Parallel Simulation of Semiconductor Devices on MIMD Machines

Don Reid, Asen Asenov, John R. Barker, Steve P. Beaumont

Nanoelectronics Research Centre,
Department of Electronics and electrical Engineering,
The University of Glasgow,
Glasgow, G12 8QQ, UK

Abstract
In this paper two unified, scalable and transportable parallel approaches to the numerical simulation of semiconductor devices are presented: Concurrent Device Simulation (CDS), and Spatial Device Decomposition (SDD). Both approaches have been developed and tested on the Parsytec Supercluster Model 64 which is a medium size transputer system. A series of examples illustrate the application of the developed parallel simulation tools.

I. Introduction

The rapid progress in device technology, which made possible the fabrication of nanometer scale band engineered structures [1], is now having a considerable impact on device simulation. New physical phenomena such as carrier heating, ballistic transport and quantisation govern the behaviour of nano-scale devices [2] and require more complex and expensive models for simulation. For nano-structure devices, full 3D simulation has becoming a necessity. The need for super computer power often restricts both the complexity of physical models involved in device simulation programs and their widespread application. In the foreseeable future, a significant low-cost improvement in computing performance will only be available through Multiple Instruction Multiple Data (MIMD) systems where necessary speed-up derives from the use of parallel processors sharing a large distributed memory. Hence the parallel implementation of device simulation codes has become essential [3]. Such fast parallel codes could be used for real device optimisation, sensitivity analysis and yield prediction in both research and industry. However, to be of practical use, parallel device simulation programs must be unified, scalable and portable.

We report on two parallel approaches to numerical simulation of semiconductor devices, developed in the Nanoelectronic Research Centre at Glasgow University: Concurrent Device Simulation (CDS), and Spatial Device Decomposition (SDD). Both approaches are implemented on the departmental transputer system - the Parsytec Supercluster Model 64 - which is a medium class MIMD system. It consist of 64 electronically reconfigurable T800 transputers with 4MB of local memory per processor.

II. Concurrent Device Simulation

Any practical simulation run can cover a large matrix of input data parameters from bias points to details of device design such as gate length, vertical layer structure, doping distribution, recess shape etc. A simple but effective form of parallel device simulation is to run several copies of the serial simulation code concurrently, implementing Single Program Multiple Data (SPMD) computational model [4]. To this end a fileserver has been designed to pass input data sets to a pipeline of processors running our serial H2F simulator [5].

The fileserver is split into two parts: the server which is system dependent and the harness which is application dependent. The server runs on a simple pipeline of processors (fig. 1). First it runs a *pre-processing* routine. Then it collects input data packets from a *send-data*

routine, passes them to a free processor on which a *process-data* routine (the simulator H2F in our case) runs, collects data from the *process-data* routine and returns them to the master processor for a *receive-data* routine. This cycle is repeated until all the data packets have been processed and then the server runs a *post-processing* routine.

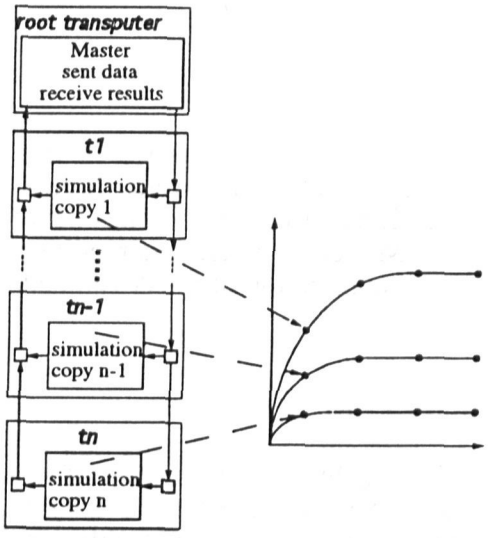

Fig. 1 Pipeline fileserver calculating in parallel an I-V characteristic

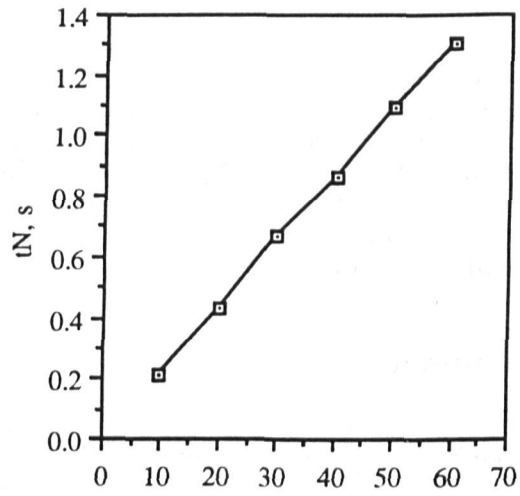

Fig. 2 Time tN required to pass set of input data to the last processor of an N-processors pipeline

The server is implemented using a simple set of programs and can easily be adapted for a number of platforms e.g. a fast serial processor; Inmos transputer boards running 3L languages; a Parsytec Multicluster 32/Supercluster 64 running Helios/Parix or a Meiko computing surface of transputers /C40s running CDL. The server links to a harness of five routines - *pre-process*, *process*, *post-process*, *send-data* and *receive-data* which are independent of the system and completely portable. For long computational jobs the fileserver has proved superior to conventional farm processing techniques supported by many of the existing parallel languages because it can be reliably scaled and is not restricted by the platform dependant variations in buffering systems.

