Asenov, A. and Brown, A.R. and Roy, S. (1998) Parallel semiconductor device simulation: from power to 'atomistic' devices. In, *International Workshop on Computational Electronics, 19-21 October 1998*, pages pp. 58-61, Osaka, Japan.

http://eprints.gla.ac.uk/3031/

Parallel Semiconductor Device Simulation: from Power to 'Atomistic' Devices

A. Asenov, A.R. Brown and S. Roy

Device Modelling Group
Department of Electronics and Electrical Engineering
University of Glasgow, Glasgow G12 8LT, UK
Tel: +44 141 330 5233, Fax: +44 141 330 4907
E-mail: A.Asenov@elec.gla.ac.uk

This paper discusses various aspects of the parallel simulation of semiconductor devices on mesh connected MIMD platforms with distributed memory and a message passing programming paradigm. We describe the spatial domain decomposition approach adopted in the simulation of various devices, the generation of structured topologically rectangular 2D and 3D finite element grids and the optimisation of their partitioning using simulated annealing techniques. The development of efficient and scalable parallel solvers is a central issue of parallel simulations and the design of parallel SOR, conjugate gradient and multigrid solvers is discussed. The domain decomposition approach is illustrated in examples ranging from 'atomistic' simulation of decanano MOSFETs to simulation of power IGBTs rated for 1000V.

1. Introduction

Computer-aided numerical modelling and simulation has become an indispensable tool in the understanding, design and optimisation of various semiconductor devices. The complex architecture of modern devices requires in many cases 3D simulation. The use of parallel processing systems is a widely accepted approach to attain the computational power and memory requirement inherent in 3D simulation [1-4]. However, considerable attention must be paid to the underlying architecture of the parallel system to ensure maximum efficiency, scalability and portability of the code.

We focus our discussion on the design of semiconductor device simulation algorithms for mesh connected MIMD platforms with distributed memory and four way connectivity. Our approach is based on finite difference or structured topologically rectangular finite element 3D grids [5]. The nature of such grids makes them amenable to partitioning over mesh connected arrays of processors using domain decomposition techniques. The relative simplicity of the corresponding parallel code design reduces the time-to-answer when new models, simulation techniques and devices are investigated.

We briefly describe the domain decomposition approach and the optimisation of the partitioning in the next section. Basic methods for generation of structured topologically rectangular 2D and 3D finite element grids are discussed in Section 3. Several aspects of the design of SOR, conjugate gradient and multigrid parallel solvers are discussed in Section 4. Finally in Secton 5 we give examples of large scale parallel semiconductor device simulation.

2. Domain decomposition

The basic idea of decomposing a 3D semiconductor device solution domain over a 2D array of NxM processors is illustrated in Fig.1 for a quarter of an IGBT cell.
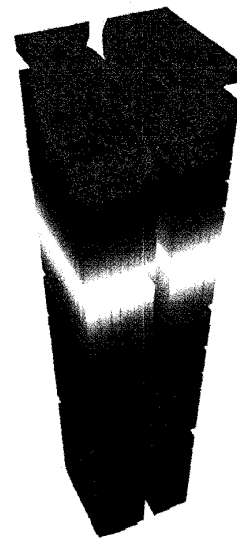

Fig.1: Partitioning of a 3D semiconductor device solution domain over a 2D array of $2x4$ processors.

The device is partitioned into $2x4$ subdomains along two of the spatial dimensions and each of the subdomains include the whole third dimension. In the above partition each processor is assigned a column of elements partitioned in one spatial plane and including all the elements in the

third direction. The partitioning must ensure that the edges of grid subdomains overlap only on neighbouring processors. For many iterative linear solvers [6], highest parallel efficiency is obtained when the largest subdomain size volume, and the largest subdomain surface are at a minimum. Ideal first order load balancing only occurs when the number of grid nodes is exactly divisible by the number of processors in each dimension of the processor array. Otherwise deep oscillations in speedup and efficiency occur (Fig.2). To improve speedup for an arbitrary grid size, an alternate partitioning can be found using simulated annealing [7] which preserves the 4-way connectivity of grid subdomains and smoothes the performance oscillations.
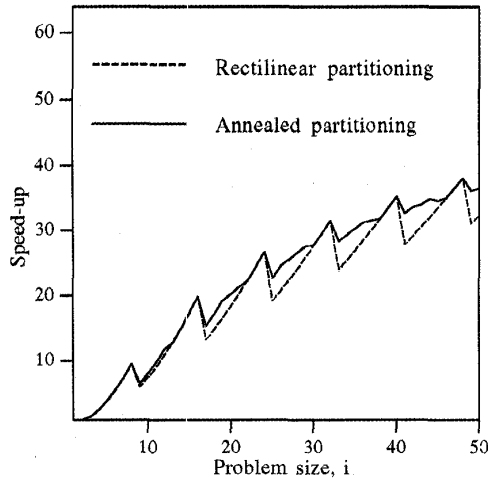

