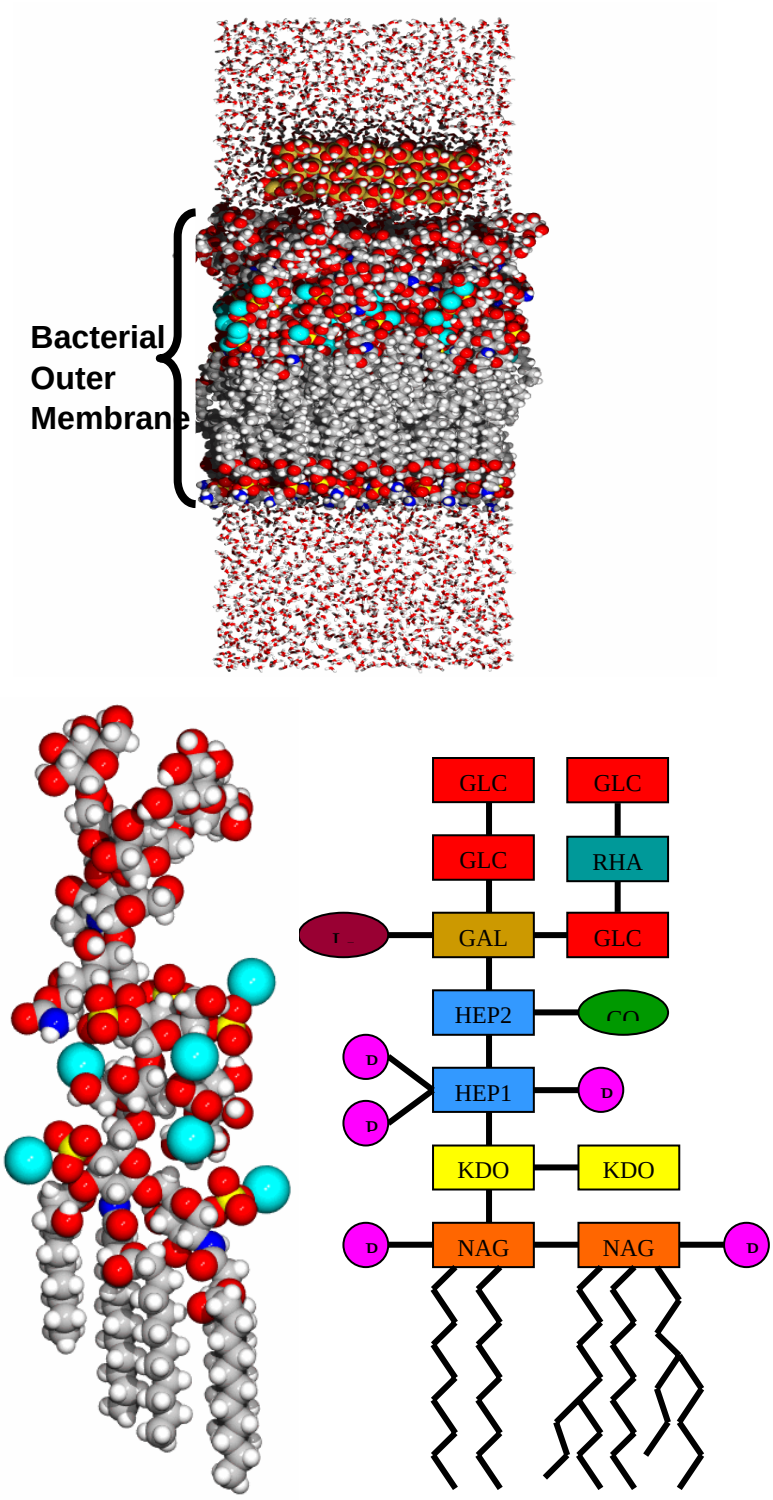


Figure 2. Molecular representation of the LPS membrane of *Pseudomonas aeruginosa* in near contact with the goethite mineral fragment (R.M. Shroll, and T.P. Straatsma, *Biophysical Journal*, 2003, *84*, 1765).

In addition to studying the detailed chemistry of mineral-water interfaces, studies of how changes in water structure at the mineral surface affect solute transport in the environment are needed. Thin films of water are responsible for contaminant migration in the unsaturated (vadose) zone. Molecular simulations at both the electronic structure and classical molecular dynamics levels provide an excellent way to examine diffusion rates of solutes as a function of mineral surface and relative humidity. This information will provide insight into the development of continuum models for unsaturated flow for the prediction of contaminant migration in areas like the proposed high-level waste reposity at Yucca Mountain, Nevada.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

There are numerous computational challenges because of the large temporal and spatial scales that must be covered in subsurface modeling. An additional "scale" that must be considered is the accuracy of the simulation. For example, time is usually considered to be linearly scaled (O(N), yet there can be a huge number of time steps depending on the length of the time step; geometry is nominally $O(N^3)$ for the three dimensions but accurate calculations can have computational scaling that is much worse. Examples of scaling greater than $O(N^3)$ scaling include appropriate sampling statistics, accurate electronic structure ($O(N^7)$), complexity of pores, numbers of chemical processes to be modeled, and system sizes and boundaries.

Substantial improvements in the different modeling approaches are needed to achieve the goals described above. For example, improvements in computational quantum chemistry, including relativistic effects, are needed for reliable predictions of solvated metal ions such as lanthanides and actinides, and their interaction with complex mineral surfaces. Substantial improvements in electronic structure methods are needed to be able to make reliable predictions of the properties and reactivities of solid-state metal oxides that are prevalent in natural systems. Substantial efforts are needed in developing approaches for the reliable prediction of reaction rates (e.g., those of reodx reactions). Substantial improvements are needed in Molecular Dynamics methods to deal with the time scale problem including improved sampling and statistical approaches. In fact, improved sampling methods are critical to many of the advances for subsurface simulation.

Current fine scale (e.g., molecular, nano, cellular compartment) models of microbial mineral interactions are extremely limited because of the complexity of the system and the lack of even experimental data with which to construct appropriate conceptual models that can be expressed in a simulation. New experimental measurements are leading to novel insights at the molecular level (e.g., the interactions of microbes with mineral surfaces). However, despite recent increases in computational power, the system size and the lack of appropriate force fields still make it difficult to do Molecular Dynamics simulations of microbial-mineral systems. In addition, current Molecular Dynamics simulations based on classical force fields do not provide ways to study redox reactions or to couple to energy processes in the cell. Computational approaches need to be developed to effectively couple genomic, proteomic and meta-bolomic data as well as biochemical network analysis into models of microbes in the subsurface.

Examples of challenges that require petascale computing are outlined below:

- *0.1 PFlops (100 TFlops).* Unraveling the impacts of surface heterogeneity on mineral surface reactivity including coupled ion adsorption and electron transfer processes. For example, this would require accurate predictions of the thermodynamics of complex mineral surfaces interacting with aqueous solutions (±1 kcal/mol leading to an error in $K < 10$).
- *1 PFlops (1000 TFlops).* Evaluating the fundamental mechanisms involved in mineral-microbe interactions including surface attachment, electron transfer, and impacts of surface coatings. For example, this would require accurate predictions of reaction rates on complex mineral surfaces. (<50 percent error in rates).

Other PRDs describe the issues that occur at the micron scale (e.g., polymer modeling, thermal fluctuations) and at the pore scale (e.g., discrete grains, continuum or discrete description of fluids: smoothed and dissipative particle hydrodynamics, lattice Boltzmann, and computational fluid dynamics). Additional scales that need to be considered include the fractured network scale, the aquifer scale (heterogeneous porous media), and the scale of geological basins.

A critical challenge will be to adapt current approaches to truly massively parallel architectures, those with >5000 processors. The planned computer architectures of the next 4-plus years will be based on technology limitations for individual chip speeds. It is not likely that chip speeds for typical processors will dramatically increase during this time period beyond 4 GHz; therefore, we will no longer benefit from the current trajectory for single processor performance. Rather, chip producers are going to multi-core technologies with lower power and cooling requirements. These machines are likely to have smaller memories per processor and very little high-performance disk for use as scratch. Thus, many algorithms and computational approaches (e.g., those in computational quantum chemistry) will need to be revised substantially to make use of these machines. For example, the use of basis sets in quantum chemistry may disappear if we are to use these new architectures effectively. An alternative may be to use quantum Monte Carlo methods (Hammond 1994; Nightingale 1999). In addition, new sampling approaches that enable a broader coverage of phase space will be needed for Molecular Dynamics simulations, because advances in single processor speed will not enable longer simulation times. Other issues dealing with distributing Molecular Dynamics simulations across large processor counts will also need to be addressed. In addition, new advances in exchange-correlation functionals for DFT are required for the reliable prediction of the electronic structure of metal oxides that are so important in geological environments.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

There are a number of computational science advances that are required for this area to be successful. Access to advanced high-performance computing resources will enable reliable predictions at individual scales and enhanced connectivity between and across scales. As noted in other PRDs, the concept of being able to up-scale and/or down-scale is critical to success in subsurface modeling. Advances in this area will translate into advances in many other fields

where the same issues of scaling in time and space are present; for example, improved phase-space sampling and statistics for moving from short to long scales. In addition, new multi-scale methods will require new mathematical/computer science approaches. However, these mathematical/computer science approaches need to be developed based on the application requirements.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

There is a substantial potential impact on subsurface science. Improved conceptual models will lead to advanced oil recovery technologies; new remediation technologies that are energy efficient, cost-effective, and safe; and reliable predictions of contaminant transport for site closure. Improved large-scale models that are based on sound scientific data will be important to stakeholders and the public. A fundamental understanding of how molecular and nanoscale processes control larger scale behavior will substantially improve large-scale models. To support the development of innovative technologies for remediating various DOE sites, we need to develop reliable models to investigate the impact of the technology and the appropriate level of risk in using the technology. These models of contaminant fate and transport in the subsurface have to be built on a detailed understanding of the binding and reaction of contaminants on soil particles as well as transport and reaction in groundwater. In addition, reliable models of the direct impact of mobile contaminants on humans and the risk of proposed remediation technologies will be critical for developing the safest and most cost-effective approaches to site cleanup and for public acceptance of the clean-up process and results. High-quality data is needed for reactive transport models, and great care must be taken to minimize the errors in the calculated underlying data used in a sophisticated environmental or chemical process model so that errors in the data do not accumulate, propagate, and ultimately invalidate the macroscopic-scale model.

TIME FRAME

There is urgent need for the data that this effort will provide. Contaminants have already been released into the environment and they are mobile (e.g., the Hanford vadose zone problem). Reliable models are needed to predict when these mobile contaminants will have a direct impact on humans so as to determine what remediation strategies need to be implemented and in what time-frame, given limited amounts of funding for the remediation effort. Interactions within the complex chemical mixtures in the waste tanks determine what methods can be used to transfer and process the wastes as well as how to safely store them for thousands of years (e.g., in the proposed waste repository at Yucca Mountain, Nevada). Such information is needed to solve DOE's cleanup problems in a comprehensive, cost-effective, permanent way. Improving existing capabilities would have an immediate as well as long-term impact.

REFERENCES

Colella, P., T. H. Dunning, Jr., W. D. Gropp, and D. E. Keyes. 2003. *A Science-Based Case for Large-Scale Simulation, Volumes 1and 2* , Office of Science, U.S. Department of Energy. See http://www.pnl.gov/scales/docs/volume1_72dpi.pdf and http://www.pnl.gov/scales/docs/SCaLeS_v2_draft_toc.pdf

Gutowski, K. E. and D. A Dixon, 2006. "Predicting the Energy of the Water Exchange Reaction and Free Energy of Solvation for the Uranyl Ion in Aqueous Solution," *J. Phys. Chem. A* **110**, 8840.

J. A. Hammond, W. A. Lester Jr., and P. J. Reynolds. 1994. Monte Carlo Methods in Ab Initio Quantum Chemistry, World Scientific, Singapore.

Koch, W., and M. C. Holthausen. 2001. *A Chemists Guide to Density Functional Theory*, 2nd Edition, Wiley-VCH, Weinheim.

Larentzos, J. and L. J. Criscenti. (2007) A Molecular Dynamics Study of Alkaline Earth Metal-Chloride Complexation in Aqueous Solution and on a Gibbsite Surface, *J. Phys. Chem. B.* in press.

Liu, W., W. Kuchle, and M. Dolg. 1995. "Ab Initio Pseudopotential and Density Functional Studies of Lanthanide and Actinide Systems," *Handbook on the Physics and Chemistry of Rare Earths*, eds. K. A. Gschneider Jr. and L. Eyring, Elsevier, Amsterdam, Vol. 22, Chapter 170.

McCurdy, W. (Ed.) 2005. *Opportunities for Discovery: Theory and Computation in Basic Energy Sciences*, Department of Energy

Nightingale, M. P., and C. J. Umrigar. 1999. *Quantum Monte Carlo Methods in Physics and Chemistry*. Eds. M. P. Nightingale and C. J. Umrigar, Kluwer Academic, Dordrecht.

Parr, R. G., and W. Yang. 1989. *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York

Reed, D. A.. 2003. "The Roadmap for the Revitalization of High End Computing," Computing Research Association, Washington D.C. see http://www.nitrd.gov/subcommittee/hec/hecrtf-outreach/ 20040112_cra_hecrtf_report.pdf

Rosso K.M. and Dupuis M. (2004) Reorganization energy associated with small polaron mobility in iron oxide. *J. Chem. Phys.*, **120**, 7050.

Rosso K.M. and Morgan J.J. (2002) Outer-sphere electron transfer kinetics of metal ion oxidation by molecular oxygen. *Geochim. Cosmochim. Acta*, **66**, 4223.

Rosso K.M., Smith D., and Dupuis, M. (2003) An ab initio model of electron transport in hematite (α-Fe_2O_3) basal planes. *J. Chem. Phys.*, **118**, 6455.

Shroll, R. M. and Straatsma, T. P. (2003) Molecular Basis for Microbial Adhesion to Geochemical Surfaces: Computer Simulation of *Pseudomonas aeruginosa*: Adhesion to Goethite.,*Biophysical Journal*, **84**, 1765.

Wilson, S. 1988. "Relativistic Effects in Atoms and Molecules," *Methods in Computational Chemistry, Vol. 2*, ed. S. Wilson, Plenum Press, New York.

Yip, S. 2005. *Handbook of Materials Modeling*, ed. S. Yip, Springer, New York

Zhan, C.-G. and Dixon, D. A. (2002) First-Principles Determination of Absolute Hydration Free Energy of the Hydroxide Ion, *J. Phys. Chem. A*, **106**, 9737.

MODELING AND SIMULATION OF SPATIAL STRUCTURES IN HETEROGENEOUS SYSTEMS

Paul Meakin, Idaho National Laboratory; Alexandre Tartakovsky, Pacific Northwest National Laboratory; Russ Detwiler, Lawrence Livermore National Laboratory

ABSTRACT

A structural model that is used in the definition of boundary and initial conditions is the essential starting point for any simulation of subsurface processes. Depending on the purpose of the simulation, the structural model can vary from a very simple geometrical model for pore space and fracture aperture geometries to complex multi-scale, multi-physics models for the processes that control the morphology of subsurface system of interest. Realistic process-based (mechanistic) models for pore-space geometry and biogeochemical heterogeneity have not been rigorously investigated because of the lack of suitable models and the computational resources needed to implement them. Petascale computing and advances in algorithm development will make a first-principles, process-based approach to microscale and pore scale structure increasingly attractive. The application of computational methods to the construction of larger scale structure/heterogeneity models based on either geophysical data or process-based simulations is a challenging problem that required better algorithms and more capable computing systems. Structural models for subsurface applications are often based on large datasets obtained from laboratory experiments (e.g., x-ray tomography) or geophysical measurements (e.g., seismic and electric data). The archiving, analysis, retrieval, and visualization of these large datasets will be challenging, and the inversion of large geophysics datasets obtained from one or more types of measurements to obtain large-scale three-dimensional heterogeneity fields will require computing systems that are more capable than those available today.

