View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Purdue E-Pubs

Purdue University
Purdue e-Pubs

Other Nanotechnology Publications | Birck Nanotechnology Center

12-15-2008

From NEMO1D and NEMO3D to OMEN: Moving Towards Atomistic 3-D Quantum Transport in Nano-scale Semiconductors

Gerhard Klimeck
Network for Computational Nanotechnology, Birck Nanotechnology Center, Purdue University, gekco@purdue.edu

Mathieu Luisier
Network for Computational Nanotechnology, Birck Nanotechnology Center, Purdue University

Follow this and additional works at: http://docs.lib.purdue.edu/nanodocs

Part of the Nanoscience and Nanotechnology Commons

Klimeck, Gerhard and Luisier, Mathieu, "From NEMO1D and NEMO3D to OMEN: Moving Towards Atomistic 3-D Quantum Transport in Nano-scale Semiconductors" (2008). *Other Nanotechnology Publications.* Paper 154.
http://docs.lib.purdue.edu/nanodocs/154

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

From NEMO1D and NEMO3D to OMEN: moving towards atomistic 3-D quantum transport in nano-scale semiconductors

Gerhard Klimeck and Mathieu Luisier

Network for Computational Nanotechnology, Birck Nanotechnology Center, Purdue University, gekco@purdue.edu

Abstract

Lessons learned in 15 years of NEMO development starting from quantitative and predictive resonant tunneling diode (RTD) to multi-million atom electronic structure modeling and the path for OMEN are laid out. The recent OMEN capabilities enable realistically large 3D atomistic nano-scale device simulation.

Introduction

The end of Moore's Law has been falsely predicted many times. Irrespective of "details" of continued scaling in heat dissipation, lithography, materials, or financial viability, the fundamental limit of transistors containing an indivisible, countable number of atoms is coming closer. In this regime the detailed atom arrangements and quantum mechanical behavior of the carriers are critical in the understanding of the device. Not only are the structures getting smaller, but they are morphing from planar to 3-D topologies and new atom species are introduced. From now on the distinction between new device and new material is blurred and transport modeling must embrace atomistic and quantum concepts.

NanoElectronic MOdeling (NEMO1D)

In the 1990's the RTD was touted as one of the candidates for the replacement of the coming end of the silicon transistor. In about 1997 Texas Instruments had assembled a technology in the first Industrial Nanoelectronic Research group, that might have been a viable replacement, if Silicon ever had run out of steam. Physics-based modeling was part of that effort and it resulted in the creation of NEMO1D. The RTD was the only 300K transport device that demanded a quantum mechanical analysis. NEMO1D was the first demonstration of an industrial strength simulator that is based on the Non-Equilibrium Green Function Formalism (NEGF) [1]. NEMO1D enabled for the first time the quantitative and predictive study of phonons, interface roughness, and alloy disorder scattering and the importance of atomistic layer representation through experimental benchmarks [2] (Fig. 1). While RTDs have no commercial applications today the technical understanding we gained serves in hindsight as a guide to what is to be expected in 3D nano devices.

The intellectual effort was focused on the understanding of the origin and reduction of the RTD valley current. Although it was widely accepted that phonon scattering is creating the dominating valley current, this turned out to not be true for high performance, high current InP-based InGaAs/InAlAs 300K RTDs. Instead the critical processes were (a) thermionic transport through excited states, which are much lower in energy than typically predicted with effective mass models and (b) scattering in the extended contact regions that thermalize the emitter. These processes are associated with two critical insights: 1) an atomistic basis representation that resolves the material variations on the 2-5nm scale includes effects such as band non-parabolicity, band-warping, band-to-band coupling, and full Brillouin zone transport; and 2) the extended quantum states reaching over 100nm are critical to the proper carrier injection into the central device.

