Towards Petascale Computing in Geosciences:

Application to the Hanford 300 Area

P.C. Lichtner$^{\alpha}$ (PI: lichtner@lanl.gov), G. Hammond$^{\beta}$, C. Lu$^{\delta}$, R. Mills$^{\gamma}$, D. Moulton$^{\alpha}$, B. Philip$^{\alpha}$, B. Smith$^{\sigma}$, A. Valocchi$^{\epsilon}$, S. Yabusaki$^{\beta}$

$^{\alpha}$Los Alamos National Laboratory, $^{\beta}$Pacific Northwest National Laboratory, $^{\delta}$University of Utah, $^{\gamma}$Oak Ridge National Laboratory, $^{\sigma}$Argonne National Laboratory $^{\epsilon}$University of Illinois at Urbana-Champaign

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 variably saturated groundwater flow and tracer 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_d 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:
  - Variably saturated flow within a large three-dimensional kilometer-scale domain with stratified geologic 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 μm-m.
  - U(VI) geochemistry requires the solution of large numbers of chemical components.

The Massively Parallel Reactive Flow and Transport Code PFLOTRAN

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.

- Key features/capabilities either implemented or currently being implemented(*) include:
  - Object-oriented data structures
  - PETSc solvers/preconditioners
  - Modular linkage to physicochemical processes
  - Collective parallel I/O with HDF5
  - Adaptive mesh refinement (AMR) and unstructured grids*
  - Multicontinuum subgrid model*
  - Multiphase flow
  - Thermal transport
  - Multicomponent reactive transport
  - Biogeochemistry
  - Colloid-facilitated transport*

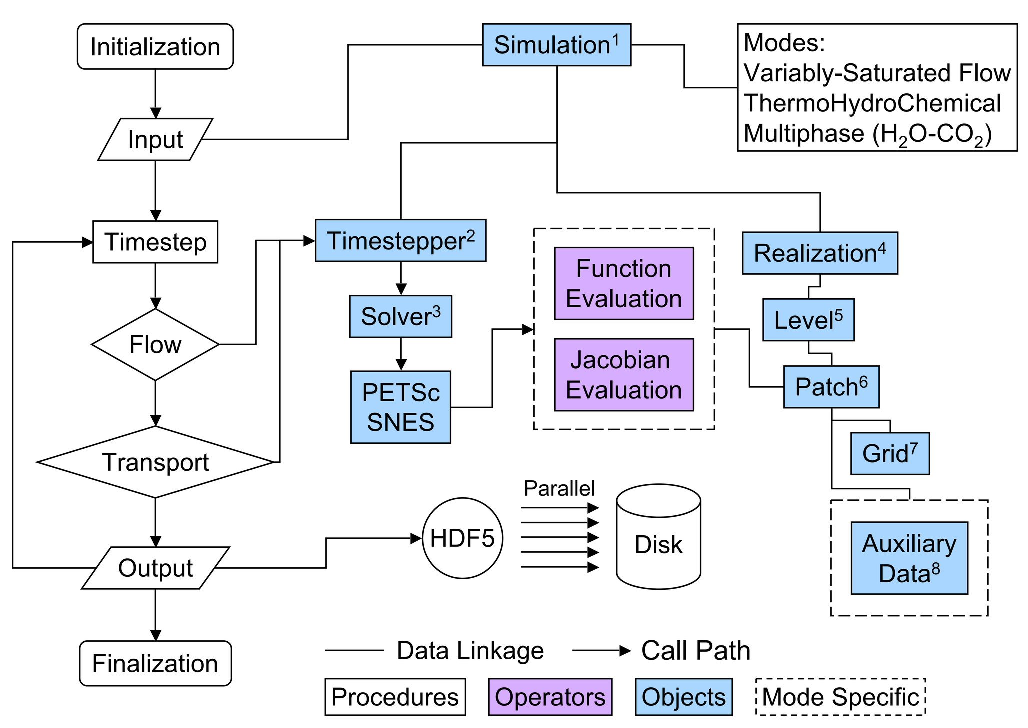

Figure 1: *PFLOTRAN flow diagram illustrating use of procedures, operators, objects, and mode specific operators and objects. Flow Chart Definitions: [1]Simulation object: Highest level data structure providing all simulation information; [2]Timestepper object: Pointer to Newton-Krylov solver and tolerances associated with time stepping; [3]Solver object: Pointer to nonlinear Newton and linear Krylov solvers (PETSc SNES/KSP/PC) and associated convergence criteria; [4]Realization object: Pointer to all discretization and field variables associated with a single realization; [5]Level object: Pointer to discretization and field variables associated with a single level of grid refinement within a realization; [6]Patch object: Pointer to discretization and field variables associated with a subset of grid cells within a level; [7]Grid object: Pointer to discretization within a patch; [8]Auxiliary Data object: Pointer field variables within a patch.*

