International Journal of High Speed Electronics and Systems
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

QUANTUM SIMULATIONS OF DUAL GATE MOSFET DEVICES: BUILDING AND DEPLOYING COMMUNITY NANOTECHNOLOGY SOFTWARE TOOLS ON NANOHUB.ORG

SHAIKH AHMED, GERHARD KLIMECK, DERRICK KEARNEY, MICHAEL MCLENNAN

Network for Computational Nanotechnology,
Purdue University, West Lafayette, IN 4790, USA
ssahmed@purdue.edu

M P ANANTRAM

Department of Electrical and Computer Engineering,
University of Waterloo, Ontario, Canada
anant@ecemail.uwaterloo.ca

Received 16 October 2006
Revised Day Month Year

Undesirable short-channel effects associated with the relentless downscaling of conventional CMOS devices have led to the emergence of new classes of MOSFETs. This has led to new and unprecedented challenges in computational nanoelectronics. The device sizes have already reached the level of tens of nanometers where quantum nature of charge-carriers dominates the device operation and performance. The goal of this paper is to describe an on-going initiative on nanoHUB.org to provide new models, algorithms, approaches, and a comprehensive suite of freely-available web-based simulation tools for nanoscale devices with capabilities not yet available commercially. Three software packages nanoFET, nanoMOS and QuaMC are benchmarked in the simulation of a widely-studied high-performance novel MOSFET device. The impact of quantum mechanical effects on the device properties is elucidated and key design issues are suggested.

Keywords: NanoHUB; MOSFETs; Quantum effects; Online simulation; NEGF.

1. Introduction

Rapid progress in nanofabrication technology accelerating the continuous and aggressive downscaling of classical CMOS devices and the associated detrimental short-channel effects have inspired the emergence of new classes of MOSFETs. To fabricate devices beyond current scaling limits, IC companies are simultaneously pushing the planar, bulk silicon CMOS design while exploring alternative gate stack materials (high-*k* dielectric[1] and metal gates), band engineering methods (using strained Si or SiGe[2,3,4,5]), and alternative transistor structures that include primarily partially-depleted (PD) and fully-depleted (FD) silicon-on-insulator (SOI) devices. However, there is a virtual consensus that the most *practically scalable* variety of all these devices, that are in the focus of many researchers' study today, are double-gate SOI MOSFETs with a sub-10 nm gate length, ultra-thin, intrinsic channels and highly doped (degenerate) bulk electrodes—see, e.g., recent reviews[6,7] (and Figure 1). In such transistors, short channel effects typical for

their bulk counterparts are minimized, while the absence of dopants in the channel maximizes the mobility and hence drive current density.

These advanced MOSFETs may be practically implemented in several ways including planar, vertical, and FinFET geometries. However, several design challenges have been identified such as a process tolerance requirement of within 10% of the body thickness and an extremely sharp doping profile with a doping gradient of 1 nm/decade. The SIA forecasts that this new device architecture may extend MOSFETs to the 22 nm node (9-nm physical gate length) by 2016. Intrinsic device speed may exceed 1 THz and integration densities will be more than 1 billion transistors/cm^2.

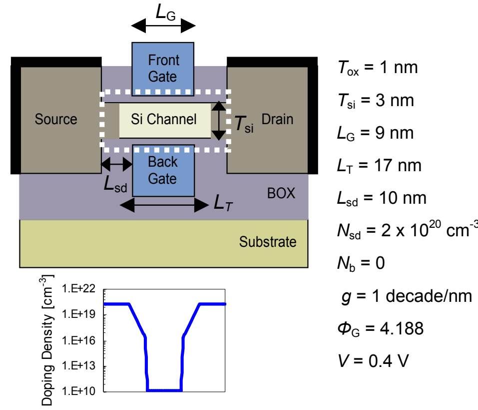

Fig. 1. The model of a widely-studied double-gate SOI MOSFET with ultra-thin intrinsic channel.