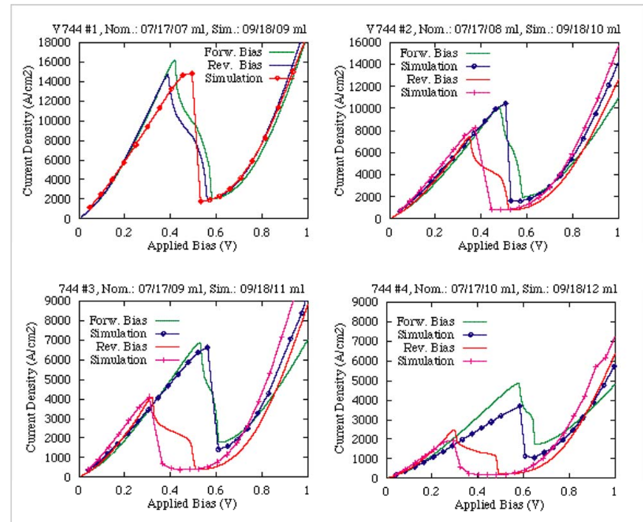

Fig. 1 Test matrix of a strained InGaAs/AlAs RTD system as described in detail in [2c]. One of the AlAs barriers is increased in thickness by one monolayer at a time introducing an asymmetric forward and reverse bias behavior. Full band sp3s* with transverse momentum integration is required to model these devices quantitatively. Band non-parabolicities lower excited states to allow valley current, complex band wrapping to the valence band in the AlAs increases the barrier transparency.

The need for atomistic modeling implies that effective mass approaches are doomed to start with and would continue to guide the community in the wrong direction. Empirical tight binding captures the critical physical effects such as band-coupling, band warping, spin, and strain for electrons [2] and holes [3].

Figure 1 depicts a comparison between an experimental testmatrix of InGaAs/AlAs RTDs as described in detail in [2c]. Forward and reverse bias performances are plotted on the same positive voltage axis. The testmatrix intentionally increased one of the barriers from device to device, causing charge accumulation in the central device in one bias direction (shifting the current peak to higher voltages, since the charge in the device pushes against the voltage drop in the collector), and charge depletion in the central device, since charges escape faster to the collector, than they are injected from the emitter side. Current peak voltage and amperage, as well as valley current and turn on-of the second resonance are faithfully reproduced.

Authorized licensed use limited to: Purdue University. Downloaded on March 11, 2009 at 23:23 from IEEE Xplore. Restrictions apply.

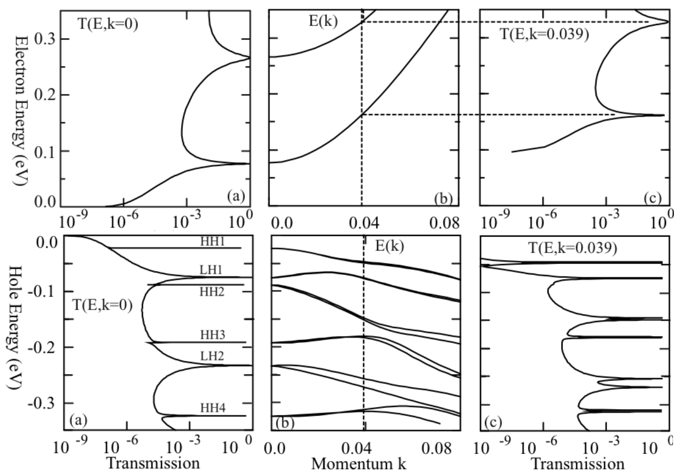

Fig. 2: Transmission and transverse dispersion curves from NEMO1D for electrons (top) and holes (bottom) in a simple GaAs/AlAs RTD from [3]. The electron dispersion is roughly parabolic for this device and translates the higher transverse momentum transmission vertically shifted in energy (not the case for high performance InGaAs devices, which show a high degree of non-parabolicity). The hole transmission shows strong features of HH and LH resonances. The dispersion shows strong HH and LH mixing, and the higher momentum transmission curve has no resemblance with the shape of the zero momentum curve and shows symmetry-breaking induced spin-splitting.

