Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

A Many-Task Parallel Approach for Multiscale Simulations of Subsurface Flow and Reactive Transport

TIMOTHY SCHEIBE
XIAOFAN YANG
KAREN SCHUCHARDT
KHUSHBU AGARWAL
JARED CHASE
BRUCE PALMER
ALEXANDRE TARTAKOVSKY

PACIFIC NORTHWEST NATIONAL LABORATORY
RICHLAND, WA

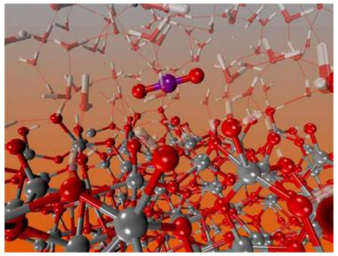
Important Modeling Applications in Environment and Energy

- remediation

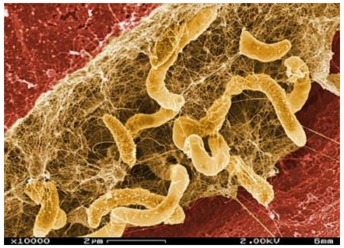
A critical challenge:

- Large-scale phenomena are driven by **coupled heterogeneous** processes at very small scales

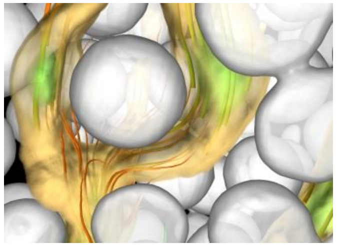
Molecules

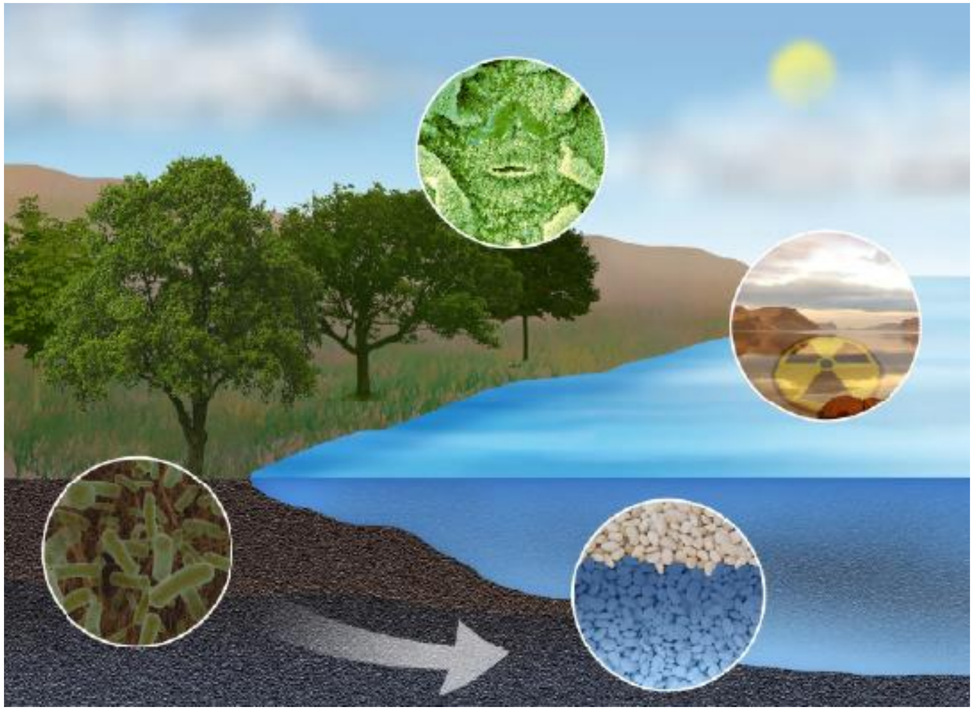


Cells



Pore Spaces





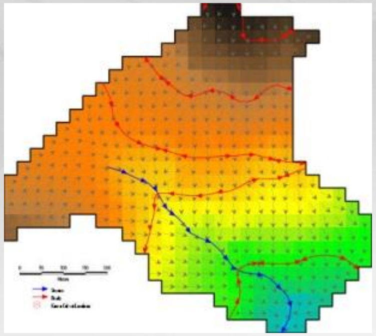
Communities, Ecosystems, Watersheds …

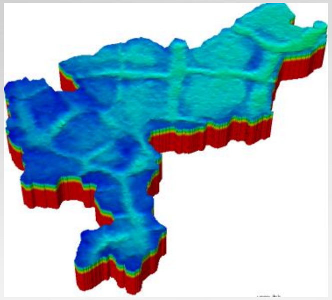
Dynamically coupled models at multiple scales

Complex Systems Models

Highly Parameterized, Coarsely Resolved (Km)

Large-scale drivers (e.g., climate change, land use)





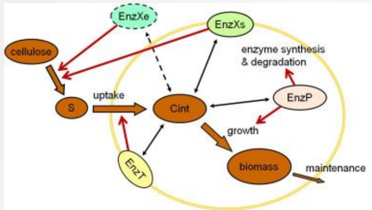
Watershed and River Hydraulics Models
Land Surface Biogeochemical Models

Super-parameterization

System Models

Parameterized, Intermediate Resolution (m)

Initial and Boundary Conditions, System-Scale Interactions

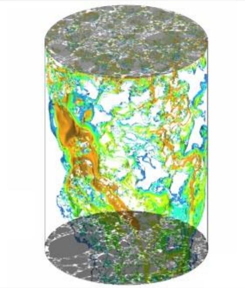


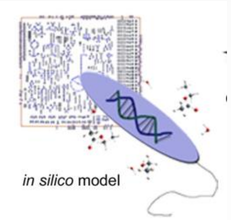
Aquifer Flow and Reactive Transport Models
Single Crop – Plot Models
Microbial community models

Process conceptualization and parameterization (static or dynamic)

Process Models

Independently Specified Characteristics, High Resolution (µm)





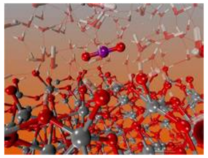
Pore-Scale Flow and Reactive Transport Models
Genome-scale Metabolic Models
Microbial Community Network Models
Individual Plant/Root Models

Inform Process Understanding, Characterization

Specify Experimental Conditions

Fundamental Theory/Experiment

***Ab initio* studies, Molecular Resolution**



Atomistic and Molecular Models

Hybrid Multiscale Simulation

Dynamically couple microscale and macroscale models where/when needed, use static models where appropriate (optimal bottom-up/top-down)

- Classifies multiscale approaches into Motifs based on the nature of the problems addressed
- Motifs contain specific algorithms/methods for multiscale coupling

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

WHY HYBRID COUPLING ?

CONTINUUM-SCALE METHODS

- Assume complete mixing below grid resolution
- Represent subsurface phenomena in an average sense, e.g. flow and transport
- Parameters may not represent intrinsic properties

PORE-SCALE METHODS

- Provide details on particle level interactions
- Adequate for modeling fractures, precipitation, etc
- Cannot scale to realistic domains which contain billions of particles

Too expensive!!!

