


# Full-band quantum simulation of electron devices with the pseudopotential method: Theory, implementation, and applications

M. G. Pala<sup>1</sup> and D. Esseni<sup>2</sup>

<sup>1</sup>*Centre de Nanosciences et de Nanotechnologies, Centre National de la Recherche Scientifique, Université Paris-Sud, Université Paris-Saclay, F-91405 Orsay, France*

<sup>2</sup>*DPIA, University of Udine, Via delle Scienze 206, I-33100 Udine, Italy*

 (Received 21 November 2017; revised manuscript received 16 January 2018; published 28 March 2018)

This paper presents the theory, implementation, and application of a quantum transport modeling approach based on the nonequilibrium Green's function formalism and a full-band empirical pseudopotential Hamiltonian. We here propose to employ a hybrid real-space/plane-wave basis that results in a significant reduction of the computational complexity compared to a full plane-wave basis. To this purpose, we provide a theoretical formulation in the hybrid basis of the quantum confinement, the self-energies of the leads, and the coupling between the device and the leads. After discussing the theory and the implementation of the new simulation methodology, we report results for complete, self-consistent simulations of different electron devices, including a silicon Esaki diode, a thin-body silicon field effect transistor (FET), and a germanium tunnel FET. The simulated transistors have technologically relevant geometrical features with a semiconductor film thickness of about 4 nm and a channel length ranging from 10 to 17 nm. We believe that the newly proposed formalism may find applications also in transport models based on *ab initio* Hamiltonians, as those employed in density functional theory methods.

DOI: [10.1103/PhysRevB.97.125310](https://doi.org/10.1103/PhysRevB.97.125310)

## I. INTRODUCTION

In nanoelectronic and nanophotonic devices semiconductor materials are frequently structured at truly nanometric dimensions. While early examples were related to high electron mobility transistors (HEMTs) employing quantum wells of III-V semiconductors, more recently also mainstream complementary metal oxide semiconductor (CMOS) technologies have introduced field effect transistors (FETs) realized in an ultrathin semiconductor film on insulator (UTB-SOI), and then Fin-FETs and multigate FETs (MuGFETs), as well as gate-all-around (GAA) nanowire transistors [1]. CMOS technologies are also exploring the use of III-V semiconductors, such as InGaAs [2–4], whose light conduction band effective masses emphasize quantum effects. The relevance of quantum effects in CMOS transistors is no longer restricted to subband splitting due to carrier confinement, in fact genuinely quantum transport phenomena have become important, such as the undesired source-drain tunneling in MOSFETs [5–7], or the band-to-band tunneling (BTBT) from the source to the channel region, which is the working principle of tunnel FETs (TFETs) [8,9].

The increasing relevance of quantum mechanical effects in CMOS devices is confirmed by recent and exciting attempts to employ silicon nanowires to implement qubits, and thus pave the way for a scalable and industrial level platform for quantum computing [10].

A quantum transport formalism including the full-band energy effects produced by atomic scale potentials has a notoriously daunting complexity, which explains the widespread use of simplified Hamiltonians based either on the effective mass approximation (EMA) [11–13], or on the  $\mathbf{k}\cdot\mathbf{p}$  approach [14–16]. Both the EMA and  $\mathbf{k}\cdot\mathbf{p}$  methods have accuracy limitations at ultraconfined scale and, most importantly, they

describe low-dimensional systems including only the states close to a given symmetry point in the Brillouin zone of the underlying semiconductor, typically the  $\Gamma$  point for  $\mathbf{k}\cdot\mathbf{p}$  models. Consequently band-to-band-tunneling effects in indirect bandgap materials (e.g., Si, Ge, SiGe) cannot be described with an EMA or a  $\mathbf{k}\cdot\mathbf{p}$  approach.

As of today, the empirical tight-binding (TB) method has been the only full-band approach used in complete quantum transport device simulations by employing the nonequilibrium Green's function (NEGF) formalism [17,18]. The empirical TB uses directly the hopping integrals and Hamiltonian matrix elements as empirical, adjustable parameters, so that an explicit form of the wave function is not available and the overall number of fitting parameters is quite large. Nonempirical TB formulations using explicit basis functions have been also explored, but their use has been limited to transport in molecules or in transistors consisting of only about 1000 atoms [19]. Density functional theory (DFT) Hamiltonians have also been reported for quantum transport using a real-space, finite difference discretization of the unit cell, and based on either a wave-function matching [20], or an NEGF formalism [21,22], but this approach has been so far limited to very small systems.

Plane-wave-based methods provide both band structure and explicit expressions for the atomistic wave function, however, their use in DFT calculations usually requires such a large basis that transport calculations for technologically relevant devices are computationally very problematic. The empirical pseudopotential (EP) method provides an alternative with a more tractable computational burden; it relies on a fairly small number of fitting parameters and it has been successfully used for band-structure calculations in nanostructured devices [23–27], whereas only a few attempts have been reported to use EP for transport calculations. EP methods have been used

to investigate the conductance of a carbon nanotube using a wave-function matching condition at the interface between the device and the lead [28], and more recently they have been employed with a quantum-transmitting-boundary approach to simulate ultra-thin-body FETs [29,30], as well as nanowire FETs with a body size of 0.39 nm and graphene nanoribbon transistors [31,32]. In effect the computational burden of EP-based transport models remains very heavy, so that simplifying assumptions have been frequently introduced, such as the envelope wave-function approximation which assumes that the *external* potential is slowly varying over a unit cell [31,32].

In this work we present a full-band, quantum transport methodology based on a plane-wave empirical pseudopotential Hamiltonian and the NEGF formalism [33,34], and the paper extends substantially our previous, brief communication in [35]. We do not assume any *a priori* simplification of the problem and we introduce a hybrid real-space/plane-wave basis that largely improves the computational efficiency compared to a full plane-wave approach. We outline our formalism for a three-, two- and one-dimensional electron gas, which makes it suitable for a wide set of electron devices ranging from *p-n* diodes to nanowire transistors. Then we present some exemplary results for complete, self-consistent device simulations including an Esaki diode, as well as ultra-thin-body MOSFETs and tunnel FETs with technologically relevant geometrical dimensions.

In order to clarify the novel contributions of our approach a few explicit remarks may be useful: (a) In most previous papers employing a pseudopotential Hamiltonian in a quantum transport framework authors have used a real space discretization of the unit cell along all spatial directions, then they introduced periodicity in some directions and repeated their calculations for the Bloch wave vectors in periodic directions [20–22]; our approach is different because we use real space along the transport direction and plane waves along the transverse directions; (b) in our hybrid basis the confinement operator cannot be simply computed by means of a Fourier transform of a confinement potential, because there is no local operator in real space able to describe confinement; consequently we developed the original procedure described in Sec. II B for band-structure calculations, and in Sec. III B for transport calculations; (c) band-structure and transport calculations according to our approach suggest definitions of the reduced zone different from the first Brillouin zone, that are thoroughly discussed in Sec. V; (d) the physical systems analyzed in most previous studies dealing with quantum transport with a pseudopotential Hamiltonian are extremely small and actually consist of only a few tens of atoms [20–22], whereas the methodology of this paper allowed us to report self-consistent, complete device simulations of the current-voltage characteristics for transistors consisting of several thousands of atoms and by using a simple parallelization on few tens of cores.

The paper is organized as follows. In Sec. II we argue that, because of the hybrid real-space/plane-wave basis employed in the transport model, we need to introduce a discretization of the real space and we thus discuss how such a discretization enters the EP formalism for both bulk semiconductors and quantum confined systems. Section III explains our approach to transport relying on the NEGF method, Sec. IV B deals with the calculation of charge and current, and with the scheme for

self-consistent simulations, while in Sec. V we provide details about the implementation and computational burden. In Sec. VI we present several results for complete, self-consistent device simulations and in Sec. VII we report our concluding remarks.

## II. PSEUDOPOTENTIAL METHOD IN A DISCRETE REAL-SPACE LATTICE

The empirical pseudopotential method for band-structure calculations in bulk semiconductors and nanostructures is usually formulated assuming a continuous real space, that is, without introducing any real space discretization. However, such a discretization becomes indispensable in the formalism for electronic transport described in Sec. III, so that we here discuss the implications of a discrete real-space lattice for EP calculations. This section also introduces part of the notation used in the rest of the paper.

Let us start by defining the real space lattice as  $\mathbf{r}_d = (x_d, y_d, z_d)$ , and assume the same discretization  $d = a_0/N_d$  in all spatial directions  $s = \{x, y, z\}$ , where  $a_0$  is the semiconductor lattice constant and  $N_d$  is an even integer number. In a volume  $V = (L_x L_y L_z)$  the number of lattice points in each direction is  $N_s = L_s/d = N_d L_s/a_0$  and the total number of points in the three-dimensional (3D) lattice is  $N_T = N_x N_y N_z$ . A plane wave in the discrete real space lattice can be written as

$$|\mathbf{K}\rangle = \frac{e^{i\mathbf{K}\cdot\mathbf{r}_d}}{\sqrt{N_T}}, \quad K_s = n_s \frac{2\pi}{L_s}, \quad n_s = 0, \pm 1, \pm 2, \dots, \quad (1)$$

where  $\mathbf{K}$  is the wave vector in the extended  $\mathbf{K}$  space and, more precisely, each  $\mathbf{K}$  component has a range set by the real-space discretization,

$$-\frac{\pi}{d} \leq K_s < \frac{\pi}{d}, \quad \frac{\pi}{d} = \frac{N_d}{2} \left( \frac{2\pi}{a_0} \right). \quad (2)$$

Each  $|\mathbf{K}\rangle$  is actually a column vector with  $N_T$  components, and such vectors are orthogonal and normalized according to  $\langle \mathbf{K}' | \mathbf{K} \rangle = \delta_{\mathbf{K}, \mathbf{K}'}$ , with  $\delta$  being the Kronecher symbol.

### A. Bulk semiconductors

For a bulk crystal the Schrödinger equation for EP can be written as [36,37]

$$\left[ -\frac{\hbar^2}{2m_0} \nabla^2 + V_L(\mathbf{r}) \right] \psi(\mathbf{r}) = E_b \psi(\mathbf{r}), \quad (3)$$

where  $m_0$  is the free electron mass,  $V_L(\mathbf{r})$  is the lattice potential energy, and  $E_b$  is the energy in the bulk crystal. Equation (3) assumes a local pseudopotential, which is in fact used in all the calculations reported in the present paper.

Given the periodicity of  $V_L(\mathbf{r})$  and assuming periodic boundary conditions, the solutions of Eq. (3) have the well-known form of Bloch functions,

$$\Phi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{\mathbf{G}} B_{n\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \quad (4)$$

where  $u_{n\mathbf{k}}(\mathbf{r})$  is the periodic part of the Bloch function,  $n$  the band index, and the coefficients  $B_{n\mathbf{k}}(\mathbf{G})$  determine completely the wave function  $\Phi_{n\mathbf{k}}(\mathbf{r})$ . In Eq. (4) we have also introduced

the wave vector  $\mathbf{k}$  in the reduced zone, such that  $\mathbf{K} = (\mathbf{k} + \mathbf{G})$  with  $\mathbf{G}$  being a reciprocal lattice vector.

By substituting Eq. (4) in Eq. (3) the Schrödinger equation for the local EP model can be written as [36]

$$\sum_{\mathbf{G}'} \mathbf{H}_{\mathbf{k}}(\mathbf{G}, \mathbf{G}') B_{\mathbf{k}}(\mathbf{G}') = E_b(\mathbf{k}) B_{\mathbf{k}}(\mathbf{G}), \quad (5)$$

where the elements of the Hamiltonian matrix in  $\mathbf{K}$  space read

$$\mathbf{H}_{\mathbf{k}}(\mathbf{G}, \mathbf{G}') = T(\mathbf{k} + \mathbf{G}) \delta_{\mathbf{G}, \mathbf{G}'} + V_L(\mathbf{G} - \mathbf{G}'), \quad (6)$$

where  $V_L(\mathbf{G})$  is a spectral component of the pseudopotential,  $T(\mathbf{K})$  is the kinetic energy operator, and  $E_b(\mathbf{k})$  is the energy dispersion of the bulk semiconductor. For a diamond semiconductor (e.g., Si or Ge) or a zinc-blende compound (e.g., GaAs, InAs) the spectral components of the lattice potential energy can be written as [36,37]

$$V_L(\mathbf{G}) = U_S(|\mathbf{G}|) \cos[\mathbf{G} \cdot \boldsymbol{\tau}] + i U_A(|\mathbf{G}|) \sin[\mathbf{G} \cdot \boldsymbol{\tau}], \quad (7)$$

where  $\boldsymbol{\tau} = (1/8)a_0(1, 1, 1)$  and  $U_S(|\mathbf{G}|)$ ,  $U_A(|\mathbf{G}|)$  are, respectively, the symmetric and antisymmetric form factors [with  $U_A(|\mathbf{G}|)$  being null for diamond materials].

For a continuous real-space model, the kinetic energy operator in  $\mathbf{K}$  space has the well-known expression [36],

$$T(\mathbf{k} + \mathbf{G}) = \frac{\hbar^2 |\mathbf{k} + \mathbf{G}|^2}{2m_0}. \quad (8)$$

For a discrete real-space lattice, however, the form of the kinetic energy operator in real space and consequently also in  $\mathbf{K}$  space depends on the discretization scheme. Within a finite difference discretization method, EP calculations require a higher than second-order discretization for the kinetic energy operator [38]. In this respect, if one describes the second derivative with a centered difference approximation of order  $2p$ , the kinetic energy operator in  $\mathbf{K}$  space takes the form,

$$T(\mathbf{k} + \mathbf{G}) = -t_0 \sum_{s=x,y,z} \left[ C_0 + 2 \sum_{r=1}^p C_r \cos[r(k_s + G_s)d] \right], \quad (9)$$

$$t_0 = \frac{\hbar^2}{2m_0 d^2},$$

where the coefficients  $C_r$  are the convolution mask for the second derivative and can be easily calculated for any order  $2p$ .

As for the numerical solution of Eq. (5) we note that, because the wave vector in extended space is  $\mathbf{K} = (\mathbf{k} + \mathbf{G})$ , then the maximum  $|K_s|$  in Eq. (2) results in

$$|G_s| \leq \frac{N_d}{2} \left( \frac{2\pi}{a_0} \right) \quad s = x, y, z, \quad (10)$$

which creates a link between the real-space discretization  $d = (a_0/N_d)$  and the maximum magnitude of the  $\mathbf{G}$  vectors used in the solution of the EP problem, Eq. (5).

In this respect it is interesting to notice that in any actual EP calculation it is inevitable to introduce a truncation of the  $\mathbf{G}$  vectors, which is typically enforced via a cutoff energy  $E_{\text{cut}}$ , such that the maximum  $|\mathbf{G}|$  is given by  $|\mathbf{G}|^2 = (2m_0 E_{\text{cut}}/\hbar^2)$  [36,37]. Consequently, even if an explicit real-space discretization is not usually introduced in EP calculations, nevertheless the maximum  $|\mathbf{G}|$  sets a minimum spacing between real-space

points where the wave function can be evaluated without incurring in aliasing problems. This observation reconciles the use of a discrete real-space lattice with the necessary truncation of the  $\mathbf{G}$  vectors used in the solution of Eq. (5).

It is also important to notice that, according to Eq. (10), the expansion volume of the  $\mathbf{G}$  vectors is defined as a cube, rather than as the sphere defined by a cutoff energy condition. In Eq. (10), in fact, the spatial discretization  $d = a_0/N_d$  is the same along  $x$ ,  $y$ ,  $z$ , which is necessary to preserve the crystal symmetries and ensure the equivalence of the three crystal orientations [100], [010], [001].

Two important remarks should be made about the discretized kinetic energy operator. First, Eq. (10) implies that, for any discretization  $d = a_0/N_d$ , the maximum value of  $(k_s + G_s)d$  in Eq. (5) is  $\pi$  and does not decrease with decreasing  $d$ , because the extended  $\mathbf{K}$  space expands proportionally to  $1/d$ . Second, the only means to enlarge the  $K_x$  range where the discretized  $T(\mathbf{K})$  agrees well with the continuous  $T(\mathbf{K})$  is in effect to increase the discretization order.

The discrepancies between the discretized and continuous kinetic energy have significant consequences on the band-structure calculations that are illustrated in Fig. 1, reporting the energy dispersion of bulk germanium obtained by solving Eq. (5) and using either the continuous or the discretized kinetic energy operator with  $d = a_0/10$ . The parameters for the local EP model used throughout this paper are reported in Table I.

Figure 1 reports the band structure for germanium, and it shows that for a second-order discretization the differences with respect to the continuous case are unacceptably large: The minimum of the conduction band is at the  $\Gamma$  rather than at the  $L$  point and the energy gap is only 0.5 eV. By increasing the discretization order, however, the results converge to the continuous case and, in particular, to the expected values for both the energy gap and the effective masses, in agreement with the results of Ref. [38].

If not otherwise stated, all the calculations in this work were obtained by using the  $T$  expression in Eq. (9) with  $d = a_0/10$  and  $2p = 8$ . It should be noticed that the order of discretization has essentially no influence on the computational burden.

## B. Nanostructures with quantum confinement effects

The confinement in nanostructures is here described using a single material approximation, where the confinement is obtained by means of a discontinuity of the conduction and valence band of the bulk semiconductor. For a semiconductor-oxide system with band discontinuities as large as some electron volts this seems a reasonable approximation, that has been already discussed and validated in some previous publications [27,39]. Hereafter we give explicit expressions for a quasi-2D electron gas, namely a gas where quantum confinement occurs along a single direction; the formalism for a one-dimensional (1D) electron gas is a quite straightforward extension and will be briefly discussed in Appendix A.

We first introduce a unitary step function  $\Theta(z)$  along the quantization direction  $z$ , such that  $\Theta(z) = 0$  for  $|z| \leq T_{\text{sct}}/2$  and  $\Theta(z) = 1$  for  $T_{\text{sct}}/2 < |z| < L_z/2$ , where  $T_{\text{sct}}$  is the thickness of the semiconductor film and  $L_z > T_{\text{sct}}$  is the periodicity length of the nanostructure in the  $z$  direction.

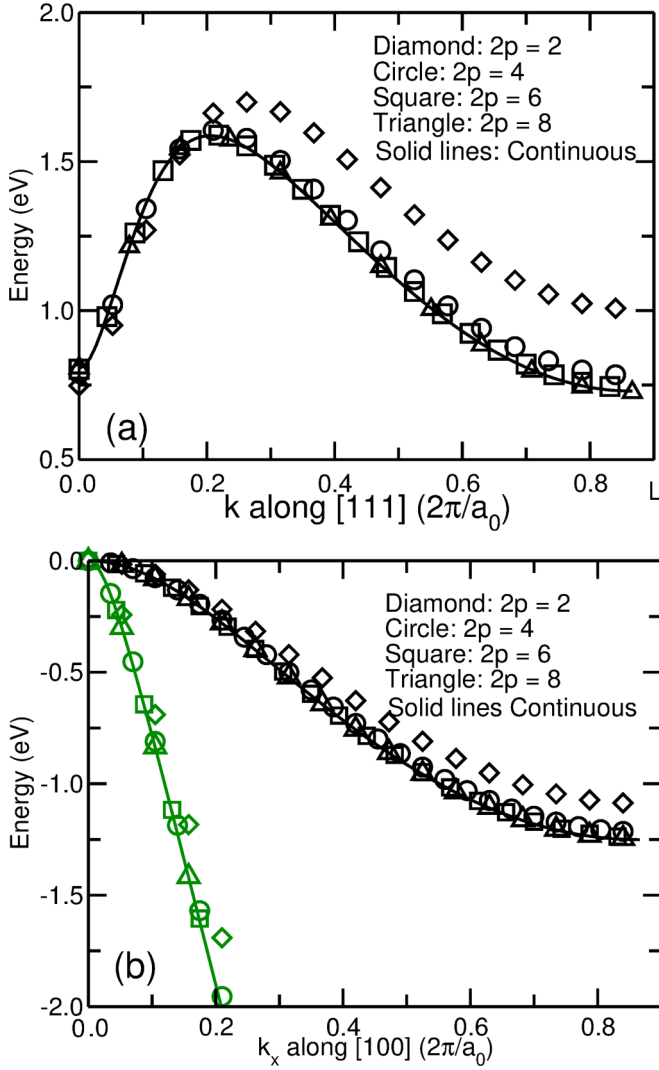


FIG. 1. Band-structure for bulk germanium as obtained from EP calculations for either a continuous model [i.e., kinetic energy operator from Eq. (8)] or a discretized real space lattice with the kinetic energy from Eq. (9)]. (a) Lowest conduction band along the [111] direction showing the minimum at the  $L$  point in  $k = |\mathbf{k}| = 0.5\sqrt{3}(2\pi/a_0)$ ; (b) valence bands along the [100] direction; spin-orbit interaction has been neglected, so that two bands are perfectly degenerate. For a discretization order larger than about  $2p = 6$  the results practically coincide with the continuous model.

The representation of  $\Theta(z)$  in  $\mathbf{K}$  space is readily given by

$$\Theta(\mathbf{K}, \mathbf{K}') = \delta_{K_x, K'_x} \delta_{K_y, K'_y} \Theta_{zF}(K_z - K'_z), \quad (11)$$

where  $\Theta_{zF}(K_z - K'_z)$  is the discrete Fourier transform along  $z$ . For a very fine spatial discretization the discrete Fourier transform approaches the analytical expression corresponding to a continuous abscissa  $z$ , namely

$$\Theta_{zF}(K_z - K'_z) \approx -\frac{1}{2\pi} \left[ \frac{\sin[(K_z - K'_z)T_{scf}/2]}{(K_z - K'_z)T_{scf}/2} \right]. \quad (12)$$

It is now important to understand that in our formalism the carrier confinement is obtained with an operator  $\mathbf{H}^{(c)}$  that, in the basis of the Bloch functions  $\Phi_{n\mathbf{k}}$  of the bulk semiconductor,

TABLE I. Pseudopotential parameters used for all calculations and expressed in Rydberg. These values are taken from Ref. [36] for Si, from Ref. [45] for Ge, and from Ref. [46] for GaAs.

	$U_S(\sqrt{3})$	$U_S(\sqrt{8})$	$U_S(\sqrt{11})$	$U_A(\sqrt{3})$	$U_A(\sqrt{4})$	$U_A(\sqrt{11})$
Si	-0.2241	0.0551	0.0724	0	0	0
Ge	-0.2768	0.0582	0.0152	0	0	0
GaAs	-0.23	0.01	0.06	0.07	0.05	0.01

can be readily written as

$$\mathbf{H}^{(c)}(n\mathbf{k}, n'\mathbf{k}') = \begin{cases} V_{cb} \Theta(n\mathbf{k}, n'\mathbf{k}') & \text{if } n' \in \text{conduction band} \\ V_{vb} \Theta(n\mathbf{k}, n'\mathbf{k}') & \text{if } n' \in \text{valence band} \end{cases}, \quad (13)$$

where  $V_{cb}$  and  $V_{vb}$  are the band discontinuity, respectively, for the conduction and the valence band, and  $\Theta(n\mathbf{k}, n'\mathbf{k}')$  is the representation of the unitary step function  $\Theta(z)$  in the Bloch function basis.

In the rest of the paper we will often refer to operators and matrices in different basis sets. Matrices will be denoted using square brackets and a subscript for the basis: for example,  $[\mathbf{H}_\Phi^{(c)}]$ ,  $[\Theta_\Phi]$  denote the matrix for, respectively, the confining operator and the unitary step function in the Bloch function basis, whereas  $[\mathbf{H}_\mathbf{K}^{(c)}]$ ,  $[\Theta_\mathbf{K}]$  denote the same matrices in the plane-wave basis. When we denote the elements of the matrices, instead, we drop the subscript and write  $\mathbf{H}^{(c)}(n\mathbf{k}, n'\mathbf{k}')$ ,  $\Theta(n\mathbf{k}, n'\mathbf{k}')$  or  $\mathbf{H}^{(c)}(\mathbf{K}, \mathbf{K}')$ ,  $\Theta(\mathbf{K}, \mathbf{K}')$ , because the symbols that indicate the elements (in this case  $n\mathbf{k}$  or  $\mathbf{K}$ ) identify univocally the basis set.

Equation (13) can be rewritten in matrix notation as

$$[\mathbf{H}_\Phi^{(c)}] = [\Theta_\Phi] [\mathbf{D}_\Phi^{(c)}], \quad (14)$$

where  $[\mathbf{D}_\Phi^{(c)}]$  is the diagonal matrix whose elements are defined as

$$\mathbf{D}^{(c)}(n\mathbf{k}, n'\mathbf{k}') = \delta_{n,n'} \delta_{\mathbf{k}, \mathbf{k}'} \times \begin{cases} V_{cb} & \text{if } n' \in \text{conduction band} \\ V_{vb} & \text{if } n' \in \text{valence band} \end{cases}. \quad (15)$$

Our goal now is to find the expression  $[\mathbf{H}_\mathbf{K}^{(c)}]$  for the confining operator in plane-wave basis  $|\mathbf{K}\rangle$ . To this purpose we consider the unitary matrix  $[\mathbf{U}_{\mathbf{K}, \Phi}]$  from Bloch functions to  $|\mathbf{K}\rangle$  basis, whose columns are the eigenvectors of the bulk crystal EP problems in Eq. (5). The form of the matrix  $[\mathbf{U}_{\mathbf{K}, \Phi}]$  depends on the shape of the reduced zone of the bulk crystal and on the sorting of the functions in both bases.

We here assume that the reduced zone of the bulk crystal is defined so that, for any two different wave vectors  $\mathbf{k} \neq \mathbf{k}'$  in the reduced zone, none of the corresponding wave vectors in the extended  $\mathbf{K}$  space have the same  $x$  and  $y$  components; namely, for any two reciprocal lattice vectors  $\mathbf{G}, \mathbf{G}'$  we must have  $(k_x + G_x, k_y + G_y) \neq (k'_x + G'_x, k'_y + G'_y)$ . Such a condition is fulfilled, for example, if we define the reduced zone of the bulk crystal according to inequalities  $-\pi/a_0 \leq k_x, k_z < \pi/a_0$  and  $-\pi/a_0 \leq k_y < \pi/a_0$ . Such a definition of the reduced zone implies that the unitary step function  $\Theta(n\mathbf{k}, n'\mathbf{k}')$  in the Bloch function basis is non-null only for Bloch functions

$\Phi_{n\mathbf{k}_{xy}k_z}$ ,  $\Phi_{n\mathbf{k}_{xy}k'_z}$  corresponding to the same  $\mathbf{k}_{xy} = (k_x, k_y)$ . Consequently the transformation of the step function from the plane-wave to the Bloch-function basis is governed by a separated unitary matrix,  $[\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]$ , for each  $\mathbf{k}_{xy} = (k_x, k_y)$  in the

$$[\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}] = \begin{bmatrix} [\mathbf{B}_{n,\mathbf{k}_{xy}k_{z,1}}(\mathbf{G})] & 0 & 0 & \cdots \\ 0 & [\mathbf{B}_{n,\mathbf{k}_{xy}k_{z,2}}(\mathbf{G})] & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [\mathbf{B}_{n,\mathbf{k}_{xy}k_{z,N_{k_z}}}(\mathbf{G})] \end{bmatrix}, \quad (16)$$

where  $[\mathbf{B}_{n,\mathbf{k}_{xy}k_z}(\mathbf{G})]$  is an  $N_G \times N_G$  matrix such that each column is the eigenvector of Eq. (5) for band  $n$  (with  $n = 1, 2, \dots, N_G$ ). The rank of  $[\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]$  is thus  $N_G N_{k_z}$ , with  $N_{k_z}$  being the number of  $k_z$  in the reduced zone of the bulk crystal.

By using  $[\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]$  the step function in the Bloch-function basis can be written as

$$[\Theta_\Phi] = [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]^\dagger [\Theta_{\mathbf{K}}] [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}], \quad (17)$$

where  $[\Theta_{\mathbf{K}}]$  is a matrix of rank  $N_G N_{k_z}$  given by the evaluation of Eq. (11) at the  $\mathbf{K} = [(\mathbf{k}_{xy}, k_z) + \mathbf{G}]$  wave vectors corresponding to the  $N_{k_z}$  EP problems defined by Eq. (5) for  $\mathbf{k} = (\mathbf{k}_{xy}, k_z)$  and  $-2\pi/a_0 \leq k_z < 2\pi/a_0$ . The confinement operator for a fixed  $\mathbf{k}_{xy}$  in the Bloch-function basis can be similarly written,

$$[\mathbf{H}_\Phi^{(c)}] = [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]^\dagger [\mathbf{H}_{\mathbf{K}}^{(c)}] [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}], \quad (18)$$

where  $[\mathbf{H}_{\mathbf{K}}^{(c)}]$  is the same operator in  $\mathbf{K}$  space, whose expression we wish to derive. We now rewrite Eq. (14) by using Eq. (17) as

$$[\mathbf{H}_\Phi^{(c)}] = [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]^\dagger [\Theta_{\mathbf{K}}] [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}] [\mathbf{D}_\Phi^{(c)}], \quad (19)$$

where  $[\mathbf{D}_\Phi^{(c)}]$  now has  $N_G N_{k_z}$  rank. If we substitute Eq. (19) in Eq. (18) and solve it for  $[\mathbf{H}_{\mathbf{K}}^{(c)}]$  we obtain

$$[\mathbf{H}_{\mathbf{K}}^{(c)}] = [\Theta_{\mathbf{K}}] [\mathbf{D}_{\mathbf{K}}^{(c)}], \quad (20)$$

where we have introduced

$$[\mathbf{D}_{\mathbf{K}}^{(c)}] = [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}] [\mathbf{D}_\Phi^{(c)}] [\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]^\dagger, \quad (21)$$

which is the expression in  $\mathbf{K}$  space of the  $[\mathbf{D}_\Phi^{(c)}]$  matrix defined in the Bloch-function basis by Eq. (15). It may be worth noting that, if the discontinuity at the valence and conduction band were the same, that is, if we set  $V_{cb} = V_{vb}$  in Eq. (15), then Eq. (20) would simplify as

$$[\mathbf{H}_{\mathbf{K}}^{(c)}] = V_{cb} [\Theta_{\mathbf{K}}]. \quad (22)$$

However, in a semiconductor-oxide system  $V_{cb}$  is positive whereas  $V_{vb}$  is negative, so that our treatment leading to Eqs. (20) and (21) is indispensable.

We reiterate that Eqs. (18)–(21) refer to a fixed  $\mathbf{k}_{xy}$ , that all matrices have an  $N_G N_{k_z}$  rank, and that the  $\mathbf{K}$  in these equations are the  $\mathbf{K} = [(\mathbf{k}_{xy}, k_z) + \mathbf{G}]$  wave vectors corresponding to the

reduced zone of the bulk crystal. If we further assume that the  $|\mathbf{K}\rangle$  functions are sorted according to the sets of  $N_G$  functions that are coupled in the EP problems in Eq. (5), then  $[\mathbf{U}_{\mathbf{K},\Phi}^{(\mathbf{k}_{xy})}]$  takes the block diagonal form,

solution of the  $N_{k_z}$  EP problems defined by Eq. (5) for  $\mathbf{k} = (\mathbf{k}_{xy}, k_z)$  and  $-2\pi/a_0 \leq k_z < 2\pi/a_0$ .

The overall Hamiltonian for a 2D electron gas in  $\mathbf{K}$  space and a given  $\mathbf{k}_{xy}$  takes thus the form,

$$\mathbf{H}_{\mathbf{k}_{xy}}(\mathbf{K}, \mathbf{K}') = T(\mathbf{k} + \mathbf{G})\delta_{\mathbf{G},\mathbf{G}'}\delta_{k_z,k'_z} + V_L(\mathbf{G} - \mathbf{G}')\delta_{k_z,k'_z} + \mathbf{H}^{(c)}(\mathbf{K}, \mathbf{K}'), \quad (23)$$

where  $\mathbf{K} = [(\mathbf{k}_{xy}, k_z) + \mathbf{G}]$ ,  $\mathbf{K}' = [(\mathbf{k}_{xy}, k'_z) + \mathbf{G}']$ , and the band structure of the 2D electron gas can be calculated solving the eigenvalue problem associated with  $\mathbf{H}_{\mathbf{k}_{xy}}(\mathbf{K}, \mathbf{K}')$  and varying  $\mathbf{k}_{xy}$  in the 2D reduced zone. All matrices in Eq. (23) have a rank  $N_G N_{k_z}$ , with  $N_{k_z}$  being the number of  $k_z$  values in the reduced zone of the bulk crystal.

Figure 2 shows the band structure for a germanium 2D electron gas obtained with the Hamiltonian in Eq. (23) and the methodology described in the present section. The calculations correspond to a square germanium quantum well with different well thicknesses  $T_{scf} = 15, 10, 7a_0$  (with  $a_0 = 0.565$  nm), the quantization direction is [001], and the bands are plotted along the [110] direction. As already mentioned the sign of the band discontinuity is different for the conduction and valence band, and in these calculations we used  $V_{cb} = 3$  eV and  $V_{vb} = -3$  eV, which result in an expected upshift of the conduction band and a downshift of the valence band compared to the bulk crystal. As it can be seen the subband quantization is the largest at the  $\Gamma$  point of the conduction band, where the effective mass is the smallest.

### III. TRANSPORT FORMALISM BASED ON THE NEGF METHOD

The transport formalism developed in this section relies on the non-equilibrium Green's function (NEGF) method. The main steps necessary for transport calculations can be summarized as follows: (a) we need an expression for the Hamiltonian matrix (both in the device and in the leads) for closed boundary conditions, as opposed to the periodic boundary conditions used in Sec. II for band-structure calculations; (b) we need to calculate the surface Green's function of the leads; (c) we need the matrices describing the coupling between the device and the leads, that eventually allow us to calculate the self-energies of the leads.

Let us now consider a device structure with a length  $L_x = N_{cx}a_0$ , with  $N_{cx}$  being the number of unit cells in the transport direction  $x$ . In the plane-wave basis and for periodic boundary

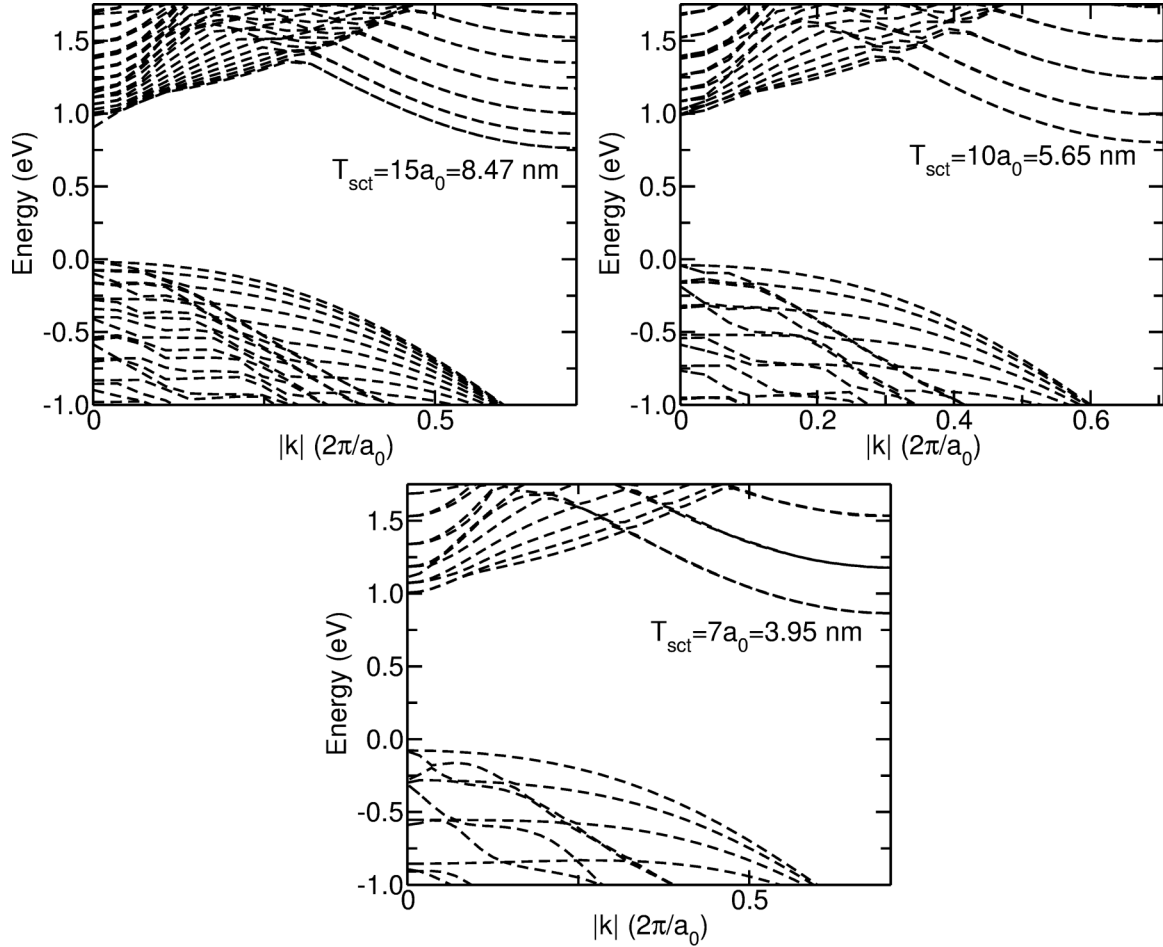


FIG. 2. Band-structure calculations obtained with the Hamiltonian in Eq. (23) for a squared germanium quantum well with different well thicknesses  $T_{sct} = 15, 10, 7a_0$ , with  $a_0 = 0.565$  nm. Quantization is along the [001] direction and the  $|\mathbf{k}|$  on the  $x$  axis runs along the [110] direction. The band edge discontinuities are  $V_{cb} = 3$  eV for the conduction band and  $V_{vb} = -3$  eV for the valence band.

conditions the Hamiltonian matrix of the device can be written by using Eq. (6), Eq. (23), or Eq. (A2), respectively, for a 3D, 2D, or 1D electron gas. For such Hamiltonians in the plane-wave basis, it is possible to describe a procedure to modify boundary conditions in the  $x$  direction from periodic to closed boundary conditions.

The main difficulty in dealing with this approach, however, is that the Hamiltonian matrices with closed boundary conditions are dense in the  $|\mathbf{K}\rangle$  basis and have no block diagonal structure. For a 2DEG, for instance, the Hamiltonian of the leads is a dense matrix whose rank is  $N_G N_{k_z} N_{k_x}^{(L)}$ , where the number  $N_{k_x}^{(L)}$  of  $k_x$  values is huge for semi-infinite leads. The rank of these matrices makes such a full plane-wave approach computationally prohibitive, so that we here propose to use a hybrid basis consisting of real space in the transport direction  $x$  and plane waves in the  $(y, z)$  directions normal to transport.

### A. Transport in a 3D electron gas

The main advantage of the hybrid basis is that the overall Hamiltonian has a block tridiagonal form. For a 3D electron gas (3DEG), for instance, the Hamiltonian in the hybrid basis

can be written,

$$[\mathbf{H}_{x\mathbf{k}_{yz}}] = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{0,1} & 0 & 0 & \cdots & 0 \\ \mathbf{H}_{0,1}^\dagger & \mathbf{H}_{2,2} & \mathbf{H}_{0,1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{H}_{0,1}^\dagger & \mathbf{H}_{N_{cx}, N_{cx}} \end{bmatrix}, \quad (24)$$

where the index  $l = 1, 2, \dots, N_{cx}$  indicates an  $a_0$  long section along  $x$  comprising  $N_d$  discretization points (see the beginning of Sec. II), and each block  $\mathbf{H}_{l,l}$ ,  $\mathbf{H}_{0,1}$  has a  $2N_G$  rank (as discussed in more detail at the end of Sec. V). Moreover, it should be noticed that  $[\mathbf{H}_{x\mathbf{k}_{yz}}]$  in Eq. (24) already incorporates closed boundary conditions along the transport direction at the first and last discretization point along the  $x$  axis.

The Hamiltonian in Eq. (24) refers to a fixed  $\mathbf{k}_{yz} = (k_y, k_z)$  in the reduced zone and the  $\mathbf{K}_{yz}$  entering Eq. (24) are the  $(y, z)$  components of the wave vectors  $\mathbf{K} = [(k_x, \mathbf{k}_{yz}) + \mathbf{G}]$  involved in the solution of the  $N_{k_x} = 2N_{cx}$  eigenvalue problems given by Eq. (5) for  $\mathbf{k} = (k_x, \mathbf{k}_{yz})$  and with  $-2\pi/a_0 \leq k_x < 2\pi/a_0$ . In this latter respect, in fact, it is important to underline that, in order for the hybrid basis to describe correctly the

transport along the  $x$  direction, it is necessary that for any  $\mathbf{K}_{yz} = [(k_y, k_z) + (G_y, G_z)]$  in Eq. (24) the corresponding  $x$  components  $K_x = k_x + G_x$  cover with no voids the entire extended  $K_x$  range  $[-(N_d/2)(2\pi/a_0), (N_d/2)(2\pi/a_0)]$ . This prescription requires that  $k_x$  in the reduced zone be in the range  $-2\pi/a_0 \leq k_x < 2\pi/a_0$ .

All off-diagonal blocks  $\mathbf{H}_{0,1}$  are the same, whereas the diagonal blocks  $\mathbf{H}_{l,l}$  differ only by the external potential energy,  $\mathbf{V}_{\text{ext}}$ , produced by doping and external biases. More specifically the  $\mathbf{H}_{l,l}$  blocks can be written as

$$\mathbf{H}_{l,l}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz}) = \mathbf{H}_{0,0}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz}) + \mathbf{V}_{\text{ext}}(\mathbf{x}_l, (\mathbf{K}_{yz} - \mathbf{K}'_{yz})), \quad (25)$$

where  $\mathbf{V}_{\text{ext}}(\mathbf{x}_l, (\mathbf{K}_{yz} - \mathbf{K}'_{yz}))$  is a matrix including the discrete Fourier transform with respect to  $(y, z)$  of the  $\mathbf{V}_{\text{ext}}$  calculated at each one of the  $N_d$  discretization points in the block  $l$ , while  $\mathbf{H}_{0,0}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz})$  is independent of the specific block  $l$ .

An inspection of the kinetic energy operator and the pseudopotential term allows us to see that  $\mathbf{H}_{0,0}$  can be written as

$$\mathbf{H}_{0,0}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz}) = V_L(\mathbf{x}_l, (\mathbf{G}_{yz} - \mathbf{G}'_{yz})) + \mathbf{T}_{0,0}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz}), \quad (26)$$

where  $V_L(\mathbf{x}_l, (\mathbf{G}_{yz} - \mathbf{G}'_{yz}))$  and  $\mathbf{T}_{0,0}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz})$  are, respectively, the pseudopotential and kinetic energy contributions, while for a 3D gas the off-diagonal blocks  $\mathbf{H}_{0,1}$  are given solely by the kinetic energy operator, that is,

$$\mathbf{H}_{0,1}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz}) = \mathbf{T}_{0,1}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz}). \quad (27)$$

The pseudopotential term  $V_L(\mathbf{x}_l, (\mathbf{G}_{yz} - \mathbf{G}'_{yz}))$  in Eq. (26) is periodic of  $a_0$  along  $x$ ; it is the same in every block  $l$  and it can be written as

$$\begin{aligned} V_L(x_i, \mathbf{G}_{yz} - \mathbf{G}'_{yz}) &= \frac{2}{N_d} \sum_{(G_x, G'_x)} V_L(G_x - G'_x, \mathbf{G}_{yz} - \mathbf{G}'_{yz}) \exp[i(G_x - G'_x)x_i], \end{aligned} \quad (28)$$

where  $\mathbf{G} = (G_x, \mathbf{G}_{yz})$ ,  $\mathbf{G}' = (G'_x, \mathbf{G}'_{yz})$  are reciprocal lattice vectors (with  $N_d/2$  being the number of  $G_x$  components in the expansion volume),  $V_L(\mathbf{G})$  is given by Eq. (7) and, with no loss of generality, we can take  $x_i = 0, d, 2d \dots (a_0 - d)$ . Moreover, by recalling the results about the discretization of the kinetic energy operator reported in Sec. II A, we see that for  $\mathbf{T}_{0,0}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz})$  in Eq. (26) we have

$$-t_0 \begin{bmatrix} (C_0 + T_{0,0}^{(yz)})\mathbf{I} & C_1\mathbf{I} & \cdots & C_p\mathbf{I} & 0 & \cdots & 0 \\ C_1\mathbf{I} & (C_0 + T_{0,0}^{(yz)})\mathbf{I} & C_1\mathbf{I} & \cdots & C_p\mathbf{I} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & C_p\mathbf{I} & \cdots & C_1\mathbf{I} & (C_0 + T_{0,0}^{(yz)})\mathbf{I} \end{bmatrix}, \quad (29)$$

where  $\mathbf{I}$  is an identity matrix of rank  $2N_G/N_d$  and  $T_{0,0}^{(yz)}(\mathbf{K}_{yz}, \mathbf{K}'_{yz})$  is given by

$$T_{0,0}^{(yz)}(\mathbf{K}_{yz}, \mathbf{K}'_{yz}) = \delta_{\mathbf{K}_{yz}, \mathbf{K}'_{yz}} \sum_{s=y,z} \left[ C_0 + 2 \sum_{r=1}^p C_r \cos[r(k_s + G_s)d] \right], \quad (30)$$

with  $K_y = k_y + G_y$ ,  $K_z = k_z + G_z$ . The off-diagonal contribution of the kinetic energy operator can instead be written as

$$\mathbf{T}_{0,1}(\mathbf{x}_l \mathbf{K}_{yz}, \mathbf{x}_l \mathbf{K}'_{yz}) = -t_0 \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ C_p\mathbf{I} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \ddots & \vdots \\ C_2\mathbf{I} & \cdots & C_p\mathbf{I} & 0 & 0 & \cdots & 0 \\ C_1\mathbf{I} & C_2\mathbf{I} & \cdots & C_p\mathbf{I} & 0 & \cdots & 0 \end{bmatrix}, \quad (31)$$

where  $\mathbf{I}$  is the same identity matrix as in Eq. (29) and the matrix in Eqs. (29) and (31) has  $N_d$  rows and columns.

### B. Transport in a 2D electron gas (2DEG)

Let us now consider a quantum well or an ultra-thin-body SOI FET with a length  $L_x = N_{cx}a_0$ , where  $N_{cx}$ ,  $N_{cz}$  are the number of unit cells, respectively, in the transport direction  $x$  and in the confinement direction  $z$ . In the hybrid basis the Hamiltonian with closed boundary conditions along  $x$  can be

written as

$$[\mathbf{H}_{x\mathbf{K}_{yz}}] = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{0,1} & 0 & 0 & \cdots & 0 \\ \mathbf{H}_{0,1}^\dagger & \mathbf{H}_{2,2} & \mathbf{H}_{0,1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{H}_{0,1}^\dagger & \mathbf{H}_{N_{cx}, N_{cx}} \end{bmatrix}, \quad (32)$$

where, differently from Eq. (24), each block  $\mathbf{H}_{l,l}$ ,  $\mathbf{H}_{0,1}$  has now a  $2N_G N_{k_z}$  rank, with  $N_{k_z} = 2N_{cz}$  being the number of  $k_z$  wave

vectors in the reduced zone of the bulk crystal (see the end of Sec. V for more details).

The Hamiltonian  $[\mathbf{H}_{x\mathbf{K}_{yz}}]$  in Eq. (32) refers to a fixed  $k_y$  in the reduced zone and, moreover, the  $\mathbf{K}_{yz}$  in Eq. (32) are the  $(y, z)$  components of the  $\mathbf{K} = [(k_x, k_y, k_z) + \mathbf{G}]$  wave vectors involved in the solution of the  $N_{k_x}N_{k_z}$  eigenvalue problems given by Eq. (5) for  $\mathbf{k} = (k_x, k_y, k_z)$  and with  $-2\pi/a_0 \leq k_x, k_z < 2\pi/a_0$ . As already mentioned in the comment to Eq. (24), the  $k_x$  range between  $-2\pi/a_0$  and  $2\pi/a_0$  ensures that, for any  $\mathbf{K}_{yz} = [(k_y, k_z) + (G_y, G_z)]$  in Eq. (32), the corresponding  $x$  components  $K_x = k_x + G_x$  cover with no voids the entire extended  $K_x$  range.

The differences in the definition of the  $\mathbf{H}_{l,l}$  and  $\mathbf{H}_{0,1}$  compared to the 3DEG case stem from the confinement Hamiltonian discussed in Sec. II B. In fact, because the confinement Hamiltonian is a nonlocal operator, it provides a contribution to both the diagonal and the off-diagonal blocks of the  $[\mathbf{H}_{x\mathbf{K}_{yz}}]$ . More specifically,  $\mathbf{H}_{l,l}$  is still expressed by Eq. (25), but  $\mathbf{H}_{0,0}$  takes now the form,

$$\mathbf{H}_{0,0}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}_l\mathbf{K}'_{yz}) = V_L(\mathbf{x}_l, (\mathbf{G}_{yz} - \mathbf{G}'_{yz})) + \mathbf{T}_{0,0}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}_l\mathbf{K}'_{yz}) + \mathbