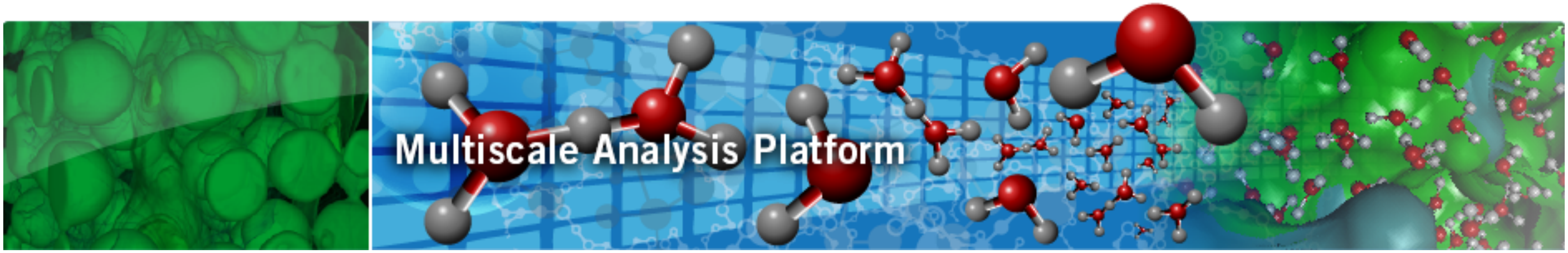
Biological and Environmental Research – Computing Needs Workshop
Rockville, MD September 11-12, 2012



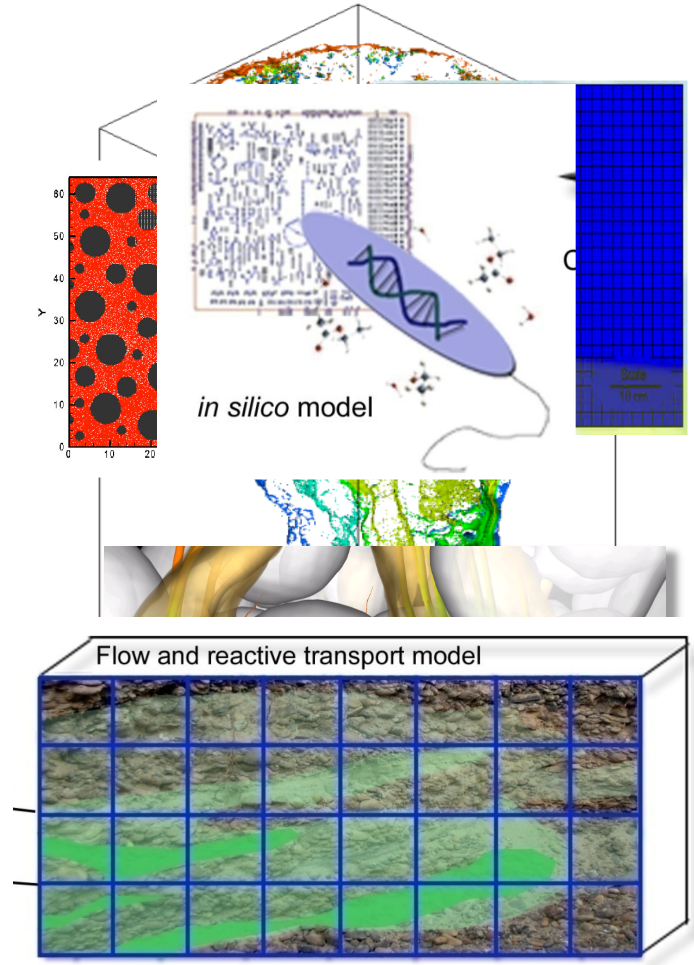
Present and Future Computing Requirements

Case Study: Subsurface Flow and Reactive Transport

Tim Scheibe
Pacific Northwest National Laboratory
tim.scheibe@pnnl.gov

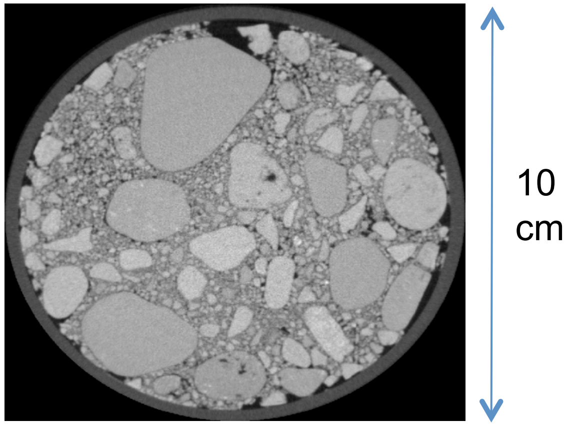
Project Description

- Multiple projects (all funded by BER / CESD / SBR):
  1. SciDAC Groundwater Science Application and SAPs (ended) – Hybrid multiscale simulation of subsurface reactive transport
  2. PNNL Subsurface Scientific Focus Area – Impact of microenvironments and transition zones
  3. University-led project (ending) – Coupling genome-scale microbial metabolism and subsurface reactive transport models (linked to Rifle Integrated Field Challenge project)

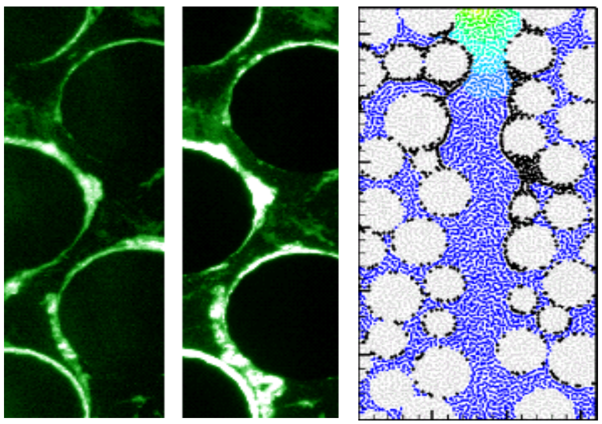


Our present focus is…

- More physics/chemistry/biology, less empiricism
  - Pore-scale and other high-resolution flow/transport modeling
  - Mechanistic biological models
- Addressing the "tyranny of scales"
  - Hybrid multiscale simulation to link pore- and continuum-scale models