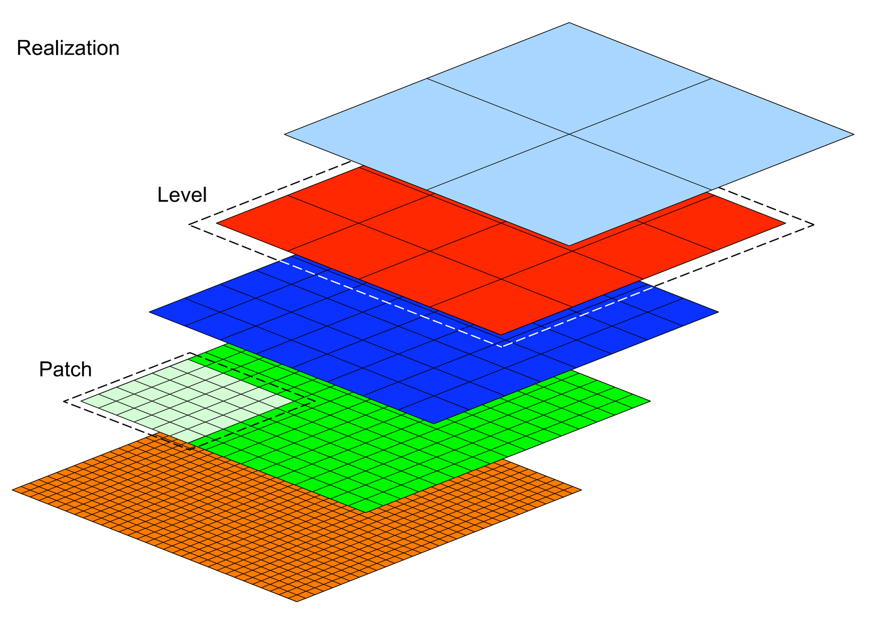

Figure 2: *Level-Patch Discretization Structure used in PFLOTRAN*

PFLOTRAN Parallel Performance

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 ($4096\times2048\times128$ = 1,073,741,824 nodes) problem for to 12,000 Jaguar processor cores (quad-core) presented in Figure 4.

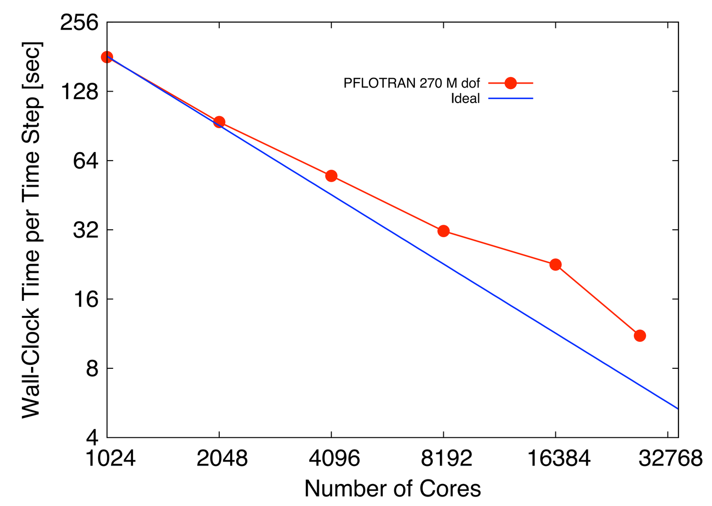

Figure 3: *Performance of PFLOTRAN running a single phase hydrologic benchmark problem on a 1350 $\times$ 2500 $\times$ 80 grid (270 million degrees of freedom).*

One-billion node performance on Jaguar

# procs	newton its	bcgs its	time(sec)
1024	16	1697	1143
2048	16	1626	697
4096	16	1720	325

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

Hanford 300 Area Conceptual Model

- The PFLOTRAN model of the Hanford 300 Area consists of a gridded domain measuring 1350 $\times$ 2500 $\times$ 20 meters (x,y,z) with orientation aligned with the Columbia River at 14° of west of north (Figure 5). 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 (x-y) $\times$ 1 meter vertical (z) (170K dof)
  - 10 meter horizontal $\times$ 0.5 meter vertical (1.35M dof)
  - 5 meter horizontal $\times$ 0.25 meter vertical (10.8M dof)

Hanford 300 Area Simulation Results

PFLOTRAN simulations using the variably-saturated Richards mode were initialized to steady state (based on 10am May 1, 1992 conditions), restarted, and run transient to 7500 hours (10am May 1, 1992 to 10am March 9, 1993).