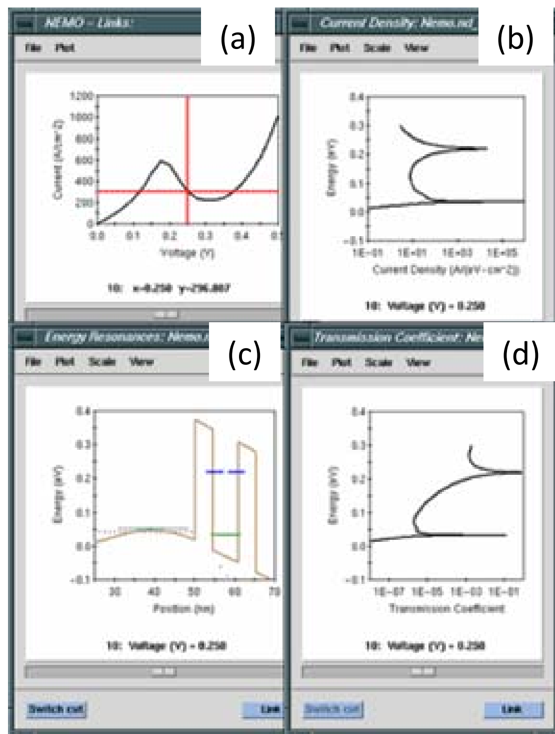

Fig. 3 NEMO1D is a full device design environment that enables interactive simulation tracking and data exploration. Data slide views enable the exploration of the data space (a) I-V, (b) J(E), (c) Ec & resonances, (d) T(E).

The behavior of holes in nanoelectronic devices under strong confinement is currently attracting a lot of interest. Figure 2 shows simulations for electrons and holes in an RTD [3]. Particularly striking are the strong hole-hole interactions which result in dramatically different transmission coefficients as a function of in-plane momentum. sp3s* and sp3d5s* tight-binding models fully capture the needed bandstructure effects at the nanometer scale under strain and disorder.

NEMO1D was built as a complete 1-D heterostructure design tool which can be used by experimental device engineers. Fig. 3 shows several screen-shots of the graphical user interface (GUI). Data selector sliders allow users to easily scan through the large output data sets. This design ultimately strongly influenced the current nanoHUB.org user interfaces based on its Rappture tool-kit (see www.rappture.org).

NEMO3D

Armed with NEMO1D insight, NEMO3D development began at NASA/JPL in 1998 [3]. Due to computational intensity atomistic NEGF transport simulations were deemed unfeasible. However, electronic structure simulations for devices containing a million atoms for quantum dots, quantum wells, and nanowires were achievable on then new Beowulf clusters in the JPL HPC group [4]. Continued work at Purdue beginning 2003 demonstrated the simulation of systems of up to 52 million atoms corresponding to $(101\text{nm})^3$ volumes [5] and resulted in quantitative modeling capabilities to explain core-shell nanowires [6], valley splitting in silicon on SiGe ultra-thin bodies [7] (Fig 3), and impurities in silicon [8]. Single impurity metrology is demonstrated in this year's IEDM paper [8].

A key NEMO1D insight had been the need for careful calibration of the tight-binding material parameters that describe the electronic structure of bulk materials including experimentally verified bandgaps, effective masses, and strain behaviors. Such parameterizations had been traditionally very tedious. The development of a genetic algorithm-based approach [9a] has helped to alleviate, but not eliminate the fitting tedium, resulting in room temperature parameterizations for common III-V and Si/Ge material systems [9].

Nano-scaled Si structures have found strong interest in the realm of quantum computing with the hope that the ability to mass-produce nano-scaled Si devices will eventually help to create a quantum computer. A critical ingredient in such a computer will be the establishment of well-defined ground states. Valley-splitting (VS) which had been studied decades earlier has therefore become of high interest again since it is responsible for the separation of the 6 different Si valleys due to quantum mechanical interactions. Deposition of strained Si onto SiGe [7] splits the valleys into 2 and 4-fold manifolds. Confinement at the nanometer scale causes symmetry breaking and splitting of the 2-fold degeneracy.