2. Nanoscale Device Simulation: Quantum Effects

Semiclassical Boltzmann transport has been the principal support in the field of modeling and simulations of semiconductor technology since its early development. To date, most *commercial* device simulations including the full-band Monte-Carlo (FBMC) method are based on the solution of the Boltzmann transport equation (BTE) and its simplified derivatives such as the hydrodynamic (HD) equations and the drift-diffusion (DD) model. However, for silicon nanoscale devices with active regions below 0.1 microns in diameter, macroscopic transport descriptions based on drift-diffusion (DD) models are clearly inadequate[8]. In the nanoscale regime, the transport is expected to be dominated by quantum effects throughout the active region. Quantum effects in the surface potential will have a profound impact on both, the amount of charge which can be induced by the gate electrode through the gate oxide, and the profile of the channel charge in the direction perpendicular to the surface (the transverse direction). Also, because of carrier confinement in the channel, mobility (or microscopically speaking, carrier scattering) will be different from the three-dimensional (3-D) case. A well known approach to study the impact of spatial confinement on mobility behavior is based on the self-consistent solution of the 2D Poisson–1D Schrödinger–2D Monte Carlo, and requires enormous computational resources as it requires storage of position dependent scattering tables that describe carrier transition between various subbands. However, in the smallest size devices, carriers experience very little or no scattering at all (*ballistic limit*). In such ballistic limit the question arises if a classical transport approach (BTE) or a full quantum mechanical approach is appropriate. Various quantum formalisms based on density matrices, Wigner functions, Feynman path integrals, and non-equilibrium Green's

functions (NEGF) have been developed and proposed. The Green's functions approach is the most exact, but at the same time appears, from the historical literature perspective, as the most difficult of all. The nanoHUB.org now features a webpage with introductory NEGF materials[9]. In contrast to, for example, the Wigner function approach (which is Markovian in time), the Green's functions method allows one to consider simultaneously correlations in space and time or space and energy, both of which are expected to be important in nanoscale devices. However, the difficulties in understanding the various terms in the resultant equations and the computational burden needed for its actual implementation are perceived as a great challenge. However, a successful utilization of the Green's function approach commercially is the NEMO (Nano-Electronics Modeling) simulator[9], which is effectively 1-D and is primarily applicable to resonant tunneling diodes. This work presented here focuses mainly on the modeling and simulations of the size-quantization effect within a fully quantum mechanical NEGF approach and a quantum-corrected Monte Carlo transport framework for 2-D MOSFET structures.

3. Community Nanoscale MOSFET Softwares

The Network for Computational Nanotechnology (NCN, http://www.ncn.purdue.edu) is a multi-university, NSF-funded initiative with a mission to lead in nanotechnology research and education as well as outreach to students and professionals by offering a set of cyber services (accessible through the nanoHUB portal www.nanoHUB.org) including interactive online simulation, tutorials, seminars, and online courses packaged using e-learning standards. In the past 12 months[a], the educational and outreach services were accessed by over 16,200 users. More than 3,500 users performed over 94,000 online simulations. Over 30 applications are available online ranging from toy models to sophisticated simulation engines not yet available commercially. All the NCN services are freely open to the public.

NanoHUB is dedicated to offer services to real users such as experimentalists and educators, not to computational scientists alone. Therefore, a user friendly GUI is required for these tools to be operated by non-experts. The tool should be available for anybody without any installation requirements. From a deployment perspective the tool development and GUI development must be streamlined, the codes must be benchmarked, and quality tested, and adequate computational resources must be available. Rappture is the new nanoHUB.org created toolkit that enables the rapid development of GUIs for applications. Two approaches can be followed: (1) The legacy application is not modified at all and wrapper script translates Rappture I/O to the legacy code. (2) Rappture is integrated into the source code to handle all I/O (see Figure 2). The first step is to declare the parameters associated with one's tool by describing Rappture objects in the Extensible Markup Language (XML). Rappture reads the XML description for a tool and generates the GUI automatically. The second step is that the user interacts with the GUI, entering values, and eventually presses the *Simulate* button. At that point,

[a] Data as of September 2006. Updated information can be found at http://nanoHUB.org/usage/

Rappture substitutes the current value for each input parameter into the XML description, and launches the simulator with this XML description as the *driver* file. The third step shows that, using parser calls within the source code, the simulator gets access to these input values. Rappture has parser bindings for a variety of programming languages, including C/C++, Fortran, MATLAB, Octave, Python, Perl, and Tcl. And finally, the simulator reads the inputs, computes the outputs, and sends the results through *run* file back to the GUI for the user to explore.

