Numerical Simulation of Groundwater Flow on MPPs*

S. F. Ashby† R. D. Falgout† T. W. Fogwell‡ A. F. B. Tompson§

Abstract

This paper discusses the numerical simulation of groundwater flow in three-dimensional heterogeneous porous media. A portable and scalable code called ParFlow is being developed for massively parallel computers to enable detailed modeling of large sites. This code uses a turning bands algorithm to generate a statistically accurate subsurface realization, and preconditioned conjugate gradients to solve the linear system that yields the flow velocity field. Preliminary numerical results for the LLNL site are presented.

1 The Need for Detailed Modeling

Groundwater contamination is a major problem at many governmental and industrial sites, including Lawrence Livermore National Laboratory (LLNL). Much of the contamination found at LLNL today originated in the 1940's, when the present site was a naval air base. Liquid chemical solvents and various petroleum products were either dumped onto the ground surface or were introduced into the subsurface via leaky storage tanks. Over time, these contaminants slowly percolated through the unsaturated zone and into the more mobile groundwaters (saturated zone). The Laboratory is obligated to track this migration and to design and execute remediation procedures. Various techniques are now being studied, tested, and implemented, including *pump-and-treat*, *dynamic steam stripping*, and *in situ biofilters* [9].

Mathematical models are often used to aid in the design and management of engineered remediation procedures, as well as to demonstrate compliance with myriad environmental regulations. Unfortunately, many of these models are based on idealized assumptions about the subsurface media and its flow behavior. For example, many codes in use today assume that the subsurface medium is 1- or 2-dimensional and homogeneous. In reality, the subsurface is three-dimensional and consists of materials that are *heterogeneous* in composition and spatial distribution. This means that some regions of the subsurface are more permeable to water flow than others, which is represented by a spatially variable property known as the *hydraulic conductivity.* These heterogeneities lead to preferential flow channels in the subsurface velocity field, which in turn lead to "fingering" in contaminant migration, that is, nonuniform dispersion over time. This behavior may dramatically impact the conclusions one draws about a given remediation procedure.

*This work was supported in part by the Laboratory Directed Research and Development program at Lawrence Livermore National Laboratory (LLNL), and by the Applied Mathematical Sciences program of the Office of Scientific Computing, Department of Energy. It was performed at LLNL under contract W-7405-ENG-48.

†Center for Computational Sciences & Engineering, LLNL, L-316, Livermore, CA 94551

‡International Technology Corporation, 4585 Pacheo Blvd., Martinez, CA 94553

§Earth Sciences Division, LLNL, L-206, Livermore, CA 94551

To adequately resolve the subsurface heterogeneities (on the order of meters) for a large site like LLNL (several square kilometers) would require upwards of one billion spatial zones. This is beyond the capability of most conventional supercomputers, and so one must employ massively parallel computing power. We are therefore developing a portable and parallel simulator called ParFlow which will enable this kind of detailed modeling of fluid flow in three-dimensional heterogeneous porous media. This code will serve as an *experimental apparatus* with which one may explore fundamental physical phenomena in realistic systems.

2 Mathematical Formulation and Solution Approach

In this paper, we consider the special case of steady state saturated (i.e., single phase) flow, which may be modeled by

$$-\nabla \cdot (K \nabla (h + z)) - Q = 0 \tag{1}$$

where h is the pressure head (from which the flow velocity field is determined), K is the hydraulic conductivity (generated via a turning bands algorithm), and Q is a source term (used to represent pumping wells, for example). At present, the problem domain is assumed to be a parallelepiped; the boundary conditions may be Dirichlet, Neumann, or mixed.

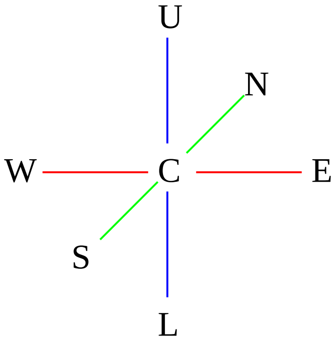
This simple problem is of practical interest because contaminant transport is most rapid in the saturated zone. Although ParFlow is designed to solve the more general multiphase flow equations, to date we have focused our attention on the single phase problem. This has allowed us to hone our linear solver technology before tackling the more demanding time-dependent nonlinear equations associated with multiphase flow.

We solve for the pressure head on a discrete mesh. In the numerical investigations to date, we have used a single uniform grid (but with different steplengths, Δx, Δy, and Δz). This makes load balancing easy. However, we have anticipated the need for local mesh refinement and adaptive gridding, and these capabilities are being designed into ParFlow. Of course, this will demand some sort of load balancing algorithm.

The spatial discretization is a standard 7-point finite difference method. In the case of saturated flow (our immediate interest), the discretized equations are linear, and so we obtain the pressure head from the solution of a system of linear equations, $A\mathbf{h} = \mathbf{f}$. The solution of this large linear system is computationally intensive and must be done efficiently and accurately. Since we are interested in detailed simulations (i.e., high resolution), we must use an iterative method.

3 Linear Solver Implementation

We have implemented preconditioned conjugate gradients [3] with diagonal scaling (DCSG), s-step Jacobi preconditioning (JsCG), and adaptive Chebyshev polynomial preconditioning [2] (PPCG). These solvers call several basic linear algebra routines, including vector update (*axpy*), inner product (*dot*), and matrix-vector multiplication (*matvec*). The representation of matrices and vectors is an important issue in the implementation of these routines. The coefficient matrix A is viewed as a stencil distributed across the processes analogous to the distribution of the spatial grid. This representation is a consequence of the underlying spatial discretization. A vector is logically similar to a grid in that it is composed of several *subvectors*, the components of which may be thought of as mesh points on a subgrid. An additional layer of *ghost* points surrounds each subvector; this will be explained below. See Figure 1.



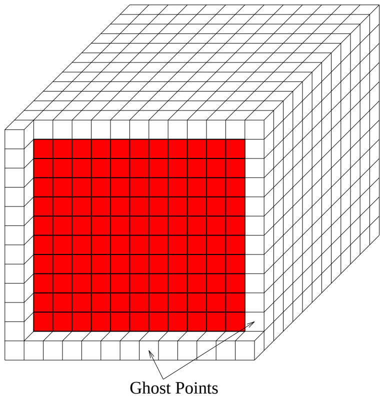


FIG. 1. *Linear solver implementation: the 7-point spatial discretization stencil is distributed across processes; subvectors are surrounded by* ghost points *to facilitate computations.*

A key operation is matrix-vector multiplication, $y = Ax$. To compute the matvec result at a given grid point (i, j, k), we "apply" the stencil to the grid: For each neighboring grid point specified by the stencil, we multiply the vector value at that point by the corresponding stencil coefficient, and then sum these products. That is, component $y_{i,j,k}$ of the matrix-vector product is given by

$$\begin{aligned} y_{i,j,k} = w_{i,j,k}x_{i-1,j,k} + e_{i,j,k}x_{i+1,j,k} + s_{i,j,k}x_{i,j-1,k} + n_{i,j,k}x_{i,j+1,k} \\ +l_{i,j,k}x_{i,j,k-1} + u_{i,j,k}x_{i,j,k+1} + c_{i,j,k}x_{i,j,k} \end{aligned}$$

which is equivalent to multiplying a row of the matrix by the vector x. By viewing the matrix-vector multiplication in this way, it is readily apparent which data needs to be communicated, namely, subgrid boundary data. We therefore exchange interprocess boundary data at the start of a matvec. Once each process has the data it needs, intraprocess matvecs are carried out in parallel. Of course, it is possible to overlap some of the communication with computation. Specifically, the intraprocess matvec can work on the process' internal mesh points first, and then update the boundary mesh points after the communication step is complete. On most machines, and for most large problems, this communication will be concluded before the internal mesh points have been updated.

The ghost points that surround subvectors are used to store the interprocess data in the matvec and other operations. This simplifies coding and leads to faster application development. Note that the single layer of ghost points is a direct result of the 7-point matrix stencil. Although we plan to investigate other discretization stencils in the future, they all can be viewed as special cases of the standard 27-point stencil, and so one layer of ghost points is sufficient.

4 Subsurface Realization Via Turning Bands

In our model, soil heterogeneity is reflected in the spatial variation of the hydraulic conductivity K at a sufficiently fine resolution. Unfortunately, property measurements are typically too sparse to specify the kind of detail required in our model. At the LLNL site, for instance, there are over 300 monitoring wells collecting data on the subsurface. Although this data can be supplemented with so-called soft data (obtained, for example, by cross-well tomography), there are still relatively few data points. Consequently, hydrologists and geologists often describe fine-scale heterogeneity with statistical interpolation techniques [4,

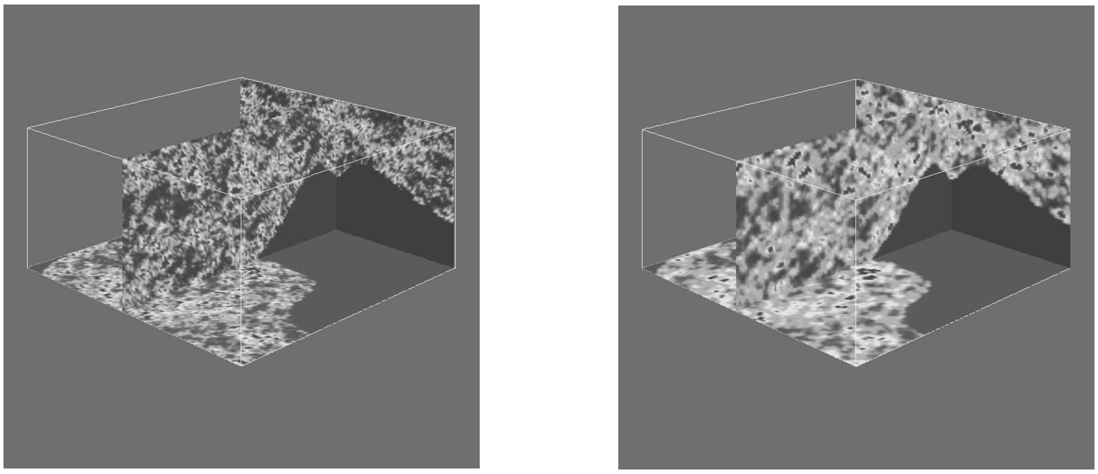


FIG. 2. *The hydraulic conductivity and flow velocity magnitude for the LLNL site.*

5, 8]. Here, measured values and statistical characteristics of a medium property may be used to produce a *realization* of the property distribution. Although such realizations cannot give the precise value of the hydraulic conductivity at an (x, y, z) coordinate, they provide a more realistic representation of property variability than simpler homogeneous models.

In order to represent K mathematically in a reasonably realistic manner, we generate hypothetical property realizations that retain specific features of the variable property [4, 5]. For example, one might assume that the variability of K can be described by a spectral random field in which the mean, variance, and spatial correlation are all known or measurable quantities. (Correlation is used as a simple measure of the spatial persistence of property values.) This type of model is frequently used to describe hydraulic conductivity distributions for both theoretical and computational investigations [1, 6, 8]. Distributions of $\ln K$ values measured at field sites (including LLNL) often appear to be Gaussian. Estimates for the mean, variance, and correlation scales of this $\ln K$ distribution may be made from preliminary data analyses.

To obtain spectral random field realizations, we are using a parallelized version of Tompson's turning bands algorithm [7]. Given certain statistical parameters that describe the hydraulic conductivity (including mean, variance, and spatial correlation), this algorithm generates a realization for the subsurface, which in turn provides the conductivity values needed in the model. It is important to remark that these values are computed in a distributed fashion.

5 Numerical Results for the LLNL Site

Our main test problem describes the LLNL site; the data was provided by colleagues in the Environmental Protection Department. The problem size is $128 \times 128 \times 64$ (over one million spatial zones) with $\Delta x = 100$, $\Delta y = 100$, and $\Delta z = 5$. The aspect ratio of 20:20:1 makes the problem difficult to solve. The Dirichlet conditions were given as hydraulic head values of 520 on the left face and 580 on the right face. Figure 2 illustrates the conductivity field used in this problem, and the resulting velocity magnitudes computed by PARFLOW. Note that a (nearly impermeable) clay region has been superposed on the turning bands realization to represent more accurately the LLNL subsurface.

The lighter grays correspond to higher (in magnitude) flow velocities. We clearly see the preferential flow channels (corresponding to high velocity magnitude) that give rise to the crucial fingering phenomenon mentioned earlier. For the most part, these channels match the regions of high conductivity, as one would expect. These channels, which critically impact contaminant migration, cannot be detected by a homogeneous code. We used J2CG with a convergence tolerance of 10^{-11} to solve the linear system in less than eight minutes on 864 processors of the nCUBE/2.

We also evaluated the parallel performance of PARFLOW. In addition to the usual speedup and efficiency curves, we measured *scaled* speedup, which measures the speedup as the problem size increases linearly with the number of processes. (The motivation for this measure is rooted in practice: the real value of massively parallel machines is that they enable the solution of larger problems.) Specifically, we fixed the problem size per process at $28 \times 28 \times 28$, and we solved larger and larger problems by adding more and more processes. The time to execute a single DSCG iteration remained nearly constant (the ideal) with a scaled efficiency of 87%.

Acknowledgements

We thank the staff of the Massively Parallel Computing Research Laboratory at Sandia National Laboratories for making available to us the nCUBE/2.

We also acknowledge the participation of Steven Smith, William Bosl, Gwen Loosmore, and John Ziagos in the PARFLOW project.

References

[1] R. ABABOU, D. B. MCLAUGHLIN, L. W. GELHAR, AND A. F. B. TOMPSON, *Numerical simulation of three-dimensional saturated flow in randomly heterogeneous porous media*, Transport in Porous Media, 4 (1989), pp. 549–565.

[2] S. F. ASHBY, T. A. MANTEUFFEL, AND J. S. OTTO, *A comparison of adaptive Chebyshev and least squares polynomial preconditioning for Hermitian positive definite linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1–29.

[3] P. CONCUS, G. H. GOLUB, AND D. P. O'LEARY, *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations*, in Sparse Matrix Computations, J. R. Bunch and D. J. Rose, eds., Academic Press, New York, 1976, pp. 309–332.

[4] N. M. JOHNSON AND S. J. DREISS, *Hydrostratigraphic interprestation using indicator geostatistics*, Water Resources Res., 25 (1989), pp. 2501–2510.

[5] A. JOURNEL AND F. ALABERT, *New method for reservoir mapping*, J. Petrol. Tech., (1990), pp. 212–218.

[6] A. F. B. TOMPSON, *Flow and transport within the saturated zone beneath Lawrence Livermore National Laboratory: Modeling considerations for heterogeneous media*, Tech. Report UCID-21828, Lawrence Livermore National Laboratory, 1990.

[7] A. F. B. TOMPSON, R. ABABOU, AND L. W. GELHAR, *Implementation of of the three-dimensional turning bands random field generator*, Water Resources Res., 25 (1989), pp. 2227–2243.

[8] A. F. B. TOMPSON AND L. W. GELHAR, *Numerical simulation of solute transport in three-dimensional, randomly heterogeneous porous media*, Water Resources Res., 26 (1990), pp. 2541–2562.

[9] U. S. DEPARTMENT OF ENERGY, *Basic research for environmental restoration*, Tech. Report DOE/ER-0482T, Office of Energy Research, Washington, DC, 1990.