A COUPLED MODEL

- Provide balance b/w computation time and model accuracy
- Use macro-scale methods to identify regions of high reactivity
- Use micro-scale methods to study “reactive” grid cells

Hybrid Multiscale is Well-Suited to a Many-Task Computing Approach

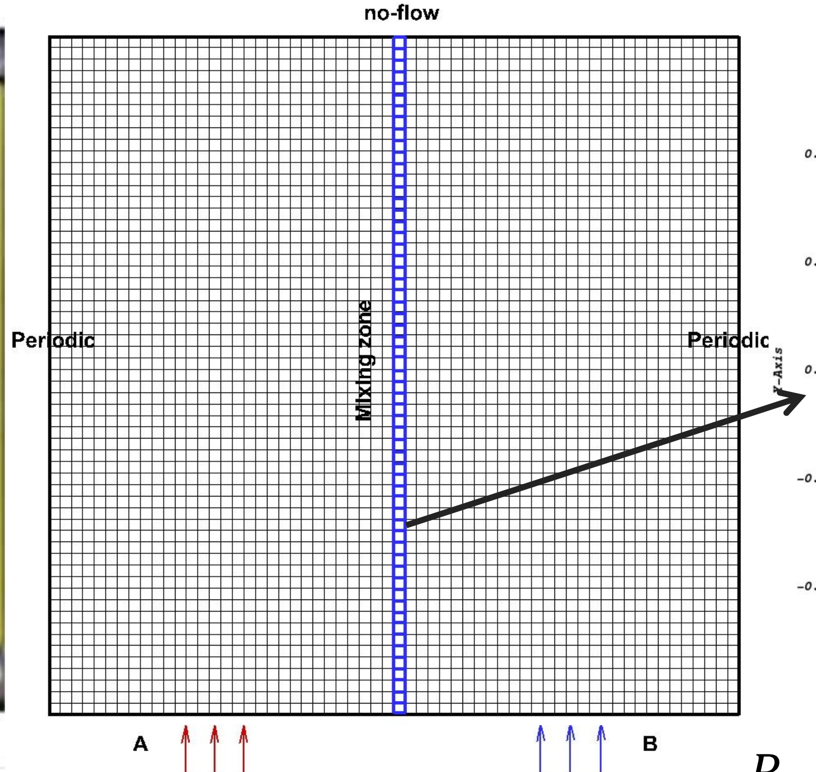
- Designed to loosely couple multiple codes at different scales through script-based workflow
- Multiple levels of concurrency
  - Each individual code can be parallelized in traditional manner (e.g., domain decomposition)
  - Multiple sub-domains are simulated at the microscale concurrently at each time step (parallel with minimal or no communication)
  - Uncertainty quantification may require multiple alternative realizations (embarrassingly parallel)
- Important to devise means of making microscale simulations independent of each other

TARGET PROBLEM

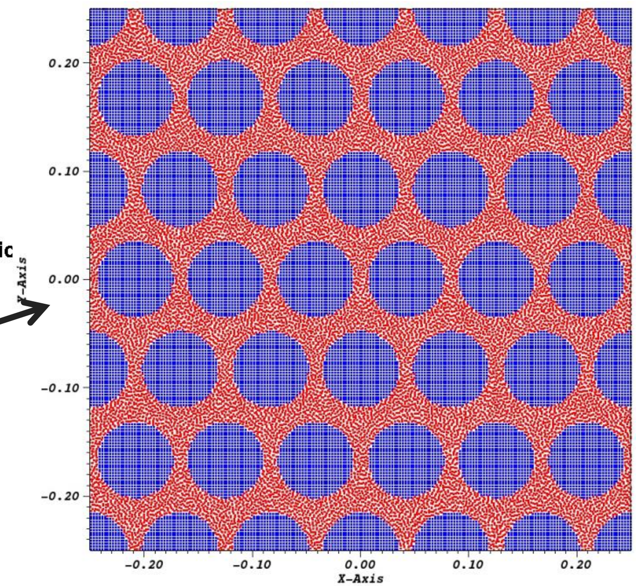
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Experiment

Hybrid Simulation



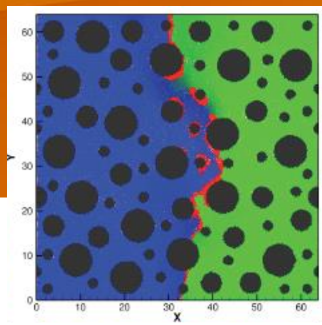
Pore-scale Simulation

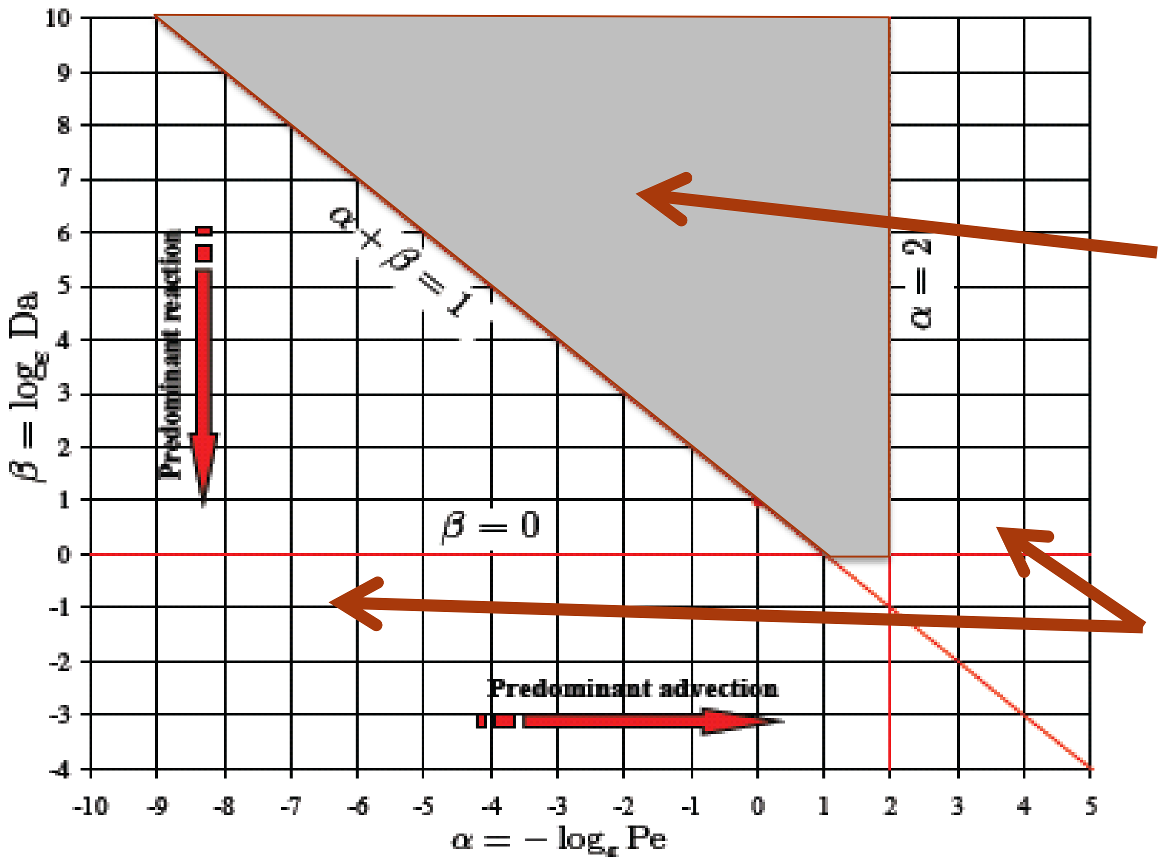


$$R_{cont} = \frac{d[C]}{dt} = k_{f,eff}[A][B]$$

- Mixing controlled kinetic precipitation reaction
  - Two solutes (A, B) are injected in a porous medium domain
  - Study formation of precipitate C
- Homogeneous medium
  - Fixed or variable number of equal size pore-scale simulations

