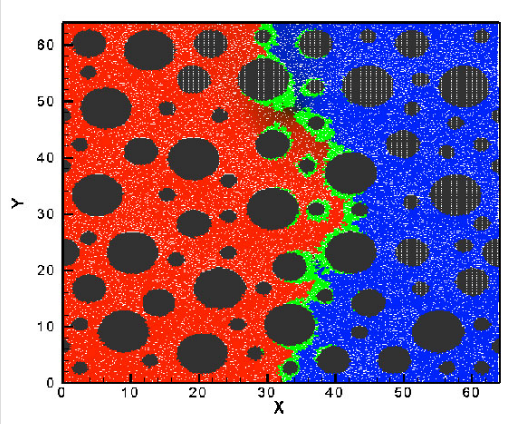
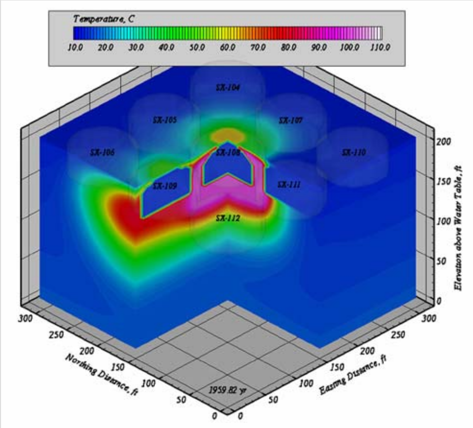
A. Project Overview

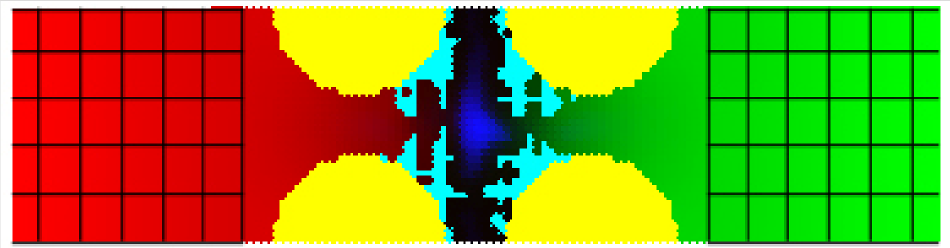
- Project Name: Hybrid Numerical Methods for Multiscale Simulations of Subsurface Biogeochemical Processes (http://subsurface.pnl.gov/)
- Sponsor: SciDAC-2 Science Application; BER / Environmental Remediation Sciences (New in SciDAC-2)
- Science goals: Develop an integrated multiscale modeling framework with the capability of directly linking different process models at continuum, pore, and sub-pore scales.
- Principal Investigator: Tim Scheibe (PNNL)
- Co-Investigators:
  - Alexander Tartakovsky, Phil Long (PNNL)
  - Daniel Tartakovksy (UC San Diego)
  - Paul Meakin, George Redden (INL)
  - Scott Brooks (ORNL)
- Project History: Started FY07 – Four year project
- Science Application Partnerships:
  - Process integration, data management, and visualization framework for subsurface sciences (Karen Schuchardt, PNNL)
  - Component-based framework for subsurface simulations (Bruce Palmer, PNNL)

B. Science Lesson

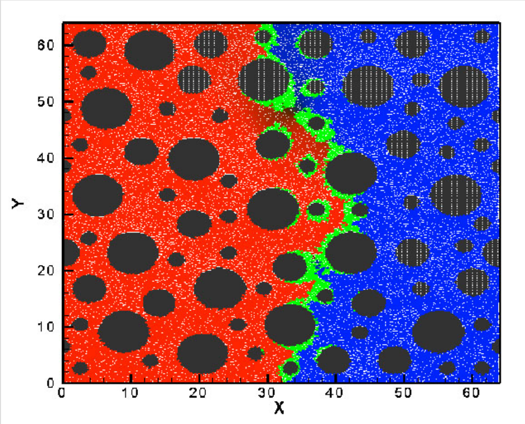
- Pore scale: Smoothed Particle Hydrodynamics (SPH).
  - Solves Navier-Stokes equations using discrete particle method (grid-free)
- Porous medium scale: STOMP Finite Volume model.
  - Finite-difference PDE approximation solved using PETSc (linear and non-linear iteration)
- Coupled pore- and continuum-scale models
  - Pore-scale only where mixing and reaction are occurring.

