Algorithms and TCAD Software using Parallel Computation

Robert W. Dutton
Stanford University
Stanford, California 94305-4055

Abstract

Progress in algorithms and simulators that exploit parallel computers are reported. Results using two generations of the Intel iPSC architecture for device analysis of 2D and 3D bipolar problems are used to illustrate substantial progress being made in parallelization. New approaches in the areas of hydrodynamic and Monte Carlo analysis are also discussed.

Introduction

Over the past decade the computational demands of TCAD have grown as have the power and availability of new parallel computers. Specifically, requirements for large 3D process and device simulations with grid numbers reaching several million have become essential---especially to analyze parasitic effects where multiple devices are involved. On the hardware side, tightly coupled shared memory machines (Cray and Alliant for example) have given way to distributed memory machines (Intel and Thinking Machines) and now experimental architectures that attempt to provide distributed and shared memory--for example the DASH Project at Stanford are being pursued. The focus of this talk is primarily aimed at summarizing experiments at Stanford in 2D and 3D device simulation made over the past several years using parallel computers. This involves a variety of algorithmic approaches including: drift-diffusion, Monte Carlo and most recently hydrodynamic formulations. The primary class of computers used in this study was the Intel iPSC series (including the commercial i860 and the experimental "Delta" machine at Caltech). Results show very promising performance and potential for achieving TeraFLOP performance using 1000s of processors within a few years.

Physical Models and Algorithms

There is a range of physical models that can be applied to semiconductor device modeling problem. Basically there are three major classes of ways to solve the Boltzmann Transport Equations using:

1. The first moment (drift-diffusion or D/D)
2. Higher order moments (hydrodynamic(HD) , energy transport (ET)...)
3. Monte Carlo (MC) as a means to directly build the statistics.

In this talk we will give examples of how each of these approaches can exploit parallel computation.

Drift-diffusion (D/D)---In the case of the classical D/D formulation, we have previously reported results on parallelization of the 3D STRIDE code. Here we report parallelization of the well-known 2D PISCES code. Figure 1a illustrates all steps needed to parallelize the PISCES code and the associated percentages of time required in the simulation. Clearly, the matrix solution part is a major but not dominant fraction. Also, one can note that parallelization of assembly of matrix elements is every bit as important in the overall effort. Figure 1b shows a profile of actual run times (wall clock time) versus the number of processors on the Intel iPSC/i860 for two problems. For comparison the times on a single CPU SUN 4/670 are also listed. Note that for the 27,600 equation problem, the speed-up factors are substantial. Also note that beyond 16 nodes for this particular problem the speed-up factor degrades due to communications limitations. Further details will be discussed in the presentation.

In previous papers we have discussed progress in 3D D/D modeling using the prototype STRIDE code [1][2]. Over the past year we have scaled the problem size and number of processors used in the analysis to much larger numbers than previously reported. Figure 2 shows the wall-clock time and grid density (two y-axes) versus the number of processors used on the Delta machine at Caltech. This is an experimental version of the Intel iPSC architecture with a backplane interconnect designed at Caltech. As can be observed from the data, we see excellent performance and efficiency even up to more than 500 processors and grid density approaching 5 Million for a bipolar example. The sustained performance was 1.7 GFlops. In this talk we will discuss some of the algorithms used with regards to preconditioning that are essential in achieving such favorable results.

Hydrodynamic (HD) and beyond---The limitations of the D/D formulation are well-known and more advanced models such as the HD or ET formulations are being used extensively. From a physical modeling point of view we continue to extend the ET approach [3]. From a numerical point of view we are investigating the HD formulation in the context of major advances in the CFD domain [4]. Specifically, we are using the time-space GLS formulation in concert with the HD carrier transport equations and developing a parallelized version of the ENSA solver code (Euler Navier-Stokes Analyzer). Since a paper on this subject has been submitted to this conference, we will only briefly highlight the results.

Monte Carlo (MC)--- is a still more complete physical approach to solving the BTE. While there are many limitations in applying MC analysis due to boundary conditions, especially across heterojunction and dielectric interfaces, there are many aspects of hot carrier phenomena that can be effectively solved using MC analysis. At SISDEP we have presented results of efforts in parallelization of MC using the University of Bologna's BEBOP code [5]. Further efforts in this direction are being reported by the Matsushita group. The basic factors affecting parallelization of MC analysis are straightforward from an algorithmic point of view. On the other hand, there are still many innovations that are emerging to use either harmonic expansions or alternate formulations for the variables, including symmetry relationships. The discussion of MC analysis will center on promising new approaches.

Discussion

The above examples illustrate the progress that is being made at three levels in developing algorithms that exploit new parallel computers. The efforts in parallelization of the classical D/D formulation show that substantial speed-ups and robust convergence on very tough bipolar problems can be achieved well into the multi-million grid domain. As for higher moments of the BTE, we see two important trends. First, the development of new algorithms such as the GLS formulation used in CFD and its parallelization should yield major benefits. Second, the opportunities to advance Monte Carlo methods are far from "out of gas" in terms of innovation and computational enhancements. A final area of discussion in this paper is the growing importance of overall support in the development and parallelization of TCAD codes. The most obvious of these needs is in the area of gridding and the domain decomposition. Other areas include: user interfaces, control of both nonlinear and linear solvers and alternative formulations. These issues are related to frameworks, standards and code development strategies.