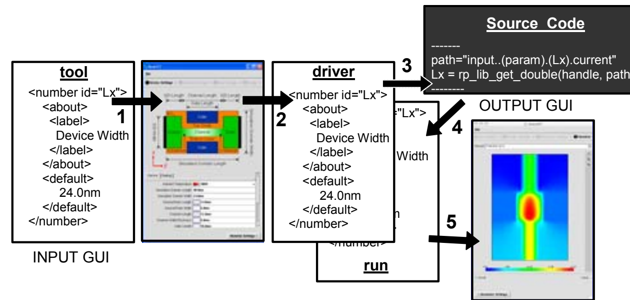

Fig. 2. Rappture: Revolutionizing tool development.

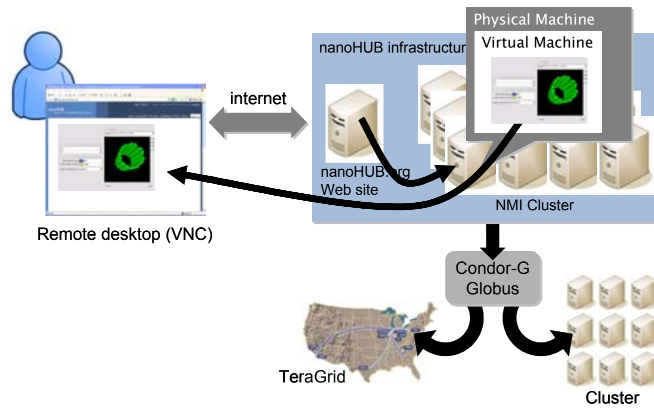

Fig. 3. Remote access to simulators and compute power.

The process of *web-based deployment* of these tools is depicted in Figure 3. A user visits the *www.nanohub.org* site and finds a link to a tool. Clicking on that link will cause our middleware to create a virtual machine running on some available CPU. This virtual machine gives the user his/her own private file system. The middleware starts an application and exports its image over the Web to the user's browser. The application looks like an Applet running in the browser. The user can click and interact with the application in real time taking advantage of high-performance distributed computing

power available on local clusters at Purdue University. Large scale calculations will soon be launched on the NSF TeraGrid or the open science grid. The following sections present three in-house software packages namely nanoFET, nanoMOS and QuaMC and benchmark these tools in the simulation of a widely-studied high-performance novel MOSFET device.

3.1 *nanoFET*

nanoFET is a newly developed tool, based on the research performed at NASA[10], that simulates quantum mechanical size quantization in the inversion layer and phase coherent and ballistic transport properties in two-dimensional MOSFET devices. The overall simulation framework consists of the *real-space* effective mass NEGF equations solved self-consistently with Poisson's equation. Solution of this set of equations is computationally intensive. Hence, nonuniform spatial grids are essential to limit the total number of grid points while at the same time resolving physical features. A novel algorithm for efficient computation of electron density without complete solution of the system of equations even in the presence of nonzero self-energies throughout the device has been used in this simulator[9,10]. The numerical problem consists in computing the diagonal elements of the matrix $G^r = \left[E\boldsymbol{I} - \boldsymbol{H} - \boldsymbol{\Sigma}\right]^{-1}$ (retarded Green's function) and $G^< = G\boldsymbol{\Sigma}^< G^\dagger$ (electron correlation Green's function), where E is the energy level, H is the device Hamiltonian matrix, and $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^<$ are self energies ($\dagger$ denotes the transpose conjugate of a matrix). The algorithmic flow is based on Dyson's equation solved through recursive Green's function approach. nanoFET has been parallelized with Message Passing Interface (MPI) and ported to various computing platforms at Purdue University. The MPI is applied in the integration procedure to calculate the charge density over the energy spectrum while the Green's function at each energy point is calculated by a serial algorithm. The resulting speed-up factor (Figure 4) shows a satisfactory scaling behavior for up to 32 processors. The reason of a declining behavior beyond 32 processors is attributed to the increased amount of time used in various MPI communication calls (right panel of Figure 4) that are implemented immediately after the parallel computation in order to exchange and gather large amount and size of data among the processors.