Fig.2: Speed-up with and without optimisation for an $8{\times}8$ processor array

3. Topologically rectangular grids

To simplify the domain decomposition and the corresponding code design we use structured 2D and 3D topologically rectangular grids. Such grids allow two or three index ordering preserving the number of grid nodes in each one of the index directions. Nodes with neighbouring indices are physically adjacent in the grid. Most finite difference grids are inherently topologically rectangular, however it is also possible to construct topologically rectangular finite element (FE) grids. Although such requirements restrict to some extent the flexibility of the FE approximation we have found that devices with rather complicated shapes may be triangulated with topologically rectangular grids.

Fig.3 illustrates the basic concepts of the topologically rectangular grid in a 2D example of finite element triangulation of a circle. Although such grids does not have the full flexibility of unstructured FE grids they allow for precise approximation of the region boundaries and local density refinement. The concept can be extended to 3D and Fig.4 illustrates the 3D FE triangulation of an etched quantum dot.
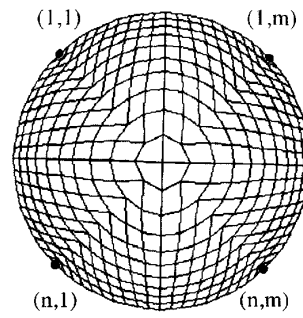

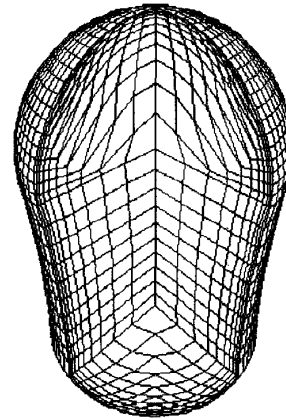Fig.3: Triangulation of a circle with a nonuniform topologically rectangular grid.

Fig.4: Triangulation of a etched quantum dot with a 3D topologically rectangular FE grid.

With some care much more complex devices such as IGBTs (Fig.5) can be triangulated in 3D simulations using topologically rectangular grids.
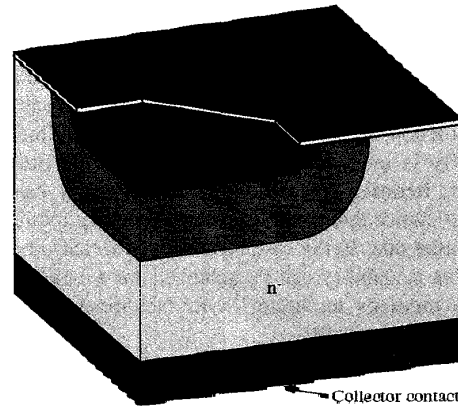

Fig.5: Schematic view of an IGBT.

As can be seen from Fig.6 that the grid conforms not only to the cellular shape of the device but also to the metallurgical pn junctions inside. Fig.7 illustrates the quality of the approximation of the complex shape of the pn^- junction deformed by implantation in the inter-cell space.
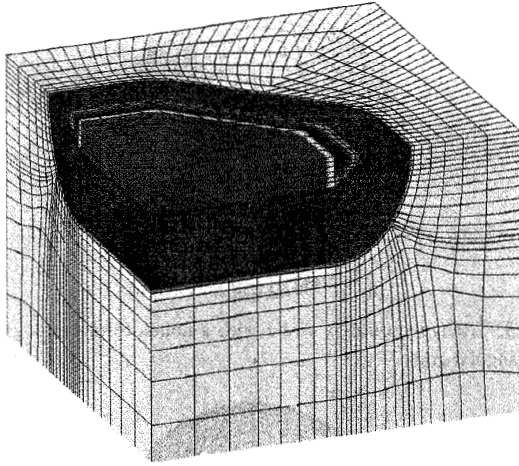
59