SUMMARY

A model for subsurface structure and biogeochemical heterogeneity is the starting point for any model for subsurface processes. A wide variety of models gave been developed, and they can be classified as geometric/object based models, statistical models, process-based models, and data-based models. However, the boundaries between these classes are fuzzy, and it is possible to combine models from two or more classes. For the most part, petascale computing is not needed for statistical and geometric/object based models. However, petascale computing would very substantially increase the resolution and accuracy of models that are used to obtain heterogeneous property fields from large geophysical date sets. In addition, computationally intense, process-based models could be used to improve the interpretation of the heterogeneous property fields obtained from data inversion in terms of structural and biogeochemical characteristics.

The availability of more powerful computing systems and better models would make pore-scale and sub-pore-scale, process-based modeling and simulation an attractive and practical approach to obtaining the structural information and spatial distribution of physical and chemical properties needed to simulate wetting behavior, fluid flow, precipitation/dissolution and adsorption/desorption on the pore scale and sub-pore scale. The process-based simulation of subsurface structure and heterogeneity require the same modeling and simulation methods and the same high-end computing capabilities as fundamental process-based models for the subsurface behavior that it critical to practical applications such as the success of oil recovery, carbon sequestration, and environmental management/remediation. In this PRD, the main emphasis is on microscopic and pore scale structure and heterogeneity. However detailed information about structure and heterogeneity is required for the accurate simulation of subsurface systems on all scales, and the challenges and opportunities associated with the use of computation based methods to obtain this essential information is also addressed. Archiving, retrieval analysis and visualization of the large datasets will be required, in addition to new models and algorithms that are effectively implemented on very large computing systems.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Simulation of the behavior of fluids in the subsurface requires information about the geometry and wetting properties of the interconnected pore volumes and fracture apertures that confine the fluids, and if the fluids react with the surrounding rock matrix or soil grains, the distribution of mineral compositions and reactivities (controlled by impurities, defects, stress, etc.) must also be taken into account. In addition, the initial distribution and identity of any microorganisms (the spatial distribution of each member of the microorganism community) must also be known for certain important applications. In practice, geological systems have distinctive structures on scales ranging from lengths on the order of nanometers (e.g., defect structures on mineral surfaces and the interiors of mineral grains and diagenetic decoration on mineral surfaces) to lengths on the order of kilometers or more (e.g., geological units and faults in oil reservoirs or aquifers). Any serious attempt to model and simulate subsurface systems for practical purposes must take these heterogeneous structures and scales into account, and a subsurface structure model is an essential component of any model that is used to simulate subsurface processes.

One of the most direct approaches to the determination of the three-dimensional pore space geometry of porous and fractured porous media is to combine a sequence of two-dimensional cross-section images to construct a three-dimensional image. This approach is very time consuming, and is being superseded by NMR imaging and x-ray tomography (Spanne et al. 1994; Coker et al. 1996). Synchrotron x-ray tomography is capable of resolutions down to about 50 nm, and three-dimensional electron density images may consist of as many as 8 billion cubic voxels ($2000 \times 2000 \times 2000$). The size of these datasets is likely to increase as the technology of synchrotron x-ray tomography continues to advance. A voxel-based representation of the pore space geometry can be obtained from the electron density map, and can be used directly in lattice based models for single- and multi-phase fluid flow (Harting et al. 2006). In some cases, if the electron density contrasts are high enough, x-ray tomography can be used to determine the

distribution of fluid phases within pore spaces and fracture apertures. This opens up the possibility of using x-ray tomography to validate multi-phase fluid flow simulations. NMR imaging could also be used in the same way with the additional advantage that fluid velocities can be measured for single-phase fluids. However, NMR has a much lower resolution than x-ray tomography. The increased brightness of third-generation synchrotron x-ray sources supports rapid sequential image acquisition for applications such as multi-phase fluid flow in porous media, and archiving, analysis and visualization of the large datasets generated in this manner will be challenging. The analysis required to convert the x-ray intensity data acquired from a charge-coupled device array detector is also computationally demanding, and parallel algorithms have been developed for this purpose. Although computationally intensive, the analysis of data from x-ray tomography currently does not require petascale computing, but such capability may be needed for the analysis of much larger datasets.

Statistical methods are widely used to generate complex disordered structures, based on statistical analysis of usually incomplete structural information from experiments or from theoretical models. The fractal models discussed below provide an example of the latter approach. Statistical measures that are often used to analyze experiments include a variety of spatial correlation functions, chord length distributions, grain/particle size distributions, and fracture length and aperture width distributions. In some cases, directional information and vector/tensor correlation functions are also required. In many cases, only two-dimensional images (cuts through a complex structure) are available, and the information obtained from analysis of these two-dimensional images can be used to "reconstruct" a three-dimensional structure (Yeong and Torquato 1998). Clearly, the reconstructed three-dimensional structure is not the same as the actual three-dimensional structure, and the uncertainties in the statistical measures used for the reconstruction process may be quite large if only a few two-dimensional images are available. However, the reconstruction procedure does allow a large number of three-dimensional structures with the same statistics to be generated. Similar methods could be used for other types of heterogeneity (e.g., biogeochemical). In most geostatistical approaches, models based on stationary statistics (position independent statistics) and low-order correlations have been employed. There is no guarantee that these relatively simple statistical models will prove to be adequate in all cases, and models based on non-stationary statistics and/or higher order correlations may be needed. Models of this type will require the collection and analysis of larger datasets to provide accurate representations of the statistical measures needed to implement the statistical models.

Geometric models are often used to generate the fracture apertures and pore geometries needed for simulations of single- and multi-phase fluid flow in the subsurface. The use of self-affine fractal models for the boundaries of fracture apertures (Thompson 1991) has become popular since Mandelbrot and Passoja (1984) showed that the rough surfaces generated by brittle fracture have a self-affine geometry, and subsequent work indicated that the Hurst Exponent, or fractal dimension, has a quasi-universal value. In addition, a variety of experimental investigations have provided evidence that the short length scale (nanometers to hundreds of nanometers) structure of pore surfaces may also be fractal. Because granular porous rocks are often major

components of oil reservoirs and aquifers, grain-packing models have been widely used to represent the structure of porous media. Most of these models are based on grains that have simple spherical or ellipsoidal shapes. In addition, a variety of simple pore network models have been used. These models consist of pore volumes connected by pore throats (channels) that have simple shapes. Very often, the pore volumes are assumed to have no effect on fluid flow other than to provide a volume of fluid which must be displaced before fluid-fluid interfaces can move from one pore throat to another. These pore-network models are often based on simple lattice geometries, but Voronoi networks and Delaunay triangulations have also been used. A number of ways of generating "topologically equivalent pore networks" from the three-dimensional geometry obtained via x-ray tomography or other methods have also been developed (Blunt et al. 2002). These geometric models provide a simplified, often quite crude representation of the pore and fracture apertures geometries, and they neglect the microscopic roughness, complex mineral overgrowth morphology and surface chemical heterogeneity that may play an important role in wetting behavior and chemical reactivity. Geometric models for subsurface structure do not typically require high capability computing systems. Most of the work on pore scale models for subsurface structure has focused on granular porous media. However, about half of the world's oil and natural gas is recovered from carbonate reservoirs (including the majority of "giant' and "super giant" reservoirs in the Middle East) in which fracture flow is dominant and the pore structure is substantially more complex.

An alternative approach, which is becoming more attractive with the increased focus on multi-scale, multi-physics modeling and simulation and the rapid growth in high-performance computing capabilities, is to simulate the biogeochemical and physical processes that contribute to the formation of complex pore space and fracture aperture geometries. This approach has been used more extensively for simulating large-scale than pore-scale structure. For example, basin fill modeling is used quite extensively in petroleum geoscience. A multi-scale, process-based model for pore scale and microscale heterogeneity might include Molecular Dynamics and kinetic Monte Carlo components to simulate the evolution of microscopic surface roughness resulting from mineral dissolution and precipitation, coupled with multi-scale models for cataclasis (the fracture and rotation of aggregates and mineral grains), pressure solution, fluid flow, solute transport. In some cases, models for biological processes, such as the growth of biofilm in organic rich contaminated systems or systems in which biological processes can be stimulated by providing nutrients are also required. Realistic process-based (mechanistic) models for pore-space geometry and biogeochemical heterogeneity have not been seriously investigated because the lack of suitable models and the computational resources needed to implement them. Molecular Dynamics and kinetic Monte Carlo models have been developed to simulate processes such as molecular beam epitaxy and catalytic reactions on surfaces and methods such as continuum phase field models with adaptive mesh refinement have been used to simulate dendritic growth (Kobayashi 1993; Provatas et al. 1998). The problem of simulating the processes that determine pore scale and microscopic structure is essentially the same as the problem of predicting the behavior of subsurface systems for a wide variety of purposes including carbon sequestration, oil recovery, and environmental remediation. The opportunities

that petascale computing can offer and the challenges that must be overcome to develop multi-scale, multi-physics codes that effectively use the capabilities of petascale computing systems are also very similar.

For large scale practical applications, a variety of hybrid approaches are employed. For example, in the oil industry, large-scale structures such as faults with displacements large enough to make them directly detectable by seismic methods can be included explicitly in reservoir simulator structural models, whereas faults with displacements that are too small to detect but large enough to have a substantial impact on fluid flow can be included by using a statistical or process-based approach. Very often, statistical information obtained from core samples is used to construct a "fine-grained" model for reservoir heterogeneity, and this fine-grained statistical representation is coarsened (up-scaled) to obtain effective properties on the scale of the reservoir simulator grid blocks (O(10 m)). This is achieved using a variety of methods including renormalization group methods (Hristopoulos and Christakos 1999) and wavelet transformation (Sahimi 2000). The construction of structural models in the oil industry usually relies heavily on processed geophysical data. Whereas this geophysical data is primarily seismic, supplemented by well-logging data, the use of electromagnetic data and potential field data in conjunction with seismic and well-logging data is increasing. A variety of "joint inversion" methods that use information from multiple sources such as electrical resistivity and seismic data are being been developed (Galardo and Maju 2003). Inverse problems are notoriously ill-posed, and they are based on optimization algorithms. "Exact" optimization algorithms are "non-polynomial complete", and cannot be used for large datasets. The inversion of geophysical datasets is computationally intensive and results in heterogeneity fields with uncertainties that are difficult to quantify. These heterogeneity fields should be considered to be, at best, "statistically representative" of the geosystem, particularly on relatively short length scales. Access to petascale computing resources would allow the uncertainties associated with the inversion of geophysical data to be substantially reduced, and this would have important practical consequences, particularly for the inversion of large datasets from complex three-dimensional systems. In this context, it is important to recognize that geophysical methods do not directly measure subsurface structure. Consequently, a multi-scale, multi-physics model that relates geophysical properties to structural and biogeochemical properties will be required to accurately interpret geophysical data. For example, the electrical resistivity of subsurface materials depends on the interactions between solute ions and mineral surfaces, the physics of the double layer, small-scale structural information, and charge transport in some minerals. Petascale computing with new multi-scale, multi-physics methods will be needed to make full use of geophysical measurements. There are also substantial computational challenges associated with the archiving, retrieval, and analysis of large geophysical datasets

The development of computer models for fractures and fracture networks is particularly challenging, and has many practical applications. For example, fractures can act as flow paths or barriers to flow in oil reservoirs, depending on details such as clay smearing or the formation of mineral deposits on fracture surfaces. In addition, the presence of difficult-to-detect fractures, the opening and closing of fractures, or ongoing fracture processes in oil reservoir or carbon

sequestration site cap rocks can have important practical consequences. A variety of approaches, including process-based, object-based, statistical and geometric models have been developed to simulate fracture networks. Many of these modeling approaches lead to structures that are not realistic. Important issues include the structure of individual fractures, fracture connectivity, the response of fractures and fracture networks to changes in geomechanical conditions during the removal and/or injection of fluids, the content of fracture apertures, and the structure of the fracture/rock matrix interface. The content of fracture apertures (gouge[1] with a wide range of particle sizes, colloids, mineral deposits) and the structure of the surrounding rock matrix, which may be chemically altered and/or mechanically damaged, adds to the challenge of developing accurate and realistic structure models. Large-scale models for subsurface processes depend on homogenization approximations based on concepts such as the representative elementary volume (REV). This concept depends on the pore scale structure, and is based on the idea that subsurface properties are scale dependent on both large and small scales, but that there is an intermediate range of scales over which the properties are scale independent. At this stage, it is not clear if homogenization approximations can be applied to systems in which fluid flow is dominated by fractures,[2] and even if homogenization approximations can be used, it is also not clear how the scale of the REV can be obtained. Because it is very difficult and expensive to characterize the structure of fracture networks in the subsurface, a computational approach to this problem might be the most practical and cost effective. This would require the development of realistic fracture network models and simulation of single-phase and multi-phase fluid flow on these networks. Multi-scale, multi-physics hybrid Molecular Dynamics/continuum mechanics (with adaptive mesh refinement) models have been developed to simulate fracture growth (Abraham et al. 1998; Rafii et al. 1998). Models of this type can be used to simulate the nanoscale structure of fracture surfaces, but the time scale of Molecular Dynamics simulations is much too short (even with accelerated Molecular Dynamics algorithms) to simulate the geometry of fractures on the pore scale and beyond.

RESEARCH DIRECTIONS

The application of process based models to generate pore-scale and sub-pore scale subsurface structure and biogeochemical heterogeneity is not completely new, but it is certainly in its infancy. The limited capability of the computing systems used by the subsurface science community combined with the lack of well-developed multi-scale, multi-physics models has discouraged this approach. The rapid increase in computing speed expected over the next 5 to 10 years and the stimulus that this will provide to the development of new algorithms that take full advantage of petascale computing systems will make this approach more attractive and practical. Larger scale models that include processes, such as the dynamics of fluvial systems, sediment transport, compaction, and sediment size segregation, have been applied more

1 Crushed and ground-up rock produced by motion of one side of a fault relative to the other.
2 This is especially true in the case of multi-phase fluid flow in fractured systems because the spatial distribution of the fluids may be inhomogeneous even for scales at which the fracture network appears to be geometrically homogeneous.

extensively. In this case, petascale computing will allow a more first-principles approach to be taken with better physics, and it will be possible to include more processes with higher spatial and temporal resolutions.

The "inversion" of geophysical data to obtain structural information has a long history. However the accuracy and level of detail in three-dimensional and four-dimensional (time dependent) heterogeneity fields obtained in this manner has also been limited by the capability of current computing systems. Faster computers, larger datasets, and better algorithms have already produced important advances. Horizontal/directional drilling and improved seismic data collection and analysis have been credited with much of the economic benefit resulting from improved oil recovery and more successful exploration. Joint inversion may also prove to be an important advance, with a wide range of applications and economic benefits.

COMPUTATIONAL SCIENCE IMPACT

New multi-scale, multi-physics models and better algorithms, implemented efficiently on high-capability (petascale) computing systems, needed to simulate subsurface structures based on first-principles, process-based methods would be directly applicable to a wide range of scientific and practical subsurface problems, including carbon sequestration and the *in situ* formation of barriers to flow for contaminant remediation applications. Extensive cross-fertilization with other areas of science such as materials science (e.g., concrete formation, ceramics, materials failure etc.) and the dynamics of granular materials (size segregation, transport and deformation) would stimulate the pace of discovery and implementation of new computational models. The archiving, retrieval analysis, and visualization of large datasets obtained from geophysical detector arrays or high-resolution, three-dimensional imaging also is a challenge that spans a very wide range of applications. In addition, the development of better data inversion/ optimization methods needed to obtain structural information from large geophysical datasets is a challenging computational/applied mathematics problem with a very wide range of applications.

SUBSURFACE SCIENCE IMPACT

Deficiencies in the structural/heterogeneity models that provides a foundation for the simulation of subsurface systems is a major (often *the* major) contributing factor to error and uncertainty. Essentially all pore-scale and sub-pore-scale models for fluid flow and reactive transport employ pore-space/fracture aperture models that do not well represent the structural complexity and biogeochemical heterogeneity found in essentially all natural systems. On larger scales, the models used to represent structural and other forms of heterogeneity are also inadequate, and it is often difficult to assess the impact of these inadequacies. The development of better models for subsurface structure and biogeochemical heterogeneity over a wide range of length scales using a combination of process-based models, structural data, and geophysical data is essential to place subsurface modeling and simulation on a firmer scientific basis, reduce uncertainties and enhance confidence in model predictions. Better models for subsurface structure and

biogeochemical heterogeneity on a wide range of length scales must go hand-in-hand with the development of multi-scale, multi-physics codes for high-performance petascale computing systems. Access to petascale computing systems would have an immediate impact on joint data inversion.

