Journal of Computational Electronics (2020) 19:1031–1046
https://doi.org/10.1007/s10825-020-01519-0

Nano-electronic Simulation Software (NESS): a flexible nano-device simulation platform

Salim Berrada[1] · Hamilton Carrillo-Nunez[1] · Jaehyun Lee[1] · Cristina Medina-Bailon[1] · Tapas Dutta[1] · Oves Badami[1] · Fikru Adamu-Lema[1] · Vasanthan Thirunavukkarasu[1] · Vihar Georgiev[1] · Asen Asenov[1]

Published online: 5 June 2020
© The Author(s) 2020

Abstract

The aim of this paper is to present a flexible and open-source multi-scale simulation software which has been developed by the Device Modelling Group at the University of Glasgow to study the charge transport in contemporary ultra-scaled Nano-CMOS devices. The name of this new simulation environment is Nano-electronic Simulation Software (NESS). Overall NESS is designed to be flexible, easy to use and extendable. Its main two modules are the structure generator and the numerical solvers module. The structure generator creates the geometry of the devices, defines the materials in each region of the simulation domain and includes eventually sources of statistical variability. The charge transport models and corresponding equations are implemented within the numerical solvers module and solved self-consistently with Poisson equation. Currently, NESS contains a drift–diffusion, Kubo–Greenwood, and non-equilibrium Green's function (NEGF) solvers. The NEGF solver is the most important transport solver in the current version of NESS. Therefore, this paper is primarily focused on the description of the NEGF methodology and theory. It also provides comparison with the rest of the transport solvers implemented in NESS. The NEGF module in NESS can solve transport problems in the ballistic limit or including electron–phonon scattering. It also contains the Flietner model to compute the band-to-band tunneling current in heterostructures with a direct band gap. Both the structure generator and solvers are linked in NESS to supporting modules such as effective mass extractor and materials database. Simulation results are outputted in text or vtk format in order to be easily visualized and analyzed using 2D and 3D plots. The ultimate goal is for NESS to become open-source, flexible and easy to use TCAD simulation environment which can be used by researchers in both academia and industry and will facilitate collaborative software development.

Keywords Integrated simulation environment · Non-equilibrium Green's function · Tight-binding · Effective mass · Kubo–Greenwood · Drift–diffusion · Variability

1 Introduction

Further down-scaling of Complementary Metal-Oxide Semiconductor (CMOS) circuits has become increasingly complex and the fundamental challenges that the semiconductor industry faces at the device level will deeply affect the design of the next-generation integrated circuits and systems [1, 2]. Silicon technology has reached the nano-CMOS era with 10 and 7 nm Fin Field-Effect Transistors (FinFETs) CMOS technologies [3–7] in mass production and 5 nm and 3 nm Gate-All-Around (GAA) NanoWire Field Effect Transistors (NWFETs) in development stage [8]. It is widely recognized that charge transport in such nanometer-scale device dimensions could be dominated by quantum mechanical effects. Moreover, nano-scaled devices are more sensitive to systematic and statistical variability, which can result in significant differences between devices on the same chip [9]. Hence, variability and quantum mechanical effects are among the main challenges which should be addressed in order to keep Moore's law alive.

✉ Salim Berrada
salim.berrada.b@gmail.com

Vihar Georgiev
Vihar.Georgiev@glasgow.ac.uk

Asen Asenov
Asen.Asenov@glasgow.ac.uk

[1] School of Engineering, University of Glasgow, Glasgow G12 8LT, UK

In order to meet the challenges of the future nano-CMOS technology above, the most-time efficient and cost-effective method is to utilize numerical simulations based on relevant theories and physical models to screen material and device architecture options and to optimize the promising solution. It is also important for such tools to be user-friendly and to be published as an open-source software to allow collaboration and co-development by both industry and academia all over the world. This will allow a collaborative effort of the electron device community to find the solutions for tomorrow CMOS circuit designs.

The main aim of this paper is to introduce the concepts and the inner working of a new nanoelectronic device computational framework—Nano-electronic Simulation Software (NESS), which is currently under development at University of Glasgow. This paper is organized as follows. Section 2 presents a general overview of NESS and the link between its different components and features. Section 3 describes the structure generator module that was developed in order to create the simulated device structures and to introduce sources of statistical variability. Section 4 is dedicated to the effective mass extractor module which was developed in order to provide the parameters to the models and the corresponding solvers implemented in the transport module, described in Sect. 5. Each subsection in Sect. 5 is dedicated to a specific functional NESS solver. Finally, conclusions are drawn in Sect. 6.

2 Overview of NESS

In this section, we provide an overview of NESS and its modular structure which is summarized in Fig. 1. There are two main modules in NESS: the structure generator and the transport solvers. The structure generator module provides flexible means to create the simulated device structures such as nanowire, FinFETs or bulk CMOS transistors considering different semiconductor materials such as Si, Ge or III–Vs materials. Furthermore, the structure generator is used to introduce the doping profile and generate the meshing of the simulation domain. It can also introduce the major sources of statistical variability such as Random Discrete Dopants (RDD), Line Edge Roughness (LER) and Metal Gate Granularity (MGG). Details of the implementation of the NESS structure generator are presented in Sect. 3.

The second main module in NESS contains the charge transport models and the corresponding solvers which can simulate using different approximations the mobility, the charge density and the current in nano-CMOS devices. The non-equilibrium Green's function (NEGF) solver is the main transport solver in NESS. It is a quantum transport module capturing quantum mechanical effects such as quantum confinement and subsequent threshold voltage shift, coherent transport and impact of scattering, and the leakage and Band-To-Band Tunneling (BTBT) currents. A detailed description of the NEGF approach implemented in NESS is given in Sect. 5.2.

The NESS computational framework has two additional modules to support the work of the structure generator and

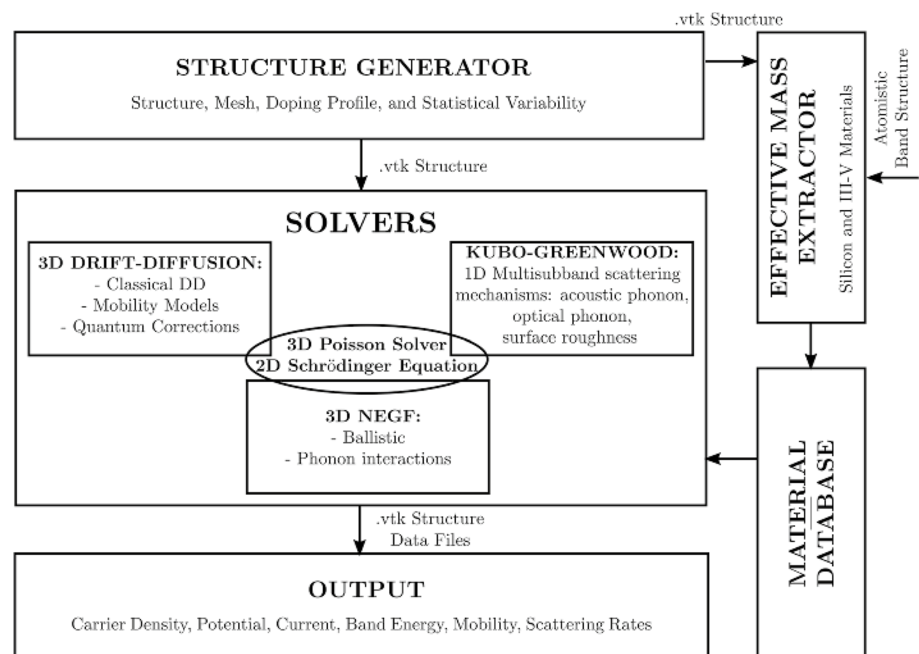


Fig. 1 Flowchart of NESS detailing its modular structure

numerical solvers. The first module is the Effective Mass (EM) extractor from atomistic simulations such as Density Functional Theory (DFT) or from semi-empirical models such as Tight-Binding (TB). The EM extractor is described in detail in Sect. 4. The other key module is material database which contains for each material all parameters relevant for the different simulation techniques and corresponding modules, e.g., dielectric constants, mobility model parameters, the parameters used in different scattering models etc. Those parameters serve as input information for the solvers.

3 Structure generator

The starting point for all NESS simulations is the creation of the simulation domain with the corresponding material regions and their parameters used by the different transport solvers. The NESS structure generator is very flexible and allows the user to create devices with various architectures, cross sections, doping profiles and variability sources. An example of a circular NWFET with LER, RDD and MGG variability sources is shown Fig. 2a, b. The methodology to generate the aforementioned sources of variability is described below:

Random Discrete Dopants In order to introduce random discrete dopants in the structure, we have adopted a rejection technique which is based on the atomic arrangement in the crystal lattice of the corresponding material [10]. The probability of finding a RDD at the ith atom site is expressed as follows:

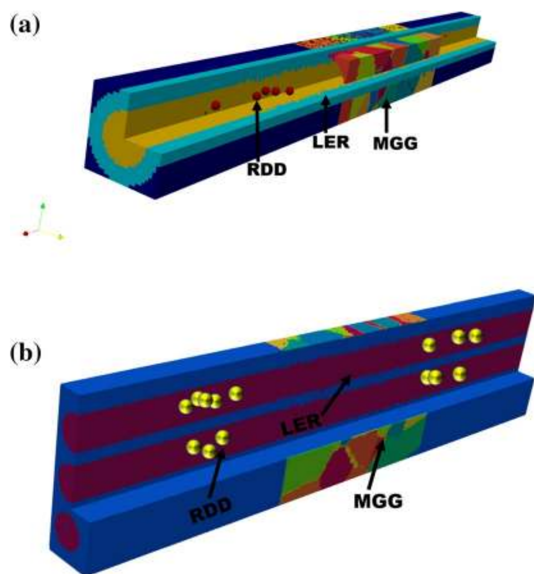


Fig. 2 **a** Circular and **b** stacked NWFETs with all variability sources that can be generated by NESS

$$P_i = V_i \cdot N_i \tag{1}$$

where V_i and N_i are the volume and the doping concentration of the ith atom. If the generated random number between 0 and 1 is smaller than P_i, the atom is replaced by the dopant. Therefore, dopants are randomly distributed based on the doping density, and the total number of dopants follows a Poisson distribution. Herein, the dopant has an elemental positive (for n-type) or negative (for p-type) charge assuming that all dopants are activated. The dopant charge is assigned to the eight surrounding nodes of the discretization grid using the cloud-in-a-cell approach.