Adaptivity Criteria





Appropriate level of complexity: **Darcy-scale simulation with effective process models and parameters**

Appropriate level of complexity: **Pore-scale resolution of flow and reactive transport**

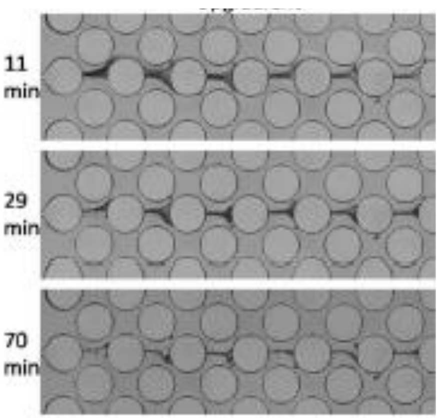
Battiato, I. and D. M. Tartakovsky, "Applicability regimes for macroscopic models of reactive transport in porous media," *Journal of Contaminant Hydrology*, 120-21(SI):18-26, 2011.
Battiato, I., D. M. Tartakovsky, A. M. Tartakovsky, and T. D. Scheibe, "Hybrid models of reactive transport in porous and fractured media," *Advances in Water Resources*, 34(9): 1140-1150, doi:10.1016/j.advwatres.2011.01.012, 2011.

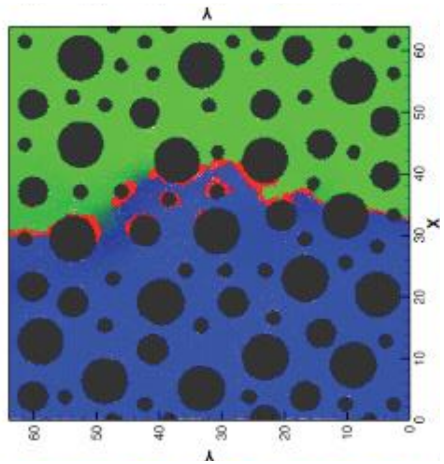
Hybrid Multiscale – Hierarchical Method

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

- For each microscale subdomain, there is a corresponding grid cell at the macroscale.
- We use a variant of Heterogeneous Multiscale Method called Dimension Reduction with Numerical Closure (Tartakovsky and Scheibe, *Advances in Water Resources*, 2011)





$$\frac{dC_i}{dt} = k_r\left\{\frac{A_i B_i}{K_{sp}} - 1\right\}$$

$$\frac{d\boldsymbol{C}}{dt} = k_r\left\{\frac{[\boldsymbol{AB}]}{K_{sp}} - 1\right\}$$

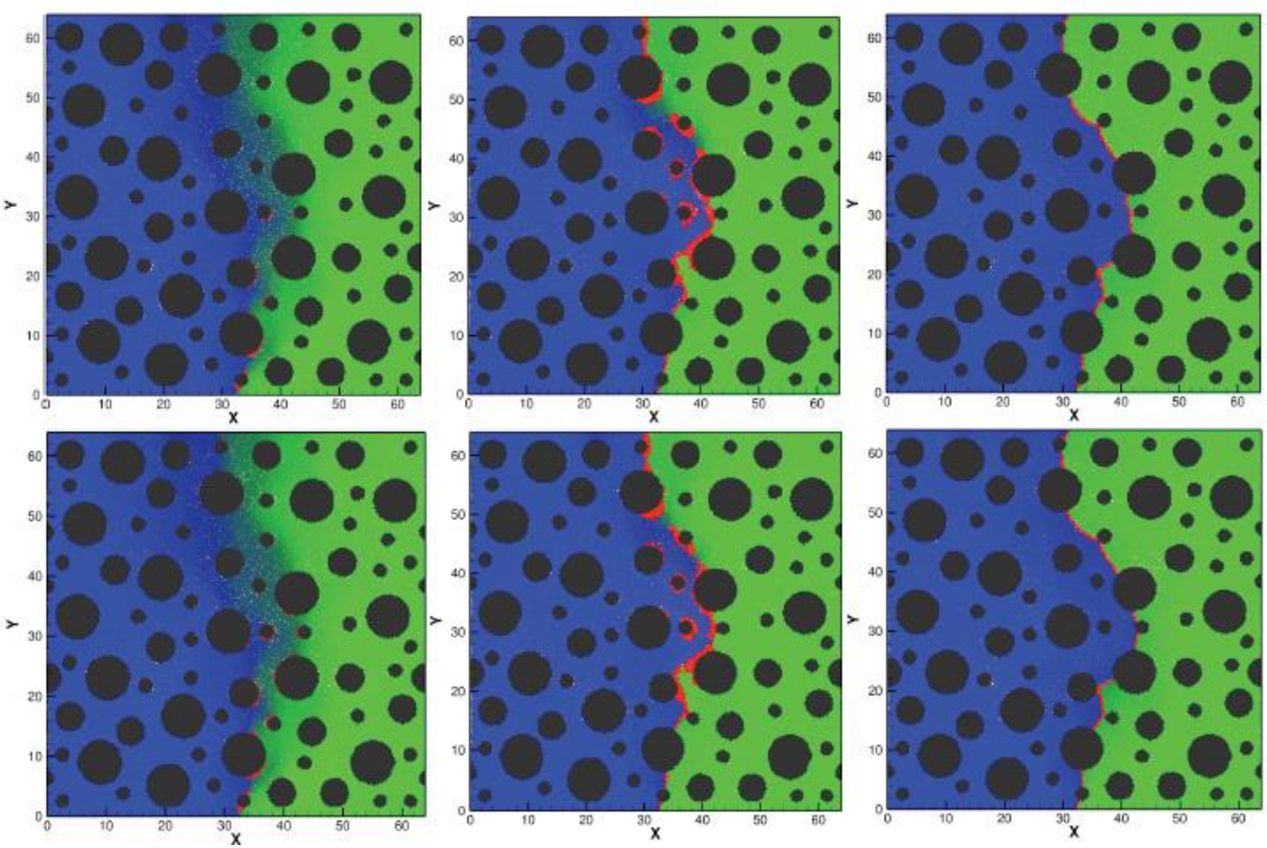
[AB] = [A][B] ?

$$m = \frac{\overline{AB}}{\bar{A}\bar{B}} = N_p \frac{\sum_{i=1}^{N_p} A_i B_i}{\sum_{i=1}^{N_p} A_i \sum_{i=1}^{N_p} B_i}$$

$$\frac{d\boldsymbol{C}}{dt} = k_r\left\{\frac{m\boldsymbol{AB}}{K_{sp}} - 1\right\}$$