TIME FRAME

Within 5 years, it should be possible to develop process-based models that predict pore structures and sub-pores-scale structures over a wide range of length scales. These models could be validated using a variety of experimental methods including x-ray tomography, which can be used to obtain detailed structural information on scales ranging from a few tens of nanometers to a few centimeters (a 10^6 range of length scales). Some systems will prove to be more challenging than others are, and we expect that it will be about ten years before this approach is broadly and routinely applicable. Access to petascale computing systems would have an immediate impact on joint data inversion. Over the next 5 to 10 years, the development of better inversion/optimization algorithms, most likely driven at least in part by applications in other areas would, in combination with advances in computing performance, allow larger datasets to be analyzed, and the accuracy of the heterogeneous property fields obtained in this manner could be substantially reduced.

REFERENCES

F. F. Abraham, J. Q. Broughton, N. Bernstein and E. Kaxiras, Spanning the length scales in dynamic simulation, Computers in Physics, **12**: 538-546 (1998)

M. J. Blunt, M. D. Jackson, Mohammad Piri and P. H. Valvante, Detailed physics, predictive capabilities and microscopic consequences for pore-network models of multi-phase flow, Advances in Water Research, 25: 1069-1089 (2002)

D. H. Coker, S. Torquato and J. H. Dunsmuir, Morphology and physical properties of Fontainbleau sandstone via tomographic analysis, Journal of Geophysical Research-Solid Earth, 101:17497-17506 (1996)

L. A. Galardo and M. A. Maju, Characterization of heterogeneous near-surface geometerials by joint 2D inversion of dc resistivity and seismic data, Geophysical Research Letters, **30**: Article no. 1638 (2003)

J. Harting, J. Chin, M. Venturoli and P. C. Coveney, Large scale lattice Boltzmann simulations of complex fluids: Advances through the use of computational Grids, Philosophical transactions of the Royal Society, **A363**: 1895-1915 (2006)

D. T. Hristopoulos and G. Christakos, Renormalization group analysis of permeability upscaling, Stochastic Environmental Research and Risk Analysis, **13:** 131-160 (1999)

R. Kobayashi, Modeling and simulation of dendritic crystal growth, Physica **D63**: 410 – 423 (1993)

B. B. Mandelbrot and D. E. Passoja, Fractal character of fracture surfaces, Nature **308**: 721-722 (1984)

N. Provatas, N. Goldenfeld and J, Dantzig, Efficient simulation of dendritic microstructure using adaptive mesh refinement, Physical Review Letters, **80**: 3308-3311 (1998)

H. Rafii, L. Hua and M. Cross, A multi-scale atomistic-continuum modeling of crack propagation in two-dimensional metal plate, Journal of Physics. Condensed Matter, 10: 2375-2387 (1998)

M. Sahimi, Fractal-wavelet neural network approach to characterization and upscaling of fractured reservoirs, Computers and Geoscience, 26:877-905 (2000)

P. Spanne, J. F. Thovert, C. J. Jacquin, W. B. Lindquist K. W. Jones and P Adler, Synchrotron computed microtomography of porous media: topology and transports, Physical Review Letters, 73: 2001-2004, 1994

M. E. Thompson, Numerical of solute transport in rough fractures, Journal of Geophysical Research - Solid Earth, **96**: 4157-4166 (1991)

C. L. Y. Yeong and S. Torquato, Reconstructing random media, Physical Review **E57**: 498-506 (1998)

COUPLED GEOPHYSICAL IMAGING FOR SUBSURFACE FLUID IDENTIFICATION AND PROCESS CHARACTERIZATION

Gregory Newman, Lawrence Berkeley Laboratory; Roelof Versteeg, Idaho National Laboratory

ABSTRACT

Meeting CO_2 sequestration measurement, monitoring, and verification (MMV) demands as well as needs for the characterization of relevant processes and properties at DOE waste legacy sites requires the ability to obtain cost-effective, volumetric, and near-time information on the temporal and spatial evolution of fluids and corresponding transport pathways within heterogeneous geological systems.

Traditional approaches to monitor and characterize such processes through point sampling sensors in boreholes are expensive and fail to provide sufficient information. The only feasible approach to this problem is a comprehensive subsurface imaging effort which effectively exploits both multi modal geophysical data (including seismic and electromagnetic wavefield data and electrical field data) and confirmatory point measurement data. Such a comprehensive subsurface imaging effort must provide for quantitative prediction of subsurface processes, including fluid identification and transport pathways. It will require novel computational and numerical tools in the development and enhancement of effective joint imaging methodologies which can use temporally and spatially dense three-dimensional and four-dimensional (i.e., time lapse) geophysical data (i.e., electrical, electromagnetic, and seismic).

Development of a coupled geophysical imaging capability will need to be based on novel inverse methodologies as well as on detailed, physics based modeling that links the micro- and meso-scale properties of the subsurface system and the processes occurring in such systems to the macroscopically observed (change in) geophysical response. The novel inverse methodologies will require faster methods for evaluating gradient, Jacobian or Hessian operators within the chosen inversion/imaging framework as well as the effective use of anticipated advancements in solvers, pre-conditioners, and multi-grid and multi-level methods that use petascale computing distributed architectures. In addition, coupled inversion will require innovative measures to link different physical models and data (both geophysical and point data) together in a way that honors method- and geometry-dependent differences in resolutions and uncertainties as well as theoretical and field-derived relationships between different physical parameters.

This development would lead to a significant breakthrough in our ability to image properties and processes in the subsurface, make quantitative predictions on flow fields and saturation levels, and obtain information on the properties that affect subsurface flow and transport at unprecedented resolution. This development is conditional on 1) the availability of extensive

and high-quality geophysical datasets over relevant field sites and well described test beds and 2) the establishment of extensive interactions and collaborations between geophysicists and computational scientists. This development will result in tools which can be coupled to advanced pore and continuum scale models in other domain areas (e.g., reactive flow and transport models, mechanical models, and models of biogeochemical processes). It will also lead to tools which would give near real-time, actionable information on subsurface processes in days instead of months to years and, thus, will be useable not only for characterization and monitoring, but even for feedback driven control.

EXECUTIVE SUMMARY

Traditional methods to monitor the lateral and volumetric extent of sequestered CO_2 and subsurface contamination and/or determine changes in the lithology, which control transport, involve the use of instrumented wells. Data from such wells provides an insufficient characterization of relevant subsurface processes and properties within heterogeneous geological media. While the information from boreholes is of very high resolution at a very fine scale in the direction along the borehole, it represents a very limited sampling of the subsurface in other spatial directions. In addition, dense temporal sampling of boreholes for all but the simplest parameters is financially infeasible. As the subsurface typically will have significant lateral and vertical variations in geology and physical properties, major features that affect the storage of CO_2 and contaminant migration and transport will often be completely missed by borehole sampling. In addition, processes will only be detected if they affect the value at the sampling point at the moment of sampling. To resolve this problem, joint geophysical imaging technologies need to be advanced and developed that can take advantage of the increasingly more common time lapse geophysical datasets. Such imaging technologies should be physics based. They should be able to provide high-resolution information on fluid processes and properties in heterogeneous geological media with sufficient resolution and confidence for both regulatory and operational needs.

SUMMARY RESEARCH DIRECTIONS

Flow fields are controlled both by transport pathways (which in turn are controlled by complex geological structures on multiple scale lengths) and changing hydrological boundary conditions. Volumetric characterization of such flow fields must be done in a setting that incorporates all factors that dictate flow and transport as well as all available data. Effective volumetric characterization can only be accomplished effectively using a joint inversion of diverse, high-quality geophysical data (i.e., seismic, electrical, and electromagnetic). Such a joint inversion will require the establishment of physics-based relationships between the parameters obtained through geophysical imaging efforts to relevant hydrological parameters. It will also require the development of novel computational schemes and approaches for such imaging.

SCIENTIFIC COMPUTATIONAL CHALLENGES

Geophysical imaging technologies can provide detailed volumetric information on subsurface properties and processes necessary to provide a detailed picture of factors controlling flow fields and transport. Several factors limit the successful application of geophysical imaging. These include 1) the ability to do true, physics-based simulation of the geophysical signature associated with a certain subsurface distribution of physical properties; 2) the ability to jointly invert for different geophysical datasets; and 3) the ability to translate the results obtained from geophysical imaging into appropriate descriptions of subsurface properties using scaling laws and rock property relationships. In addition, new approaches are needed to geophysical imaging to enhance its temporal resolving power (e.g., a better exploration of time lapse datasets).

An approach that shows significant potential is to use co-located geophysical datasets in a joint geophysical imaging approach. For example, it is known that seismic imaging is very good at mapping reservoir boundaries. However, seismic data does not allow for discrimination of the fluid type (oil, brine or gas) within the reservoir. In contrast, imaging of electromagnetic data has provided for the ability to confidently identify fluid types, however with a spatial resolution that is inferior to that of seismic wavefields. A logical approach would be to combine these methods in a physics-based joint imaging framework. For such a framework to be feasible, the time-to-solution of geophysical inverse methods needs to be decreased by one to two orders of magnitude. This then defines the two primary computational challenges associated with this PRD: development of a physics-based joint imaging framework and the development of tools and methods allowing for much faster solution times.

Joint Imaging Framework

Joint imaging of electromagnetic, electrical, and seismic data has the potential of increasing resolution above with what can now be achieved using either geophysical data by itself. It can allow for the treatment of the geophysical data arising from complex three-dimensional geological environments in a self-consistent manner. A key requirement in enabling this technology is establishment of physics based relationships between electrical conductivity and seismic wave velocities on the geophysical field scale (meters) and fluid properties in three-dimensional heterogeneous media (saturation, porosity, and permeability) on the pore scale (microns). This is of fundamental interest in subsurface science. Knowledge of such relationships could lead to a much better understanding of subsurface fluid, flow, and transport, with critical implications for environmental site characterization and remediation, exploration for oil and gas, and reservoir monitoring for safe sequestration of CO_2. Correlations between electrical and seismic properties are increasingly observed at different scales in co-located experiments; however, separate inversions of associated datasets often lead to unequivocal models. Joint inversion of such data would be a better approach that would allow one to honor the underlying physics. However, novel, robust methods need to be developed for such joint inversions.

There are two emerging approaches to joint three-dimensional inversion, which are underpinned either by the assumption that the electrical and elastic properties of the subsurface are both functions of porosity and fluid saturation (petrophysical approach) or that electrical and seismic measurements are sensing the same underlying geology, which in turn structurally controls the distribution of the petrophysical and flow properties (structural approach). The petrophysical approach is attractive, but its validity in general geological media is constantly questioned because there often is no simple or single relationship for accurately predicting the entire range of effects such as variations in clay content and shape interconnectivity of the pores on the geophysical measurements in many field situations.

Nevertheless, there are clearly cases where a relationship can be established; in the context of reservoir monitoring detailed rock property models are available and such information should be used. Often, the pertinent rock property relationships are not well characterized and in some cases, relationships do not exist. This is particularly so for the vadose or unsaturated zone, which is characterized by chaotic heterogeneous geological deposits that can be porous as well as non-porous. Here a structural approach to joint inversion would be appropriate, where petrophysical information on the geophysical measurement scale could possibly be derived from the correlated geological structure in the electrical and seismic images. This could lead to a better understanding of how transport processes occur in complex geological media (e.g., the vadose zone).

Faster solvers

One major computational obstacle to three-dimensional or four-dimensional geophysical imaging technologies is the time-to-solution. Months of processing time can be required for comprehensive three-dimensional imaging of realistic geophysical datasets. This is exacerbated by the fact that multiple solutions to the imaging problem are required to assess data quality, model uncertainty, and model uniqueness. Faster times to solutions can be achieved by both an increase in available computational resources, as well as potentially with fewer and faster non-linear inversion iterations as well as faster forward solvers. With non-linear geophysical imaging, be it seismic or electromagnetic data or both in joint imaging experiment, hundreds to thousands of forward-model evaluations are typically required. Each of these involves the solution of a partial differential equation describing the physics of the measurement. Many of these solutions are required to fine tune stabilization parameters, quantify data noise, and to assess the unique properties of the image through an appraisal process. A significant improvement in solution speed will translate into more reliable three-dimensional images of the subsurface geology, and will advance the prospect of real-time, three-dimensional earth imaging in the not too distant future, given the anticipated improvements in parallel computer hardware and algorithms.

An example of real-time imaging that has significant potential is imaging ahead of the drill bit in the exploration hydrocarbons. Here, real-time imaging could save the oil and gas industries the billions of dollars that are wasted because of well blowouts and by-passed oil. To address this

problem new pre-conditioning, multi-level methods and faster methods for evaluating gradient, Jacobian, or Hessian operators within the chosen inversion framework will be needed. Research in the physics of seismic and electromagnetic wavefield propagation and that of electrical potential fields will enable development of these fast solution methods. More importantly, these methods will need to be optimized to petaflop distributive computing architectures to reduce times-to-solution from years and months to weeks and days. Extensive interactions and collaborations will be needed between geophysicists and computational scientist to achieve this goal.

SUBSURFACE SCIENCES IMPACT

The main goal of this PRD is to provide enabling technologies for subsurface imaging of diverse geophysical data to provide information on flow fields in heterogeneous geological environments and for verification and monitoring purposes. This is made feasible by the recent advancements in geophysical acquisition systems, and resulting improvements in both data quantity and quality. The development in three-dimensional and four-dimensional joint geophysical imaging proposed here will allow us to make quantitative predictions on flow fields and saturation levels, and the factors that affect subsurface flow and transport at unprecedented resolution in a time frame acceptable to the regulatory, characterization, and exploration process.

COMPUTATIONAL SCIENCE IMPACT

The development of a joint geophysical imaging tool will advance new algorithms for simulation and imaging of multiple physical-coupled phenomena on distributed computing systems at different scales. Coupled simulations on multiple grids will be required for the different types of geophysical field measurements (electromagnetic, electrical and seismic), including the rock property models used to describe the flow fields and fluid phases along with the associated grid transfer operators. To image at the detail envisioned will require the use of dedicated high-performance Petaflop computing resources, where large-scale three-dimensional/four-dimensional geophysical imaging experiments can easily require meshes exceeding 400^3 nodes along with the enormous prerequisite data volumes; ~10's to 100's of Gbytes for high-resolution imaging. Multiple levels of parallelization will be required to solve such a problem, not only over the model space through domain decomposition techniques but also over the data space. This will lead to new parallel algorithms, including, physics-based preconditioners, multi-grid and multi-level methods, and fast schemes for evaluating gradient, Jocobian, and Hessian under the chosen optimization/imaging framework. All together, these anticipated advancements will allow earth and computer scientists to solve such problems in realistic time frame of days to weeks instead of months to years.

TIME FRAME

Substantial progress could be made in 3 to 5 years, with longer term research extending over a ten or more year period.

PORE SCALE AND MICROSCALE COMPUTATIONAL POLYMER AND COLLOIDAL SCIENCE FOR SUBSURFACE APPLICATIONS

David Trebotich, Lawrence Livermore National Laboratory; Paul Meakin, Idaho National Laboratory

ABSTRACT

Simulation-based prediction of the behavior of small particles and polymers in complex pore spaces and fracture apertures is a challenging undertaking because multiple physical processes acting on a wide range of length scales must be taken into account. New multi-scale, multi-physics codes implemented efficiently on petascale computing systems will allow computer simulations to reduce the number of experiments that are needed to develop particle and polymer based materials for subsurface applications. The same approach will also provide the information and insights needed to better predict the behavior of polymers and particles in the subsurface. Eventually pore scale models will be coupled with continuum scale models for a variety of applications including prediction of the fate and transport of toxic particles formed in radioactive waste repositories, and the impact of synthetic or biologically generated particles and polymers on oil recovery.

