Workshop on High Performance TCAD

WHPTCAD 2019

Book of Abstracts

May 24-25, 2019

Hilton Orrington Evanston, Illinois, USA

www.iue.tuwien.ac.at/iwcn2019/workshops/hptcad2019/

co-located with the
International Workshop on Computational Nanotechnology IWCN 2019

Edited by

Josef Weinbub
Paul Manstetten
Stephen Goodnick

Sponsored by

Christian Doppler Laboratory for High Performance TCAD
Institute for Microelectronics, TU Wien
http://www.iue.tuwien.ac.at/hptcad/

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, and the National Foundation for Research, Technology and Development is gratefully acknowledged.

Institute for Microelectronics, TU Wien
https://www.iue.tuwien.ac.at/

Silvaco, Inc
https://www.silvaco.com/

Welcome to the **Workshop on High Performance TCAD (WHPTCAD) 2019**. The workshop is a one day workshop split into two half-days. Technology computer-aided design (TCAD) simulations are used for designing electronic devices and circuits and for predicting their electrical characteristics. Simulations face a tremendous increase in physical and thus computational complexity which is a consequence of novel device designs using more complex device geometries as well as novel materials, both driven by ever-increasing integration densities. The drastic increase in computational complexity in certain areas requires considerable efforts in high performance computing approaches, an underdeveloped focus area within TCAD. At the same time, it is of crucial importance to address practically relevant problems and embrace the latest challenges to keep up a thorough support of simulations alongside rapidly changing technologies. The Workshop on High Performance TCAD will provide a platform to interlink experts from academia and industry with different backgrounds and application fields. Topics of interest are: computational engineering, computer graphics, materials science, semiconductor physics and chemistry, process/device/circuit design/engineering/simulation, and other related areas. The contributed abstracts were reviewed by the workshop chairs. We would like to express our gratitude to the participants of the workshop and thank our sponsors for their support. We hope that you enjoy it, as well as the host conference, and your stay in Evanston.

Josef Weinbub, Paul Manstetten, and Stephen Goodnick
Chairs of WHPTCAD 2019
May, 2019

General Chair

Josef Weinbub
Christian Doppler Laboratory for High Performance TCAD at the Institute for Microelectronics, TU Wien, Austria

Co-Chair

Paul Manstetten
Christian Doppler Laboratory for High Performance TCAD at the Institute for Microelectronics, TU Wien, Austria

Local Arrangements Chair

Stephen Goodnick
Arizona State University, USA

Friday, May 24

08:00 *Registration*

12:50 **Opening Remarks:** Josef Weinbub

Session 1: Technology Computer-Aided Design
Chair: Josef Weinbub

14:50 *Coffee break*

Session 2: Quantum Transport in Electronics
Chair: Paul Manstetten

18:00 *Reception & General Discussion*

Saturday, May 25

08:00 *Registration*

Session 3: Materials Science
Chair: Paul Manstetten

NanoElectronics Simulation Software (NESS) a Flexible Nano-TCAD Concept

Asen Asenov, Vihar Georgiev, Hamiton Carrillo-Nunez, Salim Berrada,
Cristina Medina Bailon, Oves Badami, Jaehyun Lee, Duan Meng, and Fikru Adamu Lema
Device Modeling Group, Glasgow University
asen.asenov@Glasgow.ac.uk

With 5nm CMOS technology generation in prototyping, the semiconductor industry has entered the true nanotechnology era when characteristic transistor dimensions are measured in nanometres instead of tens of nanometres. Quantum confinement and quantum transport phenomena dominate the transistor characteristics and have to be taken into account in simulations together with the charge and matter granularity and the complexity of the interface transitions. Simulation tools with different level of complexity including drift-diffusion (DD) with quantum corrections [1], 3D ensemble Monte Carlo (MC) [2], MultiSubband (MS) 2D [3] and 1D MC [4], direct Boltzmann Equation solvers [5], Non-Equilibrium Green's Function (NEGF) simulators in ballistic regime [6] and with scattering [7] have been developed by different software vendors and research groups. However, these simulators usually work in isolation and it is difficult to carry out consistent simulations for a particular transistor structure highlighting and understanding the areas of applicability and the additionality of simulation techniques with increasing complexity listed above.

In this paper, we present a flexible and open-ended simulation environment developed in Device Modelling Group at The University of Glasgow that enables simulations, with increasing complexity and physical content, to be carried out in a single nano-transistor represented by unified simulation domain.

[1] 3D Monte Carlo: Garand User Guide. https://solvnet.synopsys.com, Synopsys, inc., (2017).
[2] C. Riddet *et al.*, IEEE Trans. Nanotech. **6**, 48 (2007).
[3] C. Medina-Bailon *et al.*, IEEE Trans. Electron Dev. **64**, 3084 (2017).
[4] L. Donetti *et al.*, in Proc. SISPAD (2015).
[5] S. Jin *et al.*, IEEE Trans. Electron Dev. **55**, 727 (2008).
[6] A. Svizhenko *et al.*, J. Appl. Phys. **91**, 2343 (2002).
[7] M. Luisier and G. Klimeck, Phys. Rev. B **80**, 155430 (2009).

Electromigration in Nano-Interconnects

Lado Filipovic and Roberto Lacerda de Orio
Institute for Microelectronics, TU Wien, Austria
filipovic@iue.tuwien.ac.at

Continued device and interconnect scaling has shown to significantly contribute to the reduced lifetime of back-end-of-line copper metallization. The primary reason for this is the growing impact of grain boundaries and surface effects on the movement of electrons and atoms in the thin film. Modeling this phenomenon is very challenging since it requires a very fine definition of all boundaries and interfaces, which behave differently to bulk copper or the copper inside a single grain. To include the effects of granularity in electromigration (EM) simulations most studies apply an effective diffusivity (D_{eff}) parameter, which combines the known bulk diffusivity (D_B) with the grain boundary diffusivity (D_{GB}) using $D_{eff} = D_B + t_{GB}/D^*D_{GB}$, where t_{GB} is the thickness of the grain boundary migration channel and D is the mean grain size [1][2]. An alternative to this approach is defining the grain boundaries and material interfaces as separate materials and meshing the full structure [3][4]. This requires very fine meshes, especially to properly define points where grain boundaries and material interfaces intersect, so-called triple points. Because of the mesh size and complexity, this approach allows for small sections of an interconnect to be modeled, with few grains defined.