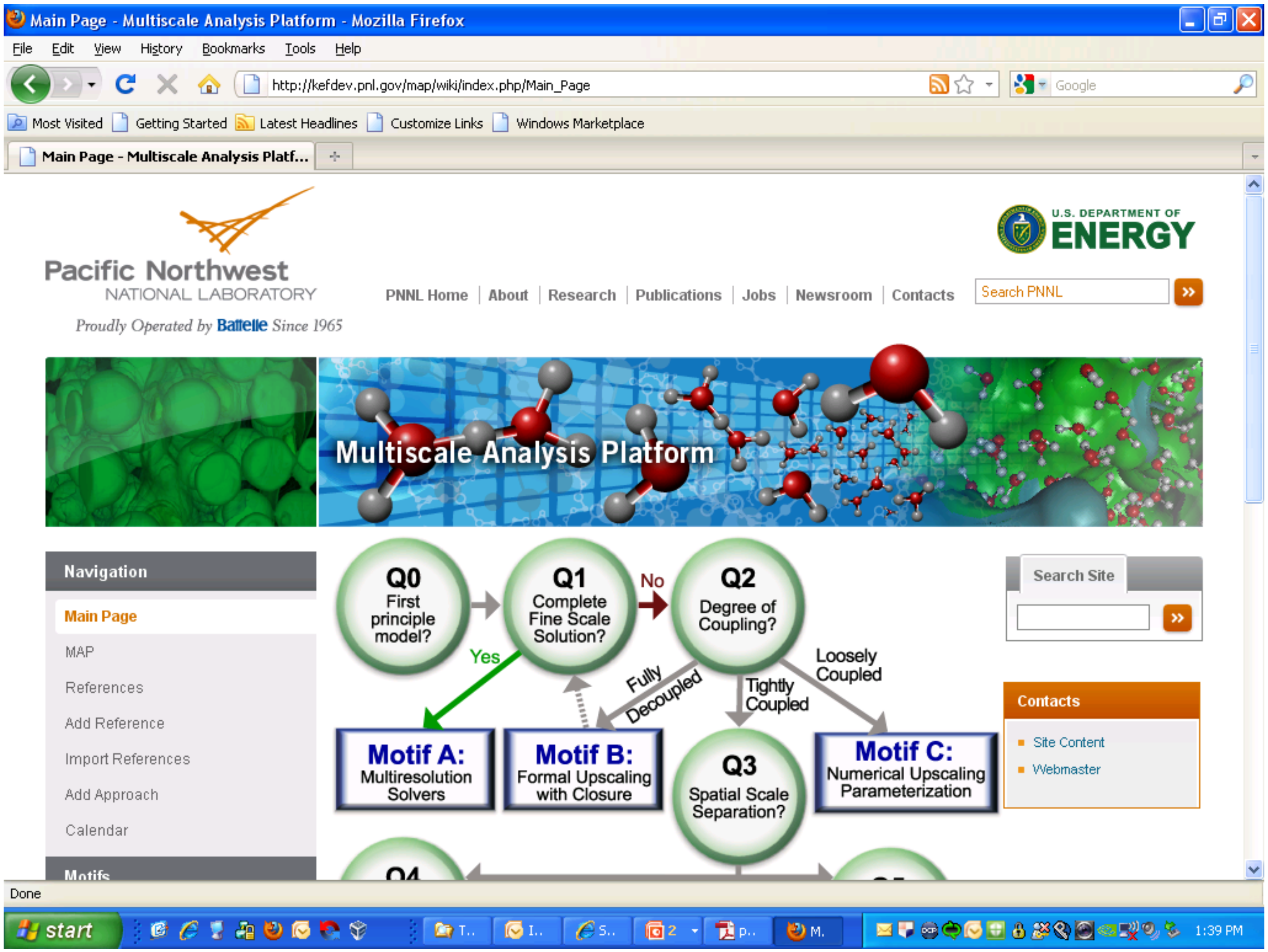
Data courtesy of John Zachara, PNNL



Tartakovsky et al., J. Porous Media, 2009
(Micromodel image: Carolyn Pearce, PNNL)

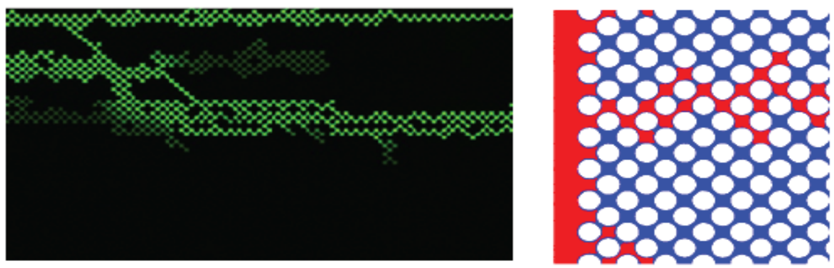
By 2017 we expect to…

- Develop fully coupled pore- and continuum-scale hybrid simulator – Next generation of subsurface simulation tools?

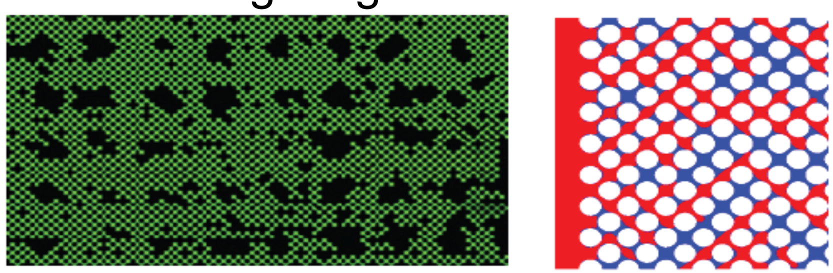


By 2017 we expect to…

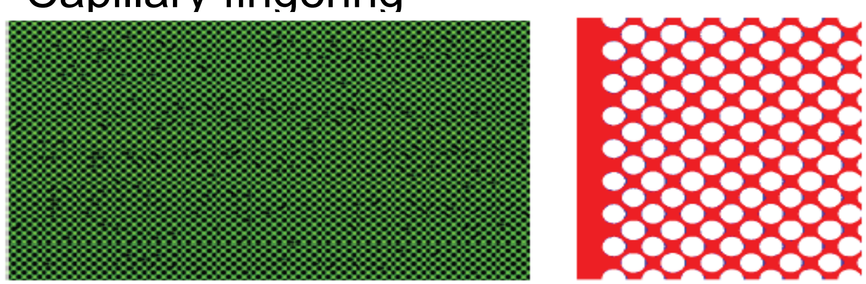
- Simulate multiphase flow, solute and energy transport, geochemical reactions, geomechanical effects, and multi-organism microbial communities