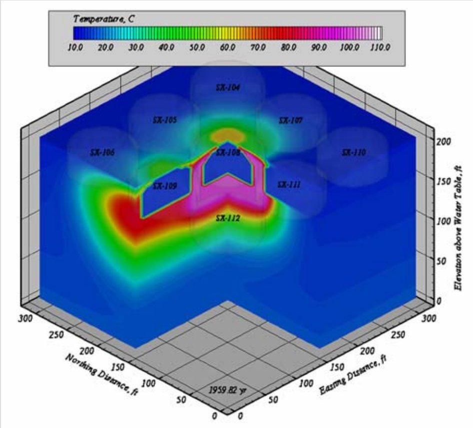


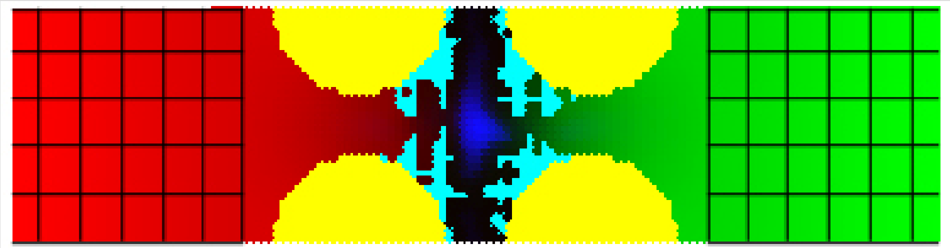




sson





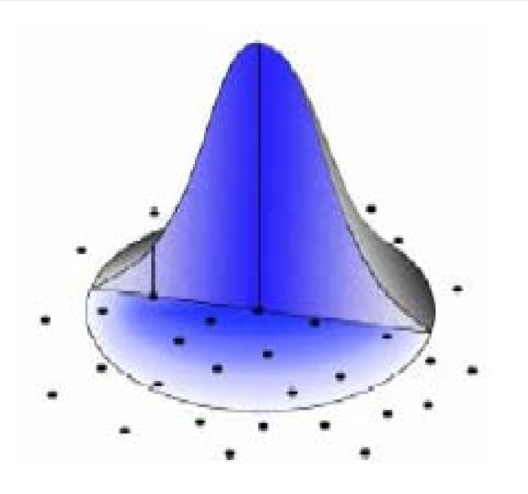


C. Parallel Programming Model

- STOMP (standalone): Preprocessor applied to directives in F90 code (joint serial/parallel code)
- STOMP (in framework): CCA components integrated with SciDAC technologies (solvers, grids)
- SPH (in framework): CCA components. Domain decomposition in space. Particles can move across domain boundaries and share information across boundaries.
- Other infrastructure: Kepler workflow environment with python scripting; linked visualization tools (SciDAC Visit) and data management (“Organizer” built on Alfresco).
- Current status:
  - Run on MPP2 at PNNL (EMSL Science Theme allocation)
  - STOMP was previously parallelized except reaction module (currently being verified)
  - CCA components being developed
  - Prototype workflow for single-scale simulation has been developed
- Future plans: Apply to benchmark problems

D. Computational Methods

- STOMP: Uses PETSc libraries for PDE numerical solution. Structured orthogonal grids, backward Euler in time with N-R iteration for nonlinearity.
- SPH: Particle method (no PDE or matrix solution). Algorithms similar to molecular dynamics. Averaging requires information from multiple nearby particles.