We provide an alternative framework which allows for a full interconnect to be modeled while taking the microstructure into consideration. The framework includes three stages: (1) Tessellation, to generate the granular structure; (2) Grid-dependent parameter assignment, which ensures a differentiation between the parameters assigned to the copper grain and those assigned to the grain boundaries and material interfaces using spatial parameter assignment [5][6][7]; and (3) Electromigration simulation in a finite element environment, including vacancy diffusion and induced stress. The EM simulation takes as input the parameter assignments from the previous step in order to calculate the vacancy diffusion and the induced hydrostatic stress. Using this method, the stress accumulation at triple points can be reproduced even with very coarse meshes. Increasing the grid spacing from 0.25nm to 1nm results in an eight-times speedup of the simulation time with only a 7.6% error introduced in the simulation.

[1] V. Sukharev and E. Zschech, J. Appl. Phys. **96,** 6337 (2004).
[2] H. Ceric *et al.*, Facta Universitatis, Series: Electronics and Energetics **27,** 1-11 (2014).
[3] V. Sukharev *et al.,* IEEE Trans. Dev. Mat. Rel. **9,** 87 (2009).
[4] H. Ceric *et al.,* Microelec. Rel. **52,** 1532 (2012).
[5] L. Filipovic *et al.,* in Proc .SISPAD (2017).
[6] L. Filipovic and R. L. de Orio, in Proc. SISPAD (2018).
[7] L. Filipovic, Microelec. Rel. **97**, 38 (2019).

High Performance Computing Aspects in Semiconductor Process Simulation

Andreas Hössinger[1], Paul Manstetten[3], Georgios Diamantopoulos[2],
Michael Quell[2], and Josef Weinbub[2]
[1]*Silvaco Europe Ltd, United Kingdom*
[2]*Christian Doppler Laboratory for High Performance TCAD at the*
[3]*Institute for Microelectronics, TU Wien, Austria*
andreas.hoessinger@silvaco.com

In recent years the gain in computing power was no longer related to a significant increase in processor core performance but was mainly related to an increase in the number of processor cores available within a CPU and to an increase in the number of CPUs available within a workstation. Nowadays the objective of any commercial software development is to make best use of this parallel computing power and obviously this also applies to semiconductor process simulation. In the past, semiconductor process simulation mainly dealt with the simulation of ion implantation and doping diffusion. With deeper penetration of 3D process technology all simulation steps dealing with the creation and modification of the geometry have received a lot more attention. The wide variety of technological applications and analysis requirements has also extended the need for an extensive hierarchy of simulation models or more precisely for simulation algorithms, because in those process simulation steps, different levels of model complexity, means different algorithms or even different data representations. Usually any such algorithms and any sub-algorithms therein requires a different approach towards optimally facilitating high performance simulation hardware.

In this work we would like to illustrate this for the case of etching simulation (embedded into a full process flow) and we would like to demonstrate recent approaches and achievements to better facilitate multi-core computing architecture. In contrast to the simulation of doping diffusion where the performance dominating steps are always the equation assembler and the linear solver, an etching process simulation has to be broken into much smaller pieces to identify the performance critical sub-modules and it is usually not only one of them which dominates. By breaking the simulation flow of an etching simulation into its main sub-modules one can also identify the various possible levels modeling for that process step. Level of modeling means that some of the sub-modules are just approximated and not simulated.

The choice of the modeling level determines the overall performance gain achievable by improving the performance of a single sub-module. Usually the requirement is to optimize the performance of all sub-modules separately, whereby every sub-module uses different and independent algorithms. The basic sub-module list for an etching simulation is:

(a) Transient loop

1. Reactor scale particle transport
2. Extraction of near surface and surface properties
3. Particle transport from the reactor scale to the feature scale and within
4. Interaction of surface with incoming particles
5. Extraction of surface velocity
6. Transient surface motion and interaction with volume properties

(b) Formation of final topology and volume data stage

We present how we have significantly improved the performance of (a-3) ("Particle transport from the reactor scale to the feature scale and within") by replacing an implicit representation of the surface, for modeling the interaction of the particles with the surface. Despite the multi-core scaling of the original implementation was already very good, we managed to further improve the total performance of that module by replacing the implicit surface representation with an explicit surface representation and by using efficiently parallel radiosity algorithms for ray surface interaction within the flux integrator [1][2]. Despite introducing overhead for implicit to explicit surface conversion the performance gain of Victory Process in typical application cases is up to a factor of 7, while maintaining the multi-core scalability.

We also managed to significantly improve performance of (a-6) ("Transient surface motion and interaction with volume properties") by developing a mesh hierarchy based parallelization and by fine tuning the data representation, of level-set re-distancing and level-set velocity extension [3][4]. Despite the inherently serial nature of some of these problems, we obtain a performance gain of Victory Process in typical application cases up to a factor of 3, as well as decent multi-core scalability.

[1] P. Manstetten *et al.*, Procedia Computer Science **108**, 245 (2017).

[2] P. Manstetten *et al.*, in Proc. SISPAD (2017).

[3] G. Diamantopoulos *et al.*, Advances in Computational Mathematics (2019), in press.

[4] G. Diamantopoulos *et al.*, in Proc. ICCSA (2017).