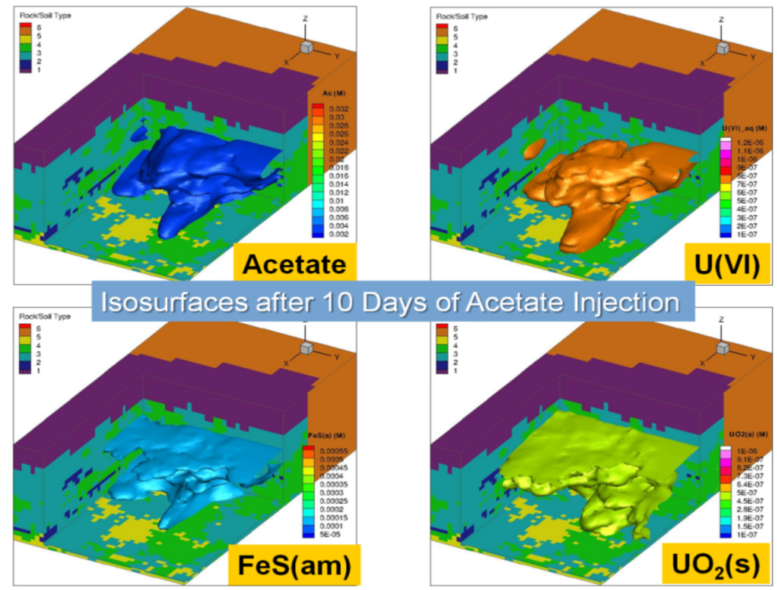


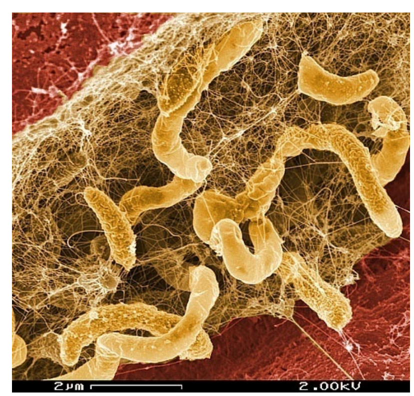
Viscous fingering



Capillary fingering







Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

By 2017 we expect to…

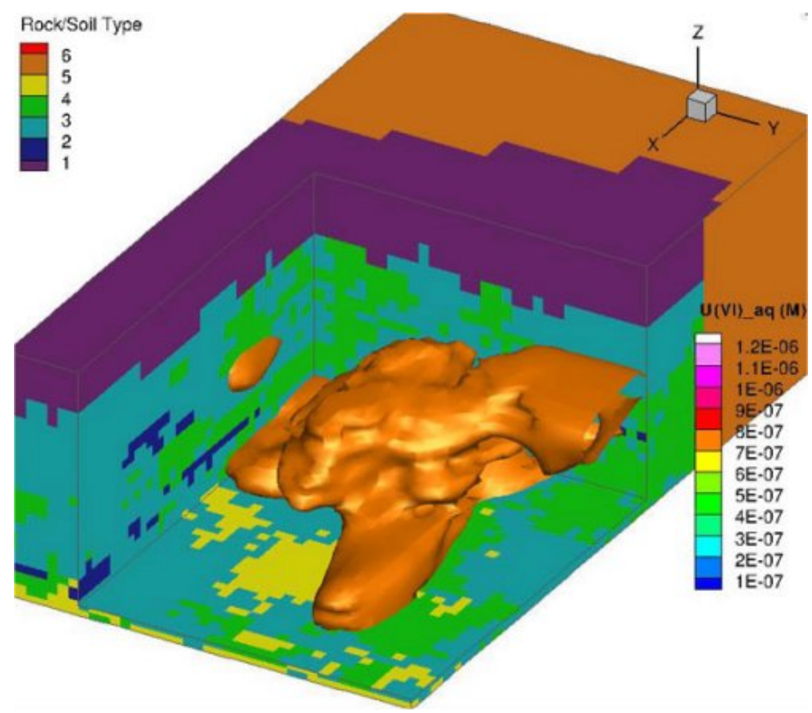
- Link subsurface models to larger-scale earth system simulations (e.g., community land model)

Codes we use are…

- eSTOMP: Continuum-scale porous media flow and reactive transport
  - Algorithms:
    - Finite difference spatial discretization
    - Newton non-linear outer loop
    - Linear inner solve
    - Operator split (reactions / transport / flow)
  - Built on Global Arrays (GA) Toolkit and PETSc
  - Parallel scaling limited by
    - Scales well to over 130,000 processors
    - Weak scaling limited by global linear system solve
    - Load balancing for reactions

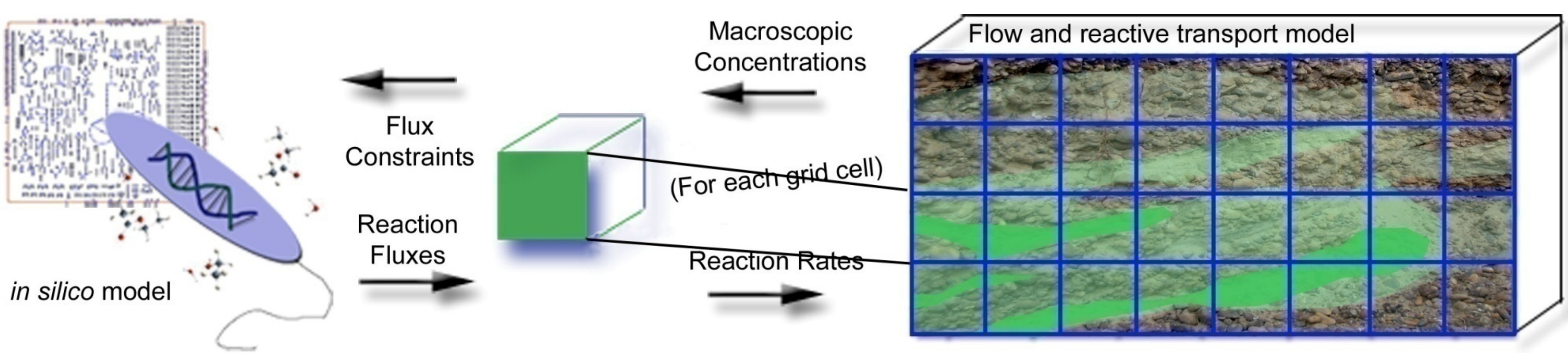
Benchmark Problem: uranium bioremediation

18m x 20m x 6.3m , 2.2M grid cells
300 time steps, 1 simulated day, checkpoint each 6 sim hours
5 lithofacies,102 biogeochemical species, 7 mineral reaction network



Codes we use are…

- eSTOMP: Continuum-scale porous media flow and reactive transport
  - Computational Challenges
    - Integrating mechanistic models of microbially-mediated reactions with complex communities of organisms
      - Small (N=500) LP solution at each iteration of each time step at each grid cell
    - Convergence issues
      - E.g., fully coupled well model in eSTOMP-CO2