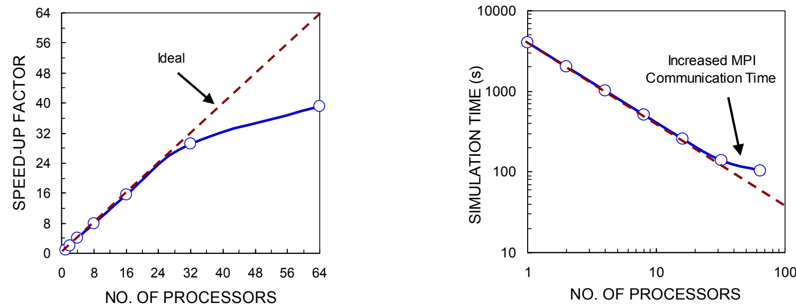

Fig. 4. (Left) nanoFET parallel performance. (Right) MPI timing diagram.

3.2 *nanoMOS*

nanoMOS is a 2-D simulator for thin body (less than 5 nm), fully depleted, double-gated *n*-MOSFETs. A choice of five transport models is available (drift-diffusion, classical ballistic, energy transport, quantum ballistic, and quantum diffusive). The transport models treat quantum effects in the confinement direction exactly and the names indicate the technique used to account for carrier transport along the channel. The quantum transport models are based on *mode-space* method within an effective mass approximation. The use of a mode space approach produces high computational efficiency that makes the quantum simulator practical for extensive device simulation and design. Scattering in the device can also be treated by a simple model that uses so-called Büttiker probes[11].

3.3 *QuaMC*

QuaMC is a quasi three-dimensional quantum-corrected semiclassical Monte Carlo transport simulator for conventional and non-conventional MOSFET devices. A parameter-free quantum field approach has been developed and utilized quite successfully in order to capture the size-quantization effects in nanoscale MOSFETs. The method is based on a perturbation theory around thermodynamic equilibrium and leads to a quantum field formalism in which the size of an electron depends upon its energy[12]. This simulator uses different self-consistent event-biasing schemes for statistical enhancement in the Monte-Carlo device simulations has been presented[13]. Regarding the Monte Carlo transport kernel, the explicit inclusion of the longitudinal and transverse masses in the silicon conduction band is done in the program using the Herring-Vogt transformation. Intravalley scattering is limited to acoustic phonons. For the intervalley scattering, both *g*- and *f*-phonon processes have been included.

4. Simulation Results and Discussion

Figure 1 shows the simulated DG FET device structure, which is similar to the device reported in Ref. 7. For quantum transport simulation purposes only the dotted portion of the device which has been termed as the *intrinsic* device is taken into considerations. The device was originally designed in order to achieve the ITRS performance specifications for the year 2016. The effective intrinsic device consists of two gate stacks above and below a thin silicon film. For the intrinsic device, the thickness of the silicon film is 3 nm. Using a thicker body reduces the series resistance and the effect of process variation but it also increases the short channel effects (SCE) due to lesser control on the channel charge. From SCE point of view, a thinner body is preferable but it is harder to fabricate very thin films of uniform thickness, and the same amount of process variation (±10%) may give intolerable fluctuations in the device characteristics. For the simulated intrinsic device, source/drain width is 3 nm. The top and bottom gate insulator thickness is 1 nm, which is expected to be near the scaling limit for SiO_2. For the gate contact, a metal gate with tunable workfunction, Φ_G, is assumed, where Φ_G is adjusted to 4.188 to provide a

specified off-current value of 4 μA/μm. The background doping of the silicon film is taken to be intrinsic, however, to take into account the diffusion of the dopant ions, the doping profile in the S/D extensions is graded with a coefficient of *g* which equals to 1 dec/nm (Figure 1). The doping gradient, *g*, affects both on-current and off-current. According to the ITRS roadmap, the high performance (HP) logic device would have a physical gate length of L_G = 9 nm at the year 2016. At this scale, 2-D electrostatics and quantum mechanical effects both play an important role. Values of all the structural parameters of the device are shown in Figure 1.