Three-Dimensional TCAD for Atomic Layer Processing

Luiz Felipe Aguinsky[1], Paul Manstetten[2], Andreas Hössinger[3],
Siegfried Selberherr[2], and Josef Weinbub[1]
[1]*Christian Doppler Laboratory for High Performance TCAD at the*
[2]*Institute for Microelectronics, TU Wien, Austria*
[3]*Silvaco Europe Ltd., United Kingdom*
aguinsky@iue.tuwien.ac.at

The continuing progress in device miniaturization requires critical dimension control to the order of nanometers. This has led to the development of advanced processing techniques, notably atomic layer deposition (ALD) and atomic layer etching (ALE) [1]. The idealized concept behind these techniques is to split the process into a series of self-limiting half-steps, enabling monolayer control. Conceptually, this should be possible due to complete surface coverage. However, the intricacies of the real chemical processes have led to the development of several models clarifying the properties of the actual processes. ALD has been studied with respect to its fundamental surface chemistry through density functional theory and lattice kinetic Monte Carlo (KMC) simulations [2]. Although invaluable, these atomistic simulations are too computationally expensive to be directly integrated within a feature scale simulator. Alternatively, feature scale voxel-based approaches have been applied to ALD [3] and ALE [4]. They are similar in concept to the lattice KMC approach, but the atomistic representation is replaced with rectangular cells containing a material mix. This approach requires the assembly of a large list of surface reactions and it cannot be straightforwardly integrated into a typical technology computer-aided design (TCAD) workflow. Analytical models are also an active area of research, having been applied to parameter estimation in certain ALD processes [5]. Conventional TCAD tools [6] often struggle with atomic layer processes. In this presentation, we will show how surfaces and their chemistry are described in process TCAD, including possible extensions. Particular attention is placed on the role of the steady-state approximation of surface coverage and how models are introduced to the transient regime.

Acknowledgments. The financial support by the Austrian Federal Ministry of Science, Research and Economy, and the National Foundation for Research, Technology and Development is gratefully acknowledged.

[1] K. Ishikawa *et al.*, Jpn. J. Appl. Phys. **56**, 06HA02 (2017).
[2] M. Shirazi and S. D. Elliott, J. Comput. Chem. **35**, 244 (2013).
[3] P. Moroz and D. J. Moroz, Jpn. J. Appl. Phys **56**, 06HE07 (2017).
[4] C. M. Huard *et al.*, J. Vac. Sci. Technol. A **35**, 031306 (2017).
[5] M. Ylilammi *et al.*, J. Appl. Phys. **123**, 205301 (2018).
[6] X. Klemenschits *et al.*, Micromachines **9**, 631 (2018).

A Parallel Computational Platform – RAPIT (Rigorous Advanced Plasma Integration Testbed)

Yun-Ming Lee[1], Meng-Hua Hu[1], Kuan-Lin Chen[2], Yi-Ming Li[3], and Jong-Shinn Wu[2]
[1]Plasma Taiwan Innovative Corp., Hsinchu County, Taiwan
[2]Department of Mechanical Engineering,
National Chiao Tung University, Hsinchu City, Taiwan
[2]Department of Electrical Engineering,
National Chiao Tung University, Hsinchu City, Taiwan
chongsin@faculty.nctu.edu.tw

In this paper, the development of a new C++ object-oriented multi-physics parallel simulation platform, named *Rigorous Advanced Plasma Integration Testbed* (RAPIT) [1][2] using unstructured meshes, is reported. Because of the design of the data structure, RAPIT can easily accommodate continuum- and/or particle-based solvers with some proper hybridization algorithms in a self-consistent way. Fig. 1 shows the architecture of the RAPIT, which is designed by the object-oriented C++ programming with embedded MPI for parallel processing. For the PDE solver, the collocated cell-centered finite-volume method is used for numerical discretization. Fig. 2 shows the architecture/procedure of parallel particle data management, in which an accompanying "Pointer Tree" is created as a pointer to the Particle Data Pool (bi-direction data pointer) with four management head pointers (Empty, Buffer, Send, Recv) and a species data pointer of number of all species under simulation. For the former, it may include, but not limited to, the Navier-Stoke (NS) equation solver for general all-spee flow modeling including advanced turbulence k-ω model and two-phase (VOF) model, the plasma fluid modeling code for modeling general low-temperature gas discharges, and the time-dependent Maxwell equation for electromagnetic wave simulation. For the latter, it may include the particle-in-cell Monte Carlo collision (PIC-MCC) and the direct simulation Monte Carlo (DSMC) solvers. Some results of DSMC, PIC-MCC, NS equation and fluid modeling solvers based on RAPIT are presented in this study. Fig. 3 shows the typical results of a capacitively coupled plasma simulation using the fluid model. In addition, one of its byproducts, named ultraMPP [1], which is a general computational platform specifically designed for the partial differential equation solvers will be also presented. More details of the study will be presented in the meeting.

[1] http://www.plasmati.tw/
[2] Y.-M. Lee *et al*., in Proc. *AIAA Paper,* 2018-2944 (2018).

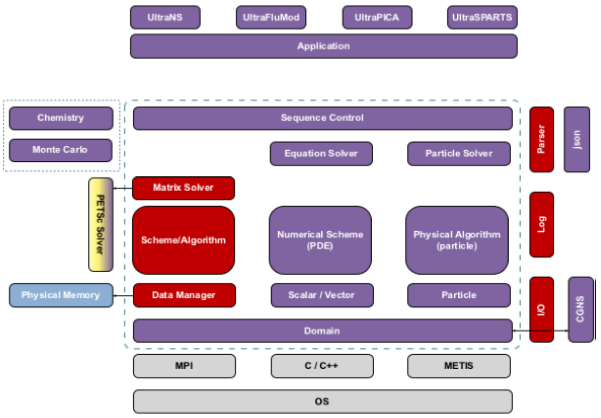

Fig.1: Software architecture of RAPIT.

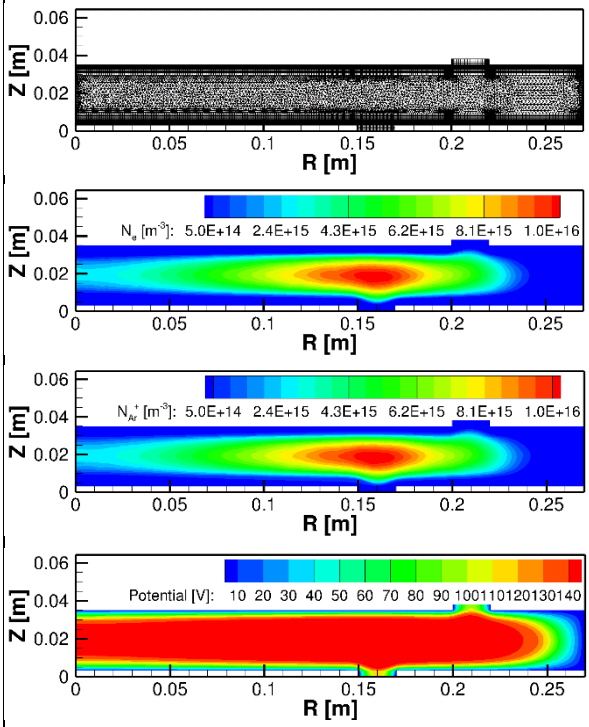

Fig.3: Averaged simulated plasma properties of an argon capacitively coupled plasma with a background pressure of 100 mtorr at 10,000th cycle.