Perfectly flat [100] Si quantum wells of 10-20 atomic monolayers can result in VS of the order of 1-10meV [7a]. Experimentally, however, the observed valley splitting is about 2 orders of magnitude smaller. This can be associated [7b] with the 2 degree slant of the experimental wafer. However, a perfectly stepped Si quantum well would result in VS that is one order of magnitude too small as compared to experiment. Atomistic disorder effects due to step disorder combined with SiGe alloy disorder raise the VS by an order of magnitude to match experimental data almost perfectly [7b]. We emphasize here that there were no particular parameter fits performed for these simulations. The Si/Ge parameters [9], Si quantum well

Authorized licensed use limited to: Purdue University. Downloaded on March 11, 2009 at 23:23 from IEEE Xplore. Restrictions apply.

size, and the SiGe buffer thicknesses, the local linear electric field and different disorder samples were the only input to the simulations. Mechanical strain and electronic structure are computed in 10 million and 2 million atom (sub-) systems.

The first simulation capability goal of NEMO3D had been self-assembled quantum dots [4,5]. Typical simulations contain 10 million atoms in the strain domain and 1-4 million atoms in the central device domain that confines electrons. Such simulations typically require 8-15 hours on 30-60 cores on a cluster [5]. Such simulation capabilities will soon be delivered on nanoHUB.org through user-friendly GUIs. However, these simulations are generally too time consuming to be used for educational purposes. Simulation times can be dramatically reduced if only a single band effective mass model, corresponding to a single "s" orbital are selected in NEMO3D. In that case the code no longer needs to solve (very slowly) for interior eigenvalue solutions, but can rapidly resolve exterior spectrum eigenvalues and deliver results in literally seconds.

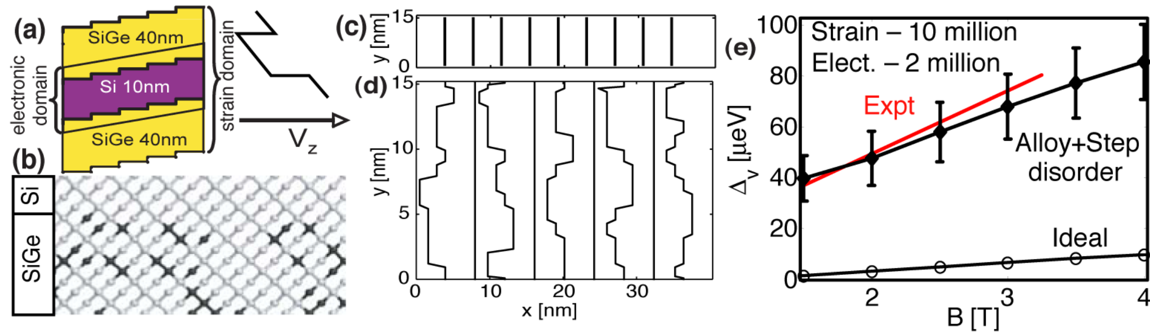

Fig. 4 . Valley splitting in a gated Si ultra thin body on a slanted SiGe substrate in a magnetic field [8b]. Disorder at the SiGe/Si interface and SiGe substrate (b) and disorder in the slanted steps (d) are critical in the quantitative agreement with the experiment (e). Ignoring the atomistic disorder underestimates the valley splitting by over an order of magnitude (ideal in (e)). Ignoring the slant of the quantum well altogether overestimates the valley splitting by 2 orders of magnitude. Thinner Si slabs can show valley splittings of the order of 10meV which will change the available states and threshold voltages of UTB devices.

