# Parallelization of the WASH123D Code—Phase II: Coupled Two-Dimensional Overland and Three-Dimensional Subsurface Flows

Jing-Ru C. Cheng (ruth.c.cheng@erdc.usace.army.mil) and
Robert M. Hunter (robert.m.hunter@erdc.usace.army.mil)
*Major Shared Resource Center, Information Technology Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS 39180, USA*

Hwai-Ping Cheng (hwai-ping.cheng@erdc.usace.army.mil) and
Hsin-Chi Lin (hsin-chi.j.lin@erdc.usace.army.mil)
*Coastal and Hydraulics Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS 39180, USA*

David R. Richards (david.r.richards@erdc.usace.army.mil)
*Information Technology Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS 39180, USA*

**Abstract.** The parallel WASH123D is designed to simulate watershed systems using a coupled system of one-dimensional (1-D) channels, 2-D overland areas, and 3-D subsurface media on parallel scalable computers. The U.S. Department of Defense High Performance Computing Modernization Program funds the parallelization of the watershed model through the Common High Performance Computing Software Initiative in order to solve one aspect of the battlespace environment, which includes space, weather, ocean, and soil, to develop a complete coupled battlespace environment. Tasks in this project include parallel algorithm development, software toolkit development, and performance studies. Currently, the parallel version, which includes coupled 2-D overland and 3-D subsurface flows, is completed and packaged for application projects, such as calibration and validation of the coupled 2-D/3-D Biscayne Bay Coastal Wetland inland flow model presented in this paper, on various computer architectures.

**Keywords:** watershed modeling, parallel computing, scientific computing, coupled application, Biscayne Bay calibration

## 1. Introduction

The U.S. Army Corps of Engineers plays a critical role in the Nation's watershed management. Watershed models have been researched and developed to simulate major hydrological processes on multiple spatial domains over varied temporal scales with interactions ranging from uncoupled to strongly coupled. Different numerical approaches for such a coupled nonlinear hydrologic processes have been proposed to be efficient and affordable, e.g., Penn State Integrated Hydrologic

Model (PIHM) (Duffy, 2004), WASH123D (Yeh et al., 2003), and HSPF (Hydrologic Simulation Program - FORTRAN). According to Yeh's review (Yeh, 2002), HSPF and WASH123D are the only models that include complete media systems, i.e., stream/rivers, overland regimes, and subsurface media, and encompass the complete suite of fluid flows and thermal, salinity, sediment, and chemical transport processes. The difference between them is that HSPF employs the parametric approach, while WASH123D is based on the first principle, physics-based approach.

The Department of Defense (DoD) Common High Performance Computing Software Initiative (CHSSI) supports the parallel WASH123D software development aiming to efficiently simulate one aspect of battlespace environment. Concurrent with this project, different ongoing application projects are included in the Comprehensive Everglades Restoration Plan (CERP). Not only is the performance demonstrated but also the correctness of modeling is studied. Facilitated by the Groundwater Modeling System (GMS) 5.1 (DoD, 2005), the mesh generation, boundary condition assignment, and results viewing can be easily accomplished visually and interactively.

The remainder of this paper is organized as follows. In Section 2, the mathematical formulation of the watershed system is briefly described. Section 3 presents the software development and the parallel implementation focused on the coupled application. In Section 4, experimental results are presented to demonstrate the correctness and performance of the implementation. Section 5 summarizes these results and discusses planned future work.

## 2. Mathematical Formulation of the Watershed System

The governing equations of two-dimensional (2-D) overland flow and three-dimensional (3-D) subsurface flow as well as the numerical approaches solving the 2- and 3-D coupled flow system are described in the following subsections. The numerical methods presented in this section are those used for demonstration in Section 4. Other numerical approaches employed in this model can be found in Yeh et al. (2003) in detail.

### 2.1. 2-D overland flow

The semi-Lagrangian finite element method (FEM) is used to solve the depth-averaged diffusive wave equation derived based on the following

2-D overland flow governing equation,

$$\frac{\partial h}{\partial t} + \nabla \cdot \mathbf{q} = S + R - E - I \quad \text{or} \quad \frac{\partial h}{\partial t} + \nabla \cdot [\mathbf{V}h] = S + R - E - I \ , \ (1)$$

where $h$ = overland water depth[L], $t$ = time[t], $\mathbf{q}$ = overland flux[L$^3$/t/L], $S$ = man-induced source[L$^3$/t/L$^2$], $R$ = rainfall rate[L/t], $E$ = evapotranspiration rate[L/t], $I$ = infiltration rate [L/t], and $\mathbf{V}$ = overland flow velocity[L/t]. Equation (1) can be written in Lagrangian form as

$$\frac{D_v h}{D\tau} + Kh = S + R - E - I \qquad \text{where} \ \ K = \nabla \cdot \mathbf{V} \ . \qquad (2)$$

Integration of (2) along its characteristic line yields

$$\left(1 + \frac{\triangle\tau}{2} {K_i}^{(n+1)}\right) {h_i}^{(n+1)} = \left(1 - \frac{\triangle\tau}{2} {K_i}^{*}\right) {h_i}^{*} + \frac{\triangle\tau}{2}\left({S_i}^{(n+1)} + {S_i}^{*}\right) + \frac{\triangle\tau}{2}\left({R_i}^{(n+1)} + {R_i}^{*}\right) - \frac{\triangle\tau}{2}\left({E_i}^{(n+1)} + {E_i}^{*}\right) - \frac{\triangle\tau}{2}\left({I_i}^{(n+1)} + {I_i}^{*}\right) (3)$$

where $\triangle\tau$ = the tracking time[t], which equals $\triangle t$ (the time interval) when the backward tracking is carried out all the way to the root of the characteristic line but is less than $\triangle t$ when the backward tracking hits the boundary before $\triangle t$ is completely consumed; ${K_i}^{(n+1)}$, ${h_i}^{(n+1)}$, ${S_i}^{(n+1)}$, ${R_i}^{(n+1)}$, ${E_i}^{(n+1)}$, and ${I_i}^{(n+1)}$ are the values of $K$, $h$, $S$, $R$, $E$, and $I$, respectively, at $\mathbf{x}_i$ at new time $t = (n+1)\triangle t$; and ${K_i}^*$, ${h_i}^*$, ${S_i}^*$, ${R_i}^*$, ${E_i}^*$, and ${I_i}^*$ are the values at ${\mathbf{x}_i}^*$, i.e., where the backward tracking ends. Equation (3) is used to compute the water depth, $h$, at all nodes except for the upstream boundary nodes, where water depth is determined by applying adequate boundary conditions. Since the flow velocity is a function of water depth as follows,

$$\mathbf{V} = \frac{-a}{n}\left[\frac{h}{1 + (\nabla Z_0)^2}\right]^{2/3} \frac{\left(\nabla H + \frac{h}{2\rho}\nabla(\Delta\rho) - \frac{\tau^s}{\rho g h}\right)}{\sqrt{\left|-\nabla H - \frac{h}{2\rho}\nabla(\Delta\rho) + \frac{\tau^s}{\rho g h}\right|}}$$

where $n$ is the Manning's roughness [tL$^{-1/3}$], $a$ is a unit-dependent factor ($a$ = 1 for SI units and $a$ = 1.49 for U.S. Customary units) to make the Manning's roughness unit-independent, $\tau^s$ is the surface shear stress [M/L/t$^2$], and $H$ is the water stage [L], (3) is used in a nonlinear iteration loop until a convergent solution is obtained.

### 2.2. 3-D subsurface flow

The governing equation of subsurface flow through saturated-unsaturated porous media is the well-known Richards equation given as follows.

$$\frac{d\theta}{dh}\frac{\partial h}{\partial t} = \nabla \cdot [\mathbf{K} \cdot (\nabla h + \nabla z)] + q \ , \tag{4}$$

where $\theta$ = moisture content[$L^3/L^3$], $h$ = pressure head [L], $\mathbf{K}$ = the hydraulic conductivity tensor[L/t], $z$ = the potential head[L], and $q$ = man-induced source[$L^3/L^3$/t]. Equation (4) is solved with the Galerkin FEM, whose details, including the discrete equation, boundary condition, linear system, etc., can be found elsewhere (Yeh et al., 2003).

### 2.3. 2- and 3-D coupling

Considering the interaction between the 2-D overland and 3-D subsurface flows, the continuity of fluxes and state variables, e.g., overland water depth and subsurface pressure head, has to impose. If the state variables exhibit discontinuity, then a linkage term is used to simulate the fluxes. Therefore, the interaction must be simulated by imposing continuity of pressures and fluxes as

$$h^o = h^s \quad \text{and} \quad Q^o = Q^s \implies I = \mathbf{n} \cdot \mathbf{K} \cdot (\nabla h^s + \nabla z) \ , \tag{5}$$

where $h^o$ is the water depth[L] in the overland if it is present, $h^s$ is the pressure head[L] in the subsurface, $Q^o$ is the flux [$L^3/L^2$/t] from the overland to the interface, $Q^s$ is the flux from the interface to the subsurface media [$L^3/L^2$/t], and $\mathbf{n}$ is an outward unit vector of the ground surface. The commonly used linkage term such as $Q^o = Q^s = k(h^o - h^s)$ is not appropriate, though it is easy to formulate and implement. This term introduces a nonphysics parameter $k$, which requires calibration to match simulation with field data and renders the coupled model *ad hoc* even though the overland and subsurface models are each individually physics-based.

Algorithm 1 lists the 2-D/3-D coupling algorithm used in WASH123D. In the algorithm, each 3-D flow-time interval may contain many 2-D flow-time intervals with smaller time-step size, which is often required for solving nonlinear 2-D overland flow when a high-resolution mesh is employed. To make computation affordable, the strongly coupled overland and subsurface flows are not adopted. The fluxes through the surface-subsurface interface are updated using (5) for 2-D/3-D in each 3-D coupling/nonlinear iteration.

ALGORITHM 1. *The 2-D/3-D Coupling Algorithm in WASH123D*

*Foreach 3D flow time step ($\triangle t_{3DF}$) do*
*Foreach 3D coupling/nonlinear iteration do*
*Foreach 2D flow time step ($\triangle t_{2DF}$) do*
*Incorporate infiltration/seepage for 2D/3D coupling*
*Foreach 2D coupling/nonlinear iteration do*
*Solve linearized 2D flow equation*
*Endfor*
*Endfor*
*Incorporate infiltration/seepage for 2D/3D coupling*
*Solve linearized 3D flow equation*
*Endfor*
*Endfor*

## 3. Parallel Implementation and Software Development

In WASH123D, different components in the coupled system are solved by different numerical approaches. Since the original serial computational kernel comprising these numerical approaches can be used by the parallel algorithms, they are left untouched to shorten the development time. In order to reach this goal, the data structure design becomes very important when object orientation, parallel implementation, software integration, and language interoperability are also factors in the software development.

### 3.1. Data structure design

The problem domain is discretized and constructed in the object `WashDomain` as sketched in Fig. 1. In the domain, there are three `WashMesh` objects to account for three systems—1-D river/stream network, 2-D overland regime, and 3-D subsurface media—that may be included. In addition, the `WashDomain` also includes (1) a coupling object named `WashCouple`, (2) the `WashGlobal` object describing the common phenomena, and (3) the `WashProcinfo` object containing the parallel environment context. Each `WashMesh` object may include `vtxDomain`, `elementDomain`, and `bdyDomain`, which are created and managed by DBuilder (Hunter and Cheng, 2004) interfacing ParMETIS (Research Team led by George Karypis, 2003) to partition each subdomain (i.e., `WashMesh`) independently and managing all the communication/synchronization tasks among them. The `WashCouple` may include the coupler for (1-D, 2-D), (1-D,

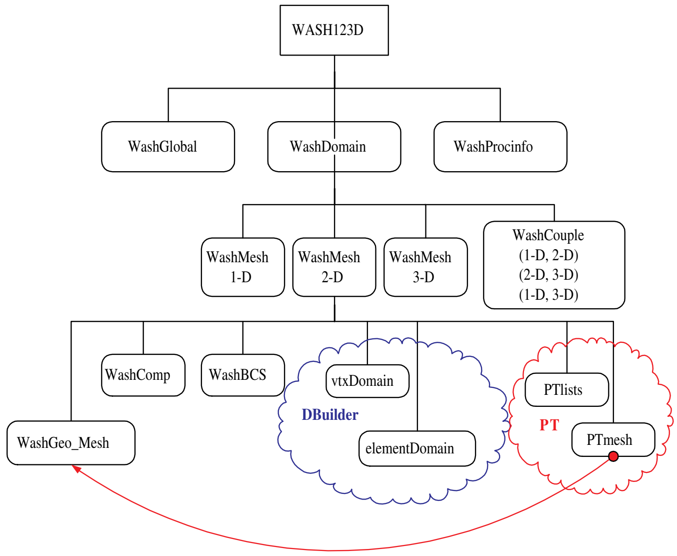


*Figure 1.* Data structure design of the parallel WASH123D

3-D), and/or (2-D, 3-D) interactions. The coupler encapsulates all the Message Passing Interface (MPI) function calls for coordination of cross subdomain (i.e., `WashMesh`); thus this software tool can easily integrate two or more applications with different physics on multidomains.

### 3.2. Software tool integration and development

The first two authors elected the approach of software tool integration and development when parallelizing the WASH123D code. In the 2-D overland subdomain, the system equations (1) are solved using particle tracking methods, with the Picard method solving the nonlinearity of the 2-D overland flow (3), to compute the total time-derivative term in (2) and by integrating the source/sink terms along the tracking path. The parallel Particle Tracking (PT) software (Cheng et al., 2004; Cheng and Plassmann, 2004b) is well developed, but it lacked the functionality of tracking particles accurately under unsteady flow fields. Thus the PT software was then facilitated with a new pathline computation kernel (Cheng and Plassmann, 2004a) and then integrated into the parallel code through a lightweight functional interface that PT provides (Cheng et al., 2004).

DBuilder (Hunter and Cheng, 2004) aims to support domain partitioning, parallel data management, coupling coordination, and parallel solver interface. This toolkit can build a vertex domain with a distributed number of vertices, an element domain with a distributed number of elements, and a boundary element domain comprised of boundary elements in the element domain, required by different numerical methods. DBuilder provides a default rule and a callback approach to users for coordination between vertices and elements required by finite element applications. DBuilder encapsulates all the MPI function calls so that application codes can instead call a set of user-interface functions that the toolkit provides to retrieve/modify parallel data. In this decade, multiphysics applications on multidomains have become a large focus, which requires the multidomain integration to integrate two or more applications. The spatial relationship between computational domains can be adjacent, partially/fully overlapped, or distinct. DBuilder allows for the building of a coupler object to avoid the dependency between meshes when partitioning. Details can be found in Hunter and Cheng (2004).

## 4. Experimental Results

### 4.1. Model calibration and validation

The parallel WASH123D code has been used in the calibration and validation of the coupled 2D/3D Biscayne Bay Coastal Wetland (BBCW) inland flow model in the CERP. The data of Years 1995-1996 (wet year) and 1999-2000 (dry year) were used in calibration, while the data of Years 1998-1999 (average year) were used for model validation.

The 3-D mesh has the 2-D mesh as its top boundary face and contains seven materials. In Fig. 2, the 3-D mesh comprises 66,712 vertices and 114,716 triangular prisms. The available observed canal stages were employed as the boundary conditions and imposed at the canal corresponding overland/subsurface vertices. The canal levees and major roads were considered to be drainage divides for the overland flow, and zero-depth boundary condition was assigned at the corresponding overland nodes. Rainfall and evapotranspiration are given based on the gauge locations. Pumping wells are considered, and the rate is distributed evenly to the vertices associated with the well.

Since most of the overland surface in the BBCW project domain was dry except for the areas close to the coastal line and for the time periods when significant storm events exist, the Manning's roughness coefficients for the four top soils were assumed based on the land-use information. Thus, the model calibration is to determine an adequate set of hydraulic conductivities for the seven materials. The groundwater observation wells (24 total) within the domain of interest are marked in Fig. 2. In total, there are 27 model runs conducted with different hydraulic conductivity associated with each material. Each model run took about 5 to 6 days of wall clock time for a 1-year simulation on a 16-way job of the Compaq SC45 machine at the Engineer Research and Development Center Major Shared Resource Center (ERDC MSRC). A comparison between the observed and the computed was made to determine the best set of hydraulic conductivity among the 27 model runs, where the mean absolute error ($\epsilon$) and the root mean squared error ($\sigma$) were used in the comparison. Table I lists $\epsilon$ and $\sigma$ of Runs 20 through 27 for these 3 years. Figure 3 shows the total head distribution at Wells G-1486 and G-1363 between observed values and simulation results of the validation case.

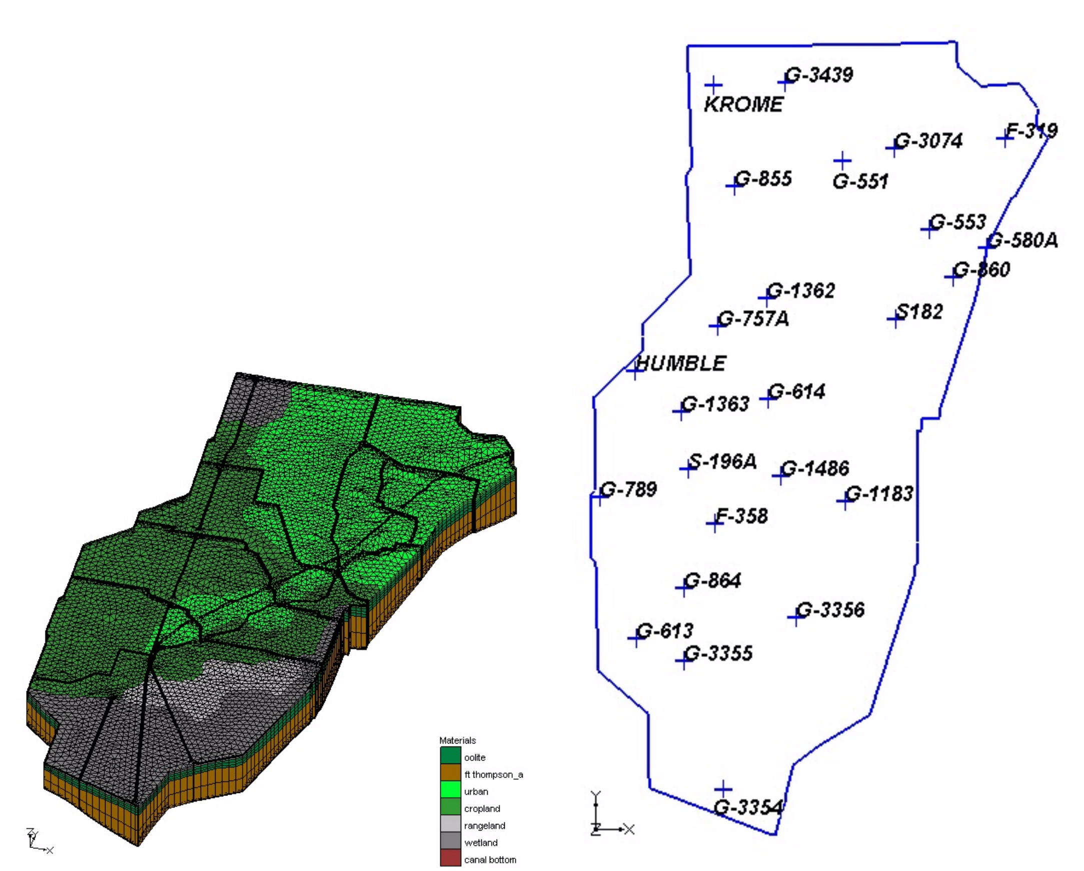


*Figure 2.* BBCW computation mesh (left) and BBCW observation wells (right)

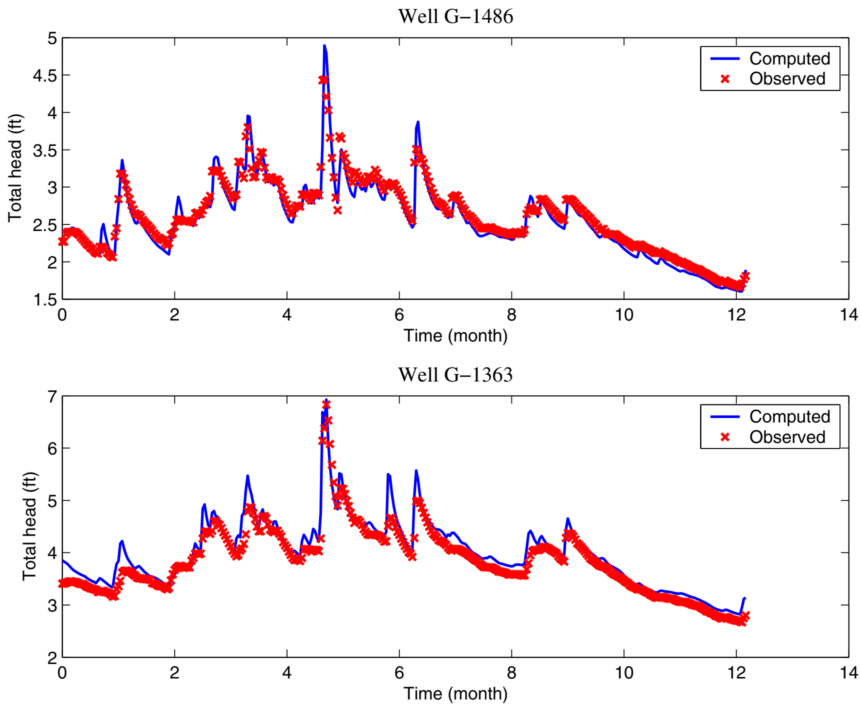


*Figure 3.* Total head distribution at Well G-1486 (top) and at Well G-1363 (bottom) of the validation case (Run 25)

### 4.2. Performance study

Fort Benning military base is located in the Upatoi River watershed, west-central Georgia south of the city of Columbus, GA, and east

Table I. Average $\epsilon$ and $\sigma$ of Runs 20 through 27

| Run ID | Wet Year | | Dry Year | | Average Year | |
|---|---|---|---|---|---|---|
| | $\epsilon$ | $\sigma$ | $\epsilon$ | $\sigma$ | $\epsilon$ | $\sigma$ |
| 20 | 0.492 | 0.608 | 0.367 | 0.474 | 0.338 | 0.403 |
| 21 | 0.461 | 0.574 | 0.370 | 0.478 | 0.339 | 0.402 |
| 22 | 0.447 | 0.561 | 0.359 | 0.465 | 0.333 | 0.395 |
| 23 | 0.260 | 0.334 | NA | NA | NA | NA |
| 24 | 0.258 | 0.335 | NA | NA | NA | NA |
| 25 | 0.257 | 0.330 | 0.293 | 0.421 | 0.226 | 0.287 |
| 26 | 0.299 | 0.385 | 0.299 | 0.428 | 0.238 | 0.306 |
| 27 | 0.335 | 0.428 | 0.322 | 0.455 | 0.271 | 0.337 |

of Phenix City, AL. Fort Benning is a primary training area for the U.S. Army Infantry. The 2-D overland domain, which covers about 450 square miles, is discretized with 103,619 vertices and 206,167 elements. The underlying 3-D domain contains 3,092,505 elements and 1,657,904 vertices.

At the ERDC MSRC, the Compaq AlphaServer SC45 machine is configured with 128 nodes. Each node has four 1-GHz processors connected by a 64-port, single-rail Quadrics high-speed interconnect switch. At the Naval Oceanographic Office MSRC, the IBM P4 machine has 168 nodes. Each node has eight 1.3-GHz CPUs and 8 GBs of memory. Nodes communicate through the IBM's Colony II switch. The linux cluster, Evolocity II, at the Army Research Laboratory (ARL) MSRC has 256 processors. Each has a 3.06-GHz CPU using Myrinet interconnect for communication. Figure 4 plots the wall clock time vs. number of processors for the coupled 2-D overland flow simulation and 3-D subsurface flow simulation on the Fort Benning site on three different architectures. Figure 4 also shows the communication overhead taken up in the total wall clock time on these three different machine architectures.

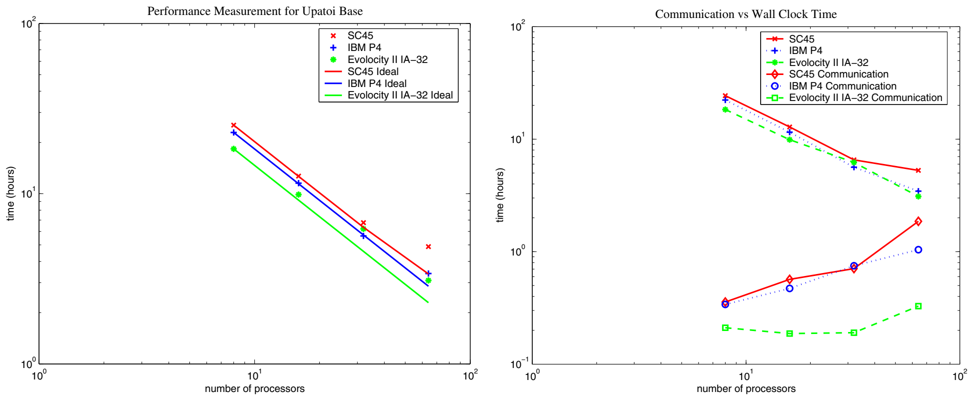


*Figure 4.* Performance measurement on three machines (left) and communication overhead compared with the total wall clock time (right)

From these figures, one can observe that ARL's linux cluster outperforms the others except when the simulation runs on 32 processors. As for the 64-processor simulation, the parallel efficiency is around 74 percent on the Evolocity II linux cluster, around 84 percent on the IBM P4, and less than 70 percent on the SC45. The main cause for such a difference is that the IBM P4 has better scalability with communication (see Fig. 4) and the Evolocity II has better CPU speed.

Figure 5 plots the high-water memory marks to show the memory scalability. The parallel implementation does impose a significant amount of memory overhead, while it can reduce a large amount of

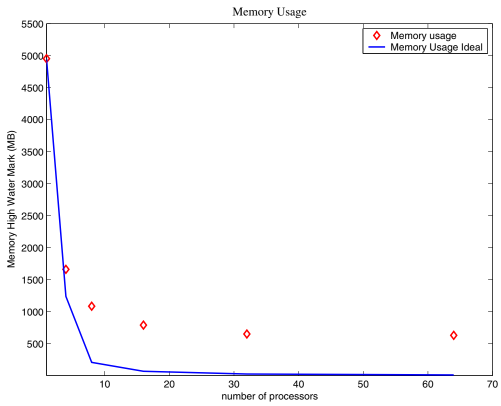


*Figure 5.* Memory usage when run on different numbers of processors

memory required on each processor. The reduction of memory can then reduce the cache miss or page swapping.

## 5. Summary and Future Plans

Current development of the parallel WASH123D has completed the 2-D overland and 3-D subsurface coupled flow system. The software tool DBuilder has successfully embedded MPI routines, reducing the application development work through leveraging of the MPI library and parallel algorithms. The result shows that the implementation in DBuilder has successfully partitioned the domain, balanced the work-load, and scaled the memory usage among processors. The experimental results show that the parallel code has been run on the production stage and promises a performance speedup. In summary of this phase, the following tasks for the software development have been completed: implementation of dynamic memory allocation, DBuilder functionality enrichment, parallel PT software integration, the parallel performance evaluation, and calibration/validation of real-world problems. Further detailed profiling such as communication and memory overhead of each component will be performed. The WASH123D team is pushing forward the development of the 1-D module and the associated coupled modules as well. Large-scale field problems are also desired for a large number of processors.

## Acknowledgments

This work was supported in part by a grant of computer time from the DoD High Performance Computing Modernization Program at the

ERDC MSRC. This work was also supported by the U.S. Army Engineer District, Jacksonville (SAJ) and the South Florida Water Management District (SFWMD).

## References

Cheng, J.-R. C., M. T. Jones, and P. E. Plassmann: June-September 2004, 'A Portable Software Architecture for Mesh Independent Particle Tracking Algorithms'. *Parallel Algorithms and Applications* pp. 145–161.

Cheng, J.-R. C. and P. E. Plassmann: 2004b, 'Parallel Particle Tracking Framework for Applications in Scientific Computing'. *The Journal of Supercomputing* **28**, 149–164.

Cheng, J.-R. C. and P. E. Plassmann: Feb. 25-27, 2004a, 'Development of Parallel Particle Tracking Algorithms in Large-scale Unsteady Flows'. In: *SIAM Conf. on Parallel Processing for Scientific Computing*. San Francisco, CA.

DoD, DoE, E.: 2005, 'GMS Web page'. http://chl.erdc.usace.army.mil/gms, Coastal and Hydrulics Laboratory, Engineer Research and Development Center, U.S. Army Corp of Engineers.

Duffy, C. J.: 2004, 'Semi-discrete dynamical model for mountain-front recharge and water balance estimation, Rio Grande of southern Colorado and New Mexico'. In: *Groundwater Recharge in a Desert Environment: The Southwestern United States. Water Science and Applications Series.* pp. 255–271, American Geophysical Union, D.C.

Hunter, R. M. and J.-R. C. Cheng: June 7-11, 2004, 'DBuilder: A Parallel Data Management Toolkit for Scientific Applications'. In: *DoD High Performance Computing Modernization Program, 2004 Users Group Conference*. Williamsburg, VA.

Research Team led by George Karypis: 2003, 'ParMETIS Parallel Graph Partitioning'. http://www-users.cs.umn.edu/ karypis/metis/parmetis/.

Yeh, G.-T., H.-P. Cheng, G. Huang, F. Zhang, H.-C. Lin, E. Edris, and D. Richards: 2003, 'A Numerical Model of Flow, Heat Transfer, and Salinity, Sediment, and Water Quality Transport in WAterSHed Systems of 1-D Stream-River Network, 2-D Overland Regime, and 3-D Subsurface Media (WASH123D: Version 2.0)'. Technical Report Draft, Engineer Research and Development Center, U.S. Army Corps of Engineers, MS.

Yeh, G.-T.: 2002, 'A rigorous treatment of interactions between various media in first principle, physics-based flow and transport modeling in watersheds'. In: *EOS Transac.*, Vol. 83. p. F436, American Geophysical Union, D.C.