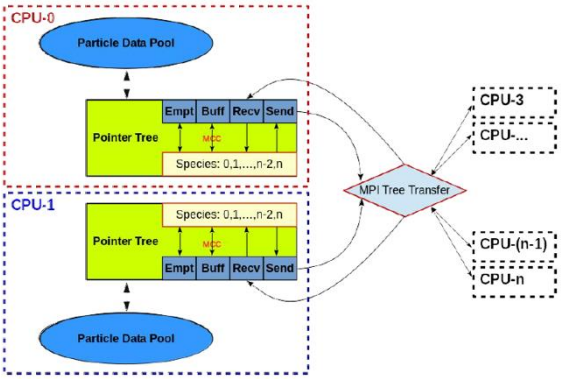

Fig.2: Parallel particle data management.

A Novel Extraction CAD Tool for Industrial Standard BSIM-CMG Models

Min-Hui Chuang and Yiming Li

Parallel and Scientific Computing Laboratory, Institute of Communications Engineering, Department of Electrical and Computer Engineering, National Chiao Tung University, Ta-Hsueh Rd. 1001, Hsinchu 300, Taiwan

ymli@faculty.nctu.edu.tw

There is presently a wide consensus in the compact modeling community that surface-potential-based models are suitable for device modeling of sub-22-nm multi-gate metal-oxide-semiconductor field effect transistors (MOSFETs), such as bulk FinFET and surrounding-gate MOSFETs. Of all the surface-potential-based MOSFETs compact models, the BSIM-CMG model [1][2] is regarded as one of the most popular models, and has been an industrial standard model for MOS transistors [3]. Device model associated with a set of proper parameters intrinsically characterizes the properties of designed and fabricated devices. Various model parameter extraction CAD tools have been reported for the BSIM-CMG model [4][5][6][7], but it has still a room to develop computationally more cost-effective extraction CAD tools for advanced device technologies. In this work, based on our earlier work [8][9][10][11], a robust electronic computer-aided-design (ECAD) tool for BSIM-CMG model parameter extraction is advanced. The configuration of ECAD prototype is with numerical and biological-inspired optimization algorithms, and possesses interface to manual adjustment, automatic execution, and hybrid operation. A unified physical-based BSIM-CMG model parameter extraction procedure is further proposed and implemented in the developed tool. Our preliminary examination shows that the functionality of ECAD is valid for both long and short channels bulk FinFET and surrounding-gate MOSFET devices. Compact modeling of different technology nodes, such as 22, 14, and 10 nm are extracted and verified with the measured data through the prototype. The newly developed ECAD features an open source software and besteads device/circuit engineers' parameter extraction of nano-CMOS devices.

Acknowledgments. This work was supported in part by Ministry of Science and Technology (MOST) of Taiwan under Contracts No. MOST-107-2221-E-009-094 and No. MOST-106-2218-E-009-149.

[1] Y.-S. Chauhan *et al.*, *FinFET Modeling for IC Simulation and Design Using the BSIM-CMG Standard* (Elsevier, 2015).

[2] BSIM Group. http://bsim.berkeley.edu/

[3] Compact Model Coalition. http://www.si2.org/cmc/

[4] BSIMProPlus, Proplus. https://www.proplussolutions.com/bsim-en.php

[5] IC-CAP Device Modeling Software, Keysight Technologies. https://www.keysight.com/en/pc-1297149/ic-cap-device-modeling-software-measurement-control-and-parameter-extraction?cc=US&lc=eng
[6] Model Builder Program, Keysight Technologies. http://www.home.agilent.com/en/pc-2112961/model-builder-program-silicon-focused-device-modeling-software?nid=33185.0.00&cc=US&lc=eng.
[7] UTMOST IV, Silvaco. https://www.silvaco.com/products/analog_mixed_signal/device_characterization_SPICE_modeling/utmost_IV.html
[8] Y. Li *et al.*, Jpn. J. Appl. Phys. **43**, 1717 (2004).
[9] Y. Li, Microelectron. Eng. **84**, 260 (2007).
[10] P.-H. Su and Y. Li, IEEE T. Semiconduct. M. **28**, 193 (2015).
[11] Y. Li and Y.-H. Tseng, Mater. Manuf. Process. **26**, 388 (2011).

Large-Scale Quantum Transport Simulations from First-Principles

Mathieu Luisier, Mohammad Hossein Bani-Hashemian, Mauro Calderara, and Sascha Brück
Integrated Systems Laboratory, ETH Zurich, 8092 Zurich, Switzerland
mluisier@iis.ee.ethz.ch

The modeling of currently fabricated nanoscale devices (transistors, memory cells, sensors, light-emitting diodes…) does not only require to go beyond classical theories such as drift-diffusion equations and to include quantum mechanical effects, but also to accurately describe the properties of each material (combination) constituting their active region. First-principles bandstructure models, for example density functional theory (DFT), lend themselves perfectly to this task. Hence, combining DFT with a quantum transport solver is often seen as the holy grail of device simulation. The basic theory behind such approaches and their first demonstrations have been around for more than 15 years [1], but hardware resources have only recently become powerful enough to allow for the consideration of large-scale structures composed of several thousands of atoms. These improvements alone are still not sufficient: they must be accompanied by the development of novel parallel numerical algorithms that can take advantage of the available computing systems, especially graphics processing units (GPUs).