Hierachical Hybrid Multiscale Method

Dimension Reduction with Numerical Closure



Complete Pore-Scale Solution

Dimension Reduction Solution

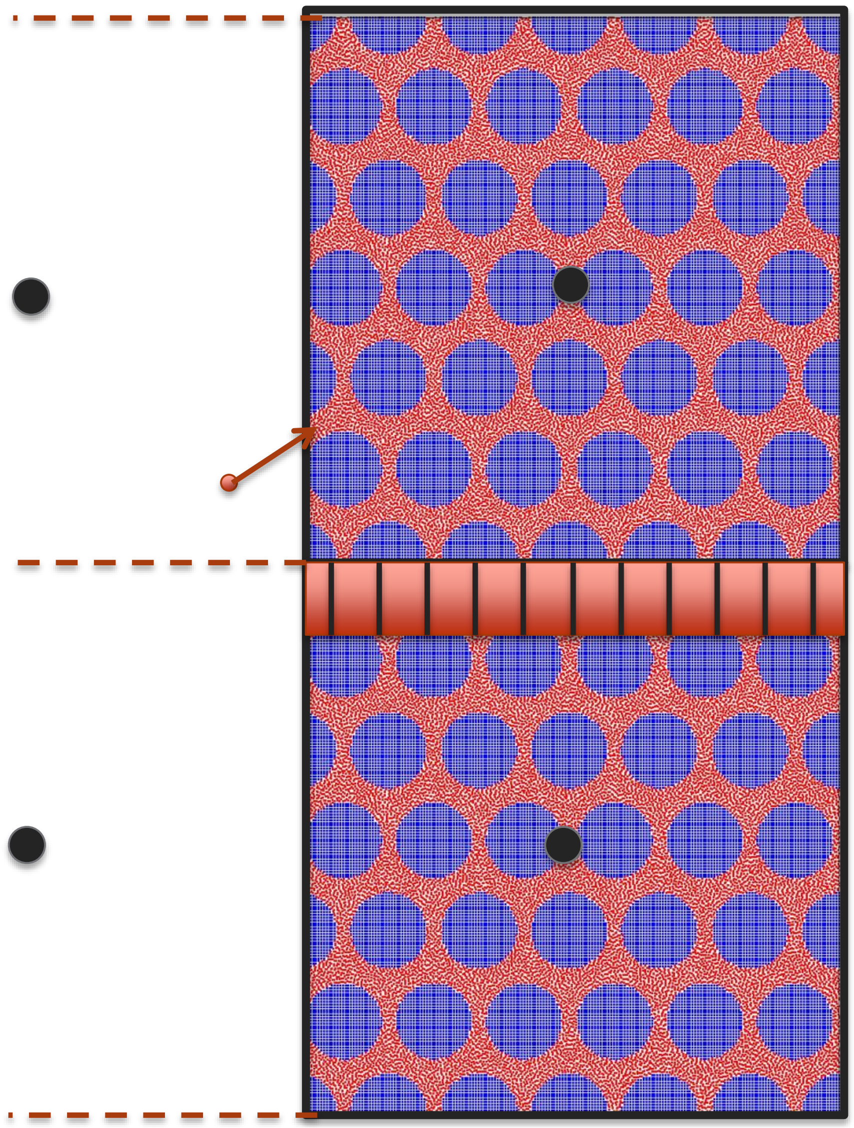
$t_d = 27.5$ $t_d = 223.3$ $t_d = 1112.5$

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

Concurrent Multiscale Simulation

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- We want each microscale subdomain to be maximally independent of other subdomains
  - Preserve loose coupling, non-invasive to code
  - Minimize or eliminate communication
- In reality, the subdomains are not independent; linked through boundary conditions



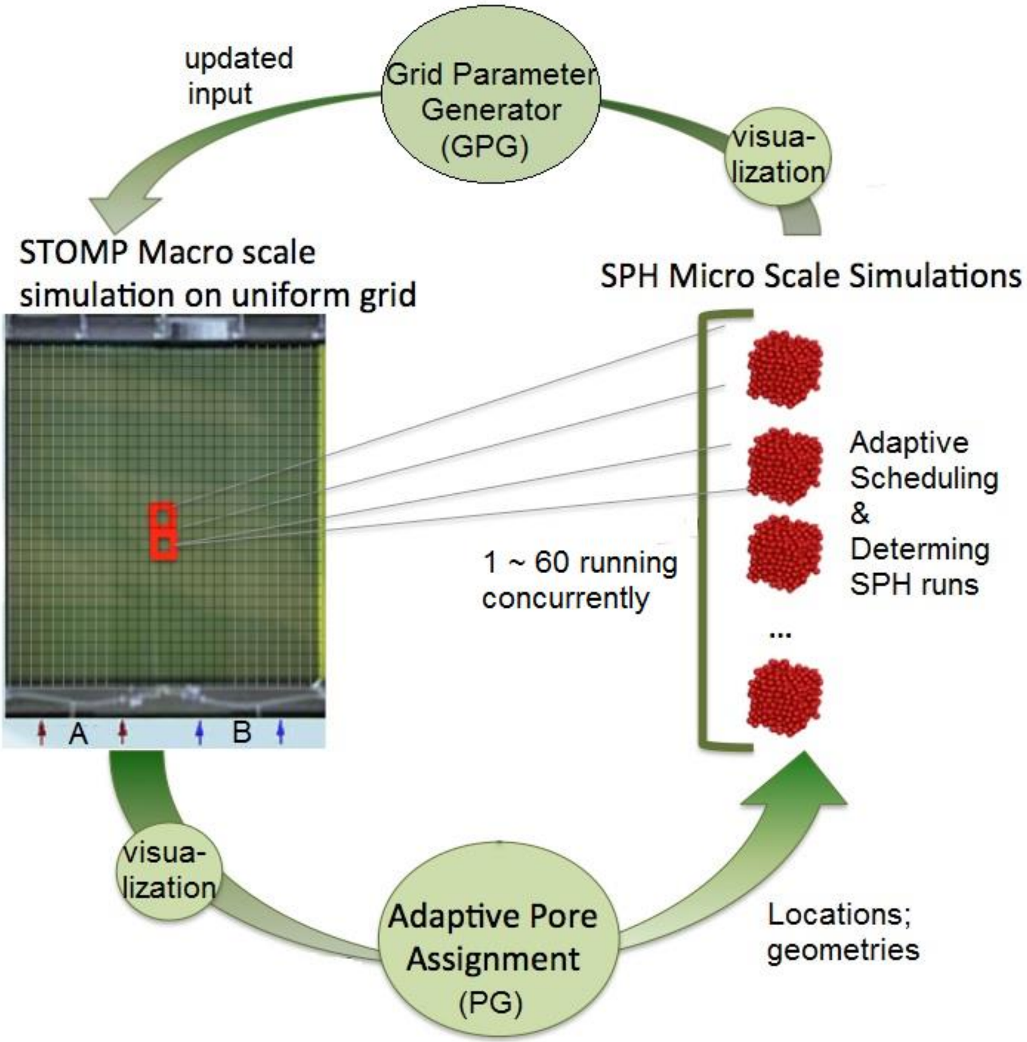
USING SWIFT

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Swift provides scripting language for distributed parallel computing
  - An inherently parallel execution model
  - Simple interface definitions
  - Internal error handling
  - Simple C-like syntax
  - Portable
- Runtime system for Swift script
  - Translate programs into task graphs
  - Schedule, monitor, execute task graphs
    on local clusters and/or distributed grid resources
  - Annotate data products with provenance metadata