Codes we use are…

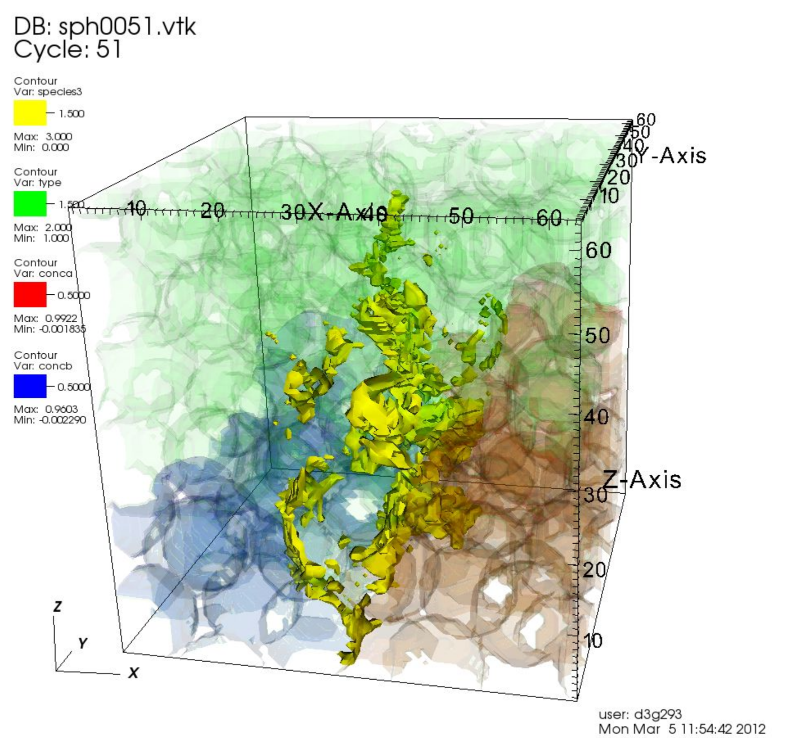
- SPH: Pore-scale porous media flow and reactive transport
  - Algorithms:
    - Smoothed Particle Hydrodynamics – lagrangian mesh-free particle method
    - No global linear matrix solve
    - Local force calculation requires tree search for neighbors
    - Reactions – system of ODEs
  - Built on Global Arrays (GA) Toolkit
  - Parallel scaling limited by
    - Had been I/O limited but this has been addressed through use of H5PART

Example Problem: mixing-controlled precipitation reaction

1 mm^3, 7 M computational particles

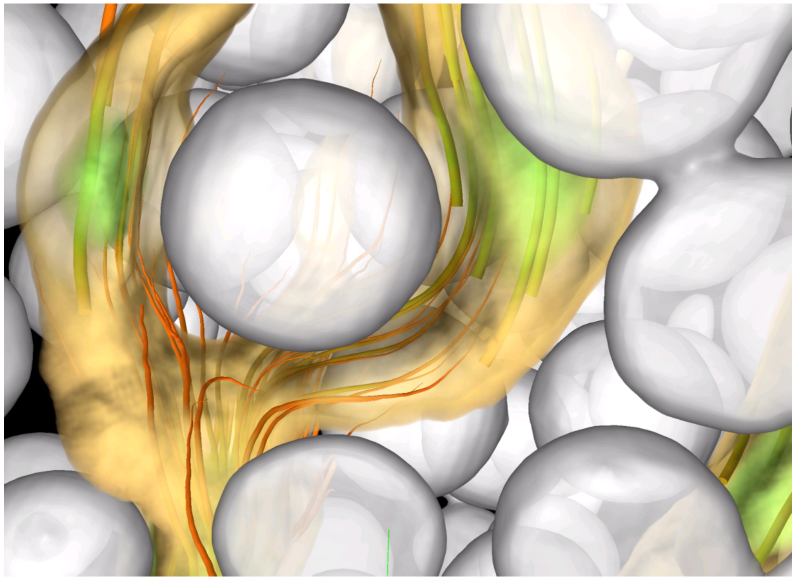
About 100 mineral grains

Two dissolved species react to form a precipitated mineral species



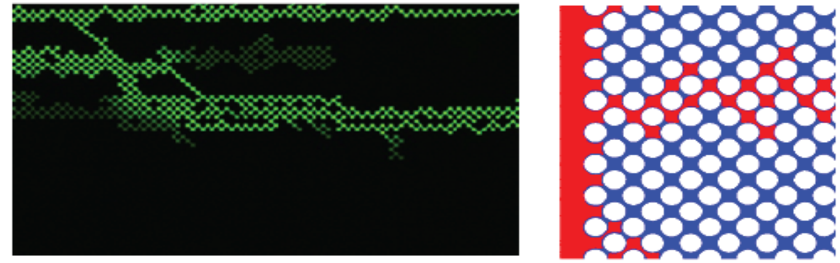
Codes we use are…

- SPH: Pore-scale porous media flow and reactive transport
  - Computational Challenges
    - Boundary conditions:
      - Periodic conditions usually used; how to deal with solute concentrations?
      - Flux-based boundary conditions had been difficult to implement
    - Time steps required for stability are typically very small
      - Strictly is for compressible flows – use for nearly incompressible fluids leads to challenges
      - Slow compared to grid-based methods for single-phase flow

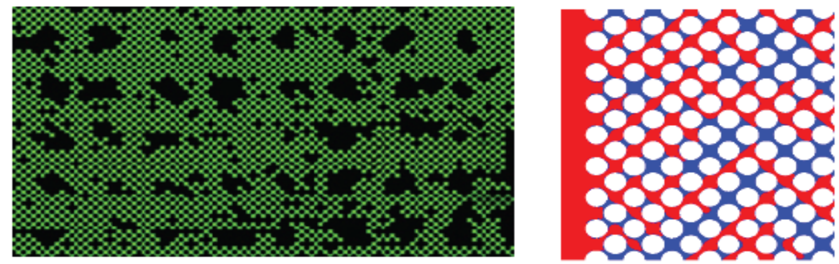


Codes we use are…

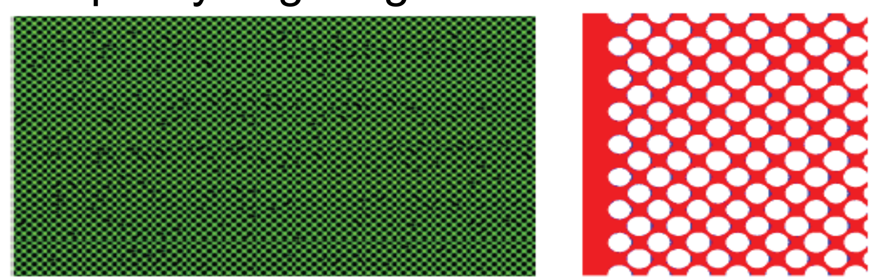
- SPH for multiphase flow
  - Can simulate surface tension and contact angle by varying particle-particle attractive forces
  - Application to new BER directions in carbon cycling within terrestrial ecosystems.
    - Currently testing 3D air-water simulations with microbial reactions for cellulose degradation



