Toward petascale computing in geosciences: application to the Hanford 300 Area

Glenn E Hammond¹, Peter C Lichtner², Richard Tran Mills³ and Chuan Lu⁴

¹Hydrology Group, Environmental Technology Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA
²Hydrology, Geochemistry, and Geology Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
³Computational Earth Sciences Group, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6015, USA
⁴Energy & Geoscience Institute, University of Utah, Salt Lake City, UT 84108, USA

E-mail: glenn.hammond@pnl.gov, lichtner@lanl.gov, rmills@ornl.gov, clu@egi.utah.edu

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.

1. Introduction

As part of the SciDAC groundwater science application area, PFLOTRAN [1–3] is being used to simulate radionuclide transport at the U.S. Department of Energy’s Hanford Site in southeastern Washington State. Significant simulation and modeling efforts were put forth in the early 1990s in an attempt to determine the rate at which uranium within the groundwater would migrate toward and discharge into the Columbia River. Unfortunately, these models lacked the sophistication necessary to accurately simulate uranium transport, predicting that the uranium would be removed or flushed out of the Hanford 300 Area subsurface by ambient
groundwater flow within a decade. Today, nearly 15 years later, uranium concentrations at the site remain relatively unchanged.

Several possibilities exist attempting to explain the discrepancies between the predicted and observed transport of uranium. These include (1) fluctuations in the stage of the Columbia river and its impact on groundwater velocities, (2) the vadose zone contribution to the source term, (3) inadequacies in the uranium transport conceptual model due to overly simplistic reactions representing the sorption of uranium on sediment grains; and, (4) uncertainty of the uranium source term mass concentrations and leaching rates.

The conceptual model for flow at the Hanford 300 Area is complicated by the highly permeable Hanford Unit and the rapidly fluctuating Columbia River shown in figure 1 producing changes in magnitude and direction of flow. Diurnal (daily) fluctuation in river stage can be up to 1.5 meters, while seasonal fluctuation can exceed 2.5 meters. The variable release of water from Priest Rapids Dam upriver is the cause of the diurnal fluctuation, while seasonal variations are due to snow melt, irrigation, and so forth. The river stage fluctuation along with the highly permeable Hanford Unit directly impacts the water table beneath the Hanford 300 Area, causing river water to flow into the area during high stage and retreat during low stage.

The soils beneath the Hanford 300 Area are composed of layers of high- and low-permeability soil, the top layer of which is the highly permeable Hanford Unit located near the water table. The Hanford Unit is composed of highly permeable cobbles, gravels, and sands. Hydraulic conductivities within the Hanford Unit are on the order of a thousand to tens of thousands of meters per day. In comparison, the Ringold Units below the Hanford Unit exhibit much lower hydraulic conductivities of 0.01 to 150 meters per day. Thus, within the Hanford Unit, where all original uranium sources are believed to have resided, groundwater has the potential of flowing rapidly with very small pressure gradients in the aquifer.

This work extends [1] to incorporate injection of a tracer into the highly variable flow field generated by the Columbia river stage fluctuations. It also incorporates the use of Richard’s equation for variably saturated flow. Results of the flow simulations are compared with well head data from the 300 Area site, and tracer transport is compared on several levels of discretization.
2. PFLOTRAN capabilities

PFLOTRAN is a parallel multiphase-multicomponent reactive flow and transport code for modeling subsurface processes. The code is written in Fortran 90 using a modular, object-oriented approach (see figure 2) and is composed of several modules for flow and transport, with the option of running the modules in coupled or decoupled mode. PFLOTRAN is capable of simulating fluid flow though porous media with fluid phases air, water, and supercritical CO₂. PFLOTRAN-generated fluid flow velocities or fluxes and phase saturation states are used by the transport module to compute solute transport. Within PFLOTRAN, transport and reaction are fully coupled. PFLOTRAN’s problem domain is discretized spatially using the integrated finite volume approach, with fully-implicit backward-Euler time differencing.

PFLOTRAN’s parallel paradigm is based on domain decomposition where the computational problem domain is divided into subdomains, with one domain assigned to each processor. PFLOTRAN leverages PETSc [4] solvers and data structures to link these subdomains within a single parallel code. Since PFLOTRAN is founded these PETSc data structures, the code is capable of utilizing the full suite of algorithms available in the library. In addition, PETSc also provides linkage to a variety of external software packages (e.g., Hypre, Trilinos, Zoltan). Within PFLOTRAN, the distributed array (DA) is the key PETSc data structure that dictates the layout and decomposition of the three-dimensional parallel domain. From the DA, PETSc generates parallel matrices (Mat) and vectors (Vec) with consistently mapped local and global indexing for the parallel communication required to solve the problem. Therefore, although an in-depth understanding of parallel communication paradigms such as the message passing interface (MPI) is helpful, it is not required since PETSc hides the communication from the end user. For example, when employing the Newton-Raphson method to solve a nonlinear system of equations in parallel using the PETSc nonlinear solver or SNES, the programmer essentially provides functions to PETSc that evaluate the residual and compute the Jacobian for the nodes that reside locally on each processor. PETSc then uses these functions to compute a solution vector, which is then returned to PFLOTRAN for boundary condition updates, I/O, and so forth. Thus, the application scientist is able to focus more on the science (in this case, subsurface physics and chemistry) rather than solvers, preconditioners, and so forth. Furthermore, because PETSc supports almost any modern computer platform and because PFLOTRAN will run anywhere that PETSc is supported, our code is highly portable.