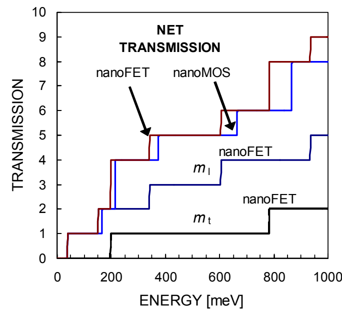

Fig. 5. Benchmarking nanoFET against nanoMOS: Transmission profile for a system with zero/flat potential applied in the device region.

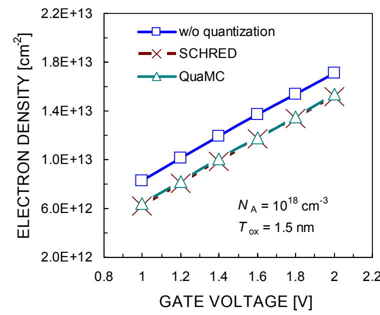

Fig. 6. Benchmarking of QuaMC against SCHRED: Variation of the sheet electron density versus gate voltage in a MOS capacitor structure.

The newly developed nanoFET simulator is first benchmarked against the nanoMOS simulator in the non-selfconsistent calculation of transmission under zero/flat potential distribution throughout the device region. The resulting transmission profile is shown in Figure 5. The curve marked as m_t on this plot corresponds to the (nanoFET derived) transmission in the primed valleys ($m_t = 0.19m_0$) and that marked as m_l corresponds to the unprimed valleys ($m_l = 0.98m_0$). The two topmost curves give the net transmission from nanoFET and nanoMOS. Noticeable is the reasonable agreement in the ground and excited state energy levels thus validating the approaches used to calculate the underlying Green's functions.

To verify the validity of the effective quantum field approach integrated in QuaMC simulator, self-consistent simulations of MOS capacitor structures have been performed and the reduction in the sheet electron density due to the band-gap widening effect is calculated. The simulation results are shown in Figure 6. Notice the excellent agreement between the SCHRED simulation data for the sheet electron density and the simulation results obtained from QuaMC. SCHRED is a 1-D Schrödinger-Poisson solver that is also available on the nanoHUB.org[14].

Next, the *intrinsic* device is simulated, within a *ballistic* approximation, to study the impact of size-quantization effects on the DG MOSFET performance. Figure 7 shows the total transmission vs. energy profile for a bias of V_G = 0 V and V_D = 0.4 V computed in

nanoFET. Note that the steps in the plot correspond to the subband levels due to channel quantization. Two main features in this plot are: (a) the integer values at the total transmission plateaus, (b) existence of a slope (not sharp) near the subbands which has significant impact on determining the current. The corresponding density of states peaks at the lateral confinement resonance level in the channel region. Figure 8 shows the current vs. energy plot where the increase in current for V_D = 0.4V (~0.49 µA/µm) as compared to that for V_D = 0.05V (~0.22 µA/µm) is due to an increase in both the tunneling (compare the shaded areas) and the thermionic components.

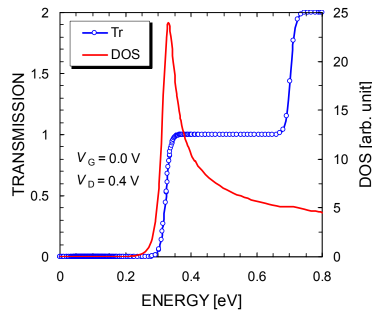

Fig. 7. nanoFET total transmission and density of states (DOS) vs. energy plot.