Here, a suite of dedicated algorithms will be presented to deal with ballistic *ab initio* quantum transport problems on CPUs and GPUs, simultaneously leveraging all these resources. The needed Hamiltonian and overlap matrices will be derived from the CP2K DFT community code that relies on contracted Gaussian-type orbitals (GTO) [2]. Based on these quantities, open boundary conditions (OBCs) will be calculated with contour integral techniques to inject electrons into the simulation domains of interest [3]. Parallel to that, the Schrödinger equation with OBCs will be solved within the so-called Wave Function approach, which produces the same numerical results as the well-known Non-Equilibrium Green's Function (NEGF) formalism, but at a higher speed for 2-D and 3-D structures [4]. Selected realistic examples will be introduced to demonstrate the computational efficiency of the assembled methods that can be extended to include various scattering mechanisms, e.g. electron-phonon interactions.

Acknowledgments. This work was supported by the Platform for Advanced Scientific Computing in Switzerland and by two grants from the Swiss National Supercomputing Centre (CSCS) under project No. s876 and pr28.

[1] M. Brandbyge *et al.*, Phys. Rev. B **65**, 165401 (2002).
[2] J. VandeVondele *et al.*, Comput. Phys. Commun. **167**, 103 (2005) .
[3] S. Brück *et al.*, J. Chem. Phys. **147**, 074116 (2017).
[4] M. Luisier *et al.*, Phys. Rev. B **74**, 205323 (2006).

Resource Control in NEMO5's Quantum Transport Calculations

Xinchen Guo, Daniel Lemus, James Charles, and Tillmann Kubis
School for Electric and Computer Engineering,
Purdue University, West Lafayette, Indiana, USA
tkubis@purdue.edu

Alongside with the continuous shrinking of semiconductor nanodevice dimensions, the computational demands in predicting their performance keeps increasing: 1) Nanodevices face pronounced quantum effects but 2) are often run in real worlds conditions such as finite temperatures and fabrication imperfections that spoil pristine quantum characteristics. Predicting both competing effects simultaneously is typically done in the framework of nonequilibrium Green's functions (NEGF) – in spite of the tremendous numerical load NEGF comes with [1].

In this work, latest approaches to limit the numerical load of NEGF in the multipurpose nanodevice simulation tool NEMO5 are presented [2]. This includes a new recursive Green's function method that allows to accurately cover long-ranged electron-phonon scattering dominant in polar materials (such as III-V and TMD semiconductors). This work enabled the introduction of various approximate treatments of nonlocal scattering, such as a predictive nonlocal compensation factors and a low-rank approximation approach [3] compatible with nonlocal recursive NEGF calculations. The semi-empirical Büttiker probe method of NEMO5 has been demonstrated on electron and phonon transport [4] as well as combined systems with mutual energy exchange between them (as needed for self-heating predictions).

On the pure numerical implementation aspect of NEMO5, great progress was achieved to scale realistic quantum transport calculations on large HPC systems, including heterogeneous computational platforms such as e.g. Xeon Phi (co-)processor architectures. The varying resources available together with a sometimes hard to predict numerical performance of quantum transport methods on heterogeneous platforms (e.g. compute clusters involving GPUs, coprocessors or even just different communication speeds due to different locations within a cluster rack) motivated a recent project of a self-adaptive resource management approach. This algorithm maintains a maximal usage of memory and CPU load over a as complete as possible runtime of NEMO5 and similar tools.

[1] J. Charles *et al.*, J. Comput. Electron. **15**, 1123 (2016).
[2] J. E. Fonseca *et al.*, J. Comput. Electron. **12**, 592 (2013).
[3] J. Z. Huang *et al.*, J. Appl. Phys. **123**, 044303 (2018).
[4] S. Sadasivam *et al.*, Phys. Rev. B **95**, 085310 (2017).

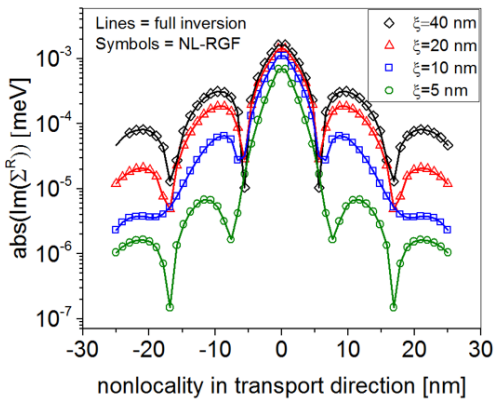

Fig.1: Comparison of the new nonlocal RGF method against a full inversion for polar optical phonon scattering self-energies in a bulk GaAs system. Different screening lengths ξ impact the nonlocality.

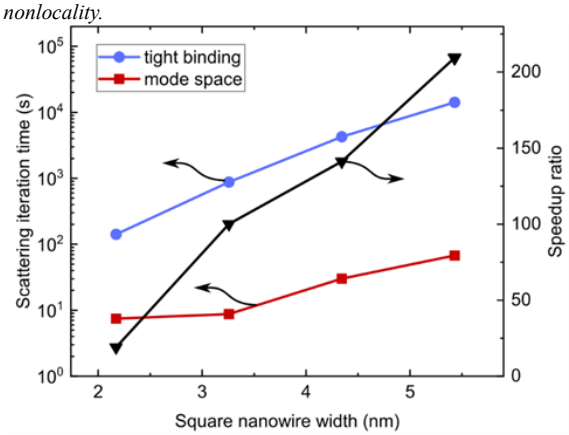

Fig.2: Time-to-solution comparison of self-consistent Born approximation scattering iterations in tight binding (original basis) and mode space (reduced basis) for various atomically resolved nanowire cross-sections.