Recently, PFLOTRAN has demonstrated a relative parallel efficiency of 79% at 12,000 cores based on a strong-scaling study performed with a 500 million node problem (Hanford 300 Area) executed on 3,000-12,000 cores. Figure 3 shows the wall-clock time versus the number of processor cores run on Jaguar at ORNL using dual-core processors. Scaling is reasonable and the slight reduction in performance is attributed to load imbalance due to the use of inactive nodes in the calculation. As proof-of-concept for petascale computing, PFLOTRAN has been run on a similar one billion node (4096×2048×128 = 1,073,741,824 nodes) problem. Timings for a single step run on 1024 to 4096 cores are presented in figure 4.
2.1. Governing equations

The governing equations that describe uranium transport at the Hanford 300 Area are based on Richards equation for fluid flow in a partially saturated medium. In this approximate formulation the gas phase is assumed to be inert. Richards equation is coupled to the advection-diffusion-reaction equation for solute transport. Richards equation may be represented in the form

$$\frac{\partial}{\partial t} (\varphi s \rho) + \nabla \cdot (\rho \mathbf{u}) = S,$$

where $\varphi$ denotes the porosity of the porous with saturation $s$ giving the fraction of pore volume filled with water, $\rho$ denotes the fluid density, $S$ refers to a source/sink term, and $\mathbf{u}$ denotes the
PFLOTRAN strong scaling

**Figure 3.** Strong scaling speedup curve for 500 million node problem run on Jaguar.

Darcy velocity defined by

\[ u = -\frac{\kappa \kappa_r}{\mu} \nabla (P - \rho g z), \]  

with fluid pressure \( P \), viscosity \( \mu \), vertical distance \( z \), saturated permeability \( \kappa \), relative permeability \( \kappa_r \), and acceleration of gravity \( g \). The relative permeability is a nonlinear function of saturation based on the van Genuchten equation.

For the purposes of this contribution, solute transport is described by the nonreactive scalar transport equation

\[ \frac{\partial}{\partial t} (\varphi s C) + \nabla \cdot (u C - \varphi s D \nabla C) = S_C, \]  

for concentration \( C \), diffusion/dispersion coefficient \( D \), and source/sink term \( S_C \). The transport equation is coupled to flow equation, equation(1), through the saturation \( s \) and Darcy flow velocity \( u \).

2.2. Numerical solution

PFLOTRAN uses a finite-volume spatial discretization combined with backward-Euler time differencing to solve the system of governing equations for subsurface flow and transport. Upwinding is used for the advective term in the transport equations. The system of nonlinear algebraic equations resulting from the discretization is solved using the highly scalable nonlinear solver framework (SNES) in PETSc. The framework allows a wide variety of solver algorithms; for the calculations described here, we used an inexact Newton-Krylov method with cubic line-search. The Jacobian systems are solved via a BiCGStab method with a block-Jacobi (domain-decomposition) preconditioner, with ILU(0) applied on each block.

3. Hanford 300 Area application

The PFLOTRAN model of the Hanford 300 Area consists of a computational 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. Three grid resolutions are simulated in this work:

- 20 meter horizontal \((x-y) \times 1\) meter vertical \(z\) = 170,000 degrees of freedom (170K dof)
- 10 meter horizontal \(\times 0.5\) meter vertical = 1,350,000 degrees of freedom (1.35M dof)
- 5 meter horizontal \(\times 0.25\) meter vertical = 10,800,000 degrees of freedom (10.8M dof)

Stratigraphy was mapped to each grid from the Hanford EarthVision database [5]. The two predominant geologic units simulated in the current conceptual model are the Hanford and Ringold units. These units are shown near the top of figure 6, a 60-meter deep representation of the problem domain viewed from the southeast. Hanford 300 Area hydraulic properties were assigned based on data provided by (author?) [6]. Transient hydrostatic and seepage face boundary conditions were assigned to the western inland and river boundaries, respectively, with time-varying datums and gradients. Specified surface recharges were assigned to the top of the model domain, while the north and south boundaries were no flow. Solute transport was simulated through an infinite rate-limited dissolution source term (normalized tracer concentration) specified near the center of the Hanford 300 Area Integrated Field-Scale Subsurface Research Challenge (IFC) site (see figure 5). Source zone region: area = 1600 m\(^2\) (107K dof), 2500 m\(^2\) (1.35M, 10.8M dof); elevation = 104–110 m.