Viscous fingering



Capillary fingering



Stable displacement

Codes we use are…

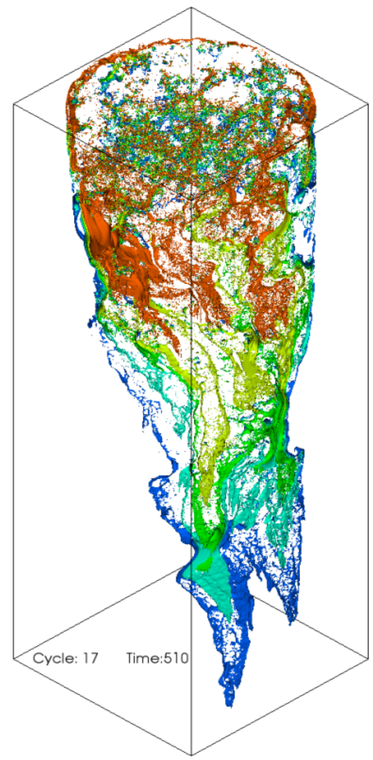
- TETHYS: Pore-scale porous media flow and transport
  - Algorithms:
    - Finite volume unstructured spatial discretization
  - Built on Global Arrays (GA) Toolkit and PETSc
  - Parallel scaling limited by
    - I/O, code structure
  - Computational challenges
    - Runs as unsteady problem to steady state – wait times in queue is limiting
    - Mesh-based approach limits application to problems with moving interfaces (e.g., multiphase flow, precipitation/ dissolution reactions, biofilms)

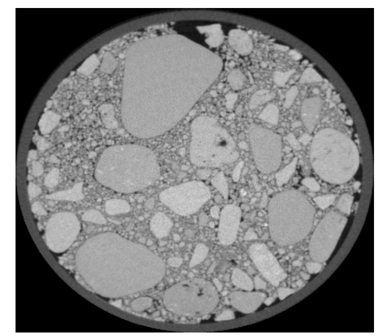
Example Problem: Navier-Stokes flow and tracer transport in a laboratory column

20 cm length, 10 cm diameter, 40 M computational nodes

50 micron spatial resolution derived from X-ray microtomography

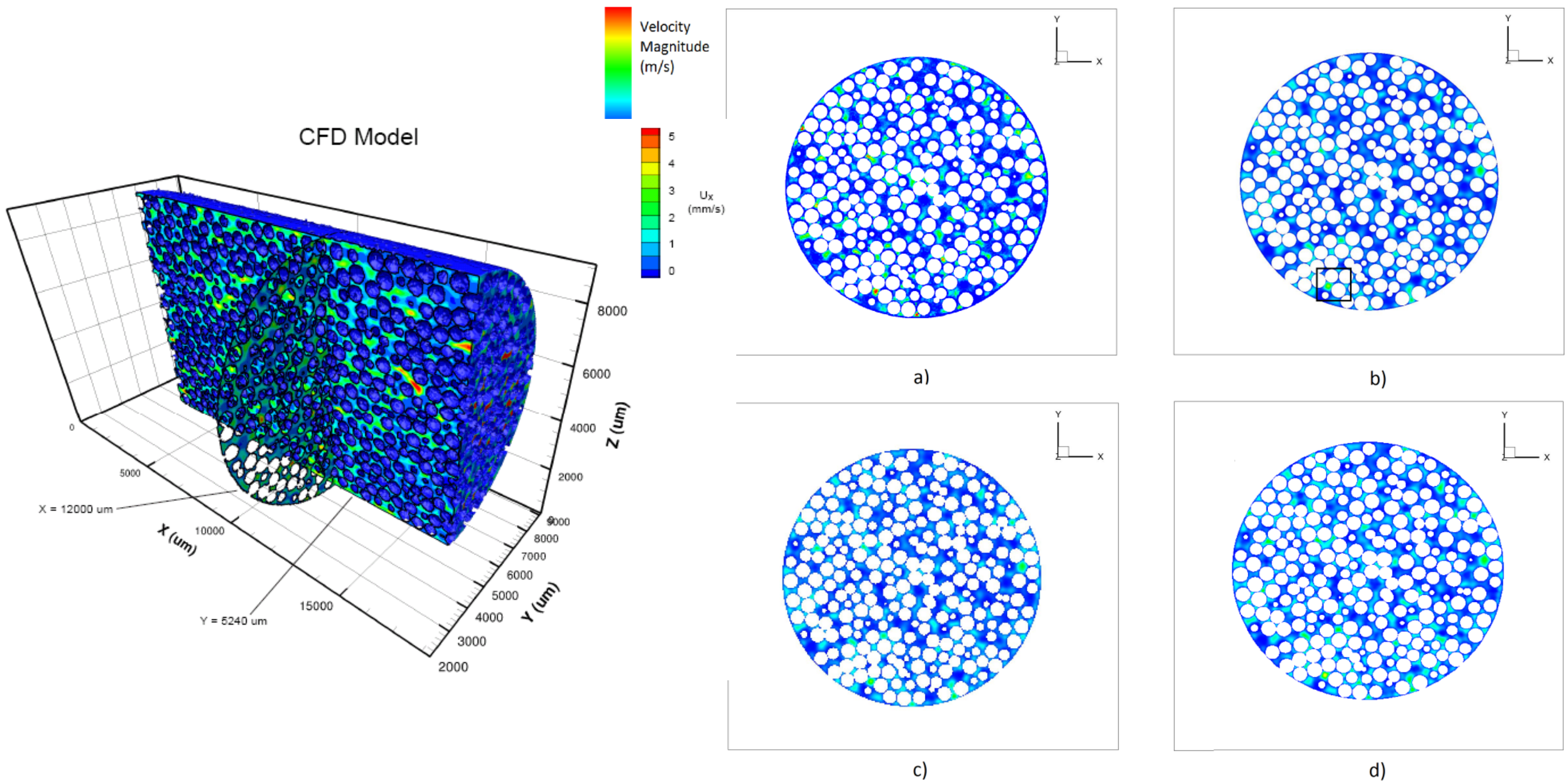
4000 cores on Hopper





Codes we use are…

- TETHYS: Pore-scale porous media flow and transport
  - Validation study with MRI



Current HPC Usage

- Machines currently used:
  - NERSC (2.5 M hours in 2012)
  - Chinook (EMSL) and Olympus (PNNL Institutional Computing) (< 1M hours in 2012)
- Concurrency, run time, # runs/year:
  - **eSTOMP:** typ. 100-1000 cores per run, O(1 day), many runs can be performed simultaneously for UQ, hundreds to thousands run/yr
  - **SPH**: typ. 1000-2000 cores per run, O(1 day), hundreds runs/yr
  - **Hybrid SPH/STOMP:** <100 cores per SPH, minutes turnaround, total allocation 1000 cores, 6 hours, < 100 runs/yr
  - **TETHYS:** 4000 cores per run, several days clock time, < 10 runs / yr

Current HPC Usage

- Data / memory requirements:
  - Data I/O and storage generally small
  - Memory requirements not limiting (or can be addressed with code efficiency)
- Necessary software, services or infrastructure
  - Workflow management tools for hybrid simulation (SWIFT)
  - Visualization (VISIT)
  - GA and PETSc