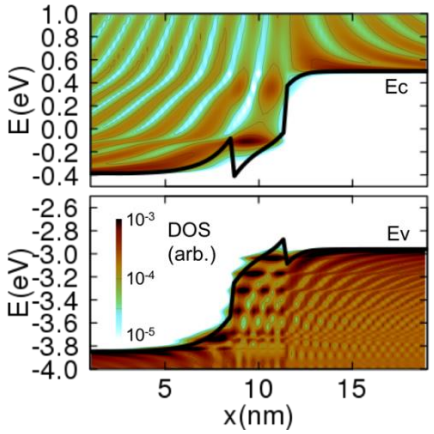

Fig.3: Energy resolved density of states (contour) and band edges (lines) of pn junction with a quantum well at Vd=3V at the Gamma point solved with NEGF and Büttiker probes. Confined states are accurately filled with the Büttiker probe method, while significantly reducing the computational load compared to self-consistent Born approximation methods.

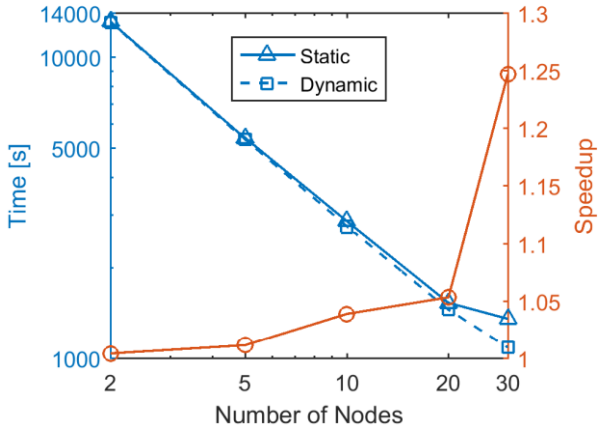

Fig.4: Speedup yielded with dynamic load distribution in quantum transmitting boundary method calculations. The speedup grows significantly with more threads per node.

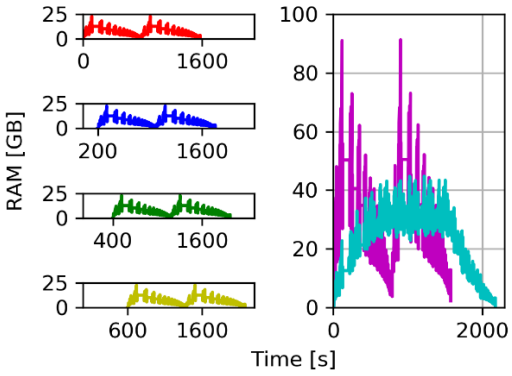

Fig.5 Significant reduction of peak memory with self-adaptive resource management for the example of 4 NEMO5 quantum transport MPI processes per node (left). The total peak memory (right) reduces when the processes are scheduled appropriately (green) instead of running simultaneously (red).

High Performance TCAD: From Simulating Fabrication Processes to Wigner Quantum Transport

Paul Manstetten[1], Georgios Diamantopoulos[2], Lukas Gnam[2], Luiz F. Aguinsky[2], Michael Quell[2], Alexander Toifl[2], Alexander Scharinger[2], Andreas Hössinger[3], Mauro Ballicchia[1], Mihail Nedjalkov[1], and Josef Weinbub[2]

[1]*Institute for Microelectronics, TU Wien, Austria*

[2]*Christian Doppler Laboratory for High Performance TCAD at the Institute for Microelectronics, TU Wien, Austria*

[3]*Silvaco Europe Ltd., United Kingdom*

weinbub@iue.tuwien.ac.at

Manufacturing integrated circuits was never a simple process. However, with scaling slowing down more significantly in the single digit nanometer regime due to skyrocketing fabrication costs the need for comparatively cheap simulation-based predictions has further increased. However, the necessary simulation tools in electronics and specifically in semiconductor product development – collectively summarized with the term technology computer-aided design (TCAD) – increasingly struggle with the added complexities. More complicated device geometries (requiring inherently challenging three-dimensional models and simulations), new materials, and arising quantum effects translate to drastically increased computational efforts, in turn resulting in progress-impeding long simulation runtimes. Consequently, reducing simulation runtimes by applying high performance computational methods is vital to prevent impeding the pace of research. Novel modeling and simulation approaches, using parallelization wherever possible, are needed, in turn requiring inter-disciplinary avenues to be explored. In this talk, an overview of the key research areas of the Christian Doppler Laboratory for High Performance TCAD will be given. In particular, key computational challenges and advancements in simulating (1) fabrication processes (e.g. etching [1][2] and oxidation in silicon carbide [3]) and (2) quantum transport in the Wigner picture [4][5][6][7].

Acknowledgments. The financial support by the Austrian Federal Ministry of Science, Research and Economy, and the National Foundation for Research, Technology and Development is gratefully acknowledged.

[1] L. Gnam *et al.*, Micromachines **9**, 550 (2018).

[2] P. Manstetten *et al.*, Proc. Comp. Sci. **108**, 245 (2017).

[3] V. Simonka *et al.*, Solid-State Electron. **128**, 135 (2017).

[4] J. Weinbub and D.K. Ferry, Appl. Phys. Rev. **5**, 041104 (2018).

[5] J. Weinbub *et al.*, Physica Stat. Sol. RRL **12**, 1800111 (2018).

[6] M. Ballicchia e*t al.*, Appl. Sci. **9**, 1344 (2019).

[7] M. Ballicchia e*t al.*, Nanoscale **10**, 23037 (2018).

Mechanisms of Degradation of Gate Dielectrics Under Bias Stress

Alexander Shluger
Department of Physics and Astronomy, University College London, UK
a.shluger@ucl.ac.uk

Multiscale modelling of structural and electrical degradation of amorphous gate oxide films under bias stress requires modelling the interplay of electronic and ionic processes including the effects of electron injection and hydrogen incorporation. Injection of excess electrons and holes into oxide films can be responsible for the mechanisms that govern the formation of conductive filaments in resistance switching memory devices, the dielectric breakdown in microelectronic devices, and the performance of photo-electrochemical and oxide fuel cells. Theoretical modelling combined with experimental observations demonstrates that structural disorder in amorphous SiO_2, Al_2O_3, and HfO_2 films creates precursor sites which can spontaneously trap up to two electrons or holes in deep states in the band gap [1][2][3]. The electron localization weakens Me–O bonds, which can be broken upon thermal activation, creating an O^{2-} interstitial ion and a neutral O vacancy [4]. O^{2-} interstitial ions can easily diffuse through the oxide and in devices are guided to the positive electrode by the electric field [5]. Inter-diffusion of hydrogen into oxide films creates new defects [6][7] which can contribute to leakage current and dielectric breakdown. Multi-scale modelling including electron injection rates, defect creation and electron hopping through created defects is used to describe oxide degradatation and dielectric breakdown in oxide films. The results demonstrate that the creation and field-driven movement of oxygen ions causes changes in oxide structure on a much larger scale than previously thought.

