CRUNCH
Multi Component Reactive Transport Simulator

Investigate how spatial variations in physical and geochemical properties of subsurface sediments affect fluid flow and transport of reactive components, including radionuclide contaminants, in unsaturated porous media.

Sponsor: DOE

Matt Rosing: (Fort Collins, Colorado): CS
Steve Yabusaki (PNNL): Lead, subsurface flow and reactive transport
Vicky Freedman (PNNL): subsurface flow and multi-component reactive transport
Andy Ward (PNNL): characterization of multiscale material property heterogeneity in the vadoze zone
Carl Steefel: (LBL):multicomponent reactive transport

Science

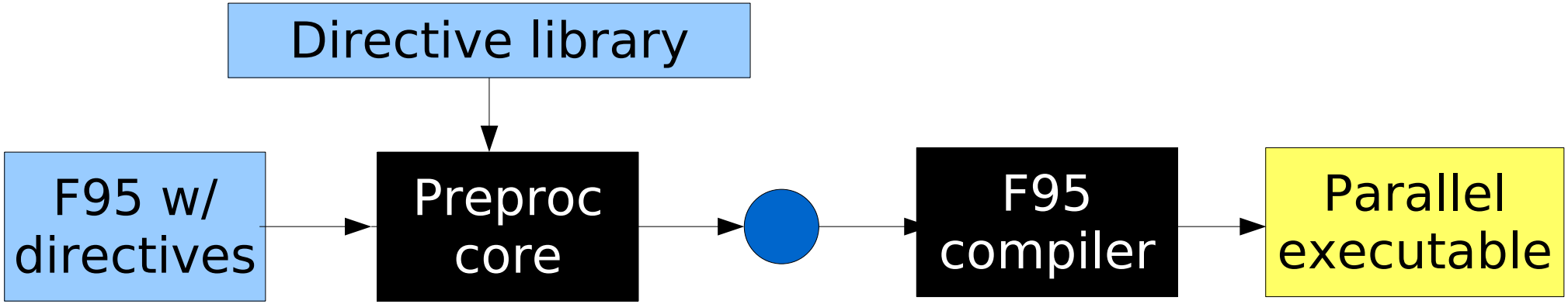
- Addresses 3-D distributions of physical and chemical heterogeneities
- Calculates solution chemistry (concentrations of multiple chemical components) that honor stoichiomentry, thermodynamics, and kinetics of multiple reactions
  - Mixed equilibrium and kinetic reactions
  - Aqueous and mineral reactions
  - Multisite surface complexation and cation exchange
- Transport aqueous components through
  - Advection
  - Diffusion
  - Dispersion

Computational Methods

- Finite volume on regular grid
- Operator split: Chemistry and Transport
  - Sequential iteration or sequential noniterative solution modes
  - Task/data parallel for chemistry/transport
- Backward Euler DAE for reactions; newton-raphson nonlinear iteration
- Transport: Advection, Diffusion, Dispersion
  - 3^{rd} order explicit TVD for advection
- PETSc global implicit 2^{nd} order Poisson Equation diffusion/dispersion

Parallel Model

- Use preprocessor:



- Source is Fortran 95 w/directives
- Intermediate: F95, MPI, 2D processor grid
- Preprocessor library, PETSC, MPI, mlib, ...

I/O Patterns and Strategy

- Input configuration at start
  - Distribute arrays before running jobs
  - 500x500x500 problem is big
- Output restart files occasionally
  - User controlled
- Currently is not an issue, could be with larger machine and problem

Visualization and Analysis

- How do we do this with more than one cell per pixel?
- Data is too big to look at in serial

Performance

- Scales nearly perfectly to 1800 processors
- Load balancing done cheap
  - Won't stay this way on larger machines
  - Looked at adaptive meshing
- **BUT** single processor performance...
  - Started at 1.5% peak, now at 20-25% peak
  - Slow process, custom code, error prone
  - Death by a thousand hacks
- Use PAPI w/preprocessor

Tools

- What tools?
  - Debuggers too slow, vtune, valgrind, ...
  - *“Let's buy the machine and worry about tools later”*
- Preprocessor
  - To parallelize
  - Debugging: compare serial to parallel
    - Write, or print*
  - Would like to rebuild preprocessor
    - Include data flow info, whole program operations
    - Matlab for programs
- PAPI – nice, but could use parallel vtune

Status and Scalability

- Scales well to 1848 processors
  - Not sure about Petsc
- Sustain 2TFlops on 11TFlop machine
- BUT this is not general
  - Specialized code, duct tape and bailing wire
  - On special data set
  - We know how to do it, but don't have tools

Road Map

- Science: always more
  - More complex systems
  - Multi-scale dynamics
  - Limiting factor is poor programmer performance
- Programming and code development
  - Software QA: for our problems and system rot
    - Legacy codes, CM-Fortran, KSR, TCGMSG, ...
    - Refactoring constantly: need assertions, tests
  - Performance (machine and programmer)
    - Dealing with memory bandwidth, load balancing
    - Need ability to tinker with code at high level
  - Manage complexity
    - libraries and tools that can be used by scientists
    - *A simple idea results in 100s of changes in the code*