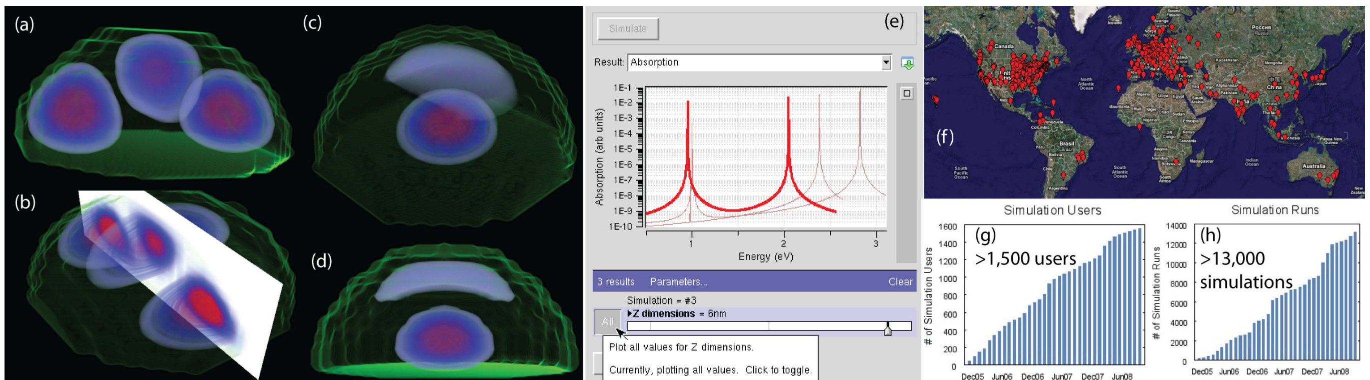

Fig 5. NEMO3D deployed in the educational application "Quantum Dot Lab" on nanoHUB.org. (a,b) show an interesting looking 4th state in a dome shaped quantum dot. (c,d) is the 7th state in the same dot. Data sliders shown in (e) allow users to interactively compare data similar to NEMO1D. (f) map of over 1,500 worldwide Quantum Dot Lab users. (g,h) Over 1,500 users have launched over 13,000 from Nov. 05 to August 08. Without any software installation users can explore different quantum dot geometries and their effects on quantum states and intra-band absorption interactively. Almost all nanoHUB tools have now this same look and feel through the new Rappture (www.rappture.org) toolkit created for nanoHUB.org.

nanoHUB.org's "Quantum Dot Lab" is powered by NEMO3D and was deployed in September 2005 primarily for educational purposes. Typical execution times are 6-10 seconds and users can select various quantum dot shapes and sizes. Fully interactive 3D rendering allows users to explore different quantum dot state symmetries as illustrated in Fig. 5. NEMO3D served over 1,500 users on nanoHUB.org as an educational tool [9] and will be deployed fully in fall of 2008.

OMEN

By 2007 the sp3s* and sp3d5s* models used in NEMO [4,5,9] are being adopted worldwide by several researchers, however, most still try to get by with effective mass or k.p models which cannot represent the crystal symmetry and its consequence on the quantum mechanical states, interfaces, and disorder. Meanwhile NEGF has become the accepted standard for quantum carrier transport, yet it remains widely assumed to be computationally too intensive to model realistically large devices, especially in an atomistic model. However, the investigation of more efficient ballistic approaches, like the Wave Function (WF) formalism, the availability of computers with tens of thousands of cores and the compatibility of NEGF and WF with multiple levels of parallelism [11] have opened the feasibility to implement OMEN [12]. OMEN comprehends the NEMO concepts and models realistically extended systems in atomistic device representations within compute times less than about an hour on parallel computers.

For nanowires and ultra-thin body (UTB) devices exhibit 3 and 4 levels of natural parallelism (Fig 6). In principle all voltage points can be considered independent, without incoherent scattering all total energies are decoupled, and the device can be spatially decomposed .UTBs have an additional degree of freedom for the transverse momentum. Large-scale parallel

Authorized licensed use limited to: Purdue University. Downloaded on March 11, 2009 at 23:23 from IEEE Xplore. Restrictions apply.

machines are becoming more and more available for researchers, not only at supercomputer centers, but also as local university resources. Figure 6b compares the OMEN scaling for computations of a 22nm UTB on 5 such machines up to 4,096 cores. Figure 6c shows that OMEN can scale to 32k processors which now results in computation times of a few minutes for formerly hero experiment simulations that used to take a half year or longer. For many device engineers such large computing resources appear to be rather out-of-reach. However, we emphasize here that within a few years we expect to have hundreds of cores available in workstations or clusters of workstations. OMEN will run efficiently on such resources.