The communication time overhead associated with the fileserver is illustrated in Fig. 2 where the measured access time to the last processor is plotted as a function of the fileserver length. Since a single bias point calculation with H2F takes approximately one hour or more, the communication overhead is negligible. However the average access time can be significantly reduced on many of the configurable transputer systems if a ternary tree is used instead of a pipeline. The effectiveness of the CDS when single bias data are calculated on each processor is also restricted by two other factors. First, because the simulation time for different bias conditions can vary significantly, the total execution time is determined by the processor with the worst combination of bias points. Secondly, the techniques of extrapolating the initial guess from previous solutions is inapplicable. The second disadvantage may be reduced by calculating sets of bias points on a each processor.

Investigations of the influence of the structure parameters on the device's performance has proved to be extremely amenable to CDS. Fig. 3 shows a set of I-V characteristics for a 100 nm gate-length δ-doped pseudomorphic HEMT, calculated in parallel with a CDS on 60

processors, with variable δ-doping and gate length. The estimated speed-up for this particular simulation is around 30 which reflects approximately 50% efficiency.

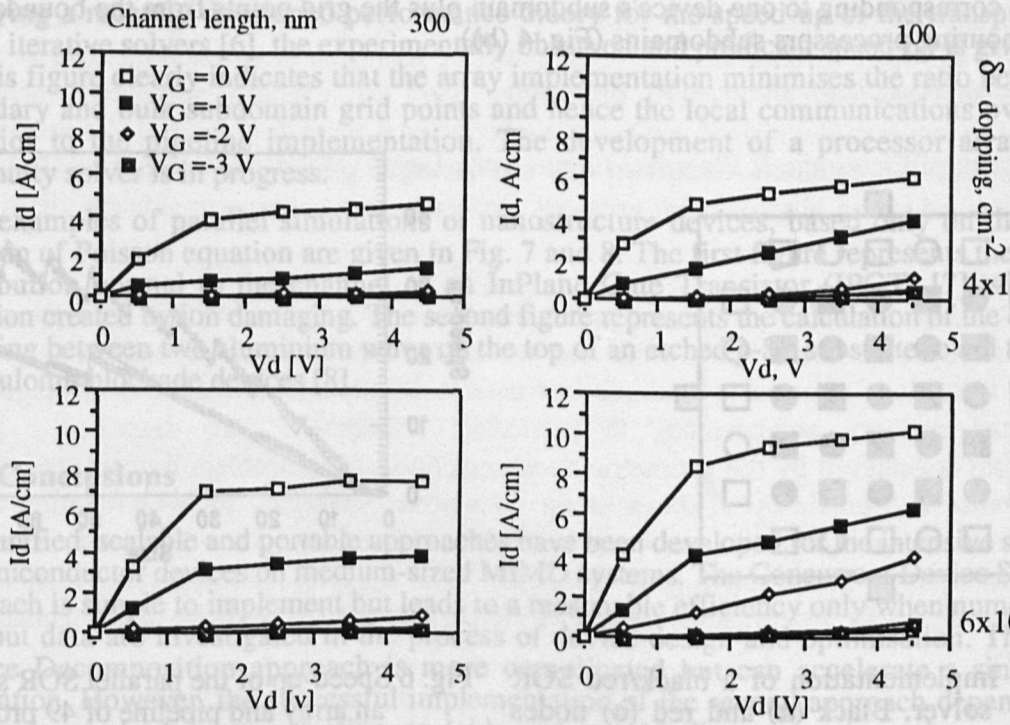

Fig. 3 Set of I-V characteristics for 4 different PHEMTs calculated simultaneously using the CDS approach

III. Spatial Device Decomposition.

The SDD approach is a powerful method of accelerating a single bias point device simulation and of overcoming the inherent memory limitation of 3D simulations. The basic idea of implementing this approach on MIMD is illustrated in Fig. 4 where the spatial device decomposition of a FET and its corresponding grid partition is sketched.

