Towards Petascale Computing in Geosciences: Application to the Hanford 300 Area

P.C. Lichtner (PI: plichtner@lanl.gov), G. Hammond, C. Lu, R. Mills, D. Moulton, B. Phillips, B. Smith, A. Valocchi, Yabusaki

Abstract

Modeling uranium transport at the Hanford 300 Area presents new challenges for high performance computing. A field-scale three-dimensional domain with an hourly fluctuating Columbia River stage coupled to flow in highly permeable sediments results in fast groundwater flow rates requiring small time steps. In this work, high-performance computing has been applied to simulate variable saturated groundwater flow and transport at the 300 Area using PFLOTRAN. Simulation results are presented for discretizations up to 10.8 million degrees of freedom, while PFLOTRAN performance was assessed on up to one billion degrees of freedom and 12,000 processor cores on Jaguar, the Cray XT4 supercomputer at ORNL.

Introduction

Historical modeling of U(VI) transport using a constant K model has significantly overestimated Hanford 300 Area U(VI) plume migration rates.

Modeling of U(VI) at the Hanford 300 Area presents several modeling challenges:
- Varily saturated flow within a large three-dimensional kilometer-scale domain with stratified geological units.
- Hourly fluctuation in Columbia River stage requires restrictive time step sizes to capture the oscillatory forcing on hydrologic units.
- Fast flow rates due to high permeabilities within the upper Hanford Unit, which is predominantly composed of cobbles and gravels (i.e. 1000 m/d).
- Multiscale physicochemical processes ranging m-m s.
- U(VI) geochemistry requires the solution of large numbers of chemical components.

PFLOTRAN utilizes a modular object-oriented, Fortran 90 implementation (see Figure 1) of the finite volume method combined with backward Euler time differencing to solve the systems of equations governing subsurface flow and transport. With a highly-scalable parallelization through tight integration of Argonne National Laboratory's PETSc library (Balay et al., 1997), PFLOTRAN runs on any computer platform supported by PETSc.

The Massively Parallel Reactive Flow and Transport Code PFLOTRAN

PFLOTRAN includes:
- Object-oriented data structures
- PETSc solvers/preconditioners
- Modular linkage to physical process libraries
- Collective parallel I/O with HDF5
- Adaptive mesh refinement (AMR) and unstructured grids
- Multicontinuum subgrid model
- Multiscale flow
- Thermal transport
- Multicomponent reactive transport
- Biogeochemistry
- Colloid-facilitated transport

Figure 1: PFLOTRAN flow diagram illustrating use of procedures, operators, objects, and mode specific operators and objects. Flow Chart Definitions: *Simulation object: highest level data structure providing all solution information.*

Simulation object: Pointer to Newton-Krylov solver and tolerances associated with time stepping.

Solver object: Pointer to nonlinear Newton and linear Krylov solvers (PETSc SNES/KSP/PC) and associated convergence criteria.

Realization object: Pointer to all discretization and field variables associated with a single realization.

Level object: Pointer to discretization and field variables associated with a single level of grid refinement within a realization.

Patch object: Pointer to discretization and field variables associated with a subset of grid cells within a level.

Grid object: Pointer to discretization within a patch.

Auxiliary Data object: Pointer field variables within a patch.

Figure 2: Level-Patch Discretization Structure used in PFLOTRAN

Figure 3 illustrates PFLOTRAN strong scaling performance on Jaguar quad-core XT4 Cray for a 270 million node problem. As proof-of-concept for petascale computing, PFLOTRAN was run with a one billion node (4096-8192-16384 nodes) problem for 12,000 Jaguar processor cores (quad-core) presented in Figure 4.

Figure 4: Jaguar performance for one time step of one billion node problem.

PFLOTRAN Parallel Performance

Figure 5: Layout of Hanford 300 Area.

Hanford 300 Area Conceptual Model

The PFLOTRAN model of the Hanford 300 Area consists of a gridded domain measuring 1350 > 2500 > 20 meters (x,y,z) with orientation aligned with the Columbia River at 14° of west of north (Figure 5). Stratigraphy used in the model is illustrated in Figure 6. The base of the model lies at 90 meters elevation above sea level.

The three grid resolutions simulated in this work include:
- 20 meter horizontal (i.e. 1 meter vertical (z)) (170K dof)
- 10 meter horizontal (i.e. 0.5 meter vertical (z)) (3.5M dof)
- 5 meter horizontal (i.e. 0.25 meter vertical (z)) (10.8M dof)

Hanford 300 Area Simulation Results

PFLOTRAN simulations using the variable-saturated Richards mode were initialized to steady state (based on 10am May 1, 1993 conditions), restarted, and run transient to 7500 hours (10am May 1, 1993 to 15am March 9, 1993).

Figure 7: 100K dof model.

Figure 8: 1.35M dof model.

Figure 9: 10.8M dof model.

Figure 10: Comparison of observed versus predicted head at well 399-3-12 with comparison to river stage.

Figure 11: Comparison of observed versus predicted head at well 399-3-12 with comparison to river stage (enlarged).

Figure 12: Magnitude of predicted pore water velocities at VIC site (enlarged) based on 26% porosity in Hanford unit.

Figure 13: Magnitude of predicted pore water velocities at VIC site (enlarged).

Conclusions

PFLOTRAN variably-saturated flow simulations utilizing data sets provided by Campbell (1994), Campbell and Newcomer (1992), Williams and Rockhold (2008) produce piezometric heads slightly higher and more oscillatory with a consistent offset than those observed in the field.

Convergence of flow variables with higher-resolution grids needs to be investigated further to rule out numerical artifacts in the solution.

High-performance computation enables the solution of large, 3D high-resolution problem domains beyond what is possible on a single processor workstation and with reasonable turnaround time, especially for calibration/optimization runs.

Acknowledgements

This research is supported under the U.S. Department of Energy SciDAC-2 program with funding provided by DOE Offices of Biological & Environmental Research (BER) and Advanced Scientific Computing Research (ASCR).

Supercomputing resources were provided by the DOE Office of Science Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program with allocations on NCCS Jaguar at Oak Ridge National Laboratory and NERSC MUF2 at Pacific Northwest National Laboratory.

We thank Mark Rockhold and Paul Thorne at Pacific Northwest National Laboratory for helpful discussion and provision of Hanford 300 Area stratigraphic and hydrologic data sets.

References