Critical insight into SiGe nanowires (Fig 7) and p- and n-doped 22nm double gate MOSFETs [13] has already been gained. An analysis of transport through a InAs/InGaAs/InAlAs alloy device is presented at IEDM this year [14].

Similar to the release of NEMO3D as a limited resource on nanoHUB.org we are currently in the process of deploying a nanowire tool that is based on OMEN that will be able to explore 3nm nanowires in a compute time of about one hour (on around 100 cores provided transparently to the user). We recently replaced the Matlab code in Bandstructure Lab [15] with OMEN resulting in significant code speed-up.

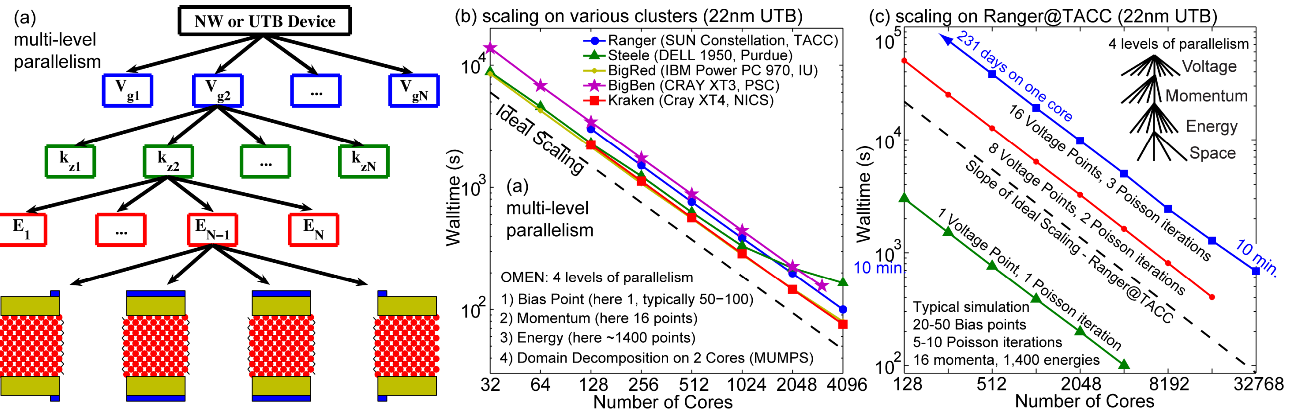

Fig. 6. (a) multi-level parallelism in voltage, momentum, energy, and space. (b) scaling on various clusters up to 4,096 cores at three supercomputer centers and 2 midwest universities (Purdue and Indian U.). Scaling only utilizes parallelism of the 3 innermost loops – course voltage parallelism not used. (c) scaling on the largest NSF-funded computational resource for US-based researchers (Ranger@TACC) utilizing all four levels of parallelism.