- Current status and future plans:
  - STOMP code recently used for 150-million element problem on 512 processors on MPP2 (20% peak flops).
  - Stretch goal: Isolate solver and grid modules from rest of the code in CCA framework
  - TOPS researchers are looking at solver issues and performance using some example problems
  - SPH code is being verified
  - Exploring collaboration with PERI to profile codes and identify bottlenecks

E. I/O Patterns and Strategy

- I/O patterns: ?
- Approximate sizes of I/O:
  - SPH: 10^6-10^9 particles with location and state variables, output at selected time planes
  - STOMP: 10^6-10^9 grid elements with parameters and state variables, output at selected time planes
- Checkpoint / restart: Restart from last saved time plane
- Current status and future plans: ?

F. Visualization and Analysis

- Data to be explored within Kepler workflow environment

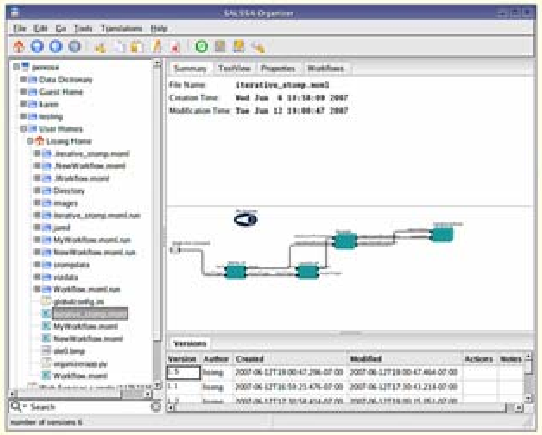
Organizer

User's view and access point to their information stored on data server:

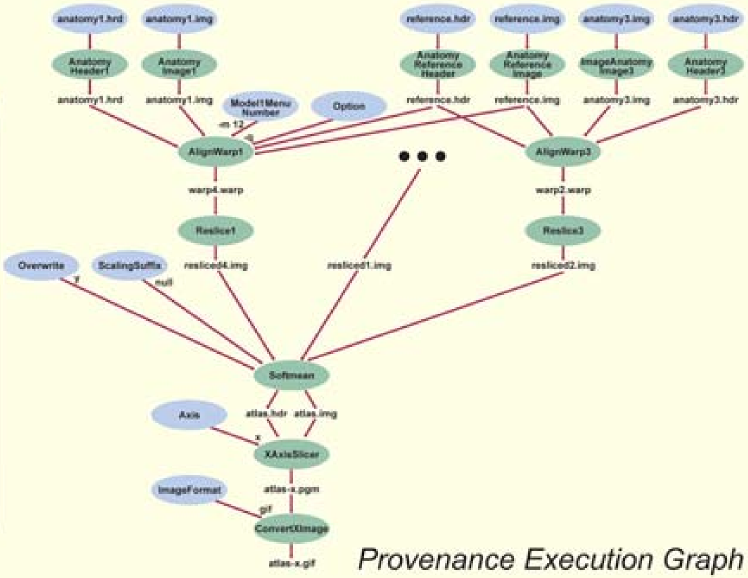
- Data
- Workflows
- Visualizations
- Versions
- Provenance

Central interface to shared knowledge base of:

- Tools
- Workflow processes
- Workflow components
- Reference data



Organizer Main Window



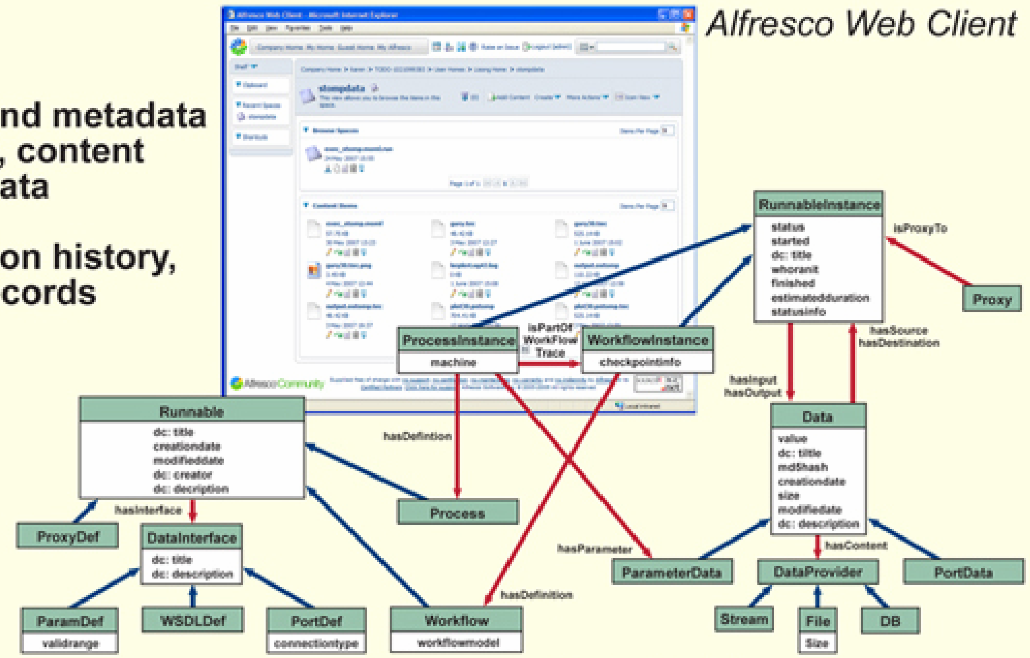
Provenance Execution Graph

F. Visualization and Analysis

- Data to be explored within Kepler workflow environment

Data Services

- Storage of raw data in native formats and metadata including ownership, user annotations, content type, and ability to add arbitrary metadata
- Support for data versioning, modification history, and workflow provenance/execution records
- Support for collaborative research as well as protecting content through authentication/authorization
- Developed using Alfresco content management system
- Sesame RDF for storing provenance
- SPARQL for querying provenance



Alfresco Web Client

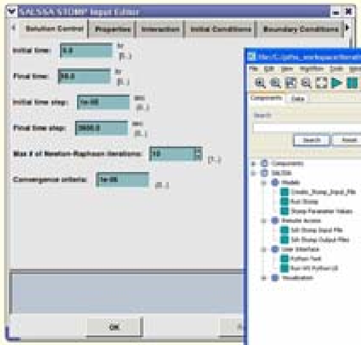
Provenance Data Model

F. Visualization and Analysis

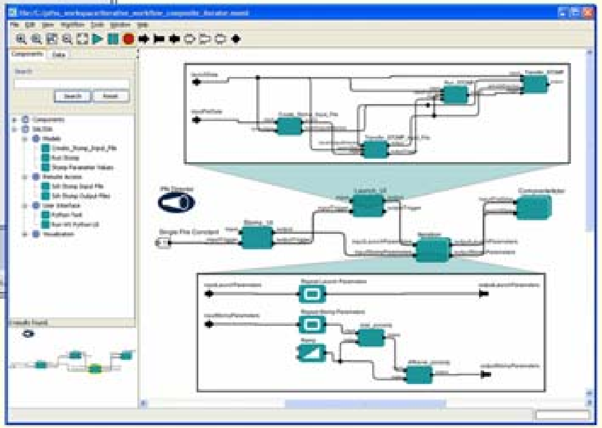
- Data to be explored within Kepler workflow environment

Workflow

- Based on Kepler
- Formally captures the steps in a process with dependencies between steps including order of execution and flow of data
- Interactive designer for creating workflow descriptions
- Save, execute, and bring back existing workflows
- Pre-defined tasks for common computational operations
- Organizer integration for saving workflows to data server
- Provenance capture for saving workflow execution steps and data
- Support for Common Component Architecture networks in workflows



wxPython GUI for Continuum Model



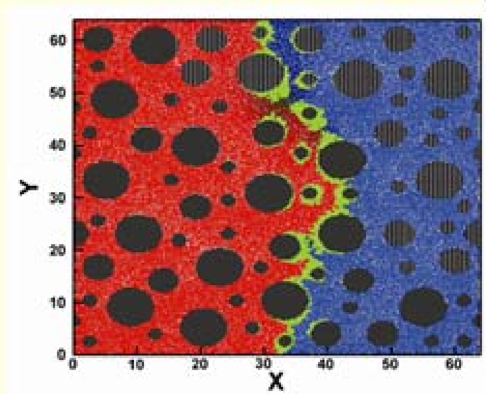
Kepler Continuum Model Workflow Showing Nested Iteration