Line edge roughness LER is generated at the interface between the channel material and the gate oxide using the same approach described in Ref. [11]. LER is thus characterized in NESS by an exponential auto-correlation function as follows:

$$C(r) = \langle \Delta(r')\Delta(r'+r) \rangle = \Delta_m^2 \mathrm{e}^{\sqrt{2}r/L_m} \tag{2}$$

where Δ_m is the root-mean-square (rms) fluctuation of the rough edge, L_m is the correlation length, r is the distance $\|\vec{r}\|$, and $\Delta(r)$ is the amplitude of LER at position $\vec{r}$.

Metal gate granularity In order to investigate the impact of MGG, grains in the metal gate have been generated by the Voronoi algorithm [12]. The work-function of each grain is assigned using a probability density specified by the user. For example, it is assumed that TiN has two grain orientations with two different values of work-function projected at the metal-insulator interface (4.4 and 4.6 eV) with a probability of 40% and 60%. The average work-function can be calculated as follows:

$$W_{\text{avg}} = \sum_i W_i \cdot \hat{P}_i \tag{3}$$

where W_i is the ith work function value and $\hat{P}_i$ is its probability. W_{avg} is the typical work function for the corresponding metal material and $\sum_i \hat{P}_i$ must be equal to one.

4 Effective mass extraction module

Currently in NESS we use the parabolic Effective Mass (EM) approximation for both electrons and holes when solving transport equations. Currently, NESS supports only diagonal effective mass tensor. The confinement effective masses (e.g., m_y and m_z in the 1-D structure) play a critical role in the device simulations as they directly impact the subband energy levels and thus the electrostatics. In NESS, the confinement effective masses $m_{y,\nu}$ and $m_{z,\nu}$ are extracted in way to match the subband minima for the lowest two subbands of each dominant valley ν obtained using empirical TB or DFT. The confinement

effective masses are obtained by solving the following set of equations for each valley using Newton-Raphson method:

$$E^{\mathrm{EM}}_{i,v}(m_y, m_z) - (E^{\mathrm{TB}}_{i,v} - E^{\mathrm{B}}_{C,v}) = 0, \quad i = 1, 2 \tag{4}$$

where v is the valley index and, $E^{\mathrm{EM}}_{i,v}$ and $E^{\mathrm{TB}}_{i,v}$ are the ith subband minima obtained from the parabolic EM approximation and TB methods, respectively, and $E^{\mathrm{B}}_{\mathrm{C}}$ is the bulk conduction band edge. The transport effective masses (m_x) are obtained from the curvature of the E−k dispersion at the minima as follows

$$m_x = \hbar^2 \left(\frac{\partial^2 E}{\partial k_x^2} \right)^{-1} \tag{5}$$

Figure 3 compares the parabolic EM dispersion with the TB band structure, which is calculated using Synopsys QuantumATK with the Boykin parameter set [13, 14]. The first two subbands of both models are in good agreement for each Δ valley especially at low energies, which are more relevant for transport. It is thus clear that at such a good agreement between the two band structure, the EM approximation can provide good estimate for the devices figures of merit at a much lower computational cost compared to the TB framework. More details about this module can be found in Ref. [15].

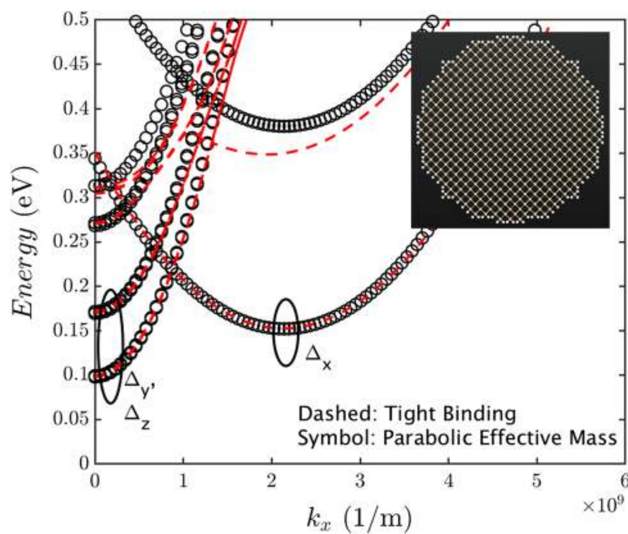


Fig. 3 Comparison of the parabolic EMA and TB band structures for a circular GAA nanowire (diameter = 5 nm). The inset shows the arrangement of silicon atoms in a cross section of the nanowire

5 Numerical solvers

5.1 Drift diffusion solver

NESS includes a 3D drift–diffusion (DD) transport solver based on the self-consistent solution of the Poisson's equation and the current continuity equation. The equations being solved are:

(a) Poisson's equation:

$$\nabla \cdot (\epsilon \nabla \psi) = -q(p - n + N_{\mathrm{D}}^{+} - N_{\mathrm{A}}^{-}) \tag{6}$$

where ϵ is permittivity, ψ is potential, q is the electron charge, n and p are the electron and hole concentrations, and N_{D}^{+} and N_{A}^{-} are the ionized donor and acceptor impurity concentrations, respectively.

(b) Continuity equation for electrons (assuming no carrier generation or recombination):

$$\nabla \cdot \mathbf{J_n} = 0 \tag{7}$$

where $\mathbf{J_n}$ is the electrons current density vector given by

$$\mathbf{J_n} = -qn\mu\nabla\psi + qD_n\nabla n \tag{8}$$

where μ is carrier mobility and $D_n = k_{\mathrm{B}}T/q$ is the diffusion coefficient. The Scharfetter–Gummel approach has been used for the discretization of the drift–diffusion equations using the Bernoulli functions [16, 17]. The current density flowing from node 1 to node 2 is given by

$$J_{n12} = \frac{qD_{12}}{h_{12}} \left[B\left(\frac{\psi_1 - \psi_2}{k_{\mathrm{B}}T/q} \right) n_1 - B\left(\frac{\psi_2 - \psi_1}{k_{\mathrm{B}}T/q} \right) n_2 \right] \tag{9}$$

where D_{12} and μ_{12} are, respectively, the diffusion coefficient and mobility at the middle of the two nodes, $n_1(n_2)$ is the electron concentration in node 1(node 2), $\psi_1(\psi_2)$ is the potential at node 1(node 2), k_{B} is the Boltzmann constant, h_{12} is the distance between the two nodes, and $B(x)$ is the Bernoulli function defined as:

$$B(x) = \frac{x}{e^x - 1} \tag{10}$$

Currently we have implemented four different models for the carrier mobility in this solver. Firstly, we have a constant mobility model that defines an isotropic low-field mobility value which is kept constant during the simulation. In addition, there are three other models to account for mobility reduction due to the impact of doping and electric fields. The Masetti model [18] has been included to capture the doping concentration dependence of mobility. This model defines a local low-field mobility dependent on the net local doping concentration N within the simulation domain. It follows the analytic function that fits empirical electron and

hole mobilities as a function of the doping material and the temperature in bulk semiconductor material [18]:

$$\mu_{\mathrm{MM}} = \mu_0 \mathrm{e}^{-p_c/N} + \frac{\mu_{\max}\left(\frac{T}{300}\right)^{-\gamma} - \mu_1}{1 + (N/C_r)^{\alpha}} - \frac{\mu_2}{1 + (C_s/N)^{\beta}} \tag{11}$$

Here, the parameter μ_0 refers to maximum mobility; $\mu_{\max}$ is the phonon limited maximum mobility with γ defining its power law temperature dependence [19]; $\mu_1 = \mu_0$ for electrons and $\mu_1 = 0$ for holes; μ_2 is a mobility parameter; N is the net doping concentration; C_r and C_s are reference concentrations; α, β and γ are fitting parameters; and $p_c = 0$ for electrons and positive for holes. Typical values for the parameters can be found in [18].

An interface mobility correction algorithm is applied after the Masetti model evaluation. This correction results in a further reduction of the mobility as the distance from the semiconductor/insulator interface increases. It is only evaluated at the nodes where the material is a semiconductor and which are located in the channel. The corrected mobility is calculated as:

$$\mu_{\mathrm{IC}} = \mu_{\mathrm{MM}} \cdot \mathrm{sFactor} \cdot \mathrm{e}^{\frac{-(\Delta y + \Delta z)}{l_{\mathrm{Decay}}}} \tag{12}$$

where μ_{IC} is the mobility after including the interface mobility correction, μ_{MM} is pre-correction mobility from the Masetti model, sFactor is the surface mobility correction factor, Δy and Δz are the distances from the interface in the confinement directions (y and z, respectively), and l_{Decay} is the exponential decay factor.

The impact of the transverse electric field is captured using the well known Yamaguchi [20] model:

$$\mu_{\mathrm{YM}} = \frac{\mu_0}{\sqrt{1 + \left(\frac{|E_Z|}{Ec_{\mathrm{YM}}}\right)^{\alpha_{\mathrm{YM}}}}} \tag{13}$$

where μ_0 is either the output of the Masetti mobility model with interface correction (if used) or simply the low-field mobility, and E_Z is the electric field in the direction normal to the transport. Ec_{YM} (critical field) and α_{YM} are fitting parameters for this model.

Finally, the impact of the longitudinal electric field (along the transport direction), E_X has been taken into account using the Caughey–Thomas model [21]:

$$\mu_{\mathrm{CTM}} = \frac{\mu_{\mathrm{YM}}}{\left[1 + \left(\frac{\mu_{\mathrm{YM}}|E_X|}{v_{\mathrm{sat}}}\right)^{\beta}\right]^{1/\beta}} \tag{14}$$

Here μ_{YM} is the transverse field dependent mobility calculated using the Yamaguchi model (Eq. 13), and v_{sat} and β are temperature dependent fitting parameters.

As an example, Fig. 4 shows the transfer characteristics obtained using the constant mobility model with the bulk mobility value of 1400 $\mathrm{cm^2/Vs}$ as well as the cumulative impact of the three mobility degradation models for a NWFET with a square cross section of $3 \times 3\,\mathrm{nm}^2$. When the mobility degradation models are switched on, the corresponding degradation of the carrier mobility is reflected in the device characteristics by a reduction of the current.

Currently, the main use of the DD module in NESS is to provide the trial potential for NEGF simulations. We provide in Sect. 5.3.3 a comparison of NEGF results to the ones obtained using DD with KG mobilities. Quantum corrections are being added to the DD solver to make it suitable for modeling devices that exhibit quantum confinement. The DD model is also useful for variability simulations after calibration to more physical models such as NEGF in the case of NESS.