Scalable	++
Typing	++
Iteration	++
Scripting	++
Dataset Mapping	+
Service Interop	+
Subflow/comp.	+
Provenance	+
Open source	+

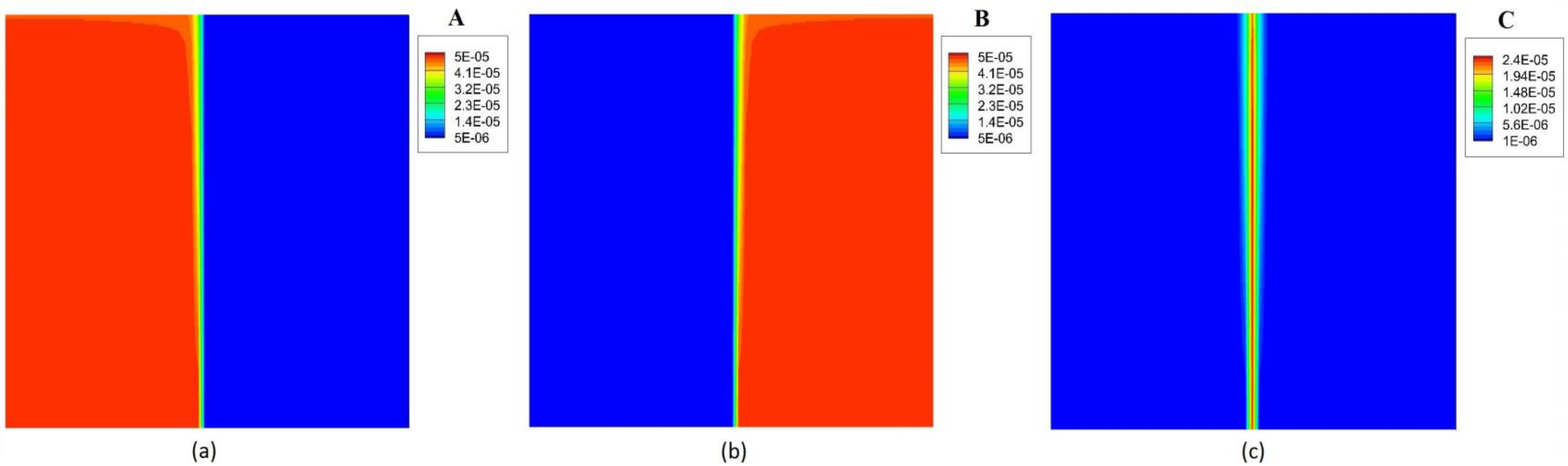
HYBRID MODEL WORKFLOW



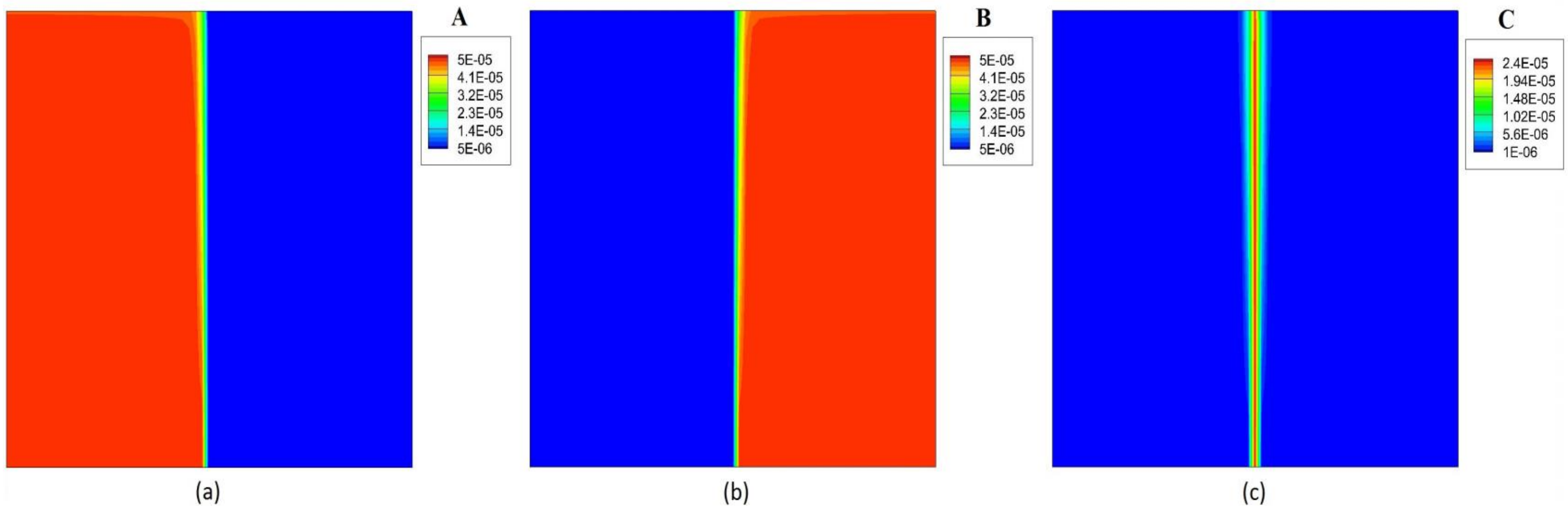
- Single STOMP/ iteration
- Multiple iterations of the coupled-model executed in the workflow
- Within each iteration, multiple independent pore-scale (SPH) simulations executed simultaneously
  - Each SPH run is itself parallel
  - Many Parallel Task Computing Paradigm
  - The number and size of SPH domains vary across iterations