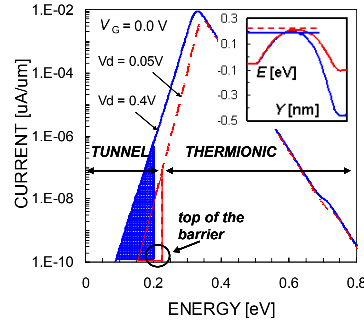

Fig. 8. nanoFET current vs. energy plot for two applied drain biases.

The device *ballistic* output characteristics are shown in Figure 9. All these simulations were performed on nanoHUB.org. For benchmarking purposes, apart from the device specifications, mentioned below is a list of key simulator parameters. (a) *nanoMOS*: A well-tempered DG MOSFET was chosen with low-field mobility = 300 cm^2/V/s, Caughy-Thomas parameter = 2, oxide-penetration was neglected, self-consistent convergence parameter = 1.0E-3, Poisson convergence parameter = 1.0E-6, all valleys were considered with 1 subband in each valley, the dual-gate and the Fermi-Dirac flags were activated. (b) *nanoFET*: nonuniform adaptive energy grid with a size (number) of 512 was used, resonance in the active region was included, full-simulation mode Poisson equation being solved self-consistently with an iteration number of 10, mass-anisotropy was included, both the $G^<$ and $G^>$ were calculated, backward recursion flag was turned on, 16 computing nodes with 1 processor/node were used. (c) *QuaMC*: Monte Carlo simulation time was 2ps with a time-step of 0.1fs, event-biasing flag was turned off, quantum correction flag was turned on for quantum-mechanical and off for classical calculations, for ballistic transport all the scattering switches were off and for dissipative transport they were turned on. As shown in Figure 9, for a device with 3nm film thickness, although at low drain biases there remains slight discrepancy in the results obtained from the nanoFET and nanoMOS simulations, they tend to overlap at higher drain voltages. The QuaMC simulations were also found to produce almost the same current as those obtained from fully quantum mechanical counterpart.

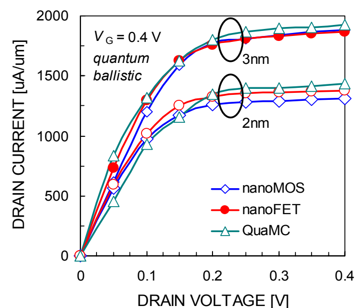

Fig. 9. Output characteristics for ballistic transport: Comparison of different simulators.

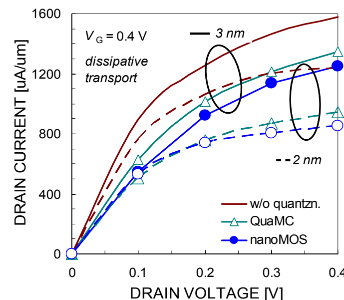

Fig. 10. Dissipative transport through nanoMOS and QuaMC: Impact of channel thickness.

Finally, simulation results for dissipative transport (a) without quantization from classical QuaMC and (b) with quantization obtained from nanoMOS and QuaMC are compared. The output characteristics are shown in Figure 10. Devices with both 3 nm and 2 nm channel thickness are used with a gate bias of 0.4 V. The salient features of this figure are: (1) Even with an undoped channel region, the devices achieve a significant improvement over the short channel effects (SCEs) as seen from the near-flat nature of the curves in the saturation region. This is due to the use of the two gate electrodes and an ultrathin SOI film which makes the gates gain more control on the channel charge. (2) Reducing the channel SOI film thickness to 2 nm further reduces the SCEs (saturation region has a reduced slope). (3) Comparing the classical and quantum mechanical simulations in both the devices, one can see that that the impact of quantization effects (percentage change in drain current) reduces with an increase in drain voltage. Higher biases primarily open all the available channels and quantum mechanical and classical calculations agree more closely. (4) Reduction in the drain current is more in 2 nm case throughout the range of the applied drain bias because of the stronger physical confinement. (5) Comparing the 3 nm and the 2 nm cases (classical or quantum mechanical), the increase in the percentage reduction in the drive current at higher drain biases is due to the series resistance effect in the thinner device. (6) Finally, the comparison between the quantum potential formalism and the NEGF approach for the device with 3 nm film thickness shows reasonable agreement which further establishes the applicability of these methods in the simulations of different technologically viable nanoscale classical and nonclassical MOSFET device structures.