EXECUTIVE SUMMARY

Computer simulations have been used for more than half a century to develop a better understanding of polymers and colloids. During this time, rapid increases in computing speed and memory capacity and speed have enabled steady progress from simple models such as the self-avoiding random walk model for polymer molecules in a good solvent to the realistic simulations of the folding of protein molecules in aqueous environments. In the past, most simulations of polymers and colloids have focused on the discovery and testing of scaling relationships, fractal dimensions, and other “universal” or generic characteristics and behaviors. The models used for this purpose have usually been based on simple particle shapes and interaction potentials—often selected to offset the effects of limited computing capability. The simulation of protein folding is a notable exception, but a detailed Molecular Dynamics simulation of protein folding, using the fastest computers available today, is possible only for small, rapidly folding, proteins. In addition, most simulations have been carried out using only a single computational method (e.g., Monte Carlo simulation, Molecular Dynamics, dissipative particle dynamics, or Brownian dynamics). Such simulations take into account only a limited subset of the physical processes that play an important role in most physical polymer and colloidal systems. In addition, the application of particle-level and molecular simulations to the behavior of polymers, colloids and nanoparticles in the subsurface has been quite limited.

A new generation of multi-scale, multi-physics (and multi-algorithm) codes that perform well on petascale computing systems will raise computational polymer and colloid science to a new level. This new capability will allow system specific simulations that include all relevant physical process to be applied to subsurface problems and applications. As computing capability grows and improved understanding and experience leads to better algorithms, the accuracy of system-specific simulations will increase, and they will play an increasingly important role in the development of polymers, colloids, and nanoparticles for subsurface applications and in predicting the behavior of anthropogenic and natural polymers and particles in the subsurface. After 10 years, we expect that fully coupled multi-scale, multi-physics methods will allow particle-level simulations to inform larger-scale subsurface flow and transport models in a fully integrated manner. We expect that advances in the application of computational colloid and polymer physics and chemistry to subsurface science and technology will benefit substantially from the efforts made by computational scientists working on a very wide range of commercially and scientifically important applications. We believe that it will be important for the computational subsurface science community to work closely with this larger community.

SUMMARY OF RESEARCH DIRECTION

Although small particles account for only a small part of the solid mass in many subsurface environments (clays and soils are important exceptions), they can play an important role in the subsurface because of their high specific surface area and potential mobility. Nanoparticles (i.e., particles with diameters on the order of nanometers) have unique physical and chemical properties (e.g., increased reactivity) that may have important consequences in the subsurface. In addition, relatively small amounts of clay or other inorganic particulate material may clog faults or borehole surfaces, and biopolymer/biocolloidal systems may clog fractured and porous media that are nutrient rich because of organic contamination or bioremediation processes. Natural polymers, such as polysaccharides, are produced by microorganisms as part of the cell wall and as extracellular and may be used in subsurface applications such as enhanced oil recovery.

Most subsurface simulation codes are based on mathematical models of single phase and/or multi-phase flow and mass transport formulated at one or more continuum scales. In these models, geomaterials with microscale and pore-scale physical and chemical heterogenities and complex pore and/or fracture geometries are treated as homogeneous “effective media,” and both fluid flow and mass transport (including mass transport mediated by colloids, nanoparticles, and polymers) through these complex media are represented by averaged flow and transport equations based on concepts such as Darcy flow, relative permeability, and averaged particle transport “trapping and release rates.” An important advance for current codes would be the capability of modeling particulate transport at the pore scale, and coupling this to continuum-scale models. Such a capability would provide a firm scientific basis for developing technologies for controlling the behavior of micron and submicron-sized particles of concern (e.g., radioactive and/or chemically toxic colloids, bacteria, or viruses) in the subsurface, and would enable a more accurate and reliable assessment of the impact of small particles and

polymers on problems such as contaminant fate and transport and applications such as CO_2 sequestration and enhanced oil recovery.

Specific applications include the fate and transport of uraninite (UO_2) particles produced by microorganisms and the use of polymers as rheology control agents to improve oil recovery (polymers are currently used in acid fracturing reservoir stimulation, and in the future, polymers may be used on a large scale for enhanced oil recovery). Prediction and control of the behavior of colloids and nanoparticles requires a better fundamental understanding of their behavior in the subsurface and the implementation of this improved understanding in software that is used to predict the behavior of subsurface systems on large scales. Both will require new mathematical models and computing systems with capabilities that are substantially greater than those that are now available to the subsurface science community. The accurate simulation of the interactions between particles and polymers and complex (geometrically and chemically heterogeneous) surfaces and pore spaces is an open problem in subsurface simulation. Similarly, effective ways of coupling small-scale simulations of particles and polymers in the subsurface with large-scale simulations for practical applications have not been developed.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Several fundamental research issues arise in the development of a new multi-scale simulation capability for the flow and transport of small particles and polymers in subsurface systems and these issues depend on the nature of the application and the system of interest. For example, the volume fraction of mobile particles involved in the transport of radioactive metals from leaking storage tanks, transfer lines, and disposal sites/repositories may be extremely small. On the other hand, the application of polymers and particles as rheology control agents in applications such as drilling muds or enhanced oil recovery requires much larger volume fractions. Although suspended small particles and dissolved polymer molecules affect fluid viscosity in a way that is well understood and relatively easy to predict when their concentration is small, the large qualitative and quantitative changes in rheology (needed for many practical applications) that occur at higher volume fractions are much harder to understand and predict. For example, solutions containing long-chain polymers behave as strongly non-Newtonian fluids on continuum scales when the polymer concentration is large enough to make the average separation between the centers of mass of adjacent polymer molecules comparable to or smaller than the diameters of the long chain molecules. In some polymer systems, polymer interactions and entanglement can result in the formation of physical gels that have non-zero elastic moduli on time scales that are important in practical applications. Similarly, concentrated particle dispersions may exhibit complex shear-thinning or shear-thickening behavior, and particle-particle interactions can result in the formation of aggregates and gel-like particle networks. The formation and breakdown of these particle assemblies can lead to complex time-dependent, temperature-dependent, and strain-rate-dependent behavior. In addition, essentially monodisperse colloids can form ordered phases that are analogous to atomic or molecular crystals. “Rigid” polymers (polymers with large persistence lengths) can undergo molecular association at low concentrations and at higher concentrations they can form anisotropic phases

and undergo phase separation. These phenomena have a large impact on the rheologies of these systems. The simulation of these phenomena is challenging, in part, because the long-range hydrodynamic interactions between particles and between the chain segments of polymer molecules (both intermolecular and intramolecular interactions) must be considered. In addition, the thermal fluctuations that lead to Brownian motion and shape fluctuations in macromolecules and particle aggregate.

A non-Newtonian continuum rheology model such as a shear-thinning power law or Bingham fluid model is one approach, but more complex models are required for materials whose behavior depends on their past history. The rheology of some polymer solutions could be represented as continuum viscoelastic fluids, but these models present their own difficulties to computing solutions because of instabilities that result from the system being described with mixed mathematical formalisms (e.g., elliptic-parabolic equations). In addition, the ratio between the particle size and the pore size is not always small enough to justify this approach. An appropriate particulate representation coupled to the fluid dynamics is needed to determine the behavior of polymers and colloids. An alternative approach would be to use a dissipative particle dynamics model to simulate both the polymer chains or particles and the fluid in which they are suspended or dissolved. In principle, a small ensemble of polymer molecules and the surrounding fluid could be simulated using classical molecular dynamics techniques. However, experience with applications such as protein folding indicates that this approach would be beyond the capabilities of petascale computing systems unless the polymer molecular weights were small. It is well known, however, that the essential behavior of polymers and colloidal-sized particles can be captured though coarse-grained models such as the Kramers' "bead-rod" representation (Kramers 1946), or the coarser Rouse "bead-spring" representation (Rouse 1953). In these models, the polymer is represented as discrete point masses (the beads) connected by rigid rods or entropic springs. The dynamics of polymers in solution is then modeled with a combination of stochastic (Brownian) forces and Stokes-law viscous drag forces (related through the fluctuation-dissipation theorem). Hydrodynamic coupling between the moving particles and the fluid is necessary to describe these systems accurately (see Figure 1). In addition, biological macromolecules are charged and chemically active, and interact through screened Coulombic interactions. These physical effects are characterized by intra-polymer, inter-polymer, and polymer-wall interaction potentials, which may be long-ranged (see Figure 2). Because of the small time scales associated with the dynamics of small particles and macromolecule chain segments, small time steps are currently required to obtain accurate results from models that couple particle dynamics and fluid flow. New multi-scale techniques for coupling will be necessary to enable long time simulations of particulate transport coupled to continuum models for subsurface applications.

The effects of moving contact lines and fluid-fluid interfaces on small particles are potentially important issues that have received little attention from a theoretical/computational point of view. Particles may be attracted or repelled from fluid-fluid interfaces, and at small scales, the stresses generated by capillary forces can be large. Consequently, moving contact lines could either mobilize or trap particles, depending on the nature of the particles, fluids and surfaces

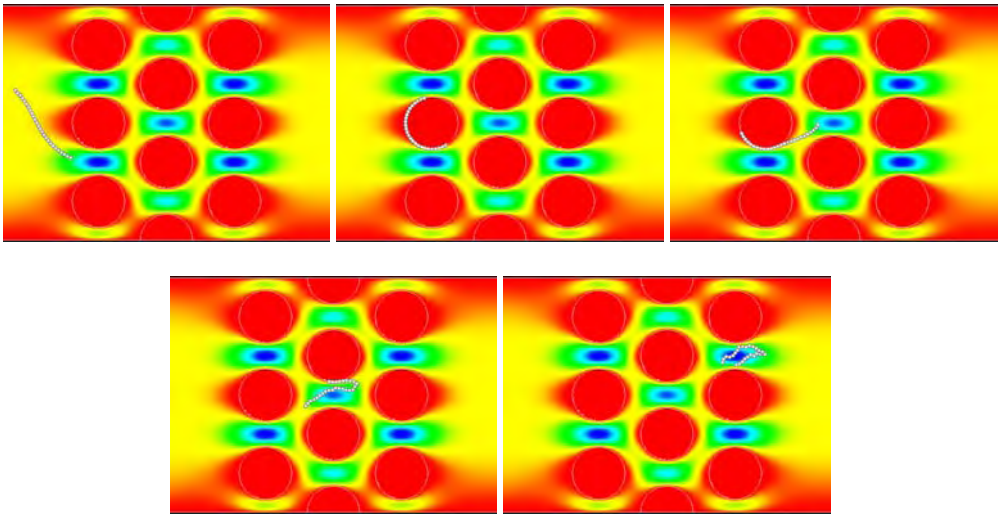


Figure 1. Time sequence of bead-rod polymer representation flowing in a microarray. Polymer enters from left in (a), then wraps around pillar in (b), is loosened by hydrodynamic and Brownian forces in (c) and is swept out of the chamber by the flow field in (d) and (e). Color map indicates underlying flow field. [Trebotich, D. et al. A Tightly Coupled Particle-Fluid Model For DNA Laden Flows in Complex Microscale Geometries. Computational Fluid and Solid Mechanics, p. 1018-1022 (2005)].

involved and the direction of motion and velocity of the contact line. In recent years, interest in the formation of particle patterns during the drying of thin liquid films and drops has increased (Deegan 1997), but this work has been mainly experimental.

The interactions between colloidal particles and solid surfaces are generally modeled using the DLVO (Derjaguin, Verwey, Landau and Overbeek) theory based on a quantum-electrodynamic treatment of the van der Waals interactions and a simplified model for the partially screened electrostatic interactions. However, this theory has had mixed success. It is based on the idea that the van der Waals and electrostatic interactions are separable, and as a consequence, it is not reliable at the ionic strengths found in biological systems (Bostrom 2001) and brine aquifers. In addition, the complications caused by the adsorption and desorption of ions on solid surfaces and/or the mobility of ions on solid surfaces are not taken into account. The calculation of the interactions between particles with complex shapes and/or surfaces that are rough or chemically heterogeneous is a difficult undertaking. New theoretical approaches and large scale computer simulations are needed to develop a better understanding of particle-surface interactions and to develop the capability of predicting particle-surface interactions for realistic subsurface systems. Similar advances are needed to better understand and predict particle-particle interactions. Additional complications arise because the molecular structure of liquids adjacent to solid surfaces is not the same as that of bulk liquid. The effect of the ordered liquid structure adjacent to liquid surfaces has been measured in sensitive measurements of the interactions between atomically smooth surfaces (Horn 1981), and this effect must be taken into account. In the presence of polymers, additional physics comes into play. Polymers may become adsorbed on

and bridge between solid surfaces, and they may also lead to repulsive interactions caused primarily by entropy effects (see Figure 2). Again the kinetics of (full or partial) adsorption and desorption, as well as migration on surfaces can have important effects. Development of a better understanding of these "colloidal" interactions and the capability of modeling them with sufficient accuracy would be important in many areas of science and technology, including basic and applied subsurface science. To predict the behavior of particles suspended in liquids flowing through porous and fractured media, it is also important to include hydrodynamic effects such as lift and drag forces. As a result of these interactions, and the size of colloidal particles, they can move more rapidly than dissolved substances through porous media, providing that they do not become trapped or adsorbed. Finally, the effects of thermal fluctuations must be taken into account.

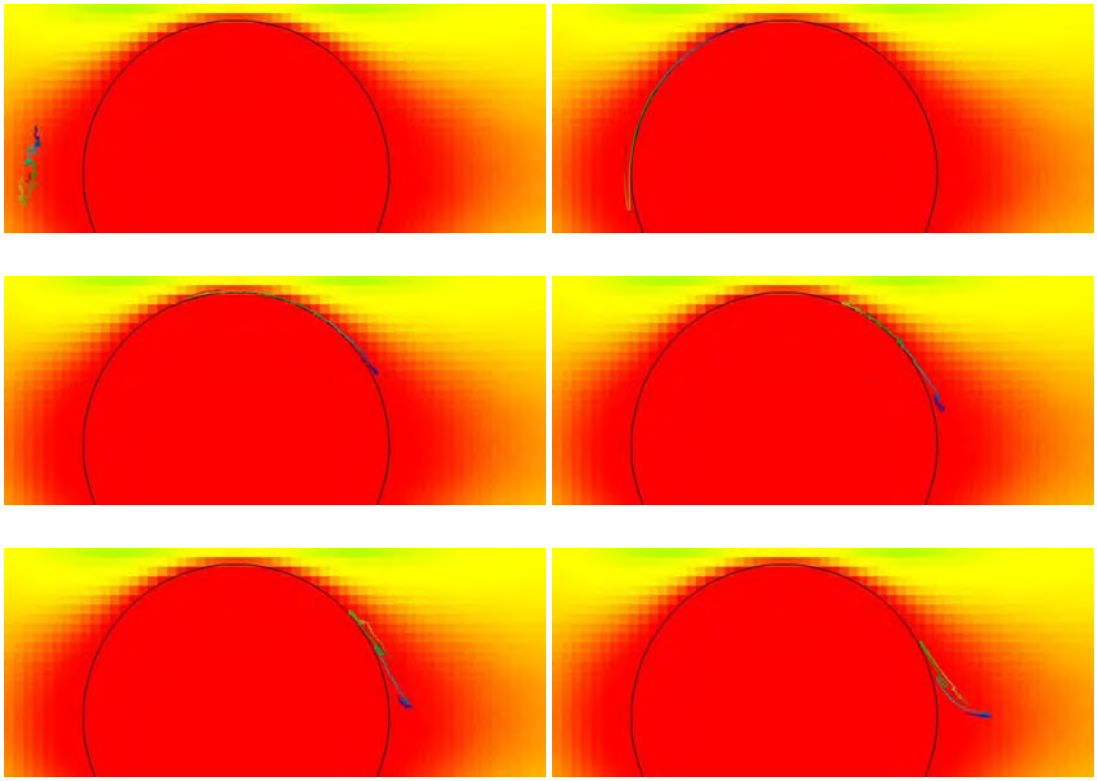


Figure 2. Time sequence of intra-polymer and polymer-structure interactions using repulsive potentials near a single post in microarray: a) Nearly entangled polymer; b) Slowing near boundary and short-range interaction; c) Acceleration around pillar caused by Brownian perturbation into velocity gradient; d) More acceleration of the tail and slowing of the head in stagnation region in wake of pillar; e) Accelerated tail catching up with stagnated head; and f) Re-entanglement in wake. Trebotich, D. et al. "A Hard Constraint Algorithm to Model Particle Interactions in DNA-laden Flows", to appear in Nanoscale and Microscale Thermophysical Engineering (2007).