PFLOTRAN simulations run to date focus on establishing and improving the accuracy of the Hanford 300 Area flow solution and the subsequent fluxes used for simulating transport. Simulations were initialized to steady state (based on 10 am May 1, 1992 conditions), restarted, and run in transient mode for 7500 hours simulation time (10 am May 1, 1992 to 10 am March 9, 1993).

Figure 7 illustrates \(x-y\) cross sections of the pressure \((z = 90\text{m})\) and tracer concentration \((z = 105\text{m})\) at 7500 hours of simulation time (10 am, Mar. 9, 1993) for the 170K, 1.35M, and 10.8M dof grids, respectively. For the 170K dof scenario, 128-processor cores on ORNL’s Jaguar supercomputer required 50 minutes to complete the 7500-hour simulation. Simulations involving 1.35M and 10.8M dofs used 1024 and 600 processor cores and ran for 2.3 and 16 hours, respectively.
Figure 6. Hanford 300 Area stratigraphy (z scale = 20x, z axis ranges 70–130 meters).

Figure 8 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. River stage is also provided as a reference, and is far more oscillatory than the observed head at the well. Figure 8(b) is a detailed view of figure 8(a) between 3250 and 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 9 for the various grid resolutions. The detailed plot (figure 9(b)) 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 converged.

4. Conclusions

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 times, especially needed for calibration/optimization runs. Higher-resolution flow simulations may be necessary to improve the accuracy of flow velocities employed in radionuclide transport simulations. Specifically with regard to the Hanford 300 Area: PFLOTRAN variably saturated flow simulations using datasets provided by [author?] [6, 7, 8] produce piezometric heads slightly higher and more oscillatory than those observed in the field. They also result in a fairly consistent offset from the the observed wells over time (the same holds for other well observation data compared to the predicted heads). Calibration of hydraulic parameters (i.e., permeability) and the inland boundary condition is needed to improve agreement between observed and predicted piezometric heads. Convergence of flow velocities with higher-resolution grids needs to be investigated further to rule out numerical artifacts in the solution.
Acknowledgments

This research is supported by the U.S. Department of Energy’s Office of Science under the Scientific Discovery through Advanced Computing (SciDAC) program. Computing time was provided by a 2008 DOE INCITE award for the project “Modeling Reactive Flows in Porous Media.” This research used resources of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

References

[1] Hammond G E, Lichtner P C and Lu C 2007 SciDAC 2007 Scientific Discovery through Advanced Computing (Journal of Physics: Conference Series vol 78) ed Keyes D (Boston, Massachusetts: IOP Publishing) p 012025

[2] Lu C and Lichtner P C 2007 SciDAC 2007 Scientific Discovery through Advanced Computing (Journal of Physics: Conference Series vol 78) ed Keyes D (Boston, Massachusetts: IOP Publishing) p 012042

[3] Mills R T, Lu C and Lichtner P C 2005 SC2005 SuperComputing 2005 (Seattle, WA: IOP Publishing)

[4] Balay S, Buschelman K, Eijkhout V, Gropp W D, Kaushik D, Knepley M G, McInnes L C, Smith B F and Zhang H 2004 PETSc users manual Tech. Rep. ANL-95/11 - Revision 2.1.5 Argonne National Laboratory

[5] Thorne P D, Bergeron M P, , Williams M D and Freedman V L 2006 Groundwater data package for Hanford assessments Tech. Rep. PNNL-14753 Rev. 1 Pacific Northwest National Laboratory, Richland, WA

[6] Williams M D, Rockhold M L, Thorne P D and Chen Y 2008 Three-dimensional groundwater models of the Hanford 300 Area Tech. Rep. PNNL-XXXX (in review) Pacific Northwest National Laboratory, Richland, WA

[7] Campbell M D 1994 Monitoring groundwater and river interactions along the Hanford Reach of the Columbia River Tech. Rep. PNL-9437 Pacific Northwest National Laboratory, Richland, WA

[8] Campbell M D and Newcomer D R 1992 Automatic measurement of water levels within the 300-FF-5 boundary Tech. Rep. PNL-7874 Pacific Northwest National Laboratory, Richland, WA
Figure 7. Comparison of tracer plume for different grid resolutions corresponding to 170K (left), 1.35M (center), and 10.8M (right) nodes at 7500 hours.

(a) Piezometric head over 7500 hour time interval.  (b) Detailed view of (a) over time interval 3250–3750 hours.

Figure 8. Comparison of observed versus predicted head at Well 399-3-12 and Columbia river stage.
(a) Pore velocity over 7500 hr time interval.  
(b) Pore velocity over time interval 3250–3750 hr.

Figure 9. Magnitude of predicted pore water velocities at IFC site for different grid resolutions.