HYBRID vs. SINGLE-SCALE MODEL

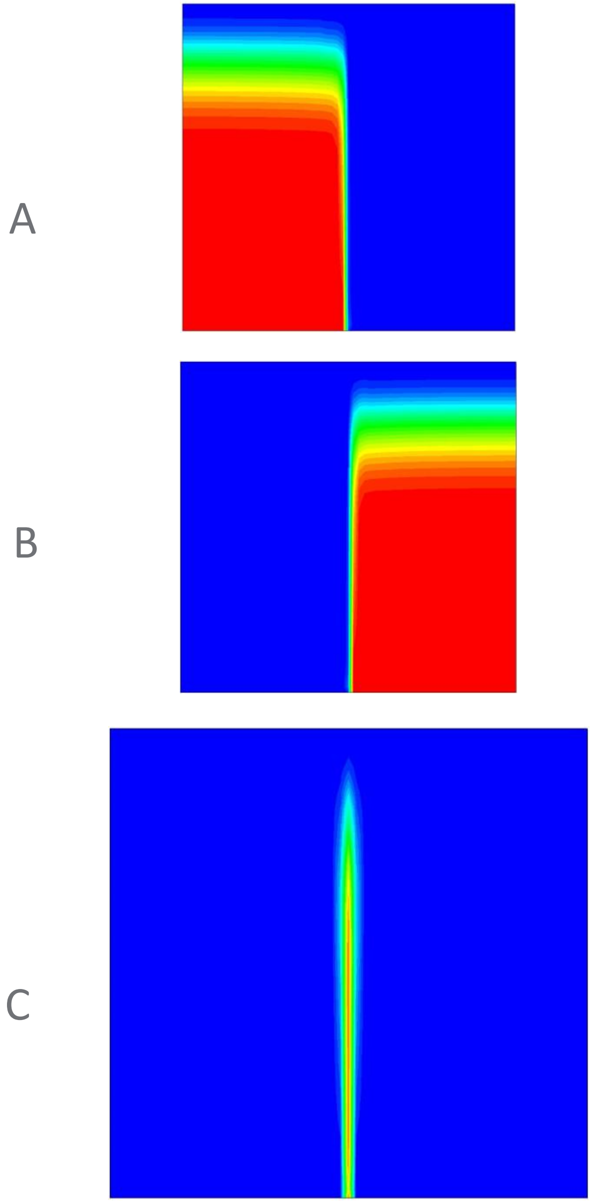
Hybrid

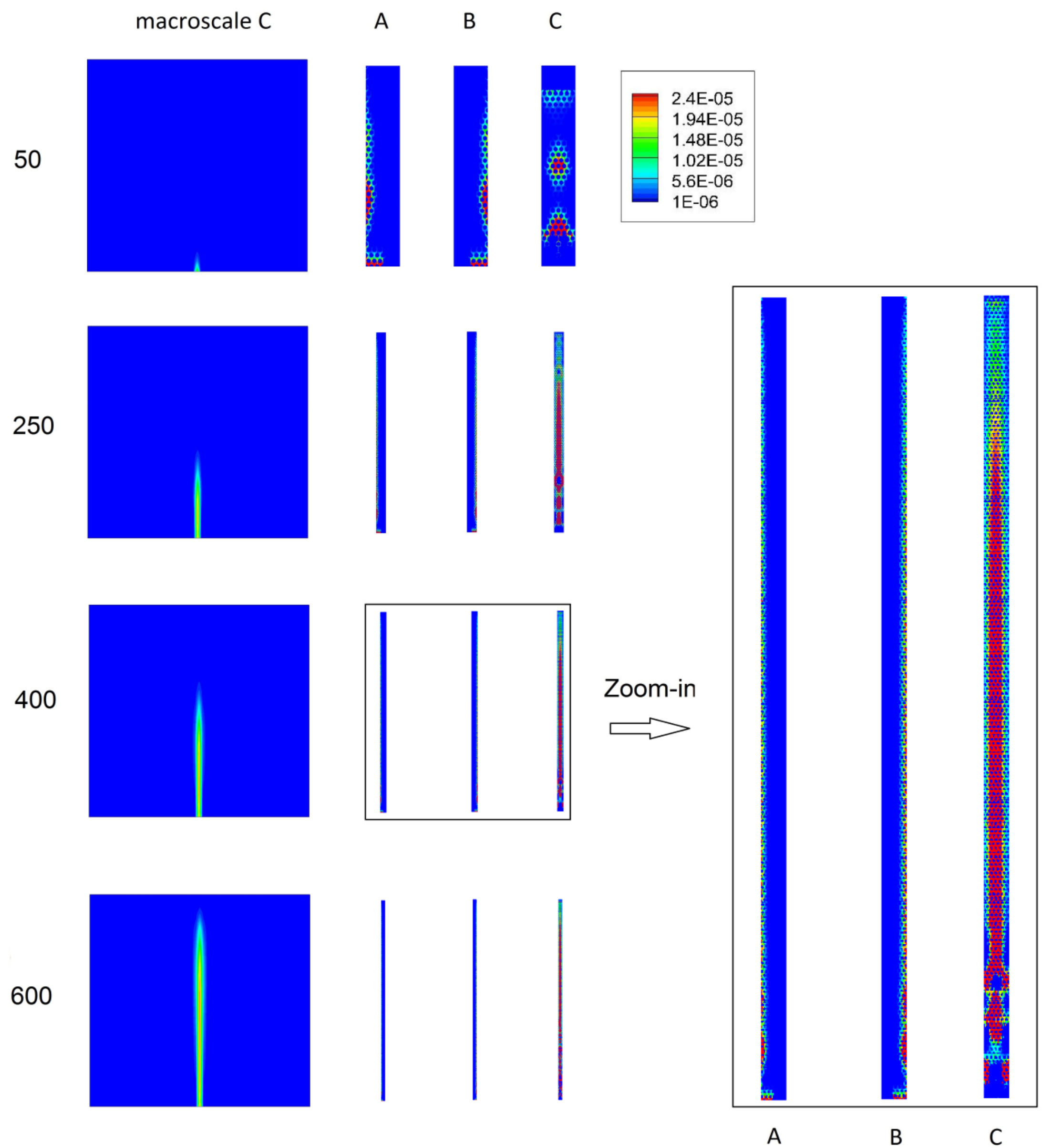


Single-scale

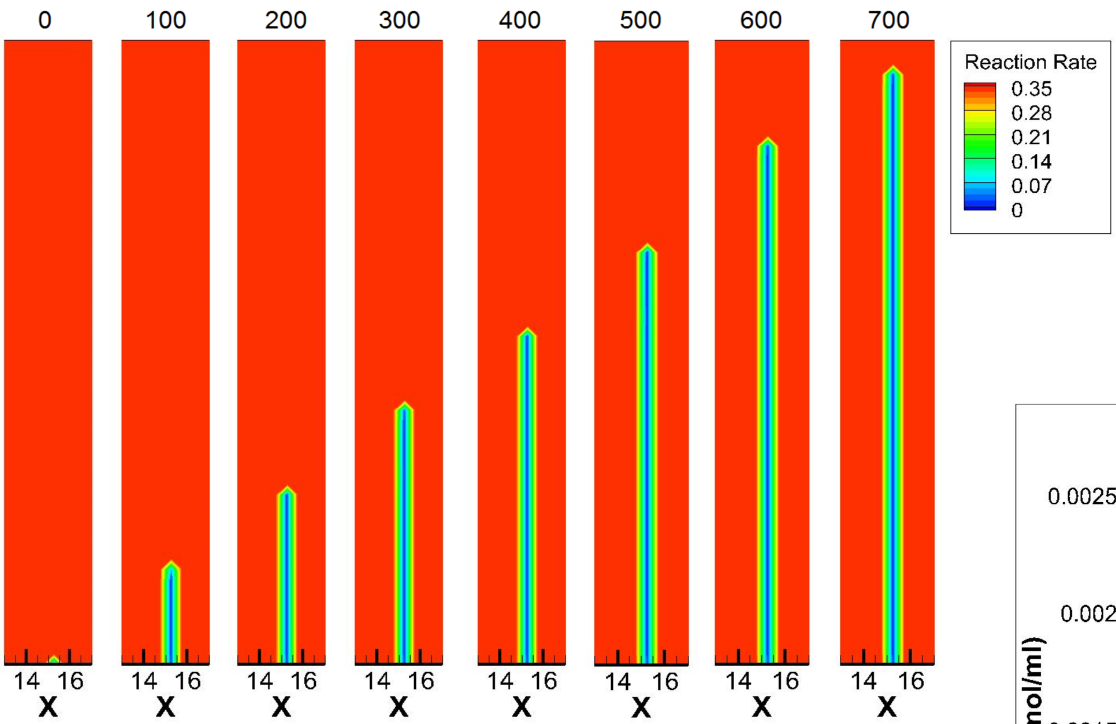


MACRO- vs. MICRO-

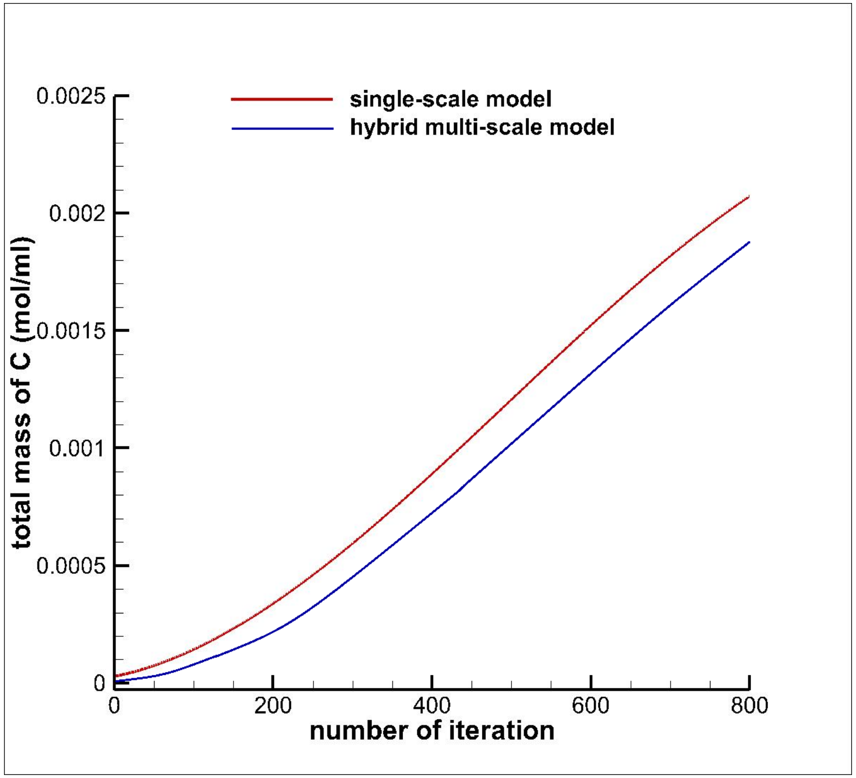




REACTION RATE & TOTAL MASS



- Due to the incorrect assumption of complete mixing at the grid scale in the single-scale model, the effective rate of reaction is too high and the amount of reaction product generated is over-estimated by ~15% percent.

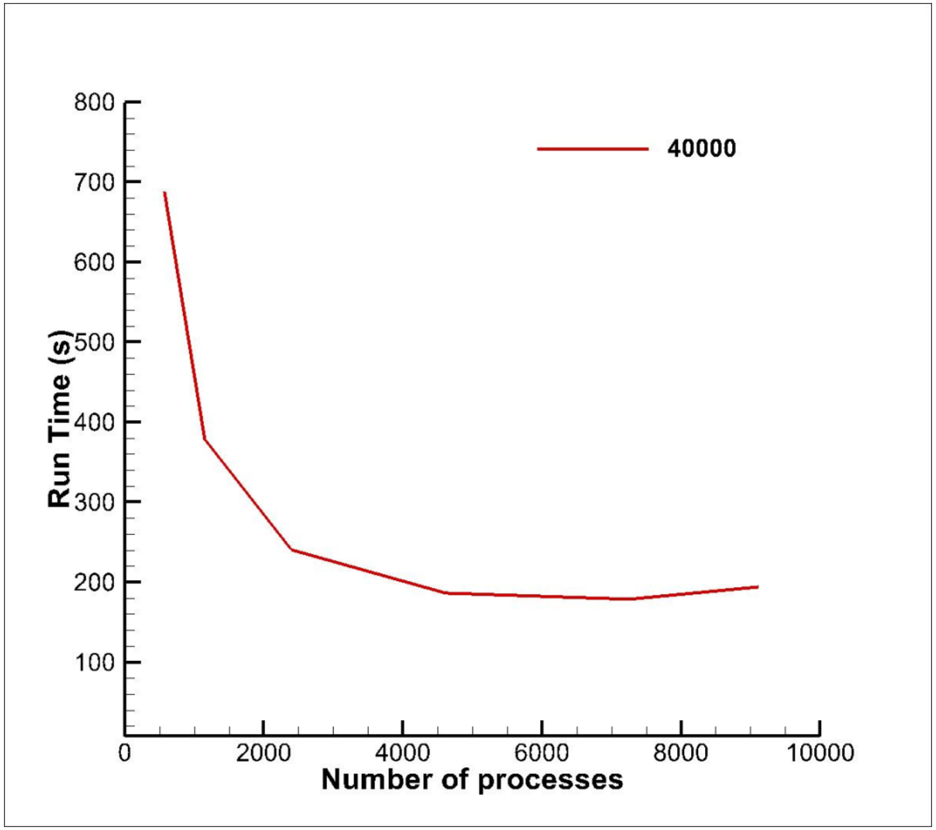


- Mixing limitations at the pore scale reduce the effective reaction rate at the macroscale significantly along the vertical mixing interface,