The development of better ways to predict the behavior of polymers and colloids in the subsurface is an important scientific challenge with important practical applications. Because

atomic scale, particle scale (typically nanometers to micrometers), and continuum physics all play a role, this will require computing capabilities that have not previously been available. Algorithms for the behavior of particles and colloids in the subsurface must be designed with the goal of facilitating multi-scale coupling. Micro-scale simulation results can be used to obtain parameters that serve as input for larger-scale models, but this requires substantial human time and intervention. The long-term grand challenge in this area of multi-scale mathematics would be to develop adaptive methods for coupling different models and scales within one framework. Adaptive algorithm refinement is an approach to coupling between different models and scales. To represent multiple length and time scales in this way, it is necessary to introduce adaptive mesh refinement (AMR) methods (Garcia 1999). This technique focuses the computing power in the areas of importance in a domain where it is necessary to resolve highly localized phenomena. AMR is a mature technology for problems without geometry, with a variety of implementations and applications for various nonlinear combinations of elliptic, parabolic and hyperbolic PDEs. AMR has also been successfully coupled to embedded boundary/volume-of-fluid algorithms to address surface geometry and interfaces. Based on an AMR framework, adaptive algorithm refinement may be used to facilitate better multi-scale coupling between modeling scales, and in particular between particles and the surrounding fluid.

Additional mathematical issues arise in these problems. The numerical accuracy of particle methods is difficult to determine in coupled fluid-particle systems. Accurate resolution and gridding of the complex geometry of typical pore scale porous media is required for realistic quantitative experimental validation studies (e.g., comparison with fluid-sensitive high resolution x-ray tomography or NMR tomography experiments). The resolution of interfaces for multi-phase flow simulations must be conservative and accurate, with well-posed boundary conditions. All of these issues and features should be approached numerically in a way that complements other algorithmic components of the entire model. For example, gridding based on embedded boundary/volume-of-fluid representations can facilitate adaptive mesh refinement and, thus, multi-scale modeling, leading to enhanced parallel scalability. This approach to irregular boundaries can work for both fixed surfaces and free interfaces. The volume of fluid representation is a natural description of multi-phase flows with resolved interfaces. The application of other ways of representing interfaces and tracking their dynamics, such as level set, phase field, and Lagrangian methods in combination with adaptive mesh refinement also need to be investigated. Finally, the meshing technique used for gridding pore scale geometry must be amenable to surface extraction from experimentally derived image data obtained from packed bed columns of porous media, and computational fluid dynamics codes should be able to directly simulate flows on the meshes obtained from tomographic imagery without loss of geometric detail in a fast and accurate manner.

COMPUTATIONAL SCIENCE IMPACT

New multi-scale, particle-fluid methods, which currently do not exist, will have a wide range of applications in other areas of science and technology. The fundamental issues that must be addressed are at the core of colloid and interface science, and they are central to polymer science,

soft condensed matter physics, and the rheo-mechanical behavior of biologic systems at the subcellular, cellular, and multicellular levels. The potential practical applications outside the subsurface area are enormous. They include coating and printing technology, micro-electromechanical systems (MEMS), health science, and water purification, to name just a few.

SUBSURFACE SCIENCE IMPACT

An important application of this capability would be to quantitatively assess the potential for micron and submicron-sized particles (such as colloids, prions, bacteria, or viruses) generated in or placed in the subsurface to move in groundwater and vadose zone systems, to be physically or chemically intercepted in host geologic media, and to become inactivated. Specific applications include the fate and transport of inorganically or biologically generated particles that are composed of radioactive and/or chemically toxic substances or can act as carriers for toxic substances and the use of polymers and colloids by the oil industry to increase oil recovery. For example, improved models would enable us to better predict how pores and fractures become clogged and how chemical microenvironments are formed. The latter is where unique geochemical and biogeochemical processes, which may dominate the bulk behavior, can occur.

TIME FRAME

Decades of research by a large community of colloid and polymer scientists, motivated by a plethora of commercially important applications and scientific challenges has provided a wealth of information and understanding that can be applied to subsurface problems and applications. New experimental techniques that allow the forces associated with the interactions between particles and surfaces to be measured directly have opened up new research opportunities. However, the inherently multi-scale (both temporal and spatial) nature of polymer and colloid systems has limited the progress that could be made using a computational approach. Petascale computing resources combined with the development of better algorithms for multi-scale problems would allow important advances to be made. Polymer and colloid science presents a hierarchy of challenges with a wide range of difficulties. Within 5 years, we expect that petascale computing and better algorithms will allow us to simulate the transport of colloids (both solid particles and polymers) fully coupled with both single-phase and multi-phase fluid flow through complex pore spaces and fracture apertures over the full range of particle volume fractions. We expect that these simulations will include particles with complex shapes, and include phenomena such as clogging of pore spaces. Within the same time scale, we expect that computational methods will be capable of providing guidance to the design of colloidal materials for specific subsurface applications and the development of surfactants and other surface modifying substances to control the behavior of colloids in the subsurface. Within a decade, we expect that it will be possible to simulate the behavior of solid particles and polymers with a wide range of properties on the multi-pore level and that the results of these simulations will play an important role in the development of better constitutive equations for field scale models. After 10 years, we expect that fully coupled multi-scale, multi-physics methods will allow particle level simulations to inform larger scale subsurface flow and transport models in a fully

integrated manner. We expect that advances in the application of computational colloid and polymer physics and chemistry to subsurface science and technology will benefit substantially from the efforts made by computational scientists working on a very wide range of commercially and scientifically important applications. We believe that it will be important for the computational subsurface science community to work closely with this larger community.

REFERENCES

Bostrom, M.; Williams, D. R. M., and Ninham, B. (2001) Specific ion effects: Why DLVO theory fails for biology and colloidal systems, *Phys. Rev. Lett.*, **87**: 168103

Deegan R. D. *et al*. (1997) Capillary flow as the cause of ring stains from dried liquid drops, *Nature* **389**: 827

Garcia, A.L., Bell, J. B., Crutchfield, W. Y., and Alder, B. J., (1999). Adaptive Mesh and Algorithm Refinement. *J. Comput. Phys*., **154**: 134.

Horn, R. G. and Israelachvili, J. N. (1981) Direct measurement of structural forces between 2 surfaces in a non-polar liquid, *J. Chem. Phys.*, **75**: 1400

Kramers, H.A., (1946) The behavior of macromolecules in inhomogeneous flow. *J. Comput. Phys.*, **14**: 415.

Rouse, P. (1953) A theory of the linear viscoelastic properties of dilute solutions of coiling polymers, *J. Chem. Phys*., **21**: 1272.

IMPROVED MEDIA PARAMETERIZATION AND RECONSTRUCTION

Mihai Anitescu, Argonne National Laboratory; George Ostrouchov, Oak Ridge National Laboratory; Laura Pyrak-Nolte, Purdue University

ABSTRACT

The present level of understanding and modeling of heterogeneous media parameterization and reconstruction is insufficient for predicting subsurface processes required for answering critical energy policy questions at a reasonable risk level. The need for accurately simulating physical processes over an unprecedented range of space and time scales requires improved media parameterizations that faithfully represent the background of the coupled physical processes that occur in the subsurface and allow for the efficient characterization of the evolution of its structure. This will allow for long-term predictions of subsurface behavior at a much lower uncertainty level than currently available and will enable technical advances in CO_2 sequestration and the extraction of oil/gas from fractured carbonate reservoirs.

EXECUTIVE SUMMARY

Critical geologic features, such as mechanical discontinuities, for any subsurface reservoir/ repository exhibit heterogeneity that spans multiple scales. Mechanical discontinuities range in scale from a few microns (micro-cracks) to centimeters-meters (fractures, joints) to kilometers (faults). They can occur singly or in sets producing heterogeneity from the pore scale to the reservoir scale. This heterogeneity will vary temporally because of participation in the hydrogeologic, geochemical, biogeochemical and tectonic cycles, in addition to any human activities/interference.

The prediction of the effect of subsurface media heterogeneity on physical phenomena at the decision level scale depends crucially on the parameterization of heterogeneity. In addition, the long time interval simulation of configurations with varying media structures needs the capability of creating samples of media with prescribed heterogeneity parameters (the inverse problem). A priority research direction is the development and validation of efficient media parameterizations that correctly capture the interaction between the various subsurface physical processes at all the relevant scales as well as the development of algorithms to solve the inverse problem. Such parameterizations must be amenable to uncertainty quantification to address the pervasive risk assessment questions that appear in mission-directed subsurface investigations. Statistical uncertainty quantification that has found its way into subsurface modeling usually treats the computational model as a black box. New approaches must bring uncertainty quantification into the model specification and result in computational models that fully incorporate uncertainty propagation rather than compute it through replication or approximately as an afterthought.

An essential component of this research direction is the experimental data-based validation of the models and algorithms that will result from this research. Exciting opportunities result from the fact at the core level (~10 cm), a wealth of experimental data can be generated that can be used for the validation of the relationship between parameterizations, microstructure, and subsurface physical phenomena.

This work will require substantial access to leadership class computational resources to solve the stochastic optimization, uncertainty quantification, parameterization selection, and design of experiments problems that arise in resolving the problem of media parameterization and reconstruction for computational subsurface science.

SUMMARY OF RESEARCH DIRECTION

To create representations of the type of heterogeneous media that is encountered in subsurface applications, we propose a multi-pronged approach that includes the discovery of novel parameterizations, the development of advanced stochastic optimization algorithms for inverse media problems, embedded uncertainty quantification in the structural description, and validation of the algorithms and methods by comparison with laboratory data.

Parameterization of heterogeneous media

The principal challenge of up-scaling techniques for complex heterogeneous media is to determine which properties on the micro-scale can be used to predict macroscopic hydraulic, mechanical and seismic properties at core- and field-scales. In turn, the up-scaling is crucially based on the parameterization of the microstructure, which must at the same time capture the correct characteristics and yet be sufficiently easy to manipulate for computations. Previous studies in the relationship between the description of random heterogeneous media and its effective physical properties do exist (Long and Hestir 1990; Torquato 2002). Nonetheless, these studies needed different parameterizations of the microstructure to characterize multi-phase flow and mechanical behavior, and the question of parameterizations that correctly represent the multi-physics of subsurface science is unanswered as of yet. This question is crucial as shown in Figure 1, which presents two numerically simulated fractures with the same porosity but with aperture distributions that have different spatial correlations. The absolute permeability (Figure 1c) of the spatially correlated fracture (Figure 1a) is twice that of the random distribution of apertures (Figure 1b). In addition, the relative permeability of gas and water (Figure 1d) for these two fractures exhibit different gas flow initiation saturations as well as residual water saturations. Therefore porosity alone will not suffice to characterize the hydraulic response. This example shows that appropriate parameterization of complex structural media and processes is key for developing predictive models that take into account the highly complex structural

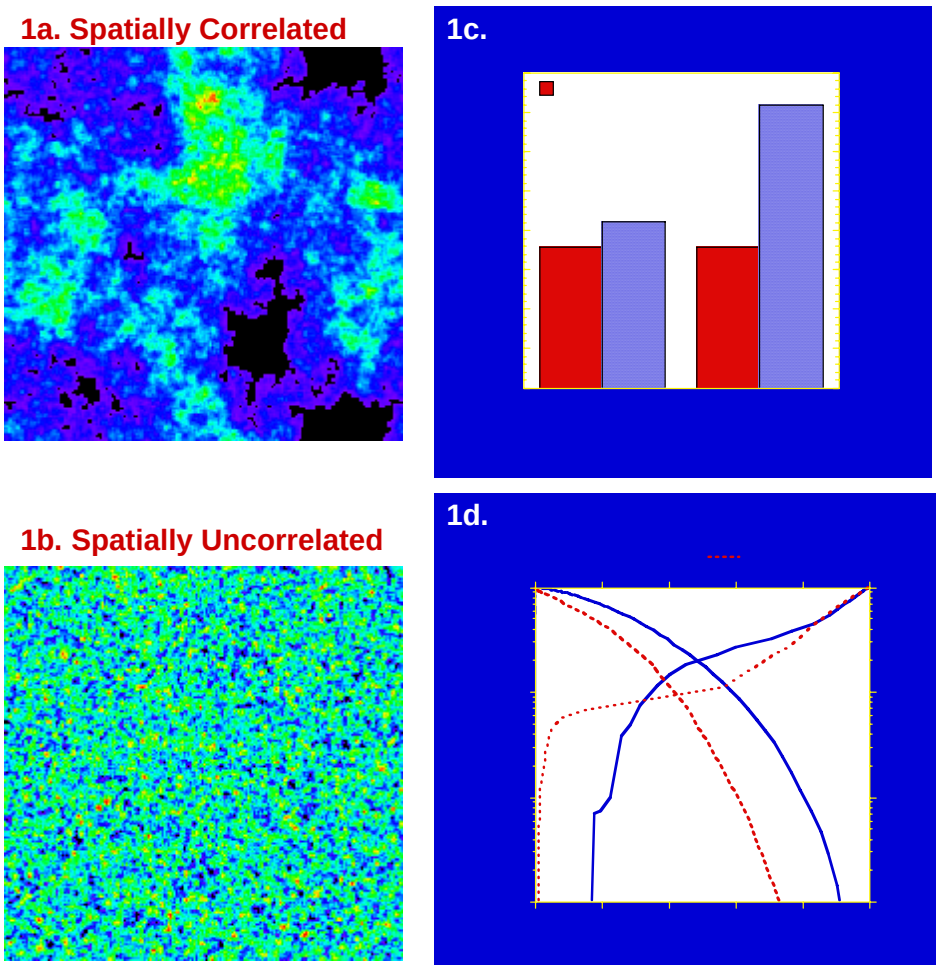


Figure 1. a) Spatially correlated and b) spatially uncorrelated fracture aperture distributions with the color representing the size of the aperture (red ~ large apertures, green-blue smaller apertures and black regions represent the contact area between the two fracture surfaces. c) The porosity, absolute permeability and d) the relative permeability of gas and water for the aperture distributions shown in 1a) and 1b).

components of the medium as well as the complex dynamic processes that occur at varying scales.

On the other hand, n-point probability functions, which may be complete representations of the heterogeneous media (Torquato 2002), are virtually intractable entities to work with, especially on the multiple scales needed in subsurface computations. Therefore, advanced parameterizations that are amenable to computations need to be developed. In the medium term, such parameterizations will likely appear by combinations of classical statistics. Although such combinations have not been considered to date, there are no conceptual obstacles towards using them. Once such combinations have been established, mechanical and hydraulical simulations will need to be conducted on multiple samples of media with a given set of parameters to establish whether these parameters faithfully characterize the effective behavior of the medium

with respect to subsurface phenomena. This investigation will require the use of massive computational power as simulations of subsurface behavior on representative domains are carried out for a given microstructure for multiple samples of microstructure.

A case to which particular attention needs to be paid is the description of the properties of media with fractures. To the best of our knowledge, spatial correlations in individual fractures within a network have only been used in one numerical model (Pyrak-Nolte and Lumsdaine 1996) for multi-phase flow through a fracture network. With this model, Morris (1999a, b) found that laboratory measurements of porosity and absolute water permeability were insufficient to predict the relative permeability behavior of a fracture network and that accurate prediction of relative permeability requires knowledge of the spatial and size distribution of apertures within the fracture network. Any model for predicting two-phase immiscible flow through a fractured subsurface reservoir (e.g., for enhanced oil recovery, contaminant transport, etc.) will be inaccurate if pore-scale structural information is not correctly parameterized for phenomenological modeling.