Fig.6: Triangulation of a 1/4 of an IGBT cell with a topologically rectangular grid.
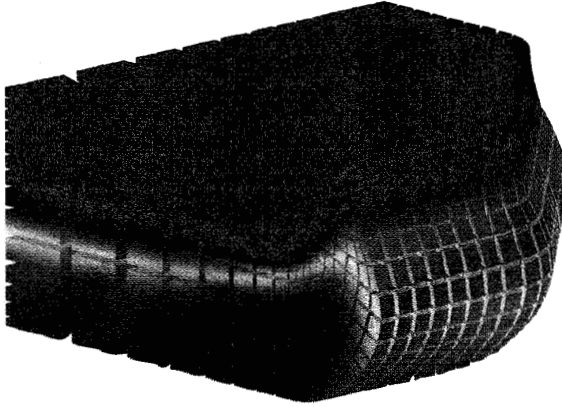

Fig.7: Detail of the grid enclosed by the pn^- junction for a stopper implanted IGBT.

4. Parallel solvers

The design of parallel linear solvers is an open area of research. The efficient parallelisation of sparse LU decomposition is extremely difficult to achieve and good scalability is even harder. In the case of 3D problems, however, iterative linear solvers are in many cases the preferred choice due to the enormous memory requirement of the direct one. In the case of mesh connected processors acceptable scalability can be achieved for a large class of iterative methods including SOR, Newton-SOR [8] and multigrid techniques [9].

Conjugate gradient (CG) type solvers are also amenable to parallelisation but the implementation of efficient and scalable preconditioning is still an issue. The incomplete Cholesky LU decomposition which is the preferable choice for single processor preconditioning of BiCGSTAB solvers is not inherently parallel. An alternative choice is to use polynomial preconditioning [10] which has a much higher degree of parallelism as it only requires the calculation of matrix-vector products. In Figs. 7 and 8 we illustrate the

effect of various degree of polynomial preconditioning on the performance of a BiCGSTAB solver for the systems of equations arising from the discretisation of the Poisson and current continuity equations respectively.
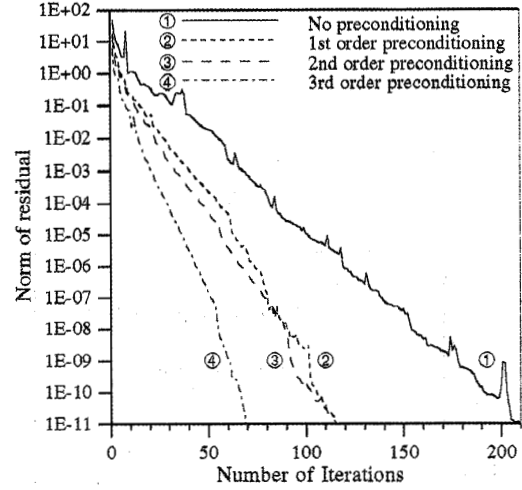

Fig.8: Convergence property of BiCGSTAB solver with polynomial preconditioning solving the system arising from the discretisation of the Poisson equation in a power diode simulation.
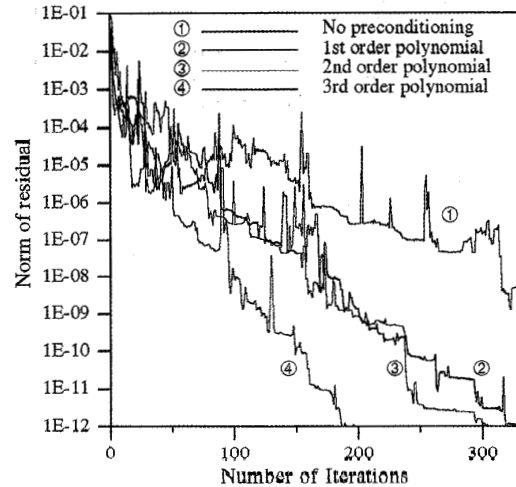

Fig.9: Convergence property of BiCGSTAB solver with polynomial preconditioning solving the system arising from the discretisation of the electron current continuity equation in a power diode simulation.

Due to the stable positive definite structure of the matrix arising from the discretisation of the Poisson equation the convergence of the BiCGSTAB solver is much faster and smoother. The ill-conditioning and large dynamic range of the variables in the current continuity case slows down the convergence. The ripples in Fig.9 are most probably associated with truncation errors in calculating the direction of descent.
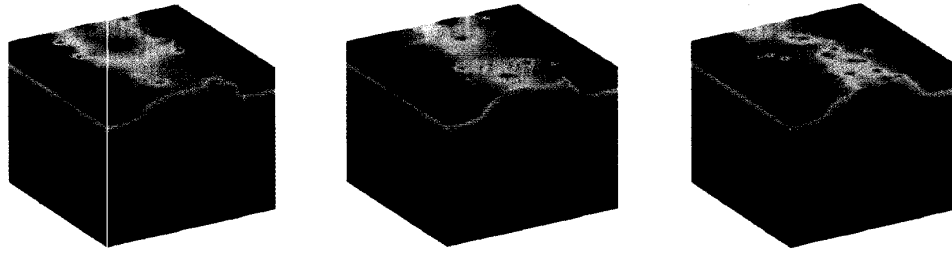
Fig.10: Potential distribution in three 30 nm MOSFETs with different microscopic arrangements of the dopants.

