Parallel Solver Study for Solving the Boltzmann Transport Equation using Spherical Harmonics Expansions on Supercomputers

Felipe Senra Ribeiro[1], Karl Rupp [1], and Tibor Grasser[1]

[1] *Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, A-1040 Wien, Austria*
{ribeiro|rupp|grasser}@iue.tuwien.com

The deterministic numerical solution of the Boltzmann Transport Equation requires Gigabytes of main memory even for spatially 2D device simulations. While detailed 3D device simulations have already been demonstrated using spherical harmonics expansions [1], it is desirable to run simulations at higher resolution - thus requiring even more memory - and to further reduce simulation time. To tackle both challenges, we have investigated how the free open-source simulator ViennaSHE [1] can be extended to compute the first scalable numerical solutions of the Boltzmann Transport Equation for semiconductors on large-scale clusters and supercomputers. In this work we focus on the linear solver stage, for which we have evaluated an algebraic multigrid solver (schematically shown in Fig. 1) [2][3][4] and a parallel sparse direct solver [5], both available through the PETSc library [6].

We have simulated the carrier distribution for a given electrostatic potential distribution in a 3D FINFET (Fig. 2) and a 2D n-channel MOSFET. Fig. 3 and Fig. 4 show our observed change of total solver time as the workload increases in proportion to the number of processes (‘weak scaling’); this is typical for simulations that are re-run on a finer grid. We observe a near-optimal performance gain proportional to the number of processes used.

When keeping the problem size constant and increasing the number of processors to reduce simulation time (‘strong scaling’), we see that the simulation time can be reduced by a factor of 2.25 when employing up to 16 processes for the MOSFET, and by a factor of 1.8 for the FINFET with 8 processes.

In both cases, algebraic multigrid has been identified as a promising candidate for scaling to hundreds of processes, while a parallel sparse direct solver provides the best performance at moderate problem sizes.

[1] K. Rupp et al., IEDM, 2011

[2] U. Trottenberg et al., Multigrid, Academic Press, 2000

[3] Howard Elman et al., Finite Elements and Fast Iterative Solvers, Oxford Science Publication,2nd Ed

[4] T. Fullenbach et al., Application of an algebraic multigrid solver to process simulation problems, SISPAD, 2000

[5] X.S.Li et al., SuperLU Users' Guide, 1999

[6] S. Balay et al., PETSc Users Manual, ANL-95/11 - Revision 3.14, 2020

Acknowledgments. *The financial support by the Austrian Science Fund (FWF) project FWF-P 29119-N35 is acknowledged. .Some of computational results presented have been achieved using the Vienna Scientific Cluster (VSC).*

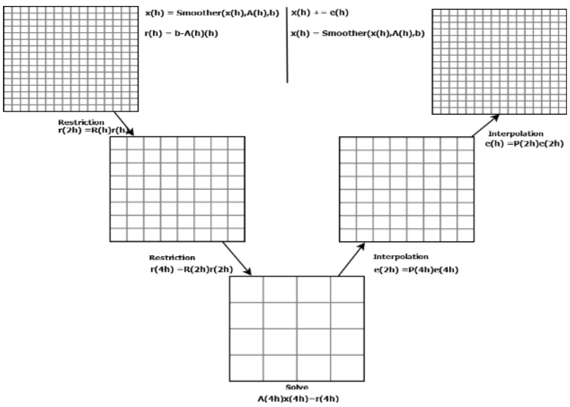

Fig. 1: A 2 level Multigrid hierarchy, where the original linear problem is Restricted two times to a coarsen problem, which is solved, and the solution is Interpolated twice to the original problem, then smoothed to remove some high frequency error.

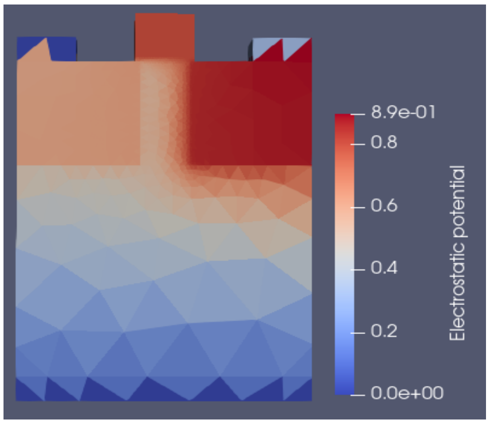

Fig. 2: Electrical Potential for the FINFET when applying 0.3 Volt between the source and the drain with 0.8 Volt at the gate.

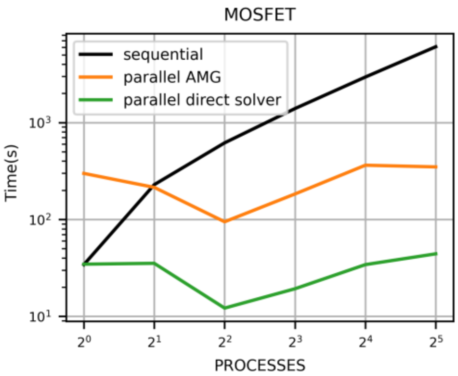

Fig .3: Weak-scaling analysis of the MOSFET simulation. The direct solver is better for small problems, while multigrid achieves slightly better weak-scalability.

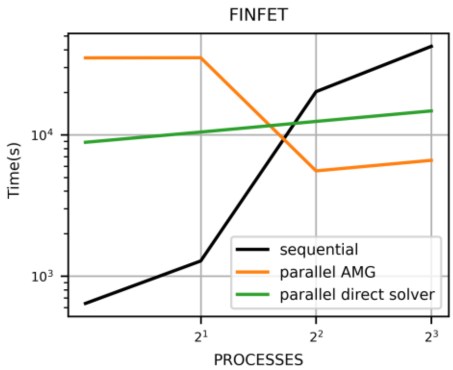

Fig. 4: Weak-scaling analysis of the FINFET simulation. The parallel solvers show large overhead when run with only one process, but are up to a factor of 2.25 faster for the larger problem sizes.

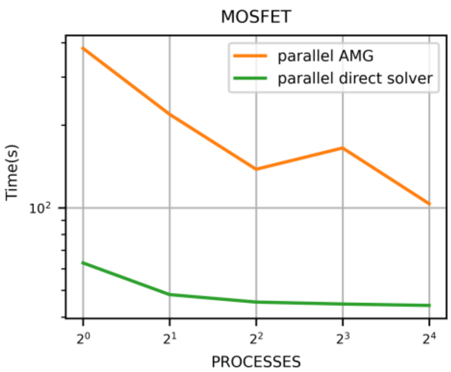

Fig. 5: Strong-scaling analysis of the MOSFET simulation. The AMG solver has higher benefit from additional processes but is still slower than the sparse direct solver overall.

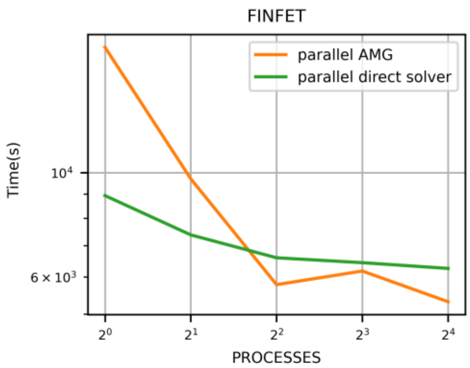

Fig. 6: Strong-scaling analysis of the FINFET simulation. Parallel AMG outperforms the parallel direct solver as soon as at least 4 processes are employed.