5. Conclusion

NCN aims at the development of new *community codes* that provide the nanoscience research community with new capabilities and lay a foundation for a new generation of CAD tools that will pave the way to ground breaking nanotechnology devices. In the past 12 months, the educational and outreach services were accessed by over 16,200 users.

More than 3,500 users performed over 94,000 online simulations. We have presented benchmarking results obtained from nanoFET, nanoMOS and QuaMC, validating the applicability of these nanoHUB.org tools in the simulations of future nanoscale MOSFETs, that show the importance of quantum effects in these novel devices.

Acknowledgments

Supported by the Indiana 21st Century Fund and National Science Foundation under Grant No. EEC-0228390. The computational resources of nanoHUB.org have been utilized. Discussions with Dragica Vasileska and Sayed Hasan are acknowledged.

References

1. W. Zhu, J. P. Han, T. P. Ma, “Mobility Measurement and Degradation Mechanisms of MOSFETs Made With Ultrathin High-k Dielectrics”, *IEEE Trans. Elect. Dev*., **51**, 98 (2004).
2. Welser, J. L. Hoyt and J. F. Gibbons, “NMOS and PMOS transistors fabricated in strained silicon/relaxed silicon-germanium structures”, *IEDM Tech. Dig*., 1000 (1992).
3. D. Vasileska and S. S. Ahmed, “Narrow-Width SOI Devices: The Role of Quantum Mechanical Size Quantization Effect and the Unintentional Doping on the Device Operation”, *IEEE Transactions on Electron Devices*, **52**, 227 (2005).
4. D. Vasileska, G. Formicone and D.K. Ferry, “Doping dependence of the mobility enhancement in surface-channel strained-Si layers,” *Nanotechnology*, **10**, 147 (1999).
5. R. Oberhuber, G. Zandler, and P. Vogl, “Subband structure and mobility of two–dimensional holes in strained Si/SiGe MOSFET's,“ *Phys. Rev. B*, **58**, 9941 (1998).
6. Thomas J. Walls, Victor A. Sverdlov, and Konstantin K. Likharev,”Nanoscale SOI MOSFETs: a comparison of two options,” *Solid–State Elect*., **48**, 857 (2004).
7. Sayed Hasan, Jing Wang, and Mark Lundstrom,” Device design and manufacturing issues for 10 nm-scale MOSFETs: a computational study,” *Solid-State Elect*., **48**, 867 (2004).
8. N. Sano, A. Hiroki, K. Matsuzawa, “Device modeling and simulation toward sub-10nm semiconductor devices,” *IEEE Trans. Nanotechnology*, **1**, 63 (2002).
9. https://www.nanohub.org/topics/negf_theory
   R. Lake, G. Klimeck, R.C. Bowen, and D. Jovanovic, J. Appl. Phys., **81**, 7845 (1997).
10. A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel and R. Venugopal, “Two Dimensional Quantum Mechanical Modeling of Nanotransistors”, *Jour. of Appl. Phys*., **91**, 2343 (2002).
    http://www.nanohub.org/simulation_tools/nanoFET_tool_information
11. Z. Ren, R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom, “nanoMOS 2.5: a two-dimensional simulator for quantum transport in double-gate MOSFETs,” *IEEE Trans. Electron. Dev*., **50**, 9, 1914, (2003).
    https://www.nanohub.org/simulation_tools/nanomos-3_tool_information
12. S. Ahmed, C. Ringhofer, D. Vasileska, “Parameter-Free Effective Potential Method for Use in Particle-Based Device Simulations”, *IEEE Transactions on Nanotechnology*, **4**, 465 (2005).
    http://www.nanohub.org/simulation_tools/quamc2d_tool_information
13. M. Nedjalkov, S. Ahmed, and D. Vasileska, “A self-consistent event biasing scheme for statistical enhancement”, *Journal of Computational Electronics*, **3**, 305 (2004).
14. SCHRED simulations performed on nanoHUB.org.
    https://www.nanohub.org/simulation_tools/schred_tool_information