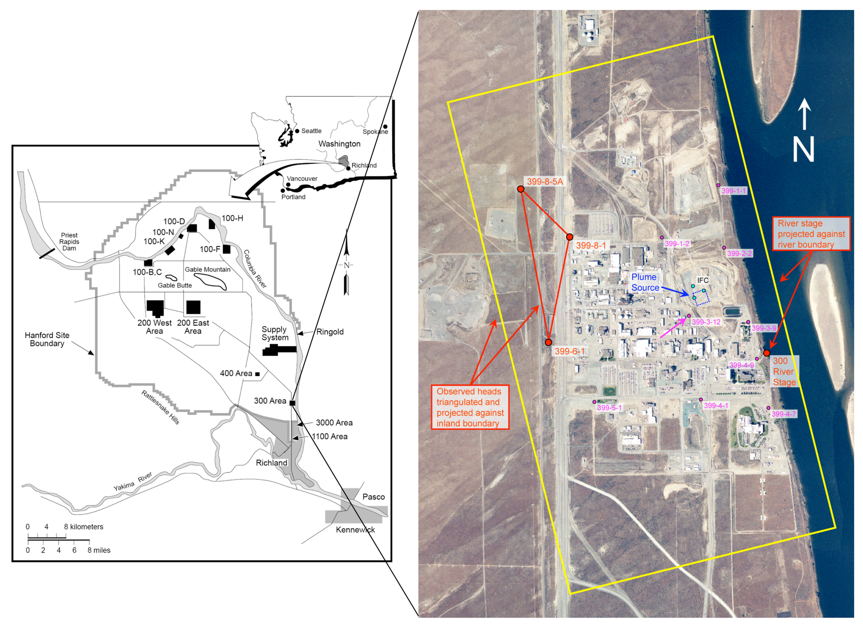

Figure 5: *Layout of Hanford 300 Area.*

- Figures 6, 7 and 8 illustrate x-y cross sections of the pressure (z = 90m) and tracer concentration (z = 105m) at 7500 hours simulation time (10am Mar. 9, 1993) for the 170K, 1.35M and 10.8M dof grids, respectively.
- Figure 9 compares the piezometric head observed at monitoring well 399-3-12 (see location in Figure 5) to the PFLOTRAN heads predicted nearest to the well for each of the simulation scenarios.
- Figure 10 is an enlarged view of Figure 9 between 3250-3750 hours. All three PFLOTRAN grid resolutions produce nearly identical piezometric heads that slightly overestimate the observed head and are more oscillatory.
- Predicted PFLOTRAN pore water flow velocities computed at the center of the IFC site are plotted in Figure 11 (enlarged Figure 12) for the various grid resolutions. The enlarged plot (hours 3250-3750) reveals that the coarse grid (170K dof) model underestimates peak velocities by as much as 60%, whereas the velocities are more consistent for the 1.35M and 10.8M dof simulations, although it is not clear that velocities have convergence.