Memory

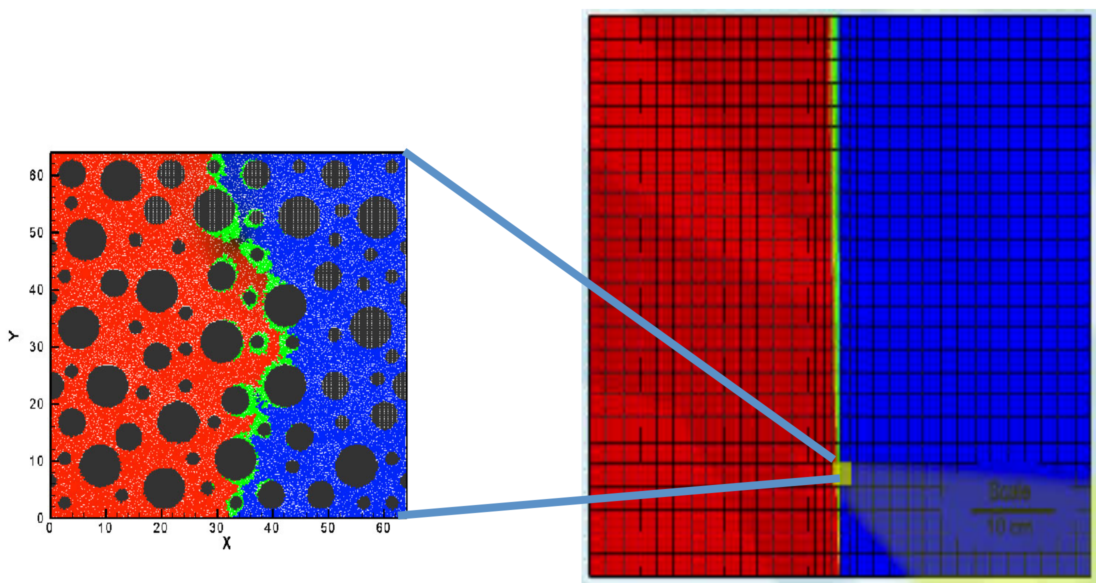
- per core memory demand limited the number of computational cores per node to 4
  - 52,800 process job had to allocate, but not use, an additional 105,600 processor cores
- software was modified to use only distributed arrays for the chemical species eliminated the temporary allocation of 102 field arrays in local memory
  - modification resulted in 1.72 GB less memory usage per core
- improved the on chip-processor utilization by a factor of three; permitted the problem to be solved with smaller processor counts.

Future HPC Usage

- At-scale codes are currently near maximum reasonable usage needs
  - Pore-scale simulation domain volumes are approaching "Darcy" scale from which macroscopic processes/parameters can be defined
  - Trying to simulate application-relevant domains with full pore-scale resolution is not a reasonable target in the foreseeable future
    - Many orders of magnitude (~10^{15}) scale gap (cm to km)
    - Couldn't meaningfully characterize at this scale anyway
  - x32 might be utilized through
    - More UQ
    - More complex microbial modeling (communities with many functional groups)
      - eSTOMP factor of 10 increase for a single in-silico species model
    - More coupling, complex processes
      - Multiphase flow, geomechanical processes
      - Larger domains (CO_2 vs. contaminant plumes)

Hybrid Multiscale Simulation

- A more interesting and potentially transformative approach is a new paradigm for subsurface modeling – directly coupling pore- and continuum-scale codes in a single simulation domain
  - Spans scale gap between fundamental process representations and applications
  - Maintains reasonable efficiency
  - Takes advantage of multiple levels of concurrency



Micromodel Experiments

- Mixing-controlled calcium carbonate precipitation (Zhang et al., ES&T 44(20), 2010).

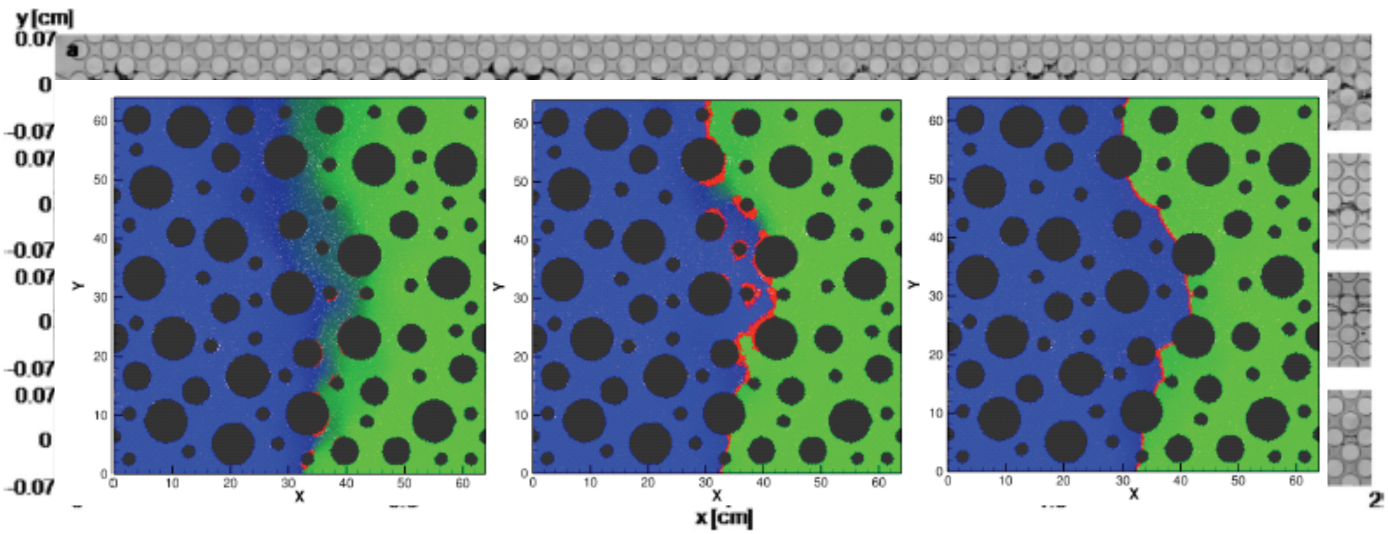
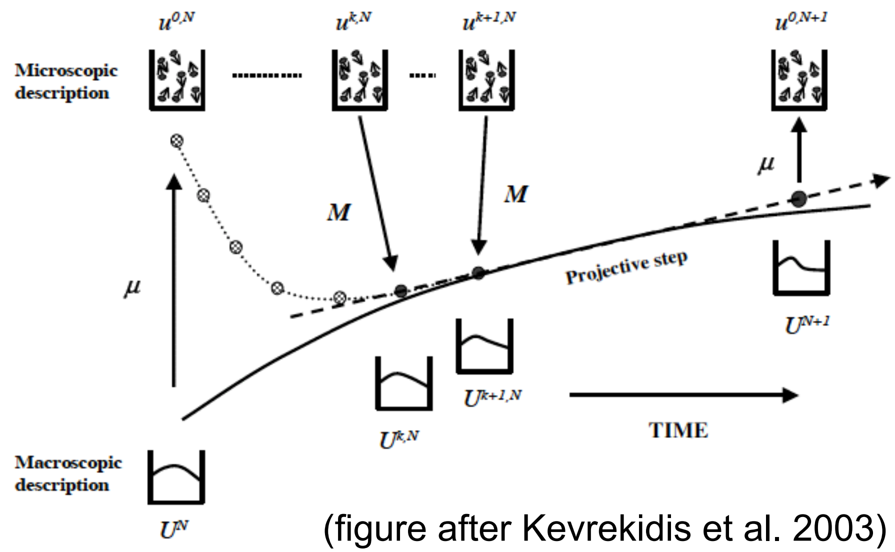


FIGURE 2. Images of center of micromodels with $CaCO_3$ precipitates formed along the mixing zone at different saturation states (a) $\Omega_c/\Omega_v = 3.4/2.8$, (b) $\Omega_c/\Omega_v = 3.8/3.1$, (c) $\Omega_c/\Omega_v = 4.6/3.9$, and (d) $\Omega_c/\Omega_v = 5.2/4.5$.