[1] A.-M. El-Sayed *et al.*, Phys. Rev. B, **89**, 125201(2014).
[2] M. Kaviani *et al.*, Phys. Rev. B, **94**, 020103 (2016).
[3] O. Dicks and A. L. Shluger, J. Phys.: Condens. Matter, **29**, 314005 (2017).
[4] D. Z. Gao *et al.*, Nanotechnology, **27**, 505207 (2016).
[5] A. Mehonic *et al.* Adv. Mater. **28**, 7486 (2016).
[6] A. –M. El-Sayed *et al.* Phys. Rev. B **92**, 014107 (2015).
[7] M. Kaviani *et al.*, Phys. Rev. B **95**, 075117 (2017).

Approaches to Atomic Scale Materials Modelling and Challenges in Using It to Understand Electronic Device Operations

Al-Moatasem El-Sayed, Filippo F. Canova, and David Z. Gao
Nanolayers Research Computing Ltd.
aelsayed@nanolayers.com

As the physical limits to scaling are approached, new strategies are constantly developed by materials scientists and engineers, such as the use of different materials and more exotic geometries, to continue improving the performance and reliability of electronic devices. Computer simulations of these materials and devices at different length scales is proving itself to be an invaluable tool to further developing these novel devices.

In this talk, I will focus on the use of atomistic scale modeling in understanding the operation of electronic devices. Strategies for generating amorphous materials and interfaces will be discussed with a focus on how to justify the use of these systems as models of realistic electronic devices. The parameters that can be extracted from these simulations will then be discussed and linked to understanding issues in electronic devices. As case studies, results of atomistic simulations will be used to provide insight into bias temperature instabilities and hot carrier degradation. Finally, these simulations can be incorporated into an automated materials simulation workflow and used to generate a large database of structure-property relationships which can then be used to train machine learning models.

Ab Initio to TCAD Workflow for Accurate and Efficient Modelling of Defects

Ignacio Martin-Bragado[1], Shela Aboud[1], Luis Agapito[1], Yong-Seog Oh[1], Yumi Park[1], Anders Blom[1], Christoph Zechner[2], Fabiano Corsetti[3], Pieter Vancraeyveld[3], Gabriele Penazzi[3], Martin Andersen[3], and Kurt Stokbro[3]

[1]*Synopsys, Inc., 690 E. Middlefield Road, Mountain View, CA 94043, USA*

[2]*Synopsys GmbH, Karl-Hammerschmidt-Str. 34-40 85609 Aschheim, Germany*

[3]*Synopsys Denmark ApS, Fruebjergvej 3, 2100 Copenhagen, Denmark*

shela.aboud @synopsys.com

Defects play an important role in all kinds of semiconductor devices, but to accurately describe them in TCAD simulations requires input parameters that traditionally are determined experimentally. As the complexity and variety of new materials introduced in novel device architectures grows, it would be desirable to instead obtain parameters such as diffusion constants and defect levels from first-principles calculations. In practice it can however be very difficult to achieve sufficient accuracy due to the complexity of the geometries and energetics of defects.

We have developed an automated approach for systematically calculating defect properties, and for transferring the results into a TCAD simulation tool. The software allows the user to automatically set up a wide range of different point defects and compute the migration paths between them. The different reaction pathways can then be transferred to a kinetic Monte Carlo program, from which the temperature-dependent diffusion constant is determined. To reduce errors in the calculated energetics due to the finite size of the computational cell, we perform calculations with different cell sizes and apply charge [1] and stress corrections, and the results can be extrapolated to the infinite cell limit. To efficiently include the entropic contribution from phonons we have developed an approximate scheme to calculate only the relevant parts of the dynamical matrix. To accurately model the defect energetics and band gaps, without incurring the very large computational cost of hybrid functionals, we use the novel SCAN [2] functional and the DFT+1/2 correction [3]. In this presentation, the methodology used and results for defect levels in Si and GaAs will be presented.

[1] C. Freysoldt *et al.*, Phys. Rev. Lett. **102,** 016402 (2009).

[2] J. Sun *et al.,* Phys. Rev. Lett. **115**, 036402 (2015).

[3] L.G. Ferreira *et al.*, Phys. Rev. B **78**, 125116 (2008).

Superconducting TCAD – Numerical Challenges

Mark E. Law, Erin Patrick, Nimesh Pokhrel, and Thomas Weingartner
Department of Electrical and Computer Engineering, University of Florida
mlaw@eng.ufl.edu

Superconducting technologies for information technology have demonstrated 370GHz performance and AC-biased AQFP (adiabatic quantum-flux-parametron) that achieves an energy dissipation near quantum limits[1]. A key enabler for wider commercial use of these technologies is the presence of design tools, including TCAD tools[2]. Like in semiconductors, there are two major components: first, simulating the manufacturing process to enable extraction of relevant material properties and shapes, and second, using that information to predict device performance.

Superconducting technology is centered on Josephson Junctions – a tiny insulator sandwiched between two superconducting metal layers. From a semiconducting processing point of view, all of the fabrication takes place in the back-end. Unlike in semiconductor back-end processing topology is frequently non-planar. Niobium cannot be polished, and there is no technology for planar via plugs. Oxide insulating layers are polished, but the metal layers have significant topology. The barrier is formed with a self-limiting oxidation and devices are passivated using anodic oxidation. Predicting this structure a key challenge and we will describe our approach to these calculations using level-set techniques[3].