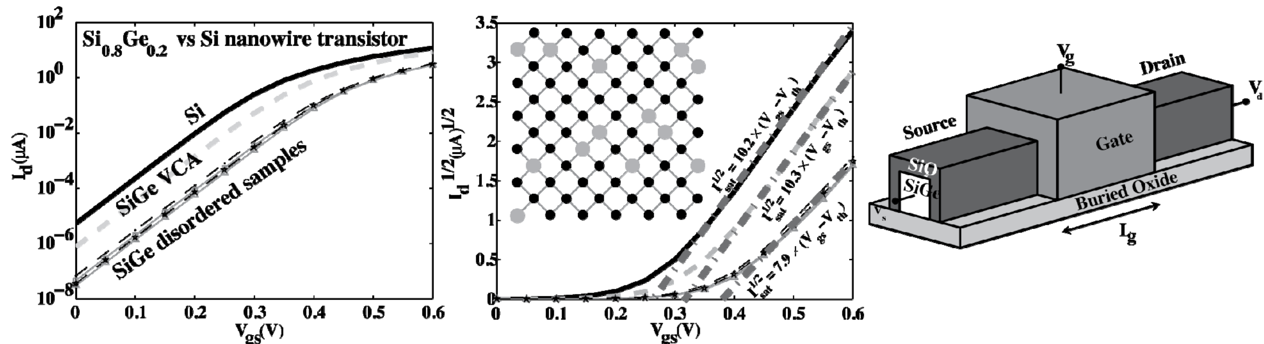

Fig. 6. Transfer characteristics of a 3.3x3.3x43.6nm^3 nanowire FET composed of 23,040 atoms. Wire is deposited on a buried oxide surrounded by a wrap around oxide of t_{ox}=1nm. The gate length is 15 nm and the source and drain are n doped at 10^{20}nm^{-3}. 4 samples of explicitly disordered SiGe material are compared to a smoothed out homogeneous (virtual crystal approximation) SiGe and to pure Si. The atomistic disorder (shown as an insert in the middle) lowers the current by over an order of magnitude. The effective channel mobility is reduced.

Acknowledgements

The NEMO and OMEN development efforts are clearly large efforts by a large number of people. We tried to cite the appropriate papers and mention Roger Lake, R. Chris Bowen, William R. Frensley, Dan Blanks, Timothy B. Boykin, Seungwon Lee, and Fabiano Oyafuso as seminal contributors.

References

[1] R. Lake, et *al*., J. of Appl. Phys. **81**, 7845 (1997)
[2] R. Lake, et *al.*, IEEE DRC, 174 (1996), R.C. Bowen, et *al*, J. of Appl. Phys. 81, 3207 (1997), G. Klimeck, et al, IEEE DRC, 92 (1997)
[3] G. Klimeck, el *al.*, Phys. Rev. B., **63**, 195310 (2001).
[4] G. Klimeck, et *al*., Computer Modeling in Eng and Sc. **3**, 601 (2002).
[5] G. Klimeck, et *al*., IEEE TED, **54**, 2079 (2007); *ibid* pg. 2090.
[6] G. Liang, et *al*., Nano Letters, **7**, 642 (2007).
[7] TB. Boykin, et *al*., Appl. Phys. Lett. **84**, 115 (2004), Phys. Rev. B. Vol. 70, 165325 (2004).N. Kharche, et *al*., Appl Phys. Lett. **90**, 092109 (2007).
[8] R. Rahman, et *al*., Phys Rev Lett, **99**, 036403 (2007); G.P. Lansbergen, et *al.*, Nature Physics, **4**, 656 (2008), G.P. Lansbergen, et al., "Transport-based dopant metrology in advanced FinFETs", IEEE IEDM 2008
[9] G Klimeck, et *al.*, Superlatt. & Microstr. **27**, 77 (2000), *ibid.*.519, [4], TB Boykin, et *al.*Phys. Rev. B **66**, 125207 (2002), TB. Boykin et *al*, Phys. Rev. B. **69**, 115201, (2004), AS Martins, et al., Phys. Rev. B. **72** 193204 (2005), TB Boykin, et *al.*, J. Phys.: Condens. Matter **19** 036203 (2007), TB Boykin, et *al.,* Phys. Rev. B **76**, 035310 (2007),
[10] G. Klimeck, et *al*., (2005), doi: 10254/nanohub-r450.3.
[11] G. Klimeck, Journal of Computational Electronics, **1**, 75 (2002).
[12] M. Luisier, et *al*., Phys. Rev. B, **74**, 205323 (2006).
[13] M. Luisier, G. Klimeck, SISPAD proceedings (2008).
[14] M. Luisier, N. Neophytou, N. Kharche, G. Klimeck, "Full-Band and Atomistic Simulation of Realistic 40 nm InAs HEMT", IEEE IEDM 2008.
[15] A. Paul, et *al.*, (2006), "Band Structure Lab," doi: 10254/nanohub-r1308.5

Authorized licensed use limited to: Purdue University. Downloaded on March 11, 2009 at 23:23 from IEEE Xplore. Restrictions apply.