F. Visualization and Analysis

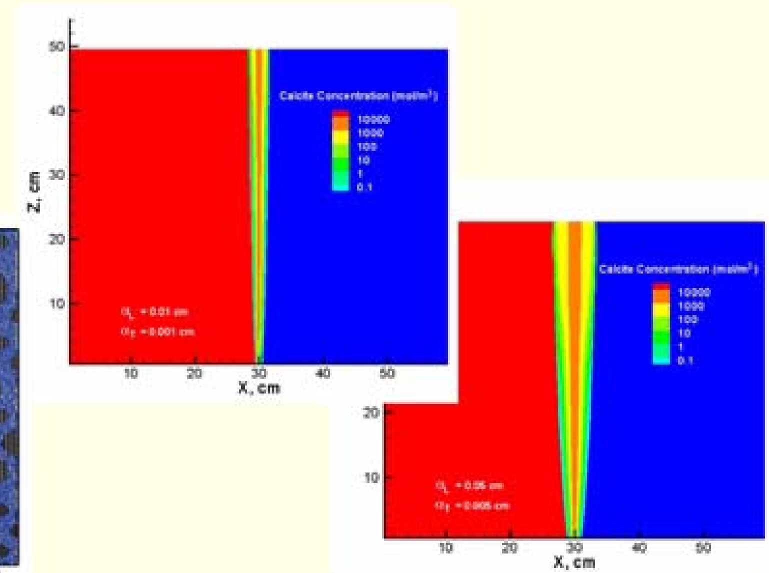
- Data to be explored within Kepler workflow environment

Visualization

- Support for a variety of visualization tools including VisIt and Tecplot for grid based data
- Organizer integration to automate multi-step data translations and visualizations
- Plan to support new hybrid visualization techniques



Pore-scale Visualization of Benchmark Problem



Continuum Model Visualizations of Benchmark Problem

F. Visualization and Analysis

- Data to be explored within Kepler workflow environment

- Current status and future plans:
  - Initial workflow developed for single-scale STOMP simulation by end of August. To be used by INL collaborator for a suite of experimental design simulations. Integrate with visualization.
  - Future:
    - Integrate automated iteration capabilities (e.g., systematic parameter variation studies)
    - Incorporate ability to execute CCA component framework and address issues on large-scale computational platforms (e.g., queuing, etc.)

G. Performance

- Tools:
  - Simple profiling and scalability tests
- Current bottleneck:
  - STOMP: Code options / logicals (need to strip down to get to 150-million). Better profiling and analysis needed.
  - SPH: Unknown
- Current status and future plans:
  - Collaboration with PERI to utilize formal performance analysis tools

H. Tools

- Debugging:
  - ?
  - Visualization, comparison to analytical solutions and/or alternative methods
- Other tools:
  - STOMP: Parallel preprocessor
  - SPH:
- Current status and future plans:
  - ?

I. Status and Scalability

- How does your application scale now?
  - Poorly understood
- Where do you want to be in a year?
  - Complete performance analysis, identify bottlenecks and strategies for addressing
  - Operable CCA components for STOMP and SPH
  - Ready to submit INCITE proposal?
- Top five pains?
- What did you change to achieve scalability?
  - See previous slides
- Current status and future plans:
  - See previous slides

J. Roadmap

- Where will your science take you over the next 2 years?
  - Algorithms developed for model hybridization and adaptivity
  - Very large datasets becoming available
- What improvements will you need to make?
  - Particle and multi-scale visualization
  - Understand and improve scalability of STOMP
  - Address solver performance
- What are your plans?
  - "Divide and conquer"
  - Use simple to complex benchmark problems
  - SAP 1: Workflow development (as before)
  - SAP 2: CCA componentization (as before)
  - Science App: Model hybridization / adaptivity approach
  - Rely heavily on SciDAC collaborations (TOPS, ITAPS, PERI, VACET/Ultra-Scale Vis, SDM, CCA).