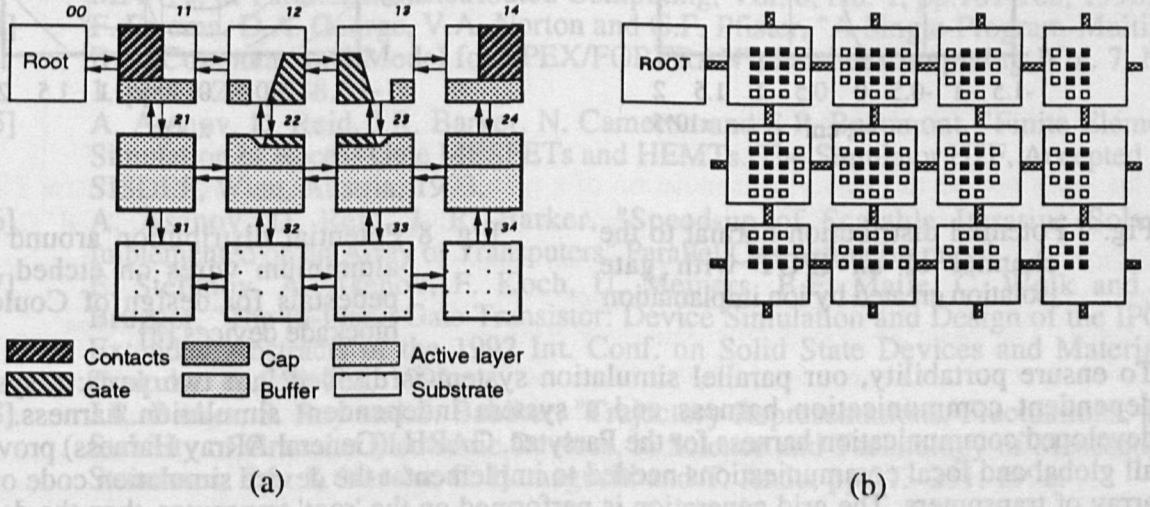

Fig. 4 Spatial device decomposition. (a) physical representation (b) grid partitioning

The solution domain of the device being simulated is spread over a large number of processors (Fig. 4. (a)). If the solution domain is topological rectangular (as in H2F [5]) the best configuration is an array of NxM processors. In this case each processor is allocated to the grid points corresponding to one device's subdomain plus the grid points from the boundary of the neighbouring processors subdomains (Fig. 4 (b)).

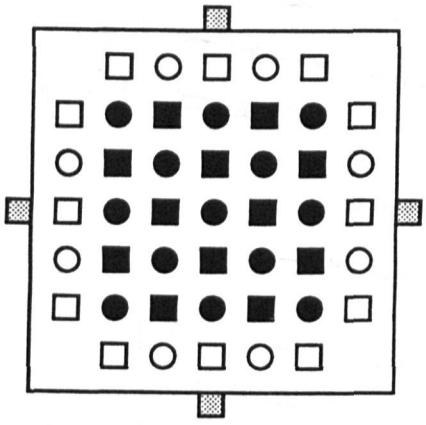

Fig. 5 Implementation of a black/red SOR solver. Black (■) and red (●) nodes are updated simultaneously

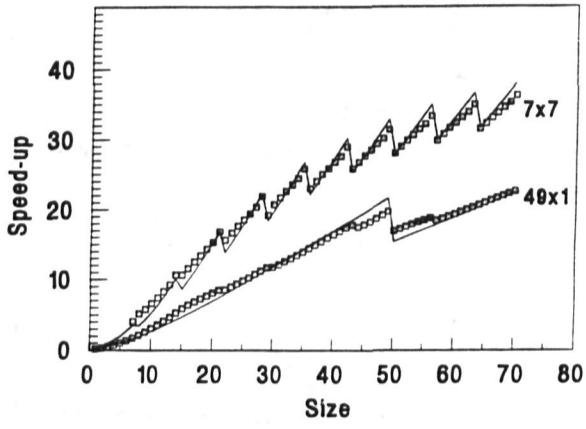

Fig. 6 Speed-up of the parallel SOR solver for an array and pipeline of 49 processors: ___ theory, □ - measurements

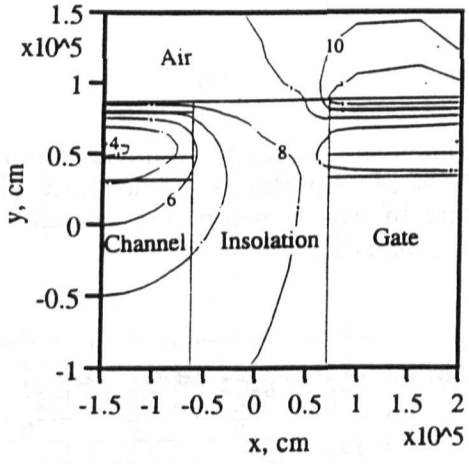

Fig. 7 Potential distribution normal to the channel of an IPGT with gate isolation created by ion implantation