- Reduced rates propagate upward through the system as the simulation proceeds

PERFORMANCE REVIEW

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

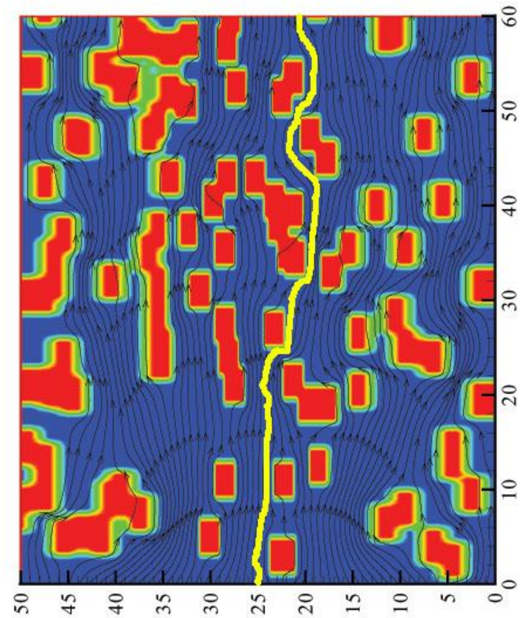
- **Computational domain**
  - 31cm x 30cm
  - 61 x 60 cells
  - 0.5 cm grid spacing
  - 40,000 SPH particles per simulation
- **Execution**
  - 1536 nodes (24 cores each) on NERSC Hopper
  - ~96 hr wall time
  - ~85% on SPH
  - Min 1152 processors (48 nodes) per SPH
  - Min 2304 processors (96 nodes) per SPH
  - ~800 iterations
  - I/O not an issue