Hybrid Multiscale Simulation

- Multiscale dimension reduction approach
  - Reduce degrees of freedom (number of time steps) solved in microscale simulation by iterating between microscale and macroscale
  - Perform numerical closure on microscale with short bursts of pore-scale simulation where insufficient general closure exists

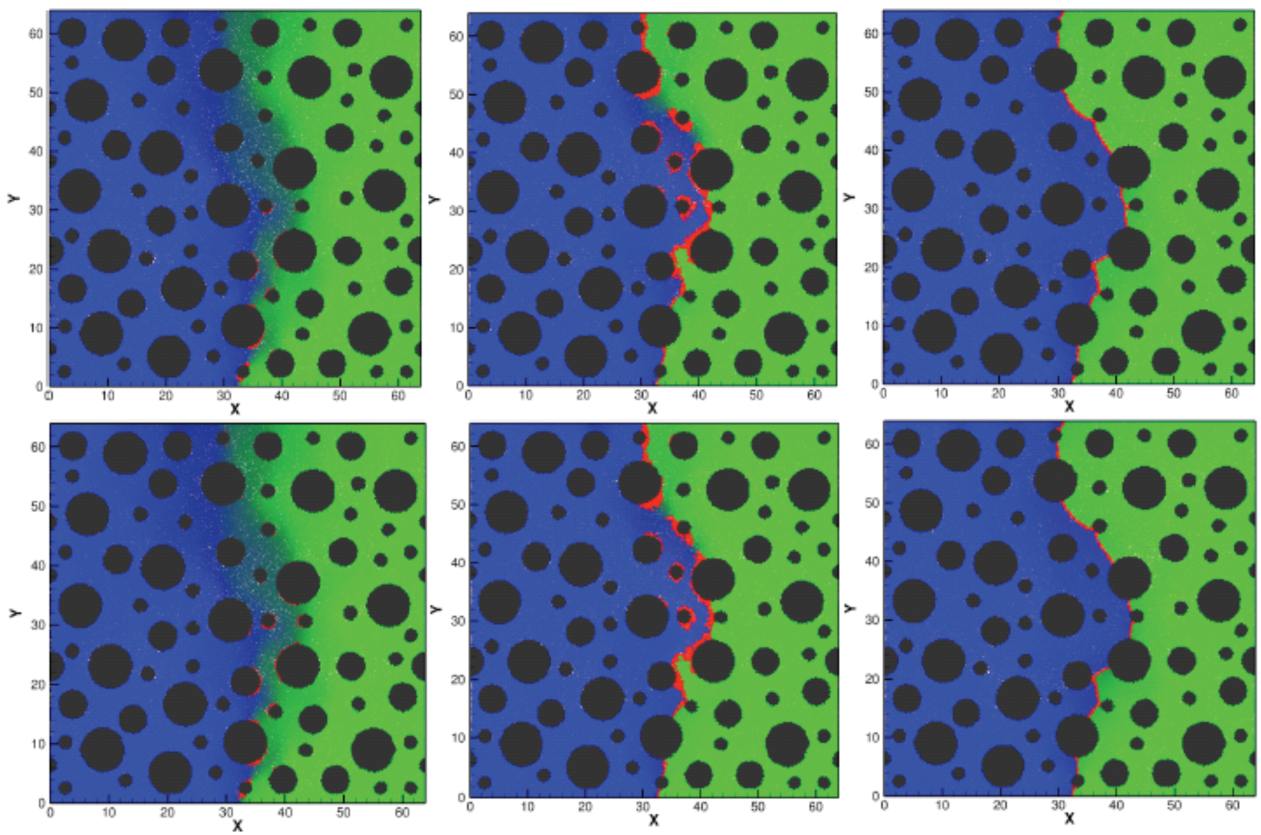


(figure after Kevrekidis et al. 2003)