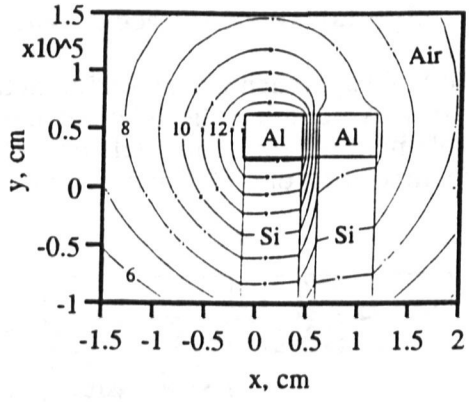

Fig. 8 Potential distribution around two aluminium wires on etched p-Si pedestals for design of Coulomb blockade devices [8]

To ensure portability, our parallel simulation system is divided into two parts: a system dependent communication harness and a system independent simulation harness. The developed communication harness for the Parsytec, GARH (General ARray Harness) provides all global and local communications needed to implementat the device simulation code on an array of transputers. The grid generation is performed on the 'root' transputer, then the device is partitioned and distributed over the complete array. All processors perform the data

initialisation and discretization in parallel. The key point is the parallel solution of the algebraic system, arising from the discretization of Poisson's and the current continuity equations. For the discretized Poisson's equation, a black/red SOR solver was implemented (Fig. 5). Applying a recently developed performance theory for the speed-up of the transputer array based iterative solvers [6], the experimentally observed and predicted speed-up is given in Fig. 6. This figure clearly indicates that the array implementation minimises the ratio between the boundary and bulk subdomain grid points and hence the local communications overhead is superior to the pipeline implementation. The development of a processor array current continuity solver is in progress.

Two examples of parallel simulations of nanostructure devices, based only on the parallel solution of Poisson equation are given in Fig. 7 and 8. The first figure represents the potential distribution normal to the channel of an InPlane Gate Transistor (IPGT) [7] with a gate isolation created by ion damaging. The second figure represents the calculation of the capacitive coupling between two aluminium wires on the top of an etched p-Si substrate to aid the design of Coulomb blockade devices [8].

IV. Conclusions

Two unified, scalable and portable approaches have been developed for the intensive simulation of semiconductor devices on medium-sized MIMD systems. The Concurrent Device Simulation approach is simple to implement but leads to a reasonable efficiency only when numerous sets of input data are investigated in the process of device design and optimisation. The Spatial Device Decomposition approach is more complicated but can accelerate a single point simulation. However, the successful implementation of the second approach depends on the development of adequate, effective, scalable and portable linear equation solvers capable of dealing with the difficulties in solving, for example, the discretized current continuity equation.

References

[1] J. Han, D.K. Ferry and P. Newman, "Ultra-Submicron Gate AlGaAs/GaAs HEMTs", IEEE Electron Dev. Lett., Vol. 11, No. 5, pp.209-211, 1990.

[2] J.R. Barker and D.K. Ferry, "On the Physics and modelling of small semiconductor devices - I", Solid-State Electron., Vol. 23, No. 6, pp.519-530, 1980.

[3] J.P. Darling, and I.D. Mayergoyz, "Parallel Algorithms for the Solution of Nonlinear Poisson Equation of Semiconductor Device Theory and Its Implementation on the MPP", J. of Parallel and Distributed Computing, Vol. 8, No. 1, pp.161-168, 1990.

[4] F. Ferema, D.A. George, V.A. Norton and G.F. Pfister, "A Single-Program-Multiple Data Computational Model for EPEX/FORTRAN", Palallel Computing, Vol. 7, No. 1, pp.11-24, 1988.

[5] A. Asenov, D. Reid, J.R. Barker, N. Cameron and S.P. Beaumont, "Finite Element Simulation of Recess Gate MESFETs and HEMTs. The Simulator H2F, Accepted for SISDEP, Wien, Austria, 1993.

[6] A. Asenov, D. Reid, J. R. Barker, "Speed-up of Scalable Iterative Solvers Implemented on an Array of Transputers" Parallel Computing (in press)

[7] E. Stefanov, A. Asenov, F. Koch, U. Meiners, B.E. Maile, C. Wolk and H. Brugger, "The In-Plane-Gate Transistor: Device Simulation and Design of the IPG", Extended Abstracts of the 1992 Int. Conf. on Solid State Devices and Materials, Tsukuba, pp.759-761, 1992.

[8] J.R. Barker, S. Roy and S. Babiker, "Trajectory Representations, Fluctuations, and Stability of Granular Electronic Devices, in Science and Technology of Mesoscopic Structures, Eds. S. Namba, C. Hamaguchi and T. Ando, pp.213-231, 1992.