A key performance metric is the Josephson critical current. Below this current, the device is fully superconducting. Above this current, the device becomes resistive. Variability of this parameter is a challenge in fabricating VLSI equivalent circuits. There are several possible origins of critical current variability and we are taking a two prong approach. We are using ab-initio techniques to compute the change of critical current with structural changes. This is limited to very small simulation areas. We are also using the sine-Gordon approximation for simulating realistic sized devices. Each poses different numerical challenges for solution on realistic structures.

[1] M. Pedram and Y. Wang, in Proc. IEEE/ACM ICCAD (2018).
[2] Fourie *et al.*, IEEE T. Appl. Supercon. **29**, 1300407 (2019).
[3] J. A. Sethian, *Level Set Methods and Fast Marching Methods* (Cambridge Univ. Press, 1999).

Parallel Implementation of the Split-Step Fourier Method for Exact Two-Electron Dynamics in Hall Interferometers

Laura Bellentani[1], Paolo Bordone[1,3], Xavier Oriols[2], and Andrea Bertoni[3]
[1]*Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università degli Studi di Modena e Reggio Emilia, Via Campi 213/A, I-41125 Modena, Italy*
[2]*Departament d'Enginyeria Electrònica, ETSE, Universitat Autònoma de Barcelona 08193, Bellaterra, Barcelona, Spain*
[3]*S3, Istituto Nanoscienze-CNR, Via Campi 213/A, 41125 Modena, Italy*
laura.bellentani@unimore.it

Edge states in the Integer Quantum Hall regime are promising candidates to implement quantum computing gates in semiconductor devices, due to their essential immunity to backscattering [1]. Their long-range coherence can be exploited in semiconductor interferometers to achieve two-qubit operations, where the necessary entanglement is generated by Coulomb interaction and/or exchange symmetry. Hence, their efficient numerical modeling has to describe the non-separable dynamics of two electrons in micrometer-size 2D spatial domains, giving a 4D simulation space. The numerical load for the allocation of such 4D real-space wave function is very challenging also for simple devices. In the pertinent literature, traditional approaches bypass the memory cost and computational burden by using time-independent scattering matrices in effective 1D schemes, which proved not to be able to fully capture the interplay between two-electron correlations and the geometry of the device, as for electron bunching [2]. On the contrary, we privileged the exact solution by developing a scalable parallel numerical solver of the time-dependent 4D Schroedinger equation for two-particle systems.

To compute the exact states of edge states, we first devise a full-scale potential that mimics the effect of top gates in a realistic sample. Then, coherently with recent protocols for single-electron injection in Hall interferometers, our qubits are initialized in Gaussian wave packets of edge states. To afford the memory load, the initial wavefunction is distributed on several computing nodes by adopting a Cartesian topology within the MPI paradigm. A parallel version of the Split-Step Fourier method [3], based on an iterative applications of the two-particle evolution operator and Fourier transform/antitransform, performs the evolution of the wavefunction. The Cartesian data distribution entail parallel 2D Fourier transforms with adequate per-row and per-column MPI communications to improve their performances.

Our numerical solver is applied to simulate two-electron bunching on a full-scale quantum point contact in the Integer Quantum Hall regime, by including exactly the interplay between exchange correlation and Coulomb interaction in the 2D geometry.

Our results shed light on the orthogonality of electron states scattered by a quantum-point contact, which can be fully captured only by wave packet approaches. Furthermore, it represents the starting point for the simulation of more sophisticated systems useful for quantum computing, as the conditional phase shifter, where the non separability of the states originates from the Coulomb interaction between electrons in two different edge channels.

[1] Y. Ji *et al.*, Nature, **422**, 415 (2003); I. Neder *et al.*, Nature, **448**, 333 (2007).

[2] D. Marian *et al.*, J Phys.: Condens. Matter, **27**, 245302 (2015).

[3] L. Bellentani *et al.*, Phys. Rev. B, **97**, 205419 (2018).

High-Performance Ray Tracing for Nonimaging Applications

Paul Manstetten[1], Luiz Felipe Aguinsky[1], Siegfried Selberherr[2], and Josef Weinbub[1]
[1]*Christian Doppler Laboratory for High Performance TCAD at the*
[2]*Institute for Microelectronics, TU Wien, Austria*
manstetten@iue.tuwien.ac.at

Today, ray tracing is most commonly associated with computer graphics applications where the objective is to render an image from a virtual scene. Such a scene is typically a collection of objects represented mathematically. The image is generated by tracing rays originating from a viewport into the scene. The intersections of these original rays with the objects in the scene commonly generate subsequent rays to approximate the rendering equation, generating an equilibrium of the radiance in the scene. The radiance towards the viewport finally defines the values of the pixels in the rendered image [1].

Nonimaging applications which use ray tracing as a computational method can potentially profit from the algorithmic advances and optimized computational performance of ray tracing engines developed for rendering, like Embree [2], Optix Prime [3], and OpenVDB [4]. Ray tracing can, for example, be applied in the simulation of radiative heat transport [5], nonimaging optics [6], [7], and particle transport [8]. In fact, for advanced semiconductor process simulations, the computational performance of a feature-scale etching or deposition process is largely determined by the ray tracing performance [9].

We assess the performance of the above-mentioned ray tracing engines using a generic test case with a spatially distributed incoherent ray workload, reproducing a common situation for nonimaging applications. Furthermore, we discuss the implications arising when a ray tracing engine is integrated into an existing simulation framework by the example of a topography simulator.

Acknowledgments. The financial support by the Austrian Federal Ministry of Science, Research and Economy, and the National Foundation for Research, Technology and Development is gratefully acknowledged.

[1] M. Pharr *et al.*, *Physically Based Rendering* (Morgan Kaufmann, 2016).
[2] Embree. https://embree.github.io
[3] Optix Prime. https://developer.nvidia.com/optix
[4] OpenVDB. https://github.com/AcademySoftwareFoundation/openvdb
[5] J. R. Mahan., *Radiation Heat Transfer: A Statistical Approach* (Wiley, 2002).
[6] OpticsStudio. https://www.zemax.com/products/opticstudio
[7] S. Scharring *et al.*, Opt. Eng. **56**, 011007 (2017).
[8] ViennaTS. https://viennats.github.io
[9] P. Manstetten *et al.*, Procedia Comp. Sci. **108**, 245 (2017).