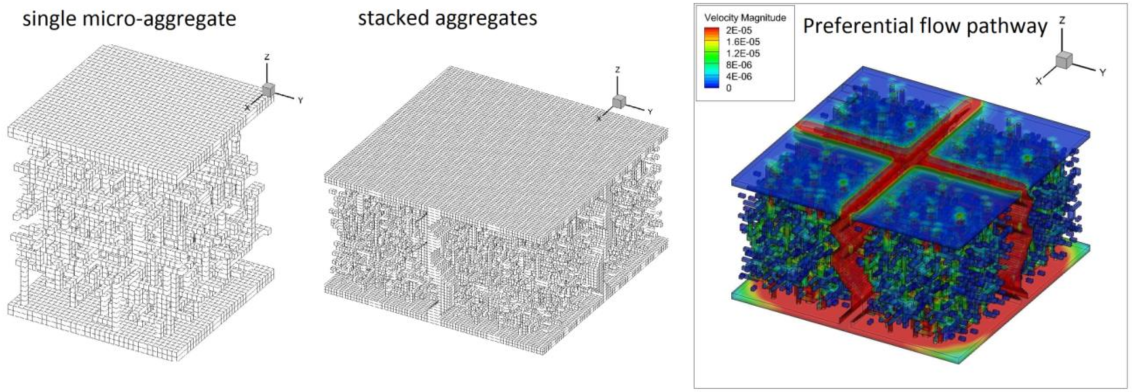
CHALLENGES

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

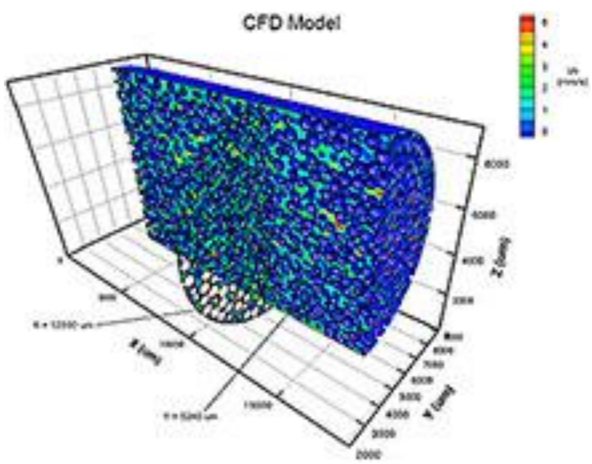
- Swift related issues
  - How to handle file collections
  - Input file staging
  - Dynamic assignment of variables
- Visualization not included inside the workflow
  - Computationally inefficient
- Next steps:
  - Full mineral precipitation/dissolution reaction
  - Better adaptivity criteria applied to inhomogeneous porous media
  - Increase efficiency of lifting operator
  - Other applications



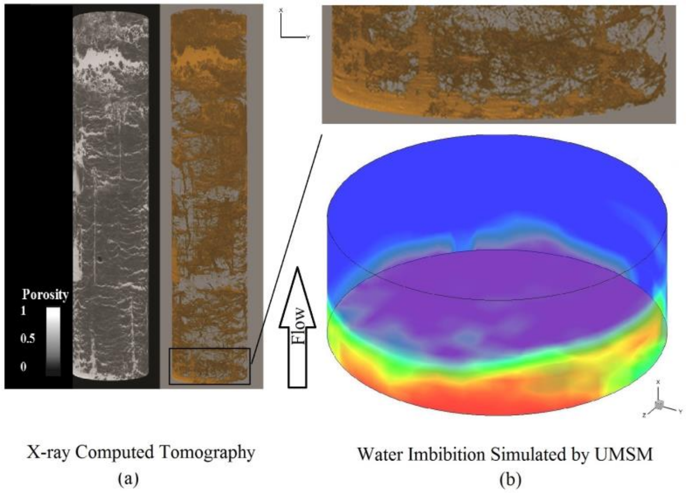
ADVANCES IN PORE-SCALE MODELING



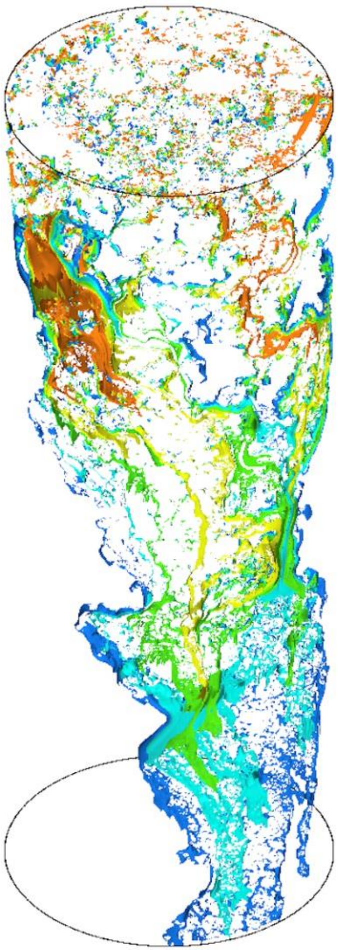
Yang, Scheibe, et al. 2014, Trans. Porous Media



Yang, Scheibe, et al. 2013, AWR



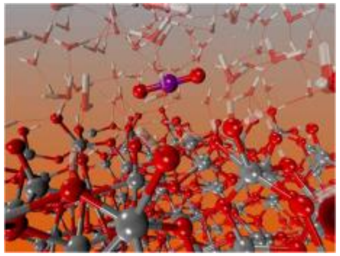
Yang, Liu, et al. 2014, SSSAJ

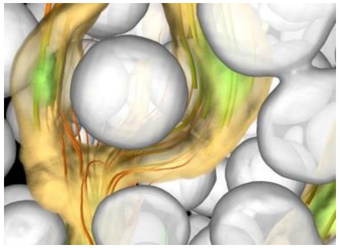


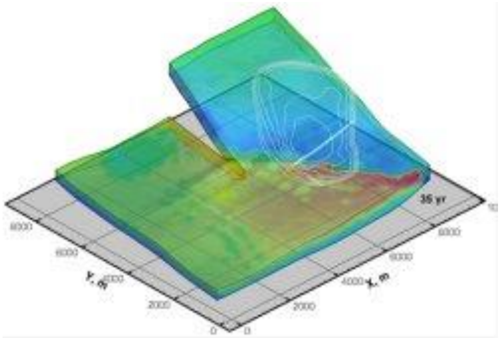
Scheibe et al., WRR, in review

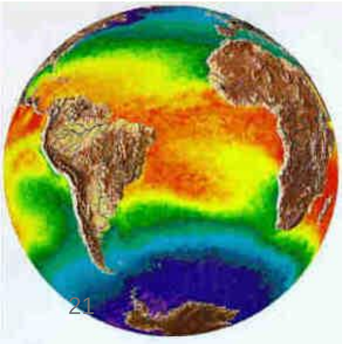
Concluding Remarks

- As fundamental process understanding at small scales increases, and need for better predictive models at large scales increases, the motivation for hybrid models that couple across scales is great.
- Multi-model coupling (hybrid multiscale modeling) is a good candidate for many-task workflow-based approaches
  - Loose coupling of model components
  - Minimally invasive
  - Multiple levels of concurrency
- We have demonstrated the efficacy of this approach for a specific application, and identified areas for future improvement









ACKNOWLEDGMENTS

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- DOE Office of Science, SciDAC
- DOE Office of Science, Biological and Environmental Research through the PNNL Subsurface Biogeochemical Research Scientific Focus Area (SFA) Project
- George Redden, Idaho National Laboratory

Battelle
The Business of Innovation

U.S. DEPARTMENT OF ENERGY | Office of Science

SciDAC
Scientific Discovery through Advanced Computing


THANK YOU