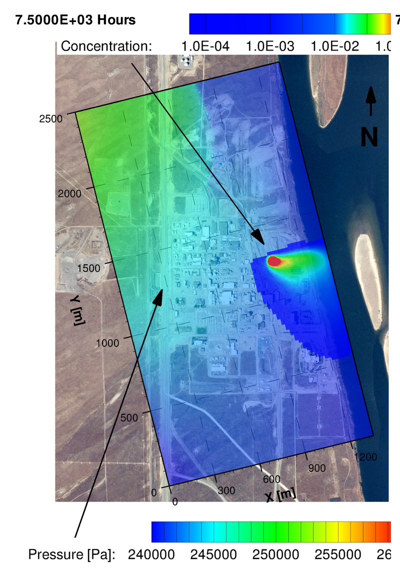

Figure 6: *170K dof model.*

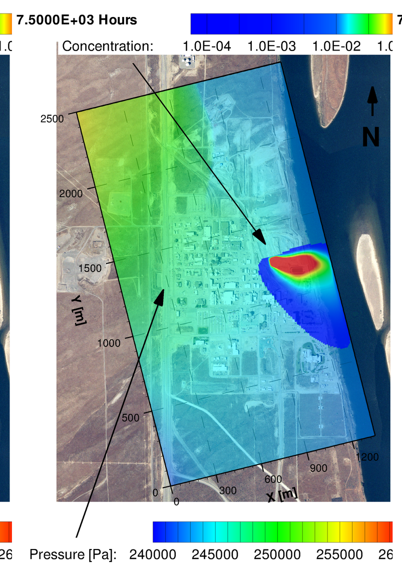

Figure 7: *1.35M dof model.*

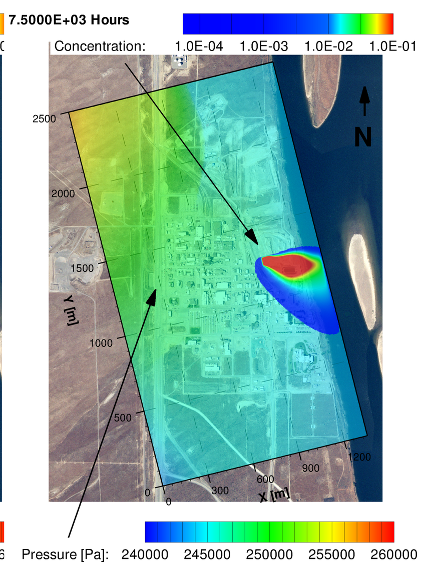

Figure 8: *10.8M dof model.*

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 velocities 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.

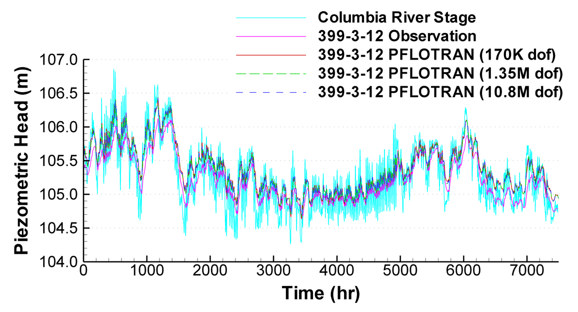

Figure 9: *Comparison of observed versus predicted head at Well 399-3-12 with comparison to river stage.*

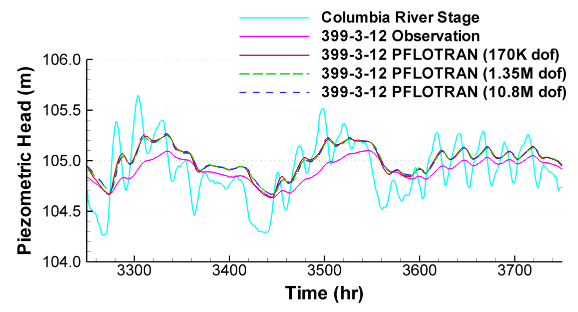

Figure 10: *Comparison of observed versus predicted head at Well 399-3-12 with comparison to river stage (enlarged).*

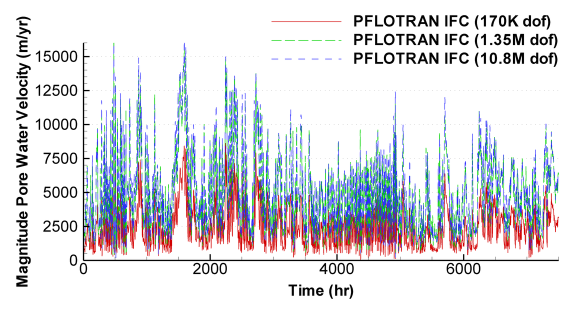

Figure 11: *Magnitude of predicted pore water velocities at IFC site (elevation = 100m) based on 20% porosity in Hanford unit.*

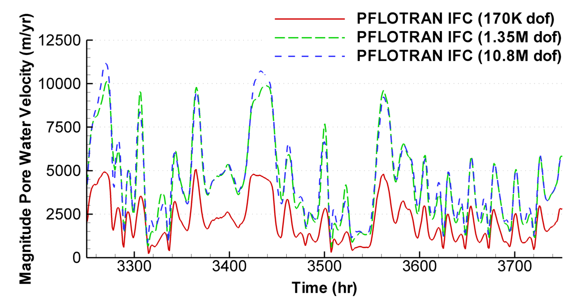

Figure 12: *Magnitude of predicted pore water velocities at IFC site (enlarged).*

References

Satish Balay, William D. Gropp, Lois Curfman McInnes, and Barry F. Smith. Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, and H. P. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 163–202. Birkhäuser Press, 1997.

M.D. Campbell. Monitoring groundwater and river interactions along the hanford reach of the columbia river. Technical Report PNL-9437, Pacific Northwest National Laboratory, Richland, WA, 1994.

M.D. Campbell and D.R. Newcomer. Automatice measurement of water levels within the 300-FF-5 boundary. Technical Report PNL-7874, Pacific Northwest National Laboratory, Richland, WA, 1992.

Mark D. Williams and Mark L. Rockhold. Three-dimensional groundwater models of the hanford 300 area. Technical Report PNNL-XXXX (in review), Pacific Northwest National Laboratory, Richland, WA, 2008.

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 EMSL MSCF MPP2 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.