Computational inverse media problems

The inverse media problem (the media reconstruction problem, or the media sampling problem; Torquato 2002) is the one of creating a heterogeneous medium with a given set of spatial statistics and parameters. While the area contains significantly more open problems compared to the many of the other inverse problems in the physical sciences, recent successes of stochastic optimization approaches have opened new directions (Rintoul and Torquato 1997). In these approaches, a merit function that matches the sought statistics to the ones of a candidate media is minimized by doing a random search in a well-chosen subset of the coordinates. Their success needs to be enhanced both in terms of scope and quality, because currently not all simple media has been successfully been reconstructed from statistics, which raises further challenges for more complicated media such as the one with mechanical discontinuities. The effect of the choice of various merit functions and the relative scaling of the different statistics considered needs to be further investigated as a means to extend the inversion capabilities to structures that have currently resisted it. A very large scale, parallel, variant of the stochastic optimization algorithm that requires minimal synchronicity and works on the modern architectures with moderate memory per processor needs to be developed. An area of fundamental investigation is deciding whether a certain choice of values of the statistics is not realizable and related to it whether the stochastic optimization algorithm is converging.

Validation by experimental data

The development of any new parameterization and statistical methods for the description of complex structural media and processes will require validation. Accessing structural information on the field scale is difficult, often spatially sparse and temporally infrequent with the conditions of the subsurface system often unknown or in a state of flux. Laboratory experiments, on the other hand, can be designed to access the data needed under controlled conditions

(e.g., pressures, temperature, material properties, fluid) and can be monitored over time to determine how the structure/system is altered. For example, the principal challenge of up-scaling techniques for multi-phase fluid dynamics in porous media is to determine which properties on the micro-scale can be used to predict macroscopic flow and the spatial distribution of fluid phases at core- and field-scales. First-principles theoretical formulations over the past decade have been derived from volume averaging theorems in which microscopic interfacial behavior is explicitly incorporated. These theories have proposed that interfacial area per volume directly affects macroscopic behavior, and that this variable may govern the observed hysteresis in the capillary pressure-saturation relationship. Direct visualization of interfaces with high resolution is only possible on the laboratory scale. Transparent two-dimensional porous media (micro-models) with a resolution of 0.6 microns per pixel edge length and total sample length of 600 microns have been used to validate the numerical model (Cheng et al. 2004a) used for flow calculations by Morris et al (1999b) and to enable visualization and quantification of the behavior of interfacial geometry while concurrently measuring fluid pressures (globally and locally) and fluid flow (Cheng et al 2004b) to test volume averaging theorem. Thus, laboratory techniques provide the best potential for validating numerical parameterization and statistical techniques for complex media over a range of length scales.

We propose the acquisition of explicit datasets connecting structural transformation on the micro-scale data with macroscopic properties (e.g., absolute permeability, stress, etc.). Data collection should be guided by statistical design of experiments (see Morris 2005 for references) to maximize efficiency and uncertainty reduction. The proposed combination of experimental methods and development of new statistical and parameterization techniques will make it possible for the first time to answer the principal question of which microscopic properties are most useful for predicting macroscopic properties of a structurally complex medium.

Statistical methods

The selection of parameterization for an interface between two physical phenomena, comes down to determining what variability in one phenomenon matters to the other phenomenon. This is a difficult problem in high dimensional settings when nonlinearities can undermine methods like canonical correlation. Recent developments in regression dimension reduction, known as central subspace methods, provide low dimensional graphical representations of (possibly non-linear) interfaces between a large set of explanatory quantities and a small set of responses (Yin and Cook 2006). These methods are computationally expensive but may be combined with fast dimension reduction methods (Ostrouchov and Samatova 2005). Such methodology would bring effective homogenization to mathematically intractable problems through statistical computation.

Discretization and simulation typically takes place in a physical representation of subsurface structure, where some process is propagated through the representation. Statistical uncertainty probability distributions on the resulting time-varying discretized field, if represented in the original physical space, are very high-dimensional (each grid point is a dimension) and

computationally intractable entities. Because of a high degree of spatial correlation between grid points, the intrinsic dimensionality (represented by a small number of coherent structures (e.g., the orthogonal vectors basis which arises from Karhunen Loeve decomposition of spatial processes) of this high-dimensional space is usually several orders of magnitude lower. Breaking away from the physical space and operating in a coherent structure space could not only drastically reduce the simulation time but also render complete uncertainty quantification tractable. The main cost of such a simulation paradigm change is the modification of physical space interaction physics into coherent structure interaction physics. However, the benefits are potentially immense because we compute only what matters and we carry full uncertainty quantification that can answer any regulatory probability questions.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Accurate descriptions are required for predictive models because mechanical discontinuities can act as "rapid" flow paths that might compromise the hydraulic and mechanical integrity of a reservoir/repository or enhance fluid recovery. This requires parameterizations that encapsulate many length scales (micrometers to kilometers), which have not been developed to date. For example, in a fracture, chemical interactions between pore fluids and the fracture walls occur locally on the sub-micron scale but affect fluid flow and wave propagation occurring along the length of the fracture, a length scale of hundreds of microns to meters or more. For a fracture partially saturated with gas and water, not only the intrinsic lengths of the fracture geometry affect the distribution of these two fluid phases but also by time-dependent processes that alter the local capillary pressure that controls the phase distribution. This issue is a part of the broader challenge of determining the appropriate structural statistics (parameterizations) for non-isotropic, heterogeneous random media that do not present scale separation that has not been resolved to date, and whose resolution will benefit areas far beyond subsurface science.

Reconstruction of non-isotropic medium (the inverse problem) is a crucial challenge for long term predictions involving time-varying structure (Wawersik et al. 2000). It is a problem that lies at the intersection of discrete geometry, three dimensional random processes, finite elements and high-performance computing. Discrete geometry problems are famously difficult to approach, as was to case for the Kepler conjecture stated in 1611 and completely proved only in 2005, using a mixture of global optimization, linear programming, and interval arithmetic (Hales 2005). Nonetheless, the importance of the topic for subsurface computations makes its pursuit an imperative, also recommended by the unprecedented availability of high-performance computing resources that will allow the development of novel stochastic optimization algorithms.

The overarching challenge is the one of providing an integrated computational framework that supports the mathematical and statistical tools and provides design of experiment and parameterization solutions for heterogeneous media structure in subsurface science. It will allow the resolution of the critical stochastic optimization, uncertainty quantification, parameterization selection, and design of experiments problems that arise in the media parameterization and

inversion for computational subsurface on the space and time scales required by the DOE mission. These processing-intensive problems are critical for further progress and require unprecedented access to leadership-class computing facilities.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Progress in parameterization and reconstruction of microstructure will result in progress for all computational characterizations and predictions of the behavior of random media. Beyond contributions to subsurface sciences, such techniques will provide new impetus to critical applications in biology, medicine, and chemical processing, among others.

POTENTIAL IMPACT ON THE SUBSURFACE SCIENCES

The proposed research will seek to answer several important questions:

- Are there a minimum number of parameters for spatially complex fracture network that enables phenomenological modeling of the potential site response?
- How does one establish and reduce the uncertainty related to this time and spatially varying fracture network?
- How does one reconstruct a media from the small data-intensive laboratory scale and the sparse data field scale to produce a model that will accurately predict reservoir scale behavior?

The results of the proposed research will have a timely impact on many of the subsurface challenges that are of critical important to energy production, carbon sequestration, water safety and radioactive waste isolation. Sequestering/extracting fluids from the subsurface can alter the underground environment by changing the local stress distribution as well as the local hydrogeology and geochemistry of a site. The site selection process for a subsurface sequestration reservoir or the licensing of a radioactive waste repository requires an understanding of the relationship among the hydraulic, mechanical, and seismic properties of a reservoir, as well as the effect of multi-scale, time-evolving media heterogeneity on this relationship. The same is true for the prediction and design of remediation strategies for contaminated groundwater in discontinuous media. The improved parameterization will aid in determining pumping designs or bio-remediation techniques by determining the most likely flow paths. The research will aid the licensing process of nuclear waste repositories by providing spatial distributions and the evolution of discontinuities with quantified uncertainties. Without a method of describing complex media, the ability to predict long-term behavior (million years) of isolation with confidence will be hindered. Over such time scales, the evolution (i.e., opening, closing, infilling, formation) of discontinuities represents a potential for catastrophic failure of an isolation site, either by compromising the hydraulic integrity of the site or through geo-mechanical failure such as slippage along faults.

TIME FRAME

Within 3 to 6 years, the results of the proposed research would generate appropriate parameterizations and inverse problems algorithms and tested them on laboratory data.

Within 10 years, the results of the proposed PRD research will enable improved sequestration of CO_2 and improve extraction of oil/gas from fractured carbonate reservoirs (which make up 40 percent to 60 percent of some corporate reservoirs) by designing isolation and recovery strategies that take advantage of the multi-scale structural heterogeneity and predicting how the structural heterogeneity is affected by stress changes and fluid extraction during production.

REFERENCES

Barton, C.C., and P.R. LaPointe, (1995). *Fractals in Petroleum Geology and Earth Processes*, Plenum Press.

Cheng, J.-T., J. P. Morris, J. Tran, A. Lumsdaine, N. J. Giordano, D. D. Nolte and L. J. Pyrak Nolte, (2004) "Single-phase flow in a rock fracture: micro-model experiments and network flow simulation, Int. J. Rock Mech. and Min. Sci.", **41/4**, 687-693

Cheng, J.-T.; Pyrak-Nolte, Laura J.; Nolte, David D.; Giordano, Nicholas J., (2004) "Linking pressure and saturation through interfacial areas in porous media", Geophys. Res. Lett., **31**, No. 8, L08502 10.1029/2003GL019282, 22

Hales, T. C (2005) "A Proof of the Kepler Conjecture." *Ann. Math.* **162**, 1065-1185, 2005.

Rintoul, M.D. and Torquato, S. (1997) "Reconstruction of the structure of dispersions". *J. Colloid Interface Sci*. **186**, 467-476.

Long, J.C.S., and K. Hestir, (1990). "Permeability of random two-dimensional fracture networks", *J. Geophys. Res*., **95**, B13, 21,256.

Morris, M.D. (2000) "Three "Technometrics" Experimental Design Classics" *Technometrics* **42**, pp. 26-27.

Morris, J. P., N. J. Giordano, J.-T. Cheng, J. Tran, A. Lumsdaine & L. J. Pyrak-Nolte, (1999b), "Fracture Geometry and Relative Permeabilities: Application to Multiphase Flow through Coal", *the International Colabed Methane Symposium held at the University of Alabama, Tuscaloosa, Alabama, May 6,1999*, paper number 9943, 377-388.

Morris, J. P. & *L. J. Pyrak-Nolte*, (1999a) "Fracture Network Geometry and Relative Permeability", *Proceedings of the Vail Rocks 1999 the 37th U.S. Rock Mechanics Syposium, Vail, Co., June 6-9, 1999*.

Ostrouchov, G. and Samatova, N.S. (2005), On FastMap and the convex hull of multivariate data: toward fast and robust dimension reduction. *IEEE T. Pattern Anal. And Machine Intelligence* **27**:1340-1343.

Pyrak-Nolte, L.J. and A. Lumsdaine, (1996) "Three-Dimensional Simulator of Fracture Networks in Coal", *Topical Report, Gas Research Institute,* GRI-96/233.

Torquato, S.(2002) *Random Heterogeneous Materials: Microstructure and Macroscopic properties*. Springer Verlag, New York.

Wawersik, W. R., Rudnicki, J. W., Orr, F. M., Pyrak-Nolte, L. J., Wilson, J. L., Ortoleva, P. J., Dove, P., Richter, F., Logan, J.. M., Wong, T.-F., Harris, J. H., and N. R. Warpinski, (2001) "Terrestrial Sequestration of CO2 - An Assessment of Research Needs", *Pure and Applied Geophysics* **43**, 97-177,.

Yin, X. and Cook, R.D. (2006), "Dimension reduction via marginal high moments in regression", *Statistics & Probability Letters*, **76 (4)**: 393-400.

MODELING PROCESSES AT THE PORE SCALE

Paul Meakin, Idaho National Laboratory; Alexandre Tartakovsky, Pacific Northwest National Laboratory; Russ Detwiler, Lawrence Livermore National Laboratory

ABSTRACT

Pore-scale modeling has been used for several decades to simulate single-phase and multi-phase fluid flow in fractured and porous media, and more recently, these models have been coupled with simple models for precipitation, dissolution, and the transport of dissolved substances. However, improved codes on more capable computing systems are needed to simulate the behavior of reactive fluids with a wide range of properties under the full range of conditions relevant to subsurface applications. Important advances could be made by efficiently implementing existing methods on petascale computing systems. However many important applications will also require the development of better single-physics algorithms and new codes that couple key physical and chemical processes on a wide range of length and time scales. Eventually, pore-scale models will be coupled with atomistic and/or continuum scale models on computing systems with capabilities that substantially exceed those of current systems. As an interim solution, pore-scale simulations will be used to provide model parameters, and to develop better constitutive equations, for field-scale models.

SUMMARY

In recent years, there has been a resurgence of interest in the application of pore-network models to a variety of subsurface problems and applications. Methods that have been developed to simulate multi-phase fluid flow by the astrophysics, statistical physics, and computational fluid dynamics communities have been modified and applied to the behavior of fluids in the subsurface. In addition, recent advances in lattice Boltzmann methods have substantially expanded the range of potential applications to subsurface processes. Despite these advances, we currently do not have algorithms that can be applied to the full range of fluids and conditions that are important in subsurface systems. In particular, large viscosity, density, and compressibility ratios are a challenge for most methods. In addition, some promising methods such a smoothed particle hydrodynamics and dissipative particle dynamics have been applied only on quite small scales (a small number of pores or fractures). Petascale computing is needed to enable the application of these methods (and other methods based on “first principles” approaches) to systems on the scale of those currently simulated using more empirical and approximate pore network models. For the first time, this will enable the gap between pore-scale and continuum-scale models to be closed using accurate methods based on first principles (at least for those systems for which the concept of a representative elementary volume [REV] is valid and the application of traditional continuum modeling approaches can be justified).

In some cases, information from or coupling with atomistic and mesoscale (i.e., between the atomistic scales and the scales on which fluids can be represented by continuum fields)

simulations will be required, while in other cases, pore-scale models will be based primarily on experimental information such as fluid viscosities and surface tensions. For example, the coupling between fluid-flow and fluid-fluid-solid contact line dynamics depends on atomistic and mesoscale physics, and complex evolving sub pore-scale morphologies (and the associated changes in surface areas and reactivities), which can change rapidly as geosystems are forced far from equilibrium must be taken into account in many applications. Although we expect that with the right resources, critical advances can be made during the next 5 to 10 years, challenges requiring new ideas, new methods and new, much more capable, computing systems will remain.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Existing field-scale models rely upon empirically derived constitutive relationships that often do not adequately represent underlying physical and chemical processes in porous and/or fractured media. Thus, the ability of these models to make scientifically defensible predictions across the ranges of spatial and temporal scales relevant to many applied problems is inherently limited. This has led to significant errors in estimates of contaminant migration and poor predictions of multi-phase flow processes in energy production applications. Similarly, large errors and uncertainties in the assessment of proposed carbon sequestration technologies are expected if currently available field-scale models are used. Pore-scale models (defined here as models for systems that consist of a number of pores—small or large—and include processes that take place within individual pores) present opportunities to systematically incorporate fundamental physical, chemical, and biological mechanisms into quantitative descriptions of complex/coupled processes such as reactive transport and/or multi-phase fluid flow.

Currently used pore-scale modeling techniques include both grid-based methods such as finite element and finite difference codes, particle-based methods such as lattice-Boltzmann simulations and particle methods such as lattice-gas models, smoothed-particle hydrodynamics (SPH), and dissipative particle dynamics (DPD). These different approaches involve inherent challenges and limitations that must be addressed before they can be used quantitatively to predict complex, non-linear, often-coupled, pore-scale processes. Furthermore, these models typically represent pore-scale processes within a small number of pores. Petascale implementation of fundamental models for pore-scale processes will allow simulations at the REV-scale, providing a firm foundation for the development of improved continuum models that are ultimately used in the decision-making process. In some cases, when the REV is very large or does not exist, direct application of pore-scale models on the REV scale will not be possible, and more innovative ways of using information from pore-scale models will be needed.