Tartakovsky and Scheibe, *Advances in Water Resources*, 2011

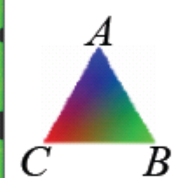
Hybrid Multiscale Simulation

- Multiscale dimension reduction approach



Complete Pore-Scale Solution

Dimension Reduction Solution

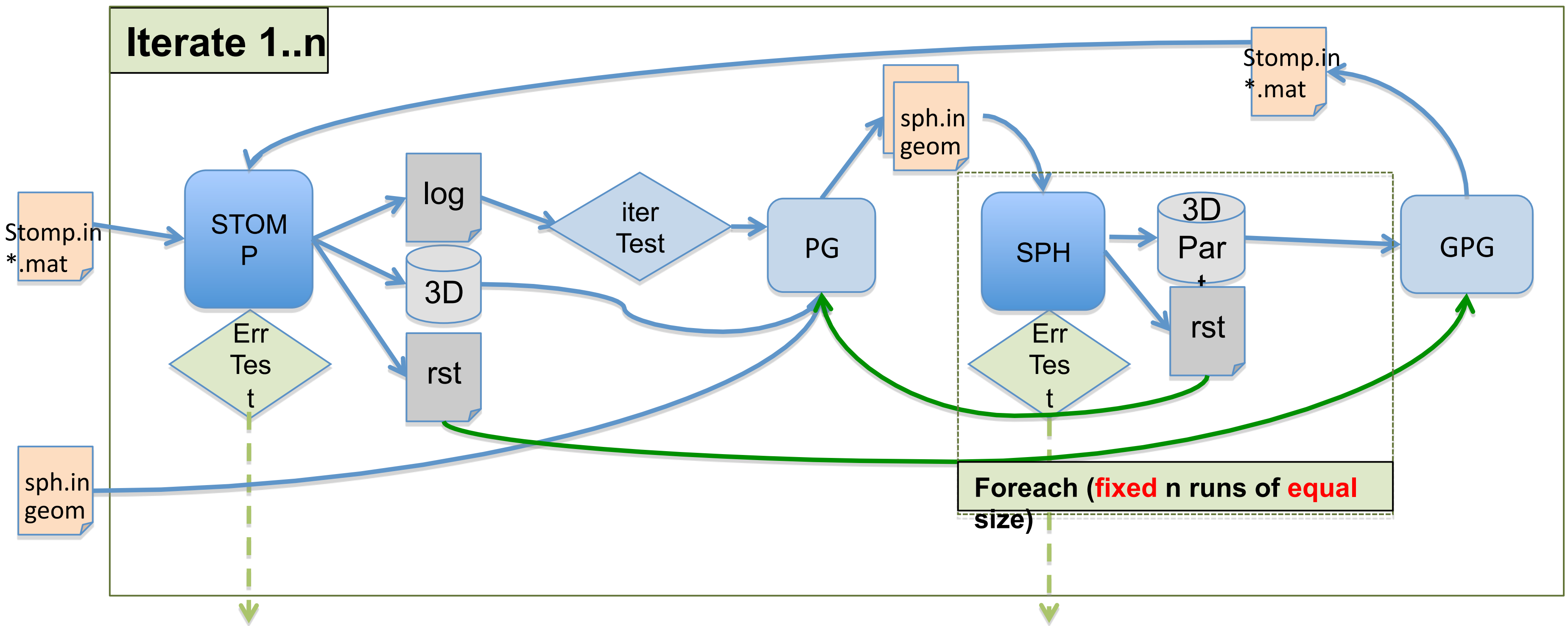


$t_d = 27.5$ $t_d = 223.3$ $t_d = 1112.5$

Tartakovsky and Scheibe, *Advances in Water Resources*, 2011

Hybrid Multiscale Simulation

- Current work: Put into the context of many possible pore-scale subdomains in a focused region with adaptivity



Uses SALSSA workflow environment and SWIFT job management tools

Future HPC Usage – Multiscale Hybrid

- Compute hours needed
  - Could effectively use x32 to make significant advances
- Changes to parallel concurrency, run time, number of runs per year
  - Multiple levels of concurrency
  - Run times and number of runs comparable, but each run would involve many “sub-runs”
- Changes to data read/written
  - I/O during simulation larger but long-term storage still small
- Changes to memory needed
  - Not significantly different
- Changes to software/services/infrastructure required
  - Workflow management tools critical
  - Visualization during simulation

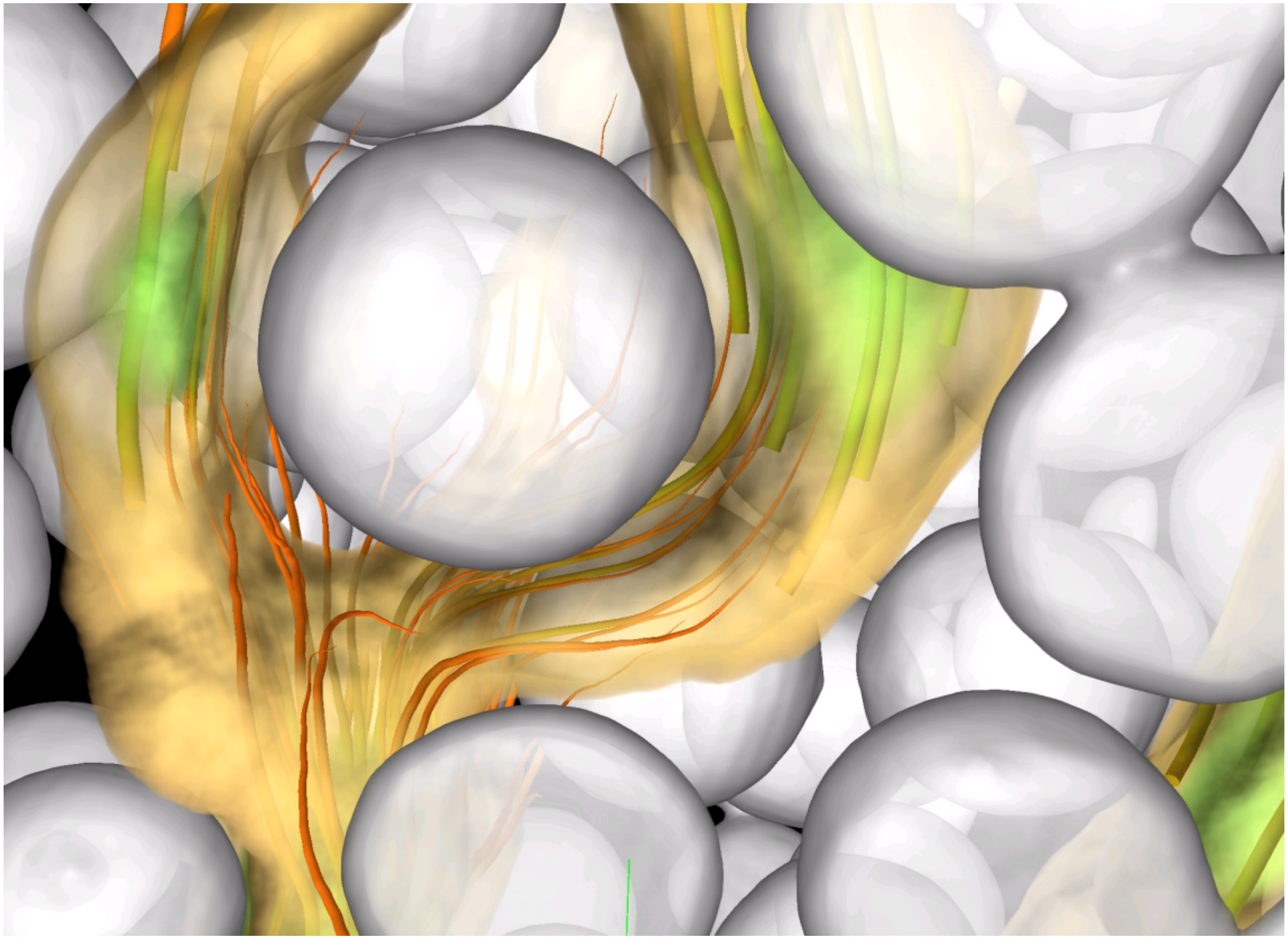
Strategies for New Architectures

- Our strategy for running on new many-core architectures (GPUs or MIC) is …
  - Poorly defined but under development
  - SPH may become more attractive under new architectures
- To date we have prepared for many core by …
  - Collaborating with computational scientists under PNNL eXtreme-Scale Computing Initative to perform testbed studies
- We are already planning to do …
- To be successful on many-core systems we will need help with
  - Updated programming models on which we heavily rely
  - E.g., will Global Arrays work well on new architectures, or be revised to do so?

Summary

- What new science results might be afforded by improvements in NERSC computing hardware, software and services?
  - New approach to multiscale simulation of subsurface processes
  - Move from parameterized phenomenological models to mechanistic process-based predictive models
- What "expanded HPC resources" are important for your project?
  - Programming models for new architectures
  - Workflow management and visualization tools

Questions?



Pacific Northwest
NATIONAL LABORATORY