5. Examples

'Atomistic' device simulation

The discrete stochastic distribution of dopants in sub 100nm MOSFETs results in 3D potential and current distributions. Study of the corresponding fluctuation effects requires 3D simulations with fine grain discretisation. Statistically significant samples of microscopically different devices have to be simulated in order to understand the trends in the variation of the parameters and to build up reliable statistics on which the IC design and optimisation should be based. This is a computationally demanding task and a good candidate for parallel simulations. Fig.10 illustrates the distribution of the potential at threshold voltage for three macroscopically identical 30nm MOSFETs with different microscopic arrangements of the dopants and completely different threshold voltages.

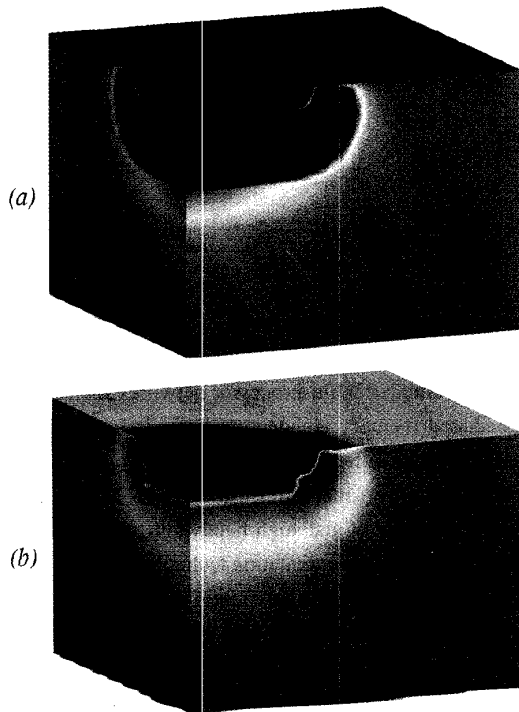### Cellular IGBT simulation

Fig.11: Distribution of electrons (a) and holes (b) in an cellular IGBT.

The distribution of electrons and holes in a cellular IGBT in on-state is shown in Fig.11 (a) and (b) respectively. In Fig.11(a) the MOSFET channel is clearly seen. On both figures significant ambipolar injection leading to conductivity modulation is seen in the low doped drift region of the device.

6. Conclusions

In this paper parallel approaches based on mesh connected arrays of processors for the purpose of semiconductor and nanostructure device simulation have been presented. The specific features of the parallel platform are accounted for in the design process which ensures the scalability and portability of the codes.

References

1. R.W. Dutton, K.H. Law, P.M. Pinsky, N. R. Aluru and B.P. Herndon: Proc. NASA Semiconductor Device Modeling Workshop (1996) 15.

2. V. K. Naik, K. Eswar, M.K. Ieong: Proc. NASA Semiconductor Device Modeling Workshop (1996) 77.

3. O. Schenk, K. Gartner, W. Fichtner: Swiss Federal Institute of Technology Zurich, Technical Report No. 97/19.

4. U.A. Ranawake, C. Huster, P.M. Lenders and S.M. Goodnik: IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems 13 (1994) 712.

5. A. Asenov, A. Brown and J.R. Barker: VLSI Design 6 (1988) 91.

6. A. Asenov, D. Reid and J.R. Barker, Parallel Computing 21 (1995) 669.

7. S. Roy, A. Asenov and J.R. Barker, Proc. Eurosim '96, eds. L. Dekker, W. Smit and J.C. Zuiderwaart, Elsevier Science B.V. (1996) 179.

8. A. Asenov, D. Reid, A. Brown and J.R. Barker: Transputer Application and Systems 1 (1993) 578

9. C.R. Arokianathan, J.H. Davies and A. Asenov: VLSI Design 8 (1998) 331

10. O.G. Johnson, C.A. Micchelli and G. Paul: Siam J. Numer. Anal. 20 (1983) 362