At present the ability of pore-scale models to accurately simulate fluid flow and chemical processes for the fluids, geomaterials, and conditions that are relevant to important problems and applications are limited. For example, none of the existing pore-scale models for multi-phase fluid flow can handle the full range of viscosities, densities, compressibilities, miscibilities, and wetting conditions that are important in subsurface applications. In addition, the reaction sites on the complex surfaces of natural pores and fractures have a broad range of reactivities and

accessibilities, which control and are controlled by adsorption and desorption processes and the evolution of complex small-scale morphologies (chemical microenvironments) generated by precipitation and dissolution. Furthermore, in many processes of interest, multi-phase fluid flow and reactive transport processes are inherently coupled leading to additional challenges to model development.

A variety of pore-network models have been developed for applications such as oil recovery, secondary migration (the migration of hydrocarbon expelled from source rock to the reservoir in which they are trapped), groundwater contamination, and more recently, CO_2 sequestration (Li 2006) and gas hydrates (Tsimpanogiannis 2006). In these models, the complex pore geometry found in rocks and soils is represented by pore volumes connected by a network of pores with simple geometries, and the physics of multi-phase fluid flow is also simplified and/or replaced by empirical rules. The limitations of these models have been recognized since their inception, but they are still the only practical approach for many applications because other approaches demand too much computing speed, and/or they cannot be used for the combination of fluid properties found in many important applications. A more first principles approach is needed to accurately simulate multi-phase fluid flow on the pore scale, for all fluid combinations of practical importance, and to bridge the gap between pore scale and continuum models.

Lattice Boltzmann simulations have provided an alternative to pore network modeling. They can be easily applied to fluid flow in complex pore geometries, they have been implemented on quite large lattices (e.g., 512×512×512; Harting 2004), and they can be easily implemented for complex pore space geometries. However, they have been restricted to quite small density ratios. New lattice Boltzmann models with improved numerical stability have very recently been developed to simulate multi-phase fluid flow with high density contrasts (at least 1000:1) (Zheng 2006; Lee 2005). At this stage, it is too soon to assess the implications that this will have for subsurface science applications.

More recently particle methods such as smoothed particle hydrodynamics (SPH) have been successfully applied to multi-phase fluid flow with large density and viscosity ratios (Tartakovsky 2006a). However, these methods are substantially less computationally efficient than lattice-Boltzmann simulations, and they have not yet been implemented on large computing systems. In addition, some fluid properties, such as the viscosity and interfacial tension, must be measured because of intrinsic particle effects. For example, momentum transport by moving particles adds to the viscosity used in the SPH equations (Hoover 1998). Another promising approach is to combine grid-based computation fluid dynamics with efficient interface tracking algorithms such as the level set, volume-of-fluid and phase field methods, and a model for the fluid-fluid-solid contact angles. Systems with large density and viscosity ratios can easily be handled, as can incompressible fluids. Currently, the subsurface science applications of this approach are quite limited (Huang 2005).

Pore-network, lattice-Boltzmann and particle methods have been used to simulate reactive transport (fluid flow coupled with dissolution, precipitation and dissolution) on the pore scale

(Kang 2006; Tartakovsky 2006b). However, such efforts are at an early stage in their development and application. The processes involved in reactive transport occur over a wide range of time and length scales. Precipitation and dissolution can cause complex alteration of submicron scale morphology, which over time lead to changes in field-scale transport properties, and reaction times may be orders of magnitude larger than the time scales associated with multi-phase fluid flow. In general, mineral growth may be reaction controlled, transport controlled, nucleation controlled or in a crossover between these limits. Broadly applicable multi-scale/multi-physics methods for reactive transport in pores and fractures are needed, but they are not currently available.

Non-linear behavior may further complicate the prediction of multi-phase fluid flow in the subsurface. The "dripping faucet" model is an important paradigm for complex behavior in simple non-linear systems (deterministic chaos, period doubling, intermittency, etc.), and there is evidence for similar behavior during multi-phase (water/air) fluid flow in fractured subsurface systems (Faybischenko 2004). The dynamics of fluid-fluid-solid contact lines and the associated phenomenon of velocity dependent contact angles also add to the complexity of multi-phase fluid flow in fractures and porous media, particularly for mineral surfaces with physical or chemical heterogeneity (Cubaud 2004). Complex Marangoni effects (Matar 1999) (flow driven by surface tension gradients) can also be expected if surfactants are used for subsurface applications such as oil recovery or contaminant remediation. At this stage, the level at which these complex phenomena must be modeled at the pore scale to reliable predict behavior at larger scales is not understood.

RESEARCH DIRECTIONS

For pore-scale models to effectively inform larger scale simulations, it is necessary to simulate processes at scales that are sufficiently large (i.e., many pores, fractured/porous medium or small fracture networks) to provide insights into and information about these processes at the "continuum" scale. Furthermore, the ability of pore-scale models to quantitatively predict physical processes in complex pore structures must be demonstrated through quantitative comparisons with the results of well controlled experiments on systems with well characterized pore geometries, wetting behaviors, and surface mineralogies. Petascale computational implementation of "first-principles," pore-scale models will provide the opportunity to extend simulations to spatial scales that are large enough to evaluate and develop constitutive relationships for complex coupled processes, and to extend pore-scale models to a much wider range of fluid properties and system parameters than can be explored experimentally. Because the computational cost of accurate pore-scale modeling approaches is currently extremely high, this will require the development of new and/or improved algorithms for pore-scale processes that are both robust and efficiently parallelized. These algorithms must overcome computational challenges associated with different pore-scale modeling approaches. These include issues such as numerical dispersion, interface tracking and parallelization difficulties associated with grid-based methods. Particle-based methods must overcome difficulties with implementing nonlinear reaction kinetics and boundary conditions, as well as issues with artificial (intrinsic particle

contributions to) interfacial tension and viscosity. Finally, developing approaches for efficiently coupling models for different processes that may occur on disparate time scales is critical to evaluating processes in systems of scientific interest and practical importance.

COMPUTATIONAL SCIENCE IMPACT

Implementation of pore-scale models at sufficiently large scales will demand the development of new/improved algorithms that capitalize on petascale computational capabilities and infrastructure. These efforts will require the development of numerical methods with improved speed and accuracy that will be generally applicable to the solution of coupled partial differential equations relevant to a broad range of problems. For example, grid-based methods will require adaptive mesh refinement techniques to adequately capture evolving interfaces and accurately simulate transport in strongly advective systems. This will require development and implementation of advanced load-balancing routines to support efficient implementation on massively parallel architectures.

Simple particle methods such as smoothed particle hydrodynamics and dissipative particle dynamics, and simple lattice Boltzmann models are local in nature, and it is expected that efficient codes that perform well on very large parallel computing systems can be developed without undue difficulty. However, the implementation of more advanced methods developed to reduce compressibility effects in particle methods, improve the stability of lattice Boltzmann models for high density contrasts and adaptive methods developed to increase computing speed and resolution may be substantially more difficult to implement efficiently on very large systems.

Advances in computational algorithms for pore scale subsurface applications will be directly relevant in many other areas including fundamental geoscience, the modeling of fluid in biological systems, micro-electromechanical systems (MEMS) and industrial processes.

SUBSURFACE SCIENCE IMPACT

Improved understanding and prediction of pore-scale processes will provide an essential foundation for predicting and controlling subsurface phenomena at all larger scales. Developing the requisite pore-scale models will lead to substantial improvements in our understanding of the limitations of currently used continuum models, thus reducing model-based uncertainty in field-scale simulations. For example, the fraction of oil recovered from a reservoir depends on the initial distribution of oil within the pores, and how this changes over time as oil is extracted. This is controlled by immiscible fluid-fluid displacements within the pore space, which under many conditions is poorly represented by the continuum models used in reservoir simulators. Currently, the oil recovery factor (oil recovered/oil originally in place) varies from an average of about 30 percent for small fields to an average of just under 50 percent for large fields, and these average recovery factors have increased significantly with time because of better technology. Assuming current world conventional recoverable oil reserves of about one trillion barrels, an

increase in the recovery factor from 40 percent to 41 percent would increase the amount of oil recovered by about 25 billion barrels, worth more than a trillion dollars (offset in part by the cost of developing and implementing new technology) at today's prices. Examples of currently pressing problems that will benefit directly from pore-scale models include the potential for fracture sealing that may slow or halt the upward migration of CO_2 in subsurface sequestration systems, and the *in situ* creation of barriers designed to restrict contaminant migration, both of which depend on pore-scale mineral growth and the resulting alteration of pore morphology. These processes cannot be adequately quantified using existing continuum models. Petascale computing applied to pore-scale models will provide a firm scientific foundation for the development of improved technologies and strategies for the remediation and/or isolation of subsurface contamination, CO_2 sequestration and hydrocarbon production.

TIME FRAME

We estimate that a concerted 5-year effort focused on the development of better codes that take full advantage of the capabilities of petascale computing systems would allow multi-phase fluid flow to be simulated in complex porous and fractured media with the spatiotemporal complexity seen in experiments. Within a decade, we expect that a new generation of multi-scale, multi-physics models and additional advances in computing technology could allow us to reliably simulate reactive transport for the full range of fluids and conditions needed for applications such as CO_2 sequestration and fossil fuel recovery. By this stage, we expect that the primary role of pore scale experiments would be validation of computer models, particularly when experiments would be difficult, dangerous or require long lead times. Although we expect that, with the right resources, critical advances can be made during the next 5 to 10 years, challenges requiring new ideas, new methods and new, much more capable, computing systems will remain.

REFERENCES

Cubaud, T. and Fermingier, M. (2004) Advancing contact lines on chemically patterned surfaces, *J. Colloid . Interface Science.* **269**: 171.

Faybischenko, B. (2004) Nonlinear dynamics in flow through unsaturated fractured porous media: status and perspectives, *Rev. Geophys.*, **42**: RG2003.

Harting, J., Venturoli M. and Coveney, P.V. (2004) Large-scale grid-enabled lattice Boltzmann simulations of complex fluid flow in porous media and under shear, *Philosophical Trans. Royal Society of London*, **A362**: 1703.

Hoover, W. G. (1998) Isomorphism linking smooth particles and embedded atoms, *Physica* **A260**:244.

Huang, H., Meakin, P., Liu, M. B. and McCreery, G. E. (2005) Modeling of multiphase fluid motion in fracture intersections and fracture networks, Geophys. Res. Lett., **32:** L19402.

Kang, Q., Lichtner, P.C. and Zhang, D. (2006) Lattice Boltzmann pore-scale model for multicomponent reactive transport in porous media, *J. Geophys. Res.*, doi:10.1029/2005JB003951.

Lee T., and Lin, C.-L. (2005) A stable discretization of the lattice Boltzmann equation for simulation of incompressible two-phase flows at high density ratio, *J.Comp. Phys.* **206**: 16

Li, L., Peters, C. A. and Celia, M. A. (2006) Upscaling geochemical reaction rates using pore scale network modeling, *Advances in Water Resources*, **29**: 1351.

Matar O. K. and Troian, S. M. (1999) The development of transient fingering during the spreading of surfactant coated films, *Physics of Fluids*, **11**: 3232.

Tartakovsky, A. and Meakin, P. (2006A) Pore scale modeling of immiscible and miscible fluid flows using smoothed particle hydrodynamics, *Advances in Water Resources*, **29**: 1464.

Tartakovsky, A.M., Meakin, P. Scheibe, T.D. and R.M. Eichler-West, (2006B) Simulations of reactive transport and precipitation with smoothed particle hydrodynamics, *J. Comp. Phys.*, doi: 10.1016/j.jcp.2006.08.013.

Tsimpanogiannis, I. N. and Lichtner, P. C. (2006) Pore-network study of methane hydrate dissociation, *Phys. Rev. E* **74**: 056303 (-1-13).

Zheng, H. W., Shu C. and Chew, Y. T. (2006) A lattice Boltzmann model for multiphase fluid flows with large density ratio, J. Comp. Phys. **218**: 353.

CONCLUSION

IMPORTANCE OF THE SUBSURFACE SCIENCES

Excellent plenary presentations and lively discussions in the panel breakouts at the workshop emphasized the importance of current and recommended future research in the subsurface sciences. We need to improve our understanding of fundamental subsurface processes related to protection of consumptive and agricultural groundwater resources, production of fossil fuels, bioremediation, geothermal energy, subsurface contaminant transport, disposal of spent nuclear fuel and waste, subsurface carbon sequestration, and developing new energy resources (such as gas hydrates). Understanding the Earth's subsurface processes is necessary for the success of the missions of the DOE and other federal and state agencies and for the overall health and safety of societies worldwide.

Groundwater is an abundant natural resource that the U.S. Geological Survey estimates makes up 30 percent of all of the freshwater in the world. By 2025, at least 3.5 billion people—nearly 50 percent of the world's population—will face water shortages. To provide a sustainable supply of water fit for human consumption, we must protect the world's aquifers from contamination and excessive groundwater withdrawals. Coastal areas present special problems, because withdrawals from freshwater aquifers often result in saltwater intrusion into the aquifer. In many regions of the world, healthy freshwater aquifers are essential to sustaining a high level of agricultural productivity

The subsurface contains both naturally occurring and human induced hazardous materials. Numerical modeling of groundwater contaminant transport is an important tool that aids in the analysis of groundwater hazard problems, both actual and potential. Accidental spills, leakage, and waste-disposal operations can lead to groundwater contamination. We need to analyze and predict the movement of solutes in groundwater systems to assess the effects of a contamination plume and properly design a waste-disposal operation or mitigation strategy. Laboratory experiments are essential to understanding geochemical reactions in the field and for obtaining the necessary reaction coefficients and rate constants used in transport models. Simulation modeling allows us to compare alternative strategies for aquifer reclamation. In some cases, the transported component of interest is thermal energy. Heat transport simulation is useful in the analysis of geothermal systems, waste heat storage systems, and some deep aquifer systems. Examples of DOE sites that rely heavily on modeling the transport of contaminants and heat are the Hanford Site in Washington State and the proposed waste repository at Yucca Mountain, Nevada.

Understanding subsurface sciences phenomena is essential for identifying, characterizing, and extracting energy and mineral resources. For example, if we are to move to a hydrogen-based energy economy, we must develop tools to convert to hydrogen fuels. All known sources of energy generation in the amounts necessary to convert hydrogen to useable fuel are geoscience-based. Without a more complete understanding of the subsurface, we may not achieve a hydrogen-based energy supply. For many decades, subsurface simulations have been used to

identify petroleum reservoirs and identify optimal petroleum extraction strategies. Today, computer simulations are playing a key role in developing new methods for exploiting the nation's oil-shale resources.

The subsurface sciences will play a large role in our efforts to mitigate global climate change. Fossil fuels will remain the main source of energy production well into the 21st century. Increased concentrations of CO_2 resulting from carbon emissions are expected unless we design energy systems that reduce the carbon emissions to the atmosphere. Roughly one third of the nation's carbon emissions come from power plants and other large point sources. To stabilize and ultimately reduce concentrations of this greenhouse gases, we need to employ carbon sequestration-carbon capture, separation, and storage or reuse.

CURRENT STATUS AND CHALLENGES

Current capabilities in computational modeling of coupled subsurface phenomena are inadequate for addressing many current and future DOE challenges. High-fidelity simulation of subsurface processes must be built on a fundamental understanding of multi-phase, multi-component flow and transport, biochemical and geochemical reactions, thermal and geo-mechanical effects, all interacting within a highly heterogeneous porous media. Sensitivity to physical processes and model parameters, and uncertainties in parameters, media properties, and the current state of the system must be taken into account. Informed decision-making will require the development of a comprehensive uncertainty/sensitivity analysis framework, within which coupled process models will be executed hundreds to thousands of times under different scenarios. The magnitude of this effort will tax even petascale computational resources.