5.2 Non-equilibrium Green's function solver

This solver is the main transport solver of NESS. It allows a quantum treatment of charge transport to capture phenomena such as tunnelling, coherence and particle interactions that strongly impact the performance of nano-scaled devices. The electrons are described in this solver by an effective mass Hamiltonian. By solving self-consistently Poisson and NEGF transport equations in coupled mode-space representation, we obtain the charge density, the potential profile and the corresponding current that flows in the device. We can either include the electron–phonon (e–ph) interactions or neglect them to study the transport in the ballistic limit [22]. It is also

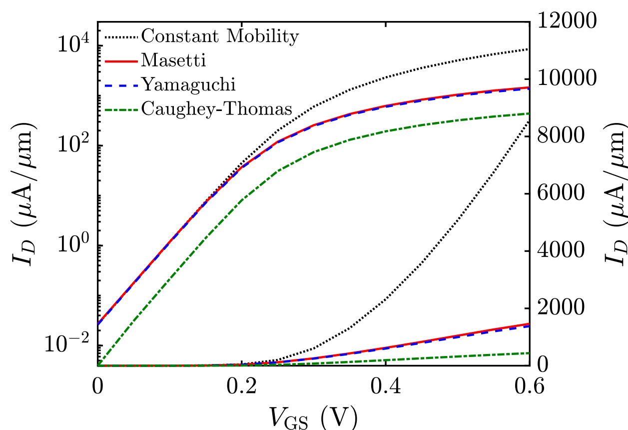


Fig. 4 Transfer characteristics of a NWFET ($W = H = 3\,\mathrm{nm}$, $L_G = 20\,\mathrm{nm}$) obtained from the classical DD module showing the impact of different mobility models. $V_{DS} = 0.05\,\mathrm{V}$

possible to consider other types of scattering explicitly such as RDD, LER or MGG.

5.2.1 Computation of the charge and the current

NEGF is a powerful quantum field theory tool that allows the computation of the time dependent quantum average of observables [23, 24]. When the system is in a steady state, one has to solve first at each energy E for the relevant components of the retarded G^{R}, the advanced G^{A} and the lesser Green's function $G^{<}$ using the following system of equations [25]:

$$G^{\mathrm{R}}(E) = \frac{1}{(E + \mathrm{i}.\eta)\cdot I - h - \Sigma^{\mathrm{R}}(E)}, \tag{15}$$

$$G^{\mathrm{A}}(E) = \left[G^{\mathrm{R}}(E)\right]^{\dagger}, \tag{16}$$

$$G^{\lessgtr}(E) = G^{\mathrm{R}}(E)\cdot \Sigma^{\lessgtr}(E)\cdot G^{\mathrm{A}}(E), \tag{17}$$

where h is the one-particle Hamiltonian, I is the identity matrix of the same dimension as h, η is an infinitesimal positive real number, and Σ^{R} and $\Sigma^{<}$ are the retarded and lesser self-energies, respectively. These self-energies take into account electrons scattering and their interaction with the contacts. The contact self-energies stem from the embedding of the contacts degrees of freedom in the active region [24]. The interaction self-energies arise from the truncation of the Martin–Schwinger hierarchy [26] to the first equation, in which a conserving approximation to the two-particle Green's function is introduced [27]. The charge at position r and the current in layer l are then obtained using the following equations [28, 29] :

$$n(r) = -\frac{i}{2\pi}\int \mathrm{d}E\, G^{<}(r, r; E), \tag{18}$$

$$j(l) = \frac{2\cdot|q|}{\hbar}\int \frac{\mathrm{d}E}{2\pi}\,\mathrm{Tr}\left[2\,\mathrm{Re}\left\{h_{l+1,l}\cdot G^{<}_{l,l+1}\right\}\right], \tag{19}$$

where $h_{l+1,l}$ are the matrix elements of the Hamiltonian between the basis states in layer $l+1$ and layer l while $G^{<}_{l,l+1}$ are the matrix elements of $G^{<}$ between the basis states of layer l and layer $l+1$. For tri-diagonal Hamiltonians and under the assumption of local scattering mechanisms in real space, one needs to compute only the upper-diagonal, the lower-diagonal and the diagonal Blocks of the lesser Green's function [22, 25]. Therefore, an efficient recursive algorithm has been proposed to solve only for these blocks [28].

5.2.2 Inclusion of the contacts as boundary conditions

The Hamiltonian representing the non interacting electron gas in the active region reads in real space:

$$\hat{H} = \int \mathrm{d}r\,\mathrm{d}r'\ \hat{\Psi}^{\dagger}(r)\ h\left(r, r'\right)\ \hat{\Psi}\left(r'\right) \tag{20}$$

where $\hat{\Psi}^{\dagger}(r)$ ($\hat{\Psi}\left(r'\right)$) is the creation (annihilation) operator described in [23, 24]. In most models, one can represent the one-particle Hamiltonian h as a succession of layers coupled to their nearest neighbors. This is true for the discretized EM Hamiltonian implemented in NESS and one obtains a tridiagonal representation when finite difference approximation is used to discretize it:

$$h\left(r, r'\right) \equiv \begin{pmatrix} h_{1,1} & h_{1,2} & 0 & \cdots & \cdots & \cdots & 0 \\ h_{2,1} & h_{2,2} & h_{2,3} & 0 & \cdots & \cdots & 0 \\ 0 & h_{3,2} & h_{3,3} & h_{3,4} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & h_{N,N-1} & h_{N,N} \end{pmatrix} \tag{21}$$

where N is the number of layers in the active region. This matrix represents only the restriction of the EM Hamiltonian to the active region. The impact of the electron exchange with the contacts is taken via the contact self-energies Σ_C. This is possible because the contacts are assumed to be invariant under a unit cell translation and in equilibrium. Therefore, it is possible to compute exactly the so-called $g^{\mathrm{R}}(C, C)$, the retarded Green's function of the contact at the interface with the device [30] and obtain the corresponding retarded self-energies for the electrons present in the device:

$$\Sigma^{\mathrm{R}}_{C} = H_{D,C}\cdot g^{\mathrm{R}}(C, C)\cdot H_{C,D}, \tag{22}$$

where $H_{D,C}$ is the matrix representing the Hamiltonian elements between the device states and the contact states and $H_{C,D}$ is its Hermitian conjugate. Thanks to the equilibrium property, the rate operator Γ and the lesser component of contact induced self-energy $\Sigma^{<}_{C}$ can be obtained from the retarded one in Eq. 22 using the fluctuation-dissipation theorem for the self-energies [24]:

$$\Gamma(E) = \mathrm{i}\left[\Sigma^{\mathrm{R}}(E) - \Sigma^{\mathrm{A}}(E)\right], \tag{23}$$

$$\Sigma^{<}_{C} = \mathrm{i}\cdot f\left(E - \mu_C\right)\Gamma(E), \tag{24}$$

where f is the Fermi distribution function and μ_C is the Fermi level of the contact. We use in NESS a fast iterative scheme to compute the contact self-energy [31].

5.2.3 The coupled-mode space approximation for effective mass Hamiltonian

The expression of the single particle Hamiltonian $h(x)$ in the EM approximation is:

$$h(r) = \underbrace{\left[-\frac{\hbar^2}{2m^*_{y,z}} \Delta_{y,z} + V(r) \right]}_{h_T} \underbrace{-\frac{\hbar^2}{2m^*_{y,z}} \frac{\partial^2}{\partial x^2}}_{h_L}. \tag{25}$$

where $\Delta_{y,z}$ is Laplace operator in the (YZ) cross-sectional plane. Therefore, $h(r)$ is the sum of a transverse part h_T describing the layers and a longitudinal part h_L describing their coupling. When Eq. 25 is discretized using a finite difference method, the Hamiltonian has the same representation as in Eq. 21 where h_T contributes to the diagonal blocks $h_{n,n}$ and h_L contributes to both diagonal and off-diagonal blocks $h_{n,n+1}$. The Coupled-Mode Space (CMS) representation is obtained by projecting each diagonal bock $h_{n,n}$ of the transverse part h_T of EM Hamiltonian on a subspace spanned by some chosen eigenmodes $\phi_i(y, z;n)$ of $h_{n,n}$. The corresponding projector is obtained by forming for each layer n a matrix $U_{n,n}$, whose columns are chosen eigenvectors of $h_{n,n}$, and performing the following transformations of the non-zero blocks of the RS Hamiltonian [32]:

$$\tilde{h}_{n,n} = U^\dagger_{n,n} h_{n,n} U_{n,n}, \tag{26}$$

$$\tilde{h}_{n,n+1} = U^\dagger_{n,n} h_{n,n} U_{n+1,n+1}. \tag{27}$$

The global transformation U is a block-diagonal matrix $U = \delta_{i,j} U_{i,j}$. The transformation U is not a unitary transformation since the transformed Hamiltonian is usually of a lower dimension than the real space one. It is unitary only in the limit where all the transverse modes are chosen and in this case the CMS Hamiltonian is simply a change of representation and is exactly equivalent to the real space Hamiltonian. However, the CMS Hamiltonian with few chosen modes reproduces by construction the exact selected EM subbands and their wavefunctions. It is therefore equal to the full rank EM Hamiltonian on the chosen subspace. Therefore, CMS offers the possibility to reproduce the effect of roughness or ionized impurities if one chooses a sufficient number of modes. The Green's functions in CMS $\tilde{G}^{R,\lessgtr}$ and in real space are related by the same transformation in Eqs. 26, 27 [32]. For instance, the matrix element between the modes i and j both located on either of the layers l and $l' = l, l \pm 1$ is given by:

$$\tilde{G}^{R,\lessgtr}(l, i;l', j;E) = \sum_{y,z} \sum_{y',z'} \phi^*_i(y, z;l) \cdot G^{R,\lessgtr}(l, y, z;l', y', z';E) \cdot \phi_j(y', z';l') \tag{28}$$

5.2.4 Inclusion of electron–phonon interactions

The interaction of the electrons with the acoustic and optical phonons is accounted for in NESS by introducing the self-energies Σ_{ac} and Σ_{op} that read in real space representation under the local approximation [33, 34]:

$$\Sigma^<_{ac,\nu}(r;E) = M_{ac} G^<_\nu(r;E), \tag{29}$$

$$\Sigma^<_{op,\nu}(r;E) = \sum_{q,\nu'} \left| M^{\nu,\nu'}_q \right|^2 \left[n_{B,q} \cdot G^<_{\nu'}(r;E - \hbar\omega_q) + (n_{B,q} + 1) \cdot G^<_{\nu'}(r;E + \hbar\omega_q) \right] \tag{30}$$

$$\Sigma^>_{op,\nu}(r;E) = \sum_{q,\nu'} \left| M^{\nu,\nu'}_q \right|^2 \left[n_{B,q} \cdot G^>_{\nu'}(r;E + \hbar\omega_q) + (n_{B,q} + 1) \cdot G^>_{\nu'}(r;E - \hbar\omega_q) \right] \tag{31}$$

where M_{ac} is the coupling constant to the acoustic phonons, ν is the electronic valley index and q refers to the optical phonon with energy $\hbar\omega_q$. $M^{\nu,\nu'}_q$ is the coupling strength of the electron–phonon interaction due to the phonons of frequency ω_q in the valley ν', whose density is given by Bose–Einstein occupation number $n_{B,q}$. We use the coupling constants obtained from the deformation potential theory [35]:

$$M_{ac} = \frac{\Xi^2 k_B T}{2\rho u_s^2}, \tag{32}$$

$$M^{\nu,\nu'}_q = \frac{\hbar (D_t K_q)^2}{2\rho\omega_q}, \tag{33}$$

where Ξ is the acoustic deformation potential, u_s is the sound velocity in the material, ρ its density and $(D_t K_q)$ the optical deformation potential corresponding to the coupling to the phonons of the valley ν'. The retarded component of the self-energy due to the e–ph interaction is given by:

$$\Sigma^R(r;E) = \frac{1}{2} \left[\Sigma^>(r;E) - \Sigma^<(r;E) \right] \tag{34}$$

Using the same notations as in Eq. 28, the self-energies due to e–ph interaction read in CMS representation [36]:

$$\tilde{\Sigma}^<_{ac}(x, i;x, j;E) = M_{ac} \sum_{k,l} F^{i,j}_{k,l}(x)\, \tilde{G}^<(x, k;x, l;E) \tag{35}$$

$$\tilde{\Sigma}^{<}_{\mathrm{op},\nu}(x,i;x,j;E) = \sum_{k,l} F^{i,j}_{k,l}(x) \sum_{q,\nu'} \left| M^{\nu,\nu'}_{q} \right|^{2} \left[\left(n_{B,q} + \frac{1}{2} \pm \frac{1}{2} \right) \tilde{G}^{<}_{\nu'}\left(x,k;x,l;E \pm \hbar\omega_q\right) \right], \quad (36)$$

$$\tilde{\Sigma}^{>}_{\mathrm{op},\nu}(x,i;x,j;E) = \sum_{k,l} F^{i,j}_{k,l}(x) \sum_{q,\nu'} \left| M^{\nu,\nu'}_{q} \right|^{2} \left[\left(n_{B,q} + \frac{1}{2} \pm \frac{1}{2} \right) \tilde{G}^{>}_{\nu'}\left(x,k;x,l;E \mp \hbar\omega_q\right) \right], \quad (37)$$

where F represents the form factors given by:

$$F^{ij}_{kl}(x) = \int \mathrm{d}y\mathrm{d}z\, \phi^{*}_{i}(y,z;x)\phi_{j}(y,z;x) \phi_{k}(y,z;x)\phi^{*}_{l}(y,z;x) \quad (38)$$

The e–ph retarded self-energy in CMS is also given by Eq. 34 where the real space self energies are replaced by the CMS ones. It is important to note that these formulae are based on an important simplification assuming that the self-energies are local in both space and time. This is a consequence of assuming that acoustic phonons are elastic and that the optical ones are dispersionless, i.e., having a well-defined energies. Despite these simplifications, this treatment of the electron–phonon interaction captures fairly accurately the impact of this scattering mechanism on the operation of real nano devices. Moreover, the deformation potentials can be tuned to get a good quantitative with experimental phonon-limited mobilities [37]. After defining the total retarded and lesser self-energies as follows:

$$\Sigma^{\mathrm{R}} = \Sigma^{\mathrm{R}}_{C} + \Sigma^{\mathrm{R}}_{\mathrm{Scat}} \quad (39)$$

$$\Sigma^{<} = \Sigma^{<}_{C} + \Sigma^{<}_{\mathrm{Scat}} \quad (40)$$

It is clear that Eqs. 15–17 and 34–38 form a non-linear system that needs to be solved self-consistently for a given potential profile. This system is solved in NESS using the Self-Consistent Born Approximation (SCBA) [22, 25].

5.2.5 Parallelization of the NEGF solver

The NEGF solver is written in C++ and is parallelized using Message Passing Interface (MPI) for C language. The main two sections of this solver which are parallelized are the diagonalization of the transverse part of the Hamiltonian in Eq. 25 and the solution of NEGF Eqs. 15–17. Therefore, the 2D Schrodinger equations for the device layers are distributed over all available cores and solved using ARPACK++ library [38], then each core communicates its results to all others before CMS NEGF calculations are started. NEGF calculation are parallelized over energy. The points in the energy grid are distributed evenly on available processors and both the Green function computation and storage is distributed over available cores. One obtains besides the calculation speed up an important decrease in the memory usage per core when the number of cores is increased. All matrix problems—except the eigenvalue problems—are solved using gmm++ library [39]. Point-to-point non-blocking communications using MPI_Issend and MPI_Irecv are used to transfer Green's function data between relevant cores. When electron–phonon interactions are considered, each core needs virtual energy points to receive unavailable $G^{\mathrm{R},\lessgtr}\left(E \pm E_{\mathrm{ph}}\right)$ matrix elements that are needed for self-energy calculations and that are stored on other cores. Assuming dE is the energy discretization step and E^{M}_{ph} is the highest phonon energy, the ratio of the virtual nodes $N_{\mathrm{ph}} = E^{M}_{\mathrm{ph}}/\mathrm{d}E$ to the sum of energy points stored on "neighboring" cores will determine the communication network topology. Based on this information, a matrix T_{ij} is defined where row "i" contains the tags to be used in the MPI_Issend calls by core "i" to send to cores "j". Conversely, core "j" uses tags in the column "j" in its MPI_Irecv calls to match all the MPI_Issend to it. This approach makes sure that deadlocks due to unmatched MPI_Issend/MPI_Irecv calls never happen.

The speed up obtained by this parallelization scheme is reported in Fig. 5 for 3×3 and 5×5 NWFETs with $L_{\mathrm{S}} = L_{\mathrm{D}} = 10\,\mathrm{nm}$ and $L_{\mathrm{G}} = 20\,\mathrm{nm}$. We used 1300 energy points for both devices and 6 subbands for the former and 10 subbands for the latter. The CPU used for this benchmark is an "Intel(R) Xeon(R) E5-2620 v4 @ 2.10 GHz" with 16 physical cores. For both examples a good speed up exceeding 12 is obtained for a parallel run with 16 cores. We have previously ran NESS NEGF calculations on hundreds of cores. However, the obtained speed up wasn't as good as the one reported here because the nodes in the used cluster were heterogeneous and connected with an Ethernet network. We believe that a good speed up can be obtained with hundreds of cores on a cluster having equivalent nodes that are connected with an infiniband network.

5.2.6 Impact of electron–phonon interaction on charge transport

The OFF-state current spectrum in the ballistic limit for a square NWFET having a $3 \times 3\,\mathrm{nm}^2$ cross section and a 10 nm gate length is presented in Fig. 6a. It shows the pseudo-particles propagation without dissipation and current conservation for each energy along the device. Moreover, one can see on the figure that a non-negligible fraction of the carriers injected from the source with energies below the top of the barrier can reach drain. This highlights the importance of considering a quantum formalism to account for the source-to-drain tunneling which is important in transistors

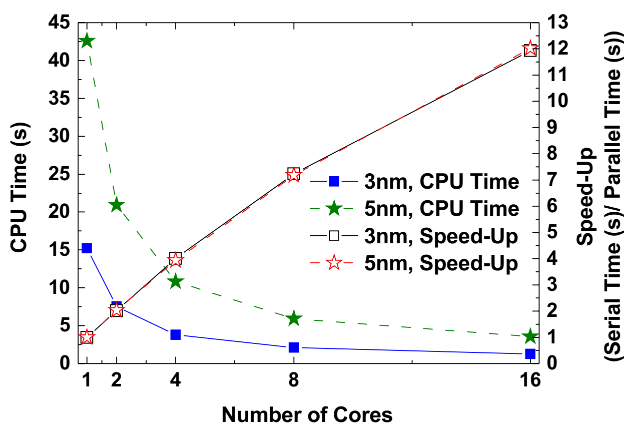


Fig. 5 Simulation time and speed-up for a $5 \times 5\,\mathrm{nm}^2$ square NWFETs with $L_S = L_D = 10\,\mathrm{nm}$ and $L_G = 20\,\mathrm{nm}$

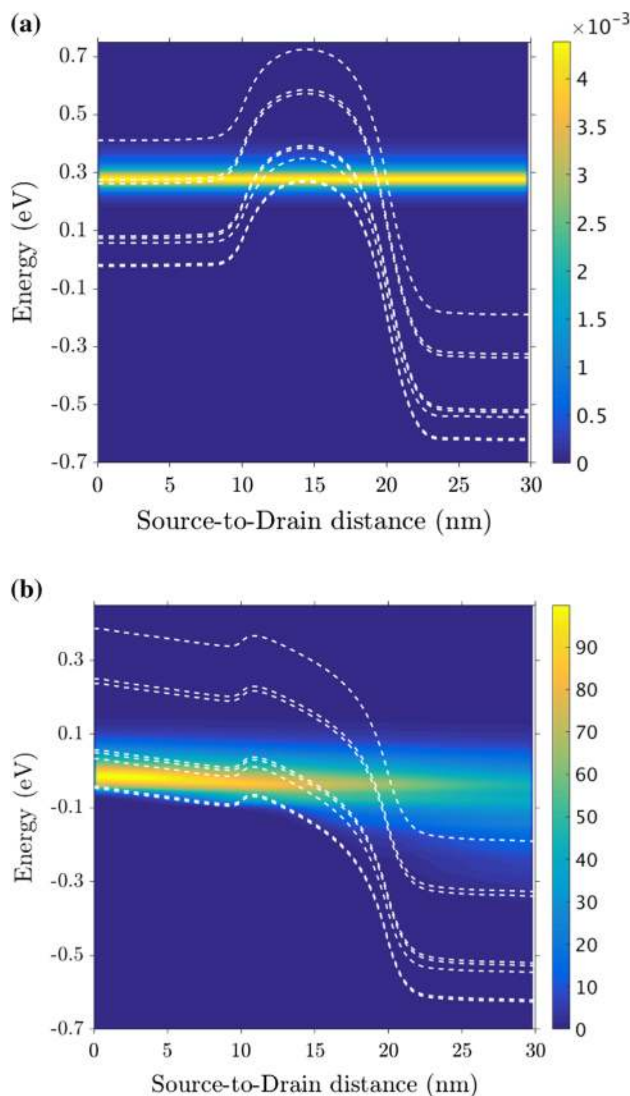


Fig. 6 The current spectra of a $3 \times 3\,\mathrm{nm}^2$ NWFET with $L_G = 10\,\mathrm{nm}$ in **a** the OFF-sate in the ballistic limit, **b** the ON-State with e–ph scattering. The reference in energy is taken at the Source Fermi level ($E_{FS} = 0\,\mathrm{eV}$) and $V_{DS} = 0.6\,\mathrm{V}$

with sub-20 nm gate lengths. The ON-state current spectrum for the same device in the presence of e–ph interaction is shown in Fig. 6b. It shows an overall current damping due to acoustic phonons and an energy relaxation of carriers as they approach the drain due to optical phonons emission. However, $J(x)$, which is obtained by integrating the current spectrum over the energy and the transverse coordinates, is still a flat function of x, the position along the channel. Figure 7 shows a comparison of the $I_D - V_{GS}$ curves for the same $3 \times 3\,\mathrm{nm}^2$ square NWFET with and without e–ph interactions considering different gate lengths. It shows a progressive reduction of the current as the gate bias increased when e–ph interactions are taken into account, reaching 48% for $L_G = 10\,\mathrm{nm}$ at $V_{DS} = 0.6\,\mathrm{V}$ and nearly 65% for $L_G = 20\,\mathrm{nm}$. It is also noticeable that both types of currents have higher values for $L_G = 10\,\mathrm{nm}$ compared to their $L_G = 20\,\mathrm{nm}$ counterparts. These observations are consistent with results previously reported in literature [22, 33].

5.2.7 Assessing of confinement and short channel effects

Shrinking the channel length to few decanonometers leads to a degradation of the electrostatic control of the gate over the electron transport in the channel [40]. The use of gate-all-around (GAA) architecture with appropriate NW cross section reduces significantly these short channel effects (SCE) [41]. However, this effect cannot be fully suppressed for sub-20 nm gate length devices because of the impact of direct source-to-drain tunneling occurring at such short channel lengths. Moreover, the effective mass of carriers in ultra-confined NW depends strongly on the NW cross section shape and size [42]. Therefore, the design rules must be established using a quantum simulator such as the NEGF solver of NESS, which can capture accurately these effects.

We have extracted in Fig. 8a, b the Drain Induced Barrier Lowering (DIBL) and Subthreshold slope (S_{th}), respectively, for different nanowire shapes at different gate lengths. The effective masses were calibrated using TB band structure to take into account their dependence on the confinement [42]. As expected, the best electrostatic control is obtained for the NWFET with the narrowest Si cross section, i.e., the circular NWFET with 3 nm diameter, followed by the square ones with 3 nm^2 and 5 nm^2 cross sections, respectively.

Moreover, all three NWFETs show rapid degradation of both the DIBL and S_{th} when the gate length is shrunk below 10 nm. However, the $I_D - V_G$ in Fig. 9 shows that improving the electrostatic control by reducing the NW cross section decreases the drive current. The use of stacked nanowires is a contemplated solution to this problem [41], and simulation tools like NESS could help the designers to co-optimize the

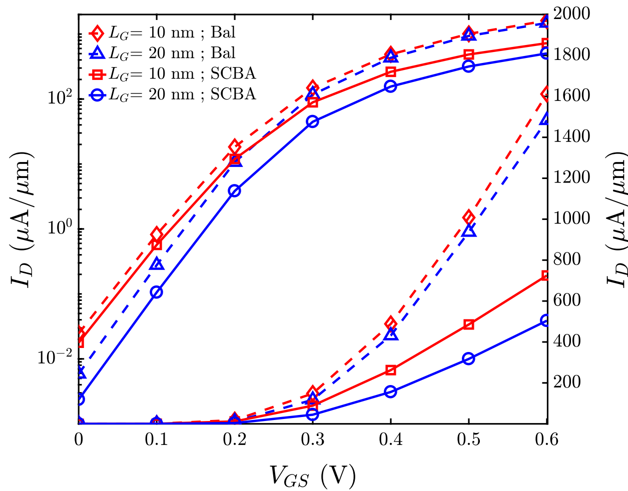
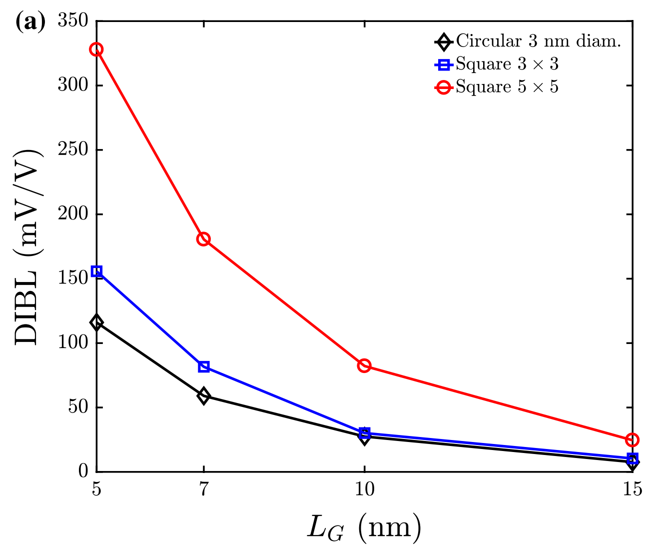


Fig. 7 I_D–V_{GS} characteristic for $3 \times 3\,\text{nm}^2$ square nanowire with and without phonons for different gate lengths



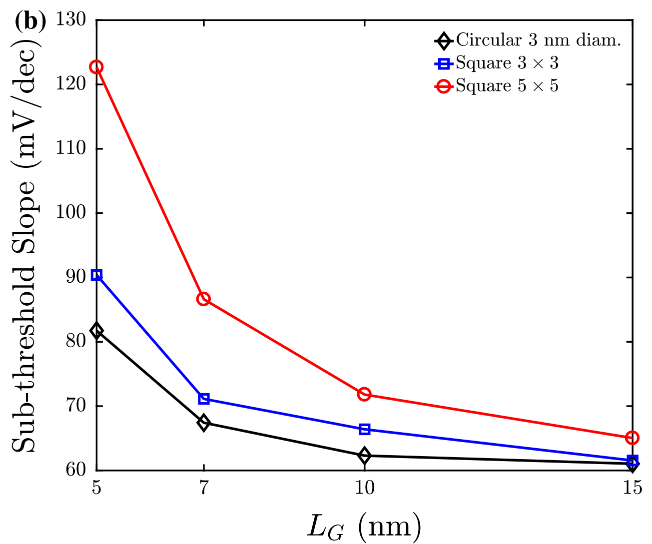


Fig. 8 **a** Drain induced barrier lowering and **b** Sub-threshold slope for 3 different gate lengths for a circular NWFET with 3 nm diameter and square ones with 3×3 and $5 \times 5\,\text{nm}^2$ cross sections

gate length, the cross section and the number of stacked NW they need to meet their performance targets.

5.2.8 Direct band-to-band tunneling model

In this section, we briefly summarize the novel procedure implemented in NESS to compute the direct band-to-band tunneling in nano-devices [43]. It is based on the coupled mode-space NEGF scheme within the EM approximation and the Flietner model of the imaginary dispersion [44, 45]. The results obtained by the BTBT model in NESS [46] show an excellent agreement with results obtained from the atomistic simulation tool OMEN [47]. The valence and conduction band edges are connected using the two-band model of the imaginary dispersion proposed by Flietner in Ref. [44]. For quantum transport simulations, the Flietner model can be rewritten as [45]:

$$\frac{\hbar^2 k^2}{2m_0} = \frac{E_g(E-E_c)(E-E_v)}{\left[\sqrt{\frac{m_0}{m_c}}(E-E_c) - \sqrt{\frac{m_0}{m_v}}(E-E_v)\right]^2} \tag{41}$$

where E_g, $E_{c(v)}$, and $m_{c(v)}$ are the band gap energy, the conduction (valence) band edge, and the conduction (valence) effective mass, respectively. The rest of the parameters take their usual meaning. Both the real conduction and valence bands in the vicinity of their extrema are correctly reproduced, and the parabolic effective mass approximation of the band-structure can be straightforwardly obtained.

Moreover, Eq. 41 allows the inclusion of an external potential $V(\mathbf{r})$ in the bands dispersion straightforwardly:

$$E \approx E_{c(v)} + V(\mathbf{r}) \pm \frac{\hbar^2 k^2}{2m_{c(v)}} + \cdots . \tag{42}$$

This sets up the appropriate envelope equation for low-dimensional semiconductors that incorporates both real and imaginary branches of the whole band structure.

The quantum transport problem for electrons and holes is then solved independently within the EM approximation using the NEGF technique in CMS representation and coupled self-consistently with the Poisson equation. Once the convergence is reached, the Valence Band (VB) and Conduction Bands (CB) are bridged through the two-band model of the imaginary dispersion proposed by Flietner and the BTBT current is computed by solving the following envelope equation:

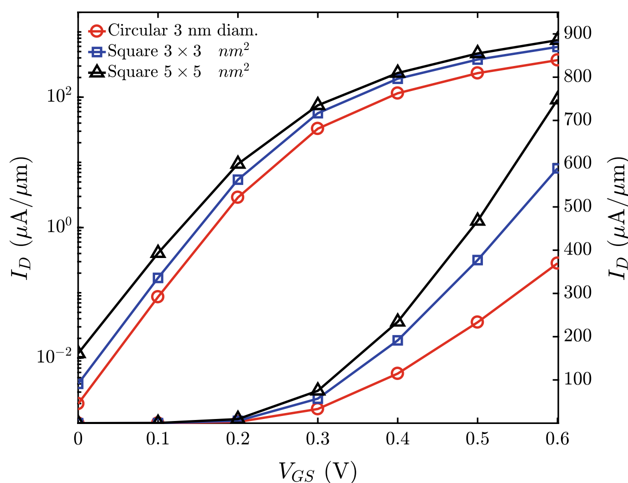


Fig. 9 $I_D - V_{GS}$ characteristic for 3 nm diameter circular NWFET and $3 \times 3\,\mathrm{nm}^2$ and $5 \times 5\,\mathrm{nm}^2$ square NWFETs for $L_G = 15\,\mathrm{nm}$

$$
-\frac{\hbar^2}{2m_0}\frac{\partial^2\chi}{\partial x^2} = \frac{E_g(E-V_c)(E-U_v)}{\left[\sqrt{\frac{m_0}{m_c}}(E-V_c)-\sqrt{\frac{m_0}{m_v}}(E-V_v)\right]^2}\chi
= F(E, U_c, V_v)\chi, \tag{43}
$$

with open boundary conditions. In the latter, $U_{c(v)}$ corresponds to the lowest conduction (highest valence) subband energy, and the coordinate x has been omitted for brevity. Finally, by defining a two-band Hamiltonian as:

$$
H = -\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial x^2} - F(E, U_c, U_v) + E, \tag{44}
$$

Equation 43 can be solved to calculate the BTBT current in nanowire transistors by means of a NEGF scheme. Figure 10b shows the current spectrum of the Tunnel Field Effect Transistors (TFET) depicted in Fig. 10a. The dashed lines represent the lowest and highest subbands of CB and VB, respectively. One can see clearly that the current flows from the VB of Si to the CB of the InAs. A detailed study of this problem has been published in Ref. [43], in which band non-parabolicity has been included for the conduction band of InAs.

5.2.9 Variability in quantum mechanical context

Variability is one of the main challenges facing the downscaling of CMOS devices. It is induced either by the fabrication process which produces variability sources such as the LER, or by statistical variability introduced by the discreteness of charge or granularity of matter as exemplified by random dopant fluctuations. The impact of these variability sources must be assessed carefully to maximize the yield.

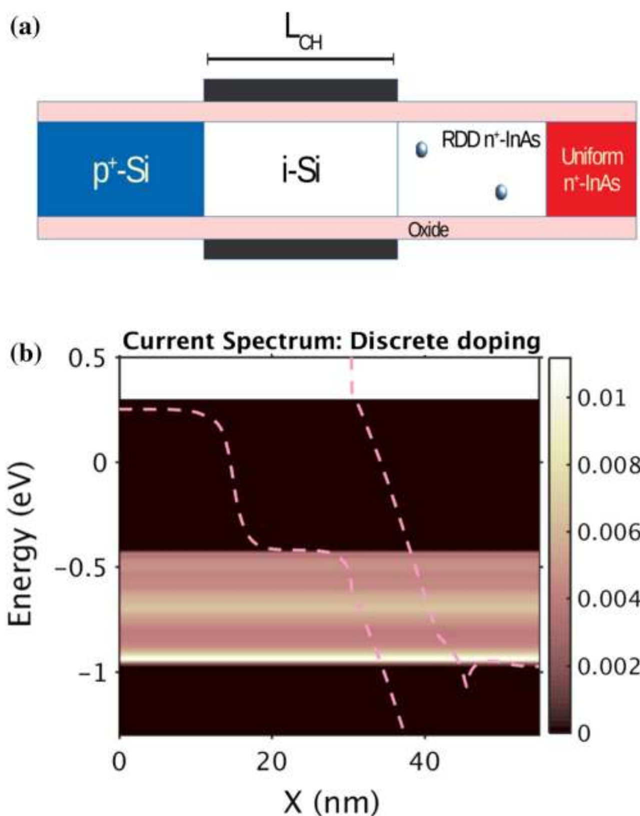


Fig. 10 **a** A schematic representation of the TFETs considered in this study, **b** the corresponding current spectrum in the ON-state in presence of random discrete dopants in the InAs section

As shown in Sect. 3, NESS comes with a powerful structure generator that enables the generation of LER, RDD and MGG. Moreover, the NEGF solver of NESS is parallelized using the Message Passing Interface (MPI), thus enabling simulations on hundreds of processing cores. Also, the Poisson solver coupled to the NEGF solver is based on a robust finite volume discretization and an efficient implementation of the self-consistency using auxiliary quasi-Fermi levels and corresponding Gummel's iteration.

All these optimizations enabled us to perform quantum mechanical variability studies employing large statistical samples. For instance, we studied the impact of dopant diffusion in the channel of Si-NWFET in Ref. [48] and we performed in Ref. [49] a simulation study of all variability sources in Si_XGe_{1-X} employing 10,000 samples. NESS can also help assessing the viability of novel transistor architectures for future nodes. For example, a study of RDD induced variability in JunctionLess Field Effect Transistors (JLFET) and TFETs confirmed that the yield would be too low for those devices to be considered for digital applications [43, 50].

5.3 The Kubo–Greenwood solver

5.3.1 Features of the KG solver

The Kubo–Greenwood (KG) solver has been implemented in NESS for the calculation of the low-field electron mobility [51, 52]. This semi-classical approach combines the quantum effects based on the 1D multi-subband scattering rates of the most relevant scattering mechanisms in NWFETs [53] and the semi-classical Boltzmann transport equation (BTE) by applying the Kubo–Greenwood formula within the relaxation time approximation [54, 55]. Moreover, it is possible to make use of the effective masses calculated by the NESS EM extractor (Sect. 4). As this solver is based on the long-channel approximation, the first step is to pre-calculate the required subband levels (E_l) and the corresponding wavefunctions (ξ_l) using a self-consistent Poisson–Schrödinger simulation in the presence of a low electric field in the transport direction. The second step is to use these quantities to compute the scattering rates whose equations are directly derived from the Fermi golden rule. The following scattering mechanisms have been implemented in NESS: (1) acoustic phonon scattering; (2) optical phonon scattering, including g-type and f-type transitions; and (3) surface roughness scattering.

In this paper, we performed a comparison of the KG mobility with the one computed by NEGF method considering acoustic phonon and g-type (intra-valley transitions) optical phonon scattering mechanisms. More details of the all the aforementioned mechanisms as well as their equations can be found in [53]. The scattering rate for the acoustic phonon mechanism has been considered to be elastic and within the short wave vector limit. Its equivalent equation from a initial subband l and a final subband l' is:

$$\Gamma_{\mathrm{Ac}}(l,k) = \frac{|D_{\mathrm{Ac}}|^2 k_{\mathrm{B}} T}{\rho \hbar u_s^2} \frac{m}{\hbar^2} \sum_{l'} \left[\int \mathrm{d}\mathbf{s} |\xi_l(\mathbf{s})|^2 |\xi_{l'}(\mathbf{s})|^2 \right] \times \theta(\epsilon(k) + \Delta E_{l'}) \left(\frac{1}{|q_1 + k|} + \frac{1}{|q_2 + k|} \right), \tag{45}$$

where ξ_l is the wavefunction of subband l, $q_{1/2} = -k \pm \sqrt{k^2 + \frac{\Delta E_{l'} 2m}{\hbar^2}}$, D_{Ac} is the acoustic deformation potential ($D_{\mathrm{Ac}} = 14.5\,\mathrm{eV}$ in this section), m is the electron effective mass in the transport direction, $\mathbf{s}$ are vectors normal to the transport direction, θ represents the step function, and $\epsilon(k)$ is the kinetic energy for a wave vector magnitude k. The rest of the parameters have their usual meaning.

We have considered fixed energy and deformation potential for the optical phonon scattering mechanism. Accordingly, the scattering rate for the g-type transitions is expressed as:

$$\Gamma_{Op}(j,l,k) = \frac{|D_{Op,j}|^2}{2\rho\omega_j} \sum_{l'} \left[\int \mathrm{d}\mathbf{s} |\xi_l(\mathbf{s})|^2 |\xi_{l'}(\mathbf{s})|^2 \right] \times \left[\frac{n_j \theta(\epsilon(k) + \Delta E^{+}_{l'j}) m}{\hbar^2} \left(\frac{1}{|q_1 + k|} + \frac{1}{|q_2 + k|} \right) + \frac{(n_j+1)\theta(\epsilon(k) + \Delta E^{-}_{l'j}) m}{\hbar^2} \left(\frac{1}{|q_3 + k|} + \frac{1}{|q_4 + k|} \right) \right] \tag{46}$$

where $q_{1/2} = -k \pm \sqrt{k^2 + \frac{\Delta E^{+}_{l'j} 2m}{\hbar^2}}$, $q_{3/4} = -k \pm \sqrt{k^2 + \frac{\Delta E^{-}_{l'j} 2m}{\hbar^2}}$, $\Delta E^{\pm}_{l'j} = E_l - E_{l'} \pm \hbar\omega_j$, n_j is the equilibrium phonon number, j refers to the phonon mode, and ω_j is the phonon energy.

Then, we present two strategies to calculate the total mobility. In the first one, the mobility associated with each particular scattering mechanism is calculated using its rate in the KG formula. The total mobility is then calculated as a function of the individual mobilities associated with each scattering mechanism using the Matthiessen rule [56]. In the second alternative, the scattering rate of both mechanisms are directly added to avoid the Matthiessen rule and thereby the total mobility for each subband is computed using the KG formalism. In general, the advantage of both semi-classical alternatives in comparison to purely quantum transport simulations is that the rates are individually computed and then combined, hence reducing dramatically the computational cost.

5.3.2 Comparison of KG and NEGF mobilities

The NEGF mobilities have been extracted using the formula [57]:

$$\mu = \left(\frac{\mathrm{d}R}{\mathrm{d}L} \right)^{-1} \frac{1}{|q|\rho_{1\mathrm{D}}}, \tag{47}$$

where q is the electron charge, $\rho_{1\mathrm{D}}$ is the 1D charge density along the NW transport direction, L is its channel length, and R is its resistance which is extracted by calculating the voltage to the current ratio. For this approximation to be valid, one must apply a very small bias of only few mV (2 mV in this section) and consider long channels to compute the resistance in the diffusive regime. We used 45 and 50 nm channel lengths to compute $\mathrm{d}R/\mathrm{d}L$. The results of the comparison for a $3 \times 3\,\mathrm{nm}^2$ square NW are shown in Fig. 11a, b. Both figures show a good agreement between the NEGF and KG mobilities for both acoustic and g-type optical phonons. The f-type optical phonons have not been implemented yet in the NEGF module of NESS and will probably be included soon in a future release. It's interesting to note from Fig. 11b that the correct phonon-limited mobility given by the curve with triangles cannot be obtained in this case by extracting the acoustic and optical phonon-limited mobility then using

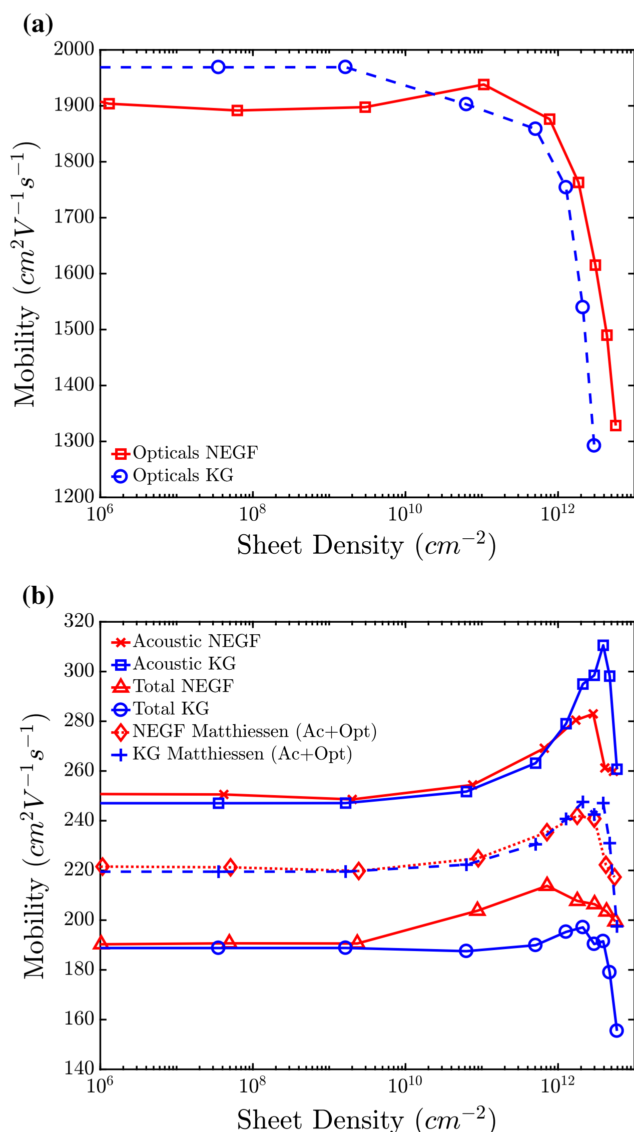


Fig. 11 Comparison of the phonon-limited mobility for a $3 \times 3\,\mathrm{nm}^2$ NWFET computed using NEGF and KG solvers for **a** the g-type Optical phonons and **b** the acoustic and acoustic+g-type optical phonons

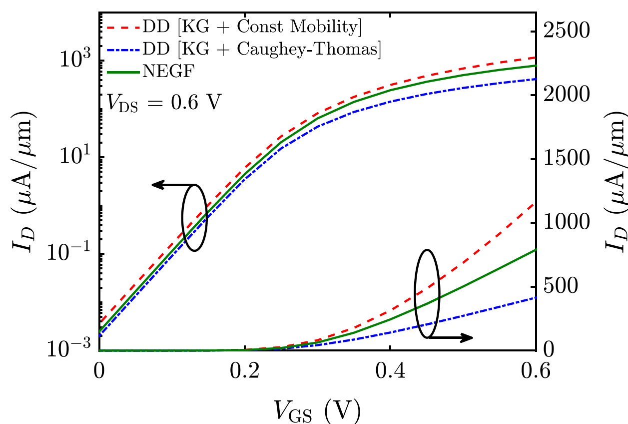


Fig. 12 I_D–V_{GS} characteristics for a $3 \times 3\,\mathrm{nm}^2$ square NWFET with $L_G = 20\,\mathrm{nm}$ obtained using NEGF with acoustic g-type e–ph interactions and classical DD with constant mobility and Caughey–Thomas model. The reference value of the mobility was obtained using the KG module. We used $V_{DS} = 0.6\,\mathrm{V}$

Matthiessen rule. This is an indication of the importance of the interference of the two scattering mechanisms considered here. It is important to highlight that in this case, KG formalism reproduces the NEGF mobility when the scattering rates are summed up rather than using Matthiessen rule. This might be an indication that even for confined devices, KG suffices to calibrate the long channel mobilities for drift–diffusion models rather that running the more expensive NEGF simulations.

5.3.3 Comparison of NEGF and DD+KG results

We present in Fig. 12 the comparison of the I_D–V_{GS} characteristic obtained using NEGF solver with the transfer characteristic obtained using DD solver for a $3 \times 3\,\mathrm{nm}^2$ square NWFET with $L_G = 20\,\mathrm{nm}$. We have used in DD a constant mobility value of $190\,\mathrm{cm}^2/\mathrm{Vs}$, which is the value extracted from Fig. 11b for the sheet densities corresponding to the applied gate bias Fig. 11b. In the OFF-state, the logarithmic plot shows a good agreement between the NEGF characteristic and the DD one with Caughey–Thomas model, and a negative V_{th} shift of DD curve with constant mobility with respect to both aforementioned curve. However, the subthreshold slopes obtained for all models are very similar. In the ON-state, the linear plots V_{th} shows that the NEGF transfer characteristic is positioned between the DD one with constant mobility and the one with Caughey–Thomas mobility degradation model.

This discrepancy in the transfer characteristics stems from the lack of quantum corrections in our DD simulations, which are crucial in the operation of a device with such a narrow cross section. Indeed, while NEGF captures the volume inversion as shown in Fig. 13a—i.e., most of the charge is located at the center of the device, the charge from the DD solution shown in Fig. 13b is mainly located at the edges and corner of the silicon body of the NWFET. This is an important shortcoming of DD that leads to inaccurate electrostatics for ultra-scaled NWFET and translates in a wrong estimate of the current. However, if quantum corrections are used in combination with mobility models to calibrate DD then a good match to NEGF results can be obtained [58]. The DD with constant mobility model and the NEGF self-consistent potential profiles in the center of

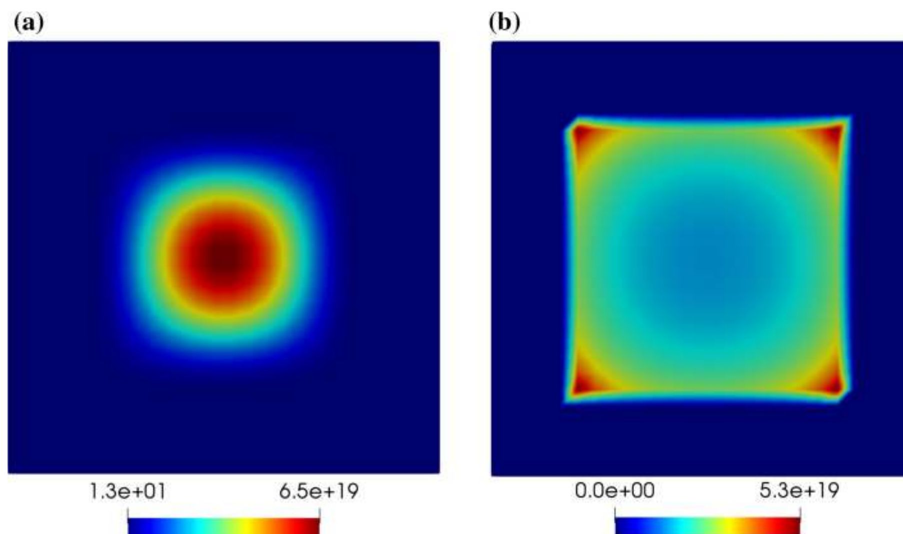


Fig. 13 Electron density profile (cm^{-3}) across a transverse cross section of a 3 nm × 3 nm square NWFET at the center of its 20 nm long channel from **a** NEGF, and **b** classical DD solvers at $V_{GS} = 0.4$ V and $V_{DS} = 0.6$ V

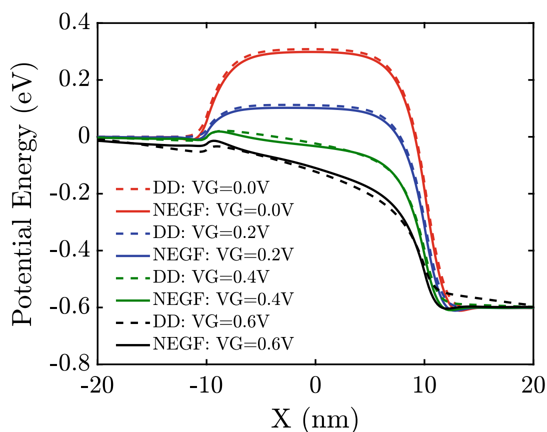


Fig. 14 Comparison of the potential profiles along the transport direction obtained from DD and NEGF self-consistent simulations at different gate biases for a $3 \times 3\,nm^2$ square NWFET with $L_G = 20$ nm at $V_{DS} = 0.6$ V. The potential was sampled at the center of each cross section.

the same device are shown in Fig. 14. Both types of potentials are quite similar, especially in the OFF-state where the NEGF simulation is usually started. Therefore, DD potential profiles are good trial potentials for NEGF self-consistent simulations despite the discrepancy in the transfer characteristics, and this can speed up considerably the NEGF solution for the first bias point.

6 Conclusion

In this paper we have presented NESS, a flexible nano-electronic device simulator under development in the Device Modeling Group of Glasgow university, and have described in details its main two modules. The first module is its structure generator that enables the generation of semiconducting devices with different architectures and can introduce the relevant sources of statistical variability in the corresponding solution domains. The second module contains the transport solvers that have been implemented so far: Drift–Diffusion, Kubo–Greenwood and non-equilibrium Green's function. All these solvers share the same simulation domain, making NESS one of the few nanodevice simulation tools that offers the possibility to compare different models for the same device and to assess their strengths and shortcomings when simulating device characteristics and extracting particular figures of merit. The results reported herein and in other cited references show that at low applied biases the NEGF and KG solvers of NESS are in good agreement with all the main results reported in the literature for NWFETs. Moreover, the MPI optimizations for the NEGF solver and the robustness of our finite volume non-linear Poisson solver enabled quantum statistical variability studies employing large statistical samples. Additionally, NESS is modular and easily extensible. NESS will be released in the summer of 2020 as an open-source software which makes it very interesting for both academia and industry in helping to address the challenges subsequent to the further down-scaling of CMOS components.

Acknowledgements This work was supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 688101 SUPERAID7. Also, this project has received funding from EPSRC UKRI under Grant Agreements No. EP/S001131/1 (QSEE) and No. EP/P009972/1 (QUANTDEVMOD).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Kuhn, K.: Chapter 1—CMOS and Beyond CMOS: Scaling Challenges: High Mobility Materials for CMOS Application, pp. 1–44. Woodhead Publishing, Sawston (2018)
2. Bohr, M.: The evolution of scaling from the homogeneous era to the heterogeneous era. In: 2011 International Electron Devices Meeting, pp. 1.1.1–1.1.6 (2011)
3. Jiang, X., Guo, S., Wang, R., Wang, Y., Wang, X., Cheng, B., Asenov, A., Huang, R.: New insights into the near-threshold design in nanoscale FinFET technology for sub-0.2V applications. In: 2016 IEEE International Electron Devices Meeting (IEDM), pp. 28.4.1–28.4.4 (2016)
4. Jurczak, M., Collaert, N., Veloso, A., Hoffmann, H, Biesemans, S.: Review of FINFET technology. In: 2009 IEEE International SOI Conference, pp. 1–4 (2009)
5. Auth, C. et al.: A 10 nm high performance and low-power CMOS technology featuring 3rd generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects. In: 2017 IEEE International Electron Devices Meeting (IEDM), pp. 29.1.1–29.1.4 (2017)
6. Ha, D. et al.: Highly manufacturable 7nm FinFET technology featuring EUV lithography for low power and high performance applications. In: 2017 Symposium on VLSI Technology, pp. T68–T69 (2017)
7. Xie, R. et al.: A 7nm FinFET technology featuring EUV patterning and dual strained high mobility channels. In: 2016 IEEE International Electron Devices Meeting (IEDM), pp. 2.7.1–2.7.4 (2016)
8. Bae, G. et al.: 3nm GAA Technology featuring multi-bridge-channel FET for low power and high performance applications. In: 2018 IEEE International Electron Devices Meeting (IEDM), pp. 28.7.1–28.7.4 (2018)
9. Lorenz, J., et al.: Process variability-technological challenge and design issue for nanoscale devices. Micromachines **10**(1), 6 (2019)
10. Frank, D., et al.: Monte Carlo modeling of threshold variation due to dopant fluctuations. In: Symposium on VLSI Technology. Digest of Technical Papers, pp. 169–170 (1999)
11. Kim, S., Luisier, M., Paul, A., Boykin, T.B., Klimeck, G.: Full three-dimensional quantum transport simulation of atomistic interface roughness in silicon nanowire FETs. IEEE Trans. Electron Devices **58**, 1371–1380 (2011)
12. Wang, X., Brown, A.R., Idris, N., Markov, S., Roy, G., Asenov, A.: Statistical threshold-voltage variability in scaled decananometer bulk HKMG MOSFETs: a full-scale 3-D simulation scaling stud. IEEE Trans. Electron Devices **58**(8), 2293–2301 (2011)
13. Boykin, T.B., et al.: Valence band effective-mass expressions in the $sp^3d^5s^*$ empirical tight-binding model applied to a Si and Ge parametrization. Phys. Rev. B **69**, 115201 (2004)
14. QuantumATK version O-2018.06, Synopsys QuantumATK. https://www.synopsys.com/silicon/quantumatk.html. Accessed 2019
15. Badami, O., et al.: Comprehensive study of cross-section dependent effective masses for silicon based gate-all-around transistors. Appl. Sci. **9**(9), 1895 (2019)
16. Scharfetter, D.L., Gummel, H.K.: Large-signal analysis of a silicon read diode oscillator. IEEE Trans. Electron Devices **16**(1), 64–77 (1969)
17. Selberherr, S.: Analysis and Simulation of Semiconductor Devices. Springer, Berlin (2012)
18. Masetti, G., Severi, M., Solmi, S.: Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon. IEEE Trans. Electron Devices **30**(7), 764–769 (1983)
19. Lombardi, C., Manzini, S., Saporito, A., Vanzi, M.: A physically based mobility model for numerical simulation of nonplanar devices. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **7**(11), 1164–1171 (1988)
20. Yamaguchi, K.: Field-dependent mobility model for two-dimensional numerical analysis of MOSFET's. IEEE Trans. Electron Devices **26**(7), 1068–1074 (1979)
21. Caughey, D.M., Thomas, R.E.: Carrier mobilities in silicon empirically related to doping and field. Proc. IEEE **55**(12), 2192–2193 (1967)
22. Luisier, M., Klimeck, G.: Atomistic full-band simulations of silicon nanowire transistors: effects of electron–phonon scattering. Phys. Rev. B **80**, 155430 (2009)
23. Mahan, G.: Many-Particle Physics. Springer, Berlin (2000). https://doi.org/10.1007/978-1-4757-5714-9
24. Stefanucci, G., Van Leeuwen, R.: Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction. Cambridge University Press, Cambridge (2013). https://doi.org/10.1017/CBO9781139023979
25. Jin, S., Park, Y.J., Min, H.S.: A three-dimensional simulation of quantum transport in silicon nanowire transistor in the presence of electron–phonon interactions. J. Appl. Phys. **99**(12), 123719 (2006)
26. Martin, P., Schwinger, J.: Theory of many-particle systems. I. Phys. Rev. **115**(6), 1342–1373 (1959)
27. Kadanoff, L.P., Baym, G.: Quantum Statistical Mechanics. W. A. Benjamin, Inc., New York (1962)
28. Svizhenko, A., Anantram, M.P., Govindan, T.R., et al.: Two-dimensional quantum mechanical modeling of nanotransistors. J. Appl. Phys. **91**(4), 2343–2354 (2002)
29. Anantram, M.P.: Modeling of nanoscale devices. Proc. IEEE **96**, 1511–1550 (2008)
30. Datta, S.: Nanoscale device modeling: the Green's function method. Superlattices Microstruct. **28**(4), 253–278 (2000)
31. Lopez Sancho, M.P., Lopez Sancho, J.M., Rubio, J.: Highly convergent schemes for the calculation of bulk and surface Green's functions. J. Phys. F Met. Phys. **15**, 851 (1985)
32. Luisier, M., Schenk, A., Fichtner, W.: Quantum transport in two- and three-dimensional nanoscale transistors: coupled mode effects in the nonequilibrium Green's function formalism. J. Appl. Phys. **100**(4), 043713 (2006)
33. Aldegunde, M., Martinez, A., Asenov, A.: Non-equilibrium Green's function analysis of cross section and channel length dependence of phonon scattering and its impact on the performance of Si nanowire field effect transistors. J. Appl. Phys. **110**(9), 094518 (2011)
34. Bescond, M., Li, C., Mera, H., et al.: Modeling of phonon scattering in n-type nanowire transistors using one-shot analytic continuation technique. J. Appl. Phys. **114**(15), 153712 (2013)
35. Jacoboni, C., Reggiani, L.: The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. Rev. Mod. Phys. **55**, 645–705 (1983)
36. Esposito, A., Martin, F., Schenk, A.: Quantum transport including nonparabolicity and phonon scattering: application to silicon nanowires. J. Comput. Electron. **8**(3), 336 (2009)
37. Niquet, Y.-M., Nguyen, V.-H., Triozon, F., et al.: Quantum calculations of the carrier mobility: methodology, Matthiessen's rule, and comparison with semi-classical approaches. J. Appl. Phys. **115**(5), 054512 (2014)
38. https://www.caam.rice.edu/software/ARPACK/arpack++.html

39. http://getfem.org/gmm/
40. Colinge, J.-P.: Multiple-gate SOI MOSFETs. Solid-State Electron. **48**, 897 (2004)
41. Barraud, S., Lapras V., Samson, M.P., et al.: Vertically stacked-NanoWires MOSFETs in a replacement metal gate process with inner spacer and SiGe source/drain. In: 2016 IEEE International Electron Devices Meeting (IEDM), pp. 17.6.1–17.6.4 (2016)
42. Badami, O., Medina-Bailon, C., Berrada, S., et al.: Comprehensive study of cross-section dependent effective masses for silicon based gate-all-around transistors. Appl. Sci. **9**(9), 1895 (2019)
43. Carrillo-Nuñez, H., Lee, J., Berrada, S., et al.: Random dopant-induced variability in Si–InAs nanowire tunnel FETs: a quantum transport simulation study. IEEE Electron Device Lett. **39**(9), 1473–1476 (2018)
44. Flietner, H.: The $E(k)$ relation for a two-band scheme of semiconductors and the application to the metal-semiconductor contact. Phys. Status Solidi B **54**(1), 201–208 (1972)
45. Carrillo-Nuñez, H., Ziegler, A., Luisier, M., et al.: Modeling direct band-to-band tunneling: from bulk to quantum-confined semiconductor devices. J. Appl. Phys. **117**(23), 234501 (2015)
46. Carrillo-Nuñez, H., Lee, J., Berrada, S., et al.: Efficient two-band based non-equilibrium Green's function scheme for modeling tunneling nano-devices. In: 2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 141–144 (2018)
47. Luisier, M., Schenk, A., Fichtner, W., et al.: Atomistic simulation of nanowires in the $sp^3d^5s^*$ tight-binding formalism: from boundary conditions to strain calculations. Phys. Rev. B **74**(20), 205323 (2006)
48. Lee, J., Berrada, S., Carrillo-Nuñez, H., et al.: The impact of dopant diffusion on random dopant fluctuation in Si nanowire FETs: a quantum transport study. In: 2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 280–283 (2018)
49. Lee, J., Badami, O., Carrillo-Nuñez, H., et al.: Variability predictions for the next technology generations of n-type Si_xGe_{1-x} nanowire MOSFETs. Micromachines **9**(12), 643 (2018)
50. Berrada, S., Lee, J., Carrillo-Nuñez H. et al.: Quantum transport investigation of threshold voltage variability in sub-10 nm JunctionlessSi Nanowire FETs. In: 2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 244–247 (2018)
51. Medina-Bailon, C., Sadi, T., Nedjalkov, M., Lee, J., Berrada, S., Carrillo-Nuez, H., Georgiev, V. P., Selverherr, S., Asenov, A.: Study of the 1D scattering mechanisms' impact on the mobility in Si nanowire transistors. In: 2018 Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS), pp. 1–4 (2018)
52. Medina-Bailon, C., Sadi, T., Nedjalkov, M., Lee, J., Berrada, S., Carrillo-Nuez, H., Georgiev, V. P., Selverherr, S., Asenov, A.: Impact of the effective mass on the mobility in Si nanowire transistors. In: 2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 297–300 (2018)
53. Sadi, T., Medina-Bailon, C., Nedjalkov, M., Lee, J., Badami, O., Berrada, S., Carrillo-Nuez, H., Georgiev, V.P., Selverherr, S., Asenov, A.: Simulation of the impact of ionized impurity scattering on the total mobility in Si nanowire transistors. Materials **12**(1), 124 (2019)
54. Ferry, D., Jacobini, C.: Quantum Transport in Semiconductors. Springer Science, Berlin (1992)
55. Jin, S., Tang, T.W., Fischetti, M.V.: Simulation of silicon nanowire transistors using Bolzmann transport equation under relaxation time approximation. IEEE Trans. Electron Devices **55**(3), 727–736 (2008)
56. Esseni, D., Driussi, F.: A quantitative error analysis of the mobility extraction according to Matthiessen rule in advanced MOS transistors. IEEE Trans. Electron Devices **58**(8), 2415–2422 (2011)
57. Luisier, M.: Phonon-limited and effective low-field mobility in n- and p-type [100]-, [110]-, and [111]-oriented Si nanowire transistors. Appl. Phys. Lett. **98**, 032111 (2011)
58. Brown, A., et al.: Use of density gradient quantum corrections in the simulation of statistical variability in MOSFETs. J. Comput. Electron. **9**(3), 187–196 (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.