Conclusions

Progress in developing TCAD algorithms and particularly device analysis codes using parallel computation is reported. Excellent efficiency and convergence behavior on 2D and 3D bipolar problems is demonstrated using production and experimental versions of the Intel iPSC architecture. Progress in solving higher moments of the BTE are reported and new algorithmic advances are discussed. Finally, challenges of overall software engineering are considered.

Acknowledgments

This work has been supported by not only government agencies (ARO and DARPA) but also through an industrial consortium (CompTech under the Department of Commerce, State of California). We especially appreciate the encouragement and ongoing support from the CSTO office of DARPA and Dr. Gil Weigand. The specific results quoted on STRIDE came from Dr. K.C. Wu, those on PISCES-MP from Mr. Bruce Herndon and the Aladdin-CAD project contributions came from Dr. Ronald Goossens.

References

[1] K. C. Wu, G. R. Chin, R. W. Dutton, "A STRIDE Towards Practical 3D Device Simulation--Numerical and Visualization Considerations," *IEEE Trans. CAD*, Vol. 10, No. 9, pp. 1132-1140, Sept. 1991.

[2] Z. Y. Wang, K. C. Wu, R. W. Dutton, "An Approach to Construct Pre-Conditioning Matrices for Block Iteration of Linear Equations," *IEEE Trans. on CAD*, Vol. 11, No. 11, pp. 1334-1343, Nov. 1992.

[3] D. Chen, E. C. Kan, U. Ravaioli, C. Shu and R. W. Dutton, "An Improved Energy Transport Model for Hot-electron Device Simulation, " *IEEE Elec. Dev. Lett.*, vol. 13, no. 1, pp. 26-28, Jan. 1992.

[4] N. R. Aluru, A. Raefsky, P. M. Pinsky, K. H. Law, R. J. G. Goossens, and R. W. Dutton, "A Finite-Element Formulation for the Hydrodynamic Device Equations," Accepted for publication in *Computer Methods in Applied Mechanics and Engineering*.

[5] S. Sugino C.S. Yao and R.W. Dutton, "Parallelization of Monte Carlo Analysis on Hypercube Multiprocessors and on a Networked EWS System," SISDEP '91 Digest, Vol. 4, pp. 275-284, Zurich, Sept. 1991.

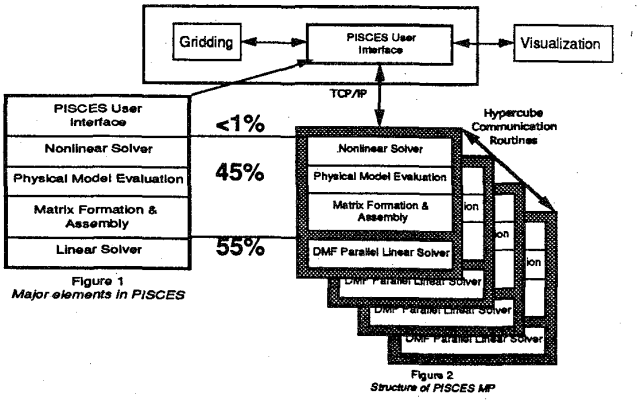

Figure 1
Major elements in PISCES

Figure 2
Structure of PISCES MP

Figure 1(a): Structure of PISCES-MP showing how different portions of the code affect execution and resulting parallelization requirements

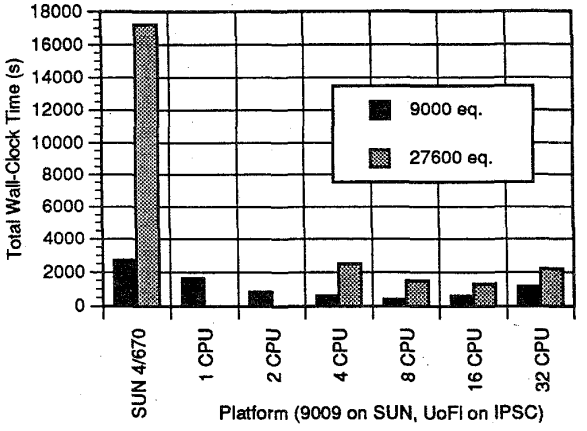

Figure 1(b): Performance of PISCES-MP showing wall-clock time versus number of CPUs on the Intel iPSC/i860 for two bipolar problems (9000 and 27,600 equations). Benchmarks for same problems on a SUN 4/670 are also shown.

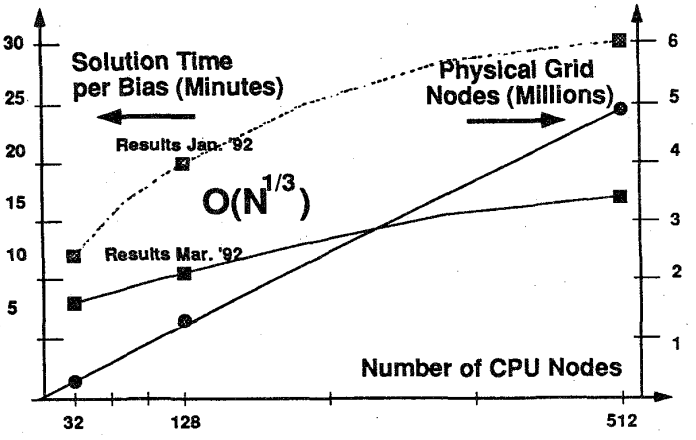

Figure 2: Solution time and number of grid in millions as a function of number of CPU nodes on the Intel Delta machine at Caltech. The problem is a 3D bipolar simulation and the code used in the experimental STRIDE program [1].