In addition to formidable computational demands, subsurface sciences research has extremely high data management requirements. This presents new challenges and opportunities for the fusion of experimental data with theory, modeling, and simulation to advance subsurface science. In many cases, this leads to new science driven informatics requiring new data management and analytic approaches to drive the discovery process. In the area of distributed data management and analysis, subsurface sciences research faces similar requirements as climate, physics, biology, and nanosciences for the analysis of large volumes of data produced at different facilities, analyzed via complex workflows, and consumed by a large and distributed user community. In addition to large data volumes, DOE experimental and computational science disciplines face challenges associated with increasingly complex data gained from different experimental and analysis techniques. Again, there are opportunities to achieve technological and methodological advances that benefit multiple disciplines. Opportunities abound for using well-established technologies to better manage data and a wide range of other types of information. These technologies include relational databases, Geographic Information Systems (GIS), and recent advances in Content Management Systems (CMS) that use rich metadata and XML schema to permit users to intelligently navigate large collections of models, algorithms, and data objects. In particular, we perceive a need for a comprehensive system for managing the large and diverse set of subsurface process simulation models and databases that

have been developed at federal laboratories, universities, and in the private sector. Data used in inverse-calibration modeling are often only used to calibrate the computational model. However, these and other data can be used to gain insight into the computational models and uncertainties. To enable optimized decision making and risk assessment, we need more powerful methods for estimating model uncertainty and model verification and validation.

A PATH FORWARD

The subsurface sciences issues discussed at the workshop and presented in this report are central to the missions of the DOE's EM, FE, and RW programmatic offices. The DOE Office of Science has unparalleled capabilities in computational science and high performance computing and is very well positioned to work with the applications offices to develop a next generation of subsurface sciences simulation tools. The investments in modeling and simulation outlined here complement the investments in fundamental environmental science by SC and remediation and monitoring programs supported by the DOE. The modeling and simulation effort will draw heavily on DOE experimental, remediation and characterization programs to provide basic insights into and data on complex environmental and subsurface systems as well as the fundamental processes that govern their behavior. The experimental programs also provide the means of validating the computational models and simulations, and validation is essential to advancing the state of the art in scientific simulation. In turn, the computational modeling and simulation will:

- Provide insights into fundamental physical, chemical, and biological processes than would otherwise be unattainable
- Maximize the return on investments in experimental facilities by more quickly and efficiently harvesting scientific results
- Provide the technical capability to design and construct innovative, new experimental and computational facilities.

As the principal U.S. federal steward**,** DOE has wide-ranging challenges in protecting groundwater resources, production of fossil fuels, geothermal energy, subsurface contaminant cleanup and containment, disposal of spent nuclear fuel and waste, and subsurface carbon sequestration. A stated goal of SC is to provide "discovery-class" tools to the scientific community to help answer the most challenging research questions of our era, attract the greatest scientific minds in the world, and enable the type of discoveries that truly change the face of science. A next generation of computational models of subsurface sciences flow and coupled process simulation will provide to the DOE new discovery-class tools to better deal with the daunting task of managing the legacy wastes from the nuclear research and weapons programs, developing new energy sources, mitigating climate change, and developing new remediation strategies. In particular, new petascale-class computationally based models will provide support to operational-level, site-specific modeling for regulatory review. This is an opportune time to

build on recent successes of the Scientific Discovery through Advanced Computing initiative to team SC computational scientists with application scientists from EM, FE, and RW to produce a new generation of subsurface simulation models.

APPENDIX – WORKSHOP PARTICIPANTS

WORKSHOP PARTICIPANTS

Aavatsmark, Ivar; Centre for Integrated Petroleum Research; ivar.aavatsmark@cipr.uib.no

Adler, Pierre; UPMC-Sisyphe; padler@ccr.jussieu.fr

Ahlfeld, David; University of Massachusetts; ahlfeld@ecs.umass.edu

Allan, Benjamin; Sandia National Laboratories; baallan@ca.sandia.gov

Allen, Myron; University of Wyoming; allen@uwyo.edu

Anderson, Robert; DOE- SC; todd.anderson@science.doe.gov

Anitescu, Mihai; Argonne National Laboratory; anitescu@mcs.anl.gov

Arbogast, Todd; University of Texas at Austin; arbogast@ices.utexas.edu

Archibald, Rick; Oak Ridge National Laboratory; 2ya@ornl.gov

Bacon, Diana; Pacific Northwest National Laboratory; diana.bacon@pnl.gov

Bayer, Paul; DOE-SC; paul.bayer@science.doe.gov

Beam, Paul; DOE-EM; paul.beam@em.doe.gov

Bell, John; Lawrence Berkeley National Laboratory; jbbell@lbl.gov

Bennethum, Lynn; University of Colorado; Lynn.Bennethum@cudenver.edu

Berkman, Clarissa; Pacific Northwest National Laboratory; clarissa.berkman@pnl.gov

Berlyand, Leonid; Pennsylvania State University; berlyand@math.psu.edu

Bolton, Edward; Yale University; edward.bolton@yale.edu

Brown, David; Lawrence Livermore National Laboratory; dlb@llnl.gov

Bylaska, Eric; Pacific Northwest National Laboratory; eric.bylaska@pnl.gov

Cahill, Tony; Texas A&M University; tcahill@civilmail.tamu.edu

Carle, Steven; Lawrence Livermore National Laboratory; carle1@llnl.gov

Cecchetti, Ann; Pacific Northwest National Laboratory; acecchetti@aol.com

Chen, Jinsong; Lawrence Berkeley National Laboratory; jchen@lbl.gov

Clement, Prabhakar; Auburn University; clement@auburn.edu

Colella, Phillip; Lawrence Berkeley National Laboratory; pcolella@lbl.gov

Collis, Scott; Sandia National Laboratories; sscoll@sandia.gov

Criscenti, Louise; Sandia National Laboratories; ljcrisc@sandia.gov

Curtis, Gary; U.S. Geological Survey; gpcurtis@usgs.gov

D'Azevedo, Ed; Oak Ridge National Laboratory; e6d@ornl.gov

Darema, Frederica; National Science Foundation; darema@nsf.gov

Dattoria, Vince; DOE-SC; vince.dattoria@science.doe.gov

Dawson, Clint; University of Texas at Austin; clint@ices.utexas.edu

Deane, Anil; DOE-ASCR; deane@er.doe.gov

Debusschere, Bert; Sandia National Laboratories; bjdebus@sandia.gov

Detwiler, Russell; Lawrence Livermore National Laboratory; detwiler@llnl.gov

Diachin, Lori; Lawrence Livermore National Laboratory; diachin2@llnl.gov

Diodato, David; U.S. Nuclear Waste Technical Review Board; diodato@nwtrb.gov

Dixon, David; University of Alabama; dadixon@bama.ua.edu

Dobson, Patrick; DOE-SC; Patrick.Dobson@science.doe.gov

Douglas, Jim; Purdue University; douglas@math.purdue.edu

Drake, Richard; Sandia National Laboratories; rrdrake@sandia.gov

Dyer, Russ; DOE-Yucca Mountain Project; russ_dyer@ymp.gov

E, Weinan; Princeton University; weinan@math.princeton.edu

Eddebbarh, Al Aziz; Los Alamos National Laboratory, aeddebba@lanl.gov

Efendiev, Yalchin; Texas A&M University; efendiev@math.tamu.edu

Egbert, Gary; Oregon State University; egbert@coas.oregonstate.edu

Eggert, Kenneth; Los Alamos National Laboratory; kge@lanl.gov

Engel, David; Pacific Northwest National Laboratory; dave.engel@pnl.gov

Engquist, Bjorn; University of Texas at Austin; engquist@math.utexas.edu

Estep, Donald; Colorado State University; estep@math.colostate.edu

Ewing, Richard; Texas A&M University; richard-ewing@tamu.edu

Fang, Yilin; Pacific Northwest National Laboratory; yilin.fang@pnl.gov

Fann, George; Oak Ridge National Laboratory; fanngi@ornl.gov

Felmy, Andy; Pacific Northwest National Laboratory; ar.felmy@pnl.gov

Flach, Greg; Savannah River National Laboratory; gregory.flach@srnl.doe.gov

Fogwell, Thomas; Energy Solutions (Fluor Hanford PHMC); thomas_fogwell@rl.gov

Garven, Grant; Johns Hopkins University; garven@jhu.edu

Ghattas, Omar; University of Texas at Austin; omar@ices.utexas.edu

Gilbertson, Mark; DOE-EM; mark.gilbertson@em.doe.gov

Glynn, Pierre; U.S. Geological Survey; pglynn@usgs.gov

Goldsztein, Guillermo; Georgia Institute of Technology; ggold@math.gatech.edu

Goodwin, David; DOE- SC; dave.goodwin@science.doe.gov

Gunzburger, Max; Florida State University; gunzburg@scs.fsu.edu

Hammond, Glenn; Pacific Northwest National Laboratory; glenn.hammond@pnl.gov

Hammond, Steve; National Renewable Energy Laboratory; steven_hammond@nrel.gov

Helland, Barbara; DOE-SC; helland@er.doe.gov

Helton, Jon; Sandia National Laboratories; jchelto@sandia.gov

Heroux; Michael; Sandia National Laboratories; maheroux@sandia.gov

Illangasekare, Tissa; Colorado School of Mines; tissa@mines.edu

James, Scott; Sandia National Laboratories; scjames@sandia.gov

Johnson, Fred; DOE-SC; fjohnson@er.doe.gov

Johnson, Gary; DOE-SC; garyj@er.doe.gov

Johnson, James; Lawrence Livermore National Laboratory; jwjohnson@llnl.gov

Jones, Wesley; National Renewable Energy Laboratory; wesley_jones@nrel.gov

Juanes, Ruben; Massachusetts Institute of Technology; juanes@mit.edu

Keating, Elizabeth; Los Alamos National Laboratory; ekeating@lanl.gov

Khaleel, Moe; Pacific Northwest National Laboratory; moe.khaleel@pnl.gov

Kitanidis, Peter; Stanford University; peterk@stanford.edu

Kouzes, Richard; Pacific Northwest National Laboratory; rkouzes@pnl.gov

Kreisman, Norman; DOE-SC; norman.kreisman@science.doe.gov

Lesmes, David; DOE-SC; david.lesmes@science.doe.gov

Li, Xiaolin; State University of New York at Stony Brook; linli@ams.sunysb.edu

Lichtner, Peter; Los Alamos National Laboratory; lichtner@lanl.gov

Lindquist, Brent; State University of New York at Stony Brook; lindquis@ams.sunysb.edu

Lipnikov, Konstantin; Los Alamos National Laboratory; lipnikov@lanl.gov

MacKinnon, Robert; Sandia National Laboratories; rjmacki@sandia.gov

Mahinthakumar, Kumar; North Carolina State University; gmkumar@ncsu.edu

Manchester, Kenneth; MSE Technology Applications; ken.manchester@mse-ta.com

Meakin, Paul; Idaho National Laboratory; paul.meakin@inl.gov

Mills, Richard; Oak Ridge National Laboratory; rmills@ornl.gov

Mishra, Srikanta; INTERA; smishra@intera.com

Moridis, George; Lawrence Berkeley National Laboratory; gjmoridis@lbl.gov

Moulton, David; Los Alamos National Laboratory; moulton@lanl.gov

Ndousse, Thomas; DOE; tndousse@er.doe.gov

Newman, Gregory; Lawrence Berkeley Laboratory; ganewman@lbl.gov

Nicholson, Thomas; U.S. Nuclear Regulatory Commission; tjn@nrc.gov

Nur, Amos; Stanford University; amos.nur@stanford.edu

O'Leary, Patrick; Idaho National Laboratory; patrick.oleary@inl.gov

Ostrouchov, George; Oak Ridge National Laboratory; ostrouchovg@ornl.gov

Ott, Bill; U.S. Nuclear Regulatory Commission; wro1@nrc.goV

Palmer, Carl; Idaho National Laboratory; carl.palmer@inl.gov

Pannala, Sreekanth; Oak Ridge National Laboratory; pannalas@ornl

Parashar, Manish; Rutgers University; parashar@caip.rutgers.edu

Parker, Jack; Oak Ridge National Laboratory; parkerjc@ornl.gov

Pascucci, Valerio; Lawrence Livermore National Laboratoryt/University of California at Davis; pascucci@acm.org

Pawar, Rajesh; Los Alamos National Laboratory; rajesh@lanl.gov

Peszynska, Malgorzata; Oregon State University; mpesz@math.oregonstate.edu

Peters, Catherine; Princeton University; cap@princeton.edu

Petersson, Anders; Lawrence Livermore National Laboratory; andersp@llnl.gov

Polansky, Walter; DOE-SC; walt.polansky@science.doe.gov

Pyrak-Nolte, Laura; Purdue University; ljpn@physics.purdue.edu

Qin, Guan; Texas A&M University; gqin@vprmail.tamu.edu

Rajaram, Harihar; University of Colorado, Boulder; hari@colorado.edu

Roche, Kenneth; Oak Ridge National Laboratory; rochekj@ornl.gov

Rodgers, Arthur; Lawrence Livermore National Laboratory; rodgers7@llnl.gov

Russell, Thomas; National Science Foundation; trussell@nsf.gov

Rustad, James; University of California at Davis; jrrustad@ucdavis.edu

Scheibe, Tim; Pacific Northwest National Laboratory; tim.scheibe@pnl.gov

Sears, Mark; DOE/SC; sears@mics.doe.gov

Sekine, Yukiko; DOE-SC; yukiko.sekine@science.doe.gov

Shoemaker, Christine; Cornell University; cas12@cornell.edu

Smith, Barry; Argonne National Laboratory; bsmith@mcs.anl.gov

Stewart, Terri; Pacific Northwest National Laboratory; terri.stewart@pnl.gov

Strayer, Michael; DOE-SC; michael.strayer@science.doe.gov

Tartakovsky, Alexandre; Pacific Northwest National Laboratory; alexandre.tartakovsky@pnl.gov

Tartakovsky, Daniel; University of California at San Diego/Los Alamos National Laboratory; dmt@ucsd.edu

Thompson, K. Michael; DOE Richland Operations Office; K_M_Mike_Thompson@RL.gov

Tompson, Andrew; Lawrence Livermore National Laboratory; afbt@llnl.gov

Travis, Bryan; Los Alamos National Laboratory; bjtravis@lanl.gov

Trebotich, David; Lawrence Livermore National Laboratory; treb@llnl.gov

Valocchi, Al; University of Illinois at Urbana-Champaign; valocchi@uiuc.edu

Van Luik, Abraham; DOE-Yucca Mountain Project; abe_vanluik@ymp.gov

Varadarajan, Uday; DOE –SC; uday.varadarajan@science.doe.gov

Versteeg, Roelof; Idaho National Laboratory; roelof.versteeg@inl.gov

Wang, Hong; University of South Carolina; hwang@math.sc.edu

Wang, Junping; National Science Foundation; jwang@nsf.gov

Weare, John; University of California at San Diego; jweare@ucsd.edu

Wheeler, Mary; The University of Texas at Austin; mfw@ices.utexas.edu

White, Mark; Pacific Northwest National Laboratory; mark.white@pnl.gov

Wolfsberg, Andrew; Los Alamos National Labortory; awolf@lanl.gov

Woodward, Nick; DOE-SC; nick.woodward@science.doe.gov

Yabusaki, Steve; Pacific Northwest National Laboratory; yabusaki@pnl.gov

Ye, Ming; Florida State University; mingye@scs.fsu.edu

Yeh, George; University of Central Florida; gyeh@mail.ucf.edu

Yetter, Christopher; DOE, Office of the Under Secretary for Science; christopher.yetter@science.doe.gov

Yotov, Ivan; University of Pittsburgh; zeus.math.pitt.edu/~yotov

Zachmann, Dave; Colorado State University; dzach@mesanetworks.net

Zhang, Dongxiao; University of Oklahoma; donzhang@ou.edu

Zheng, Chunmiao; University of Alabama; czheng@ua.edu

Zheng, Li; Los Alamos National Laboratory; lzheng@lanl.gov

Zyvoloski, George; Los Alamos National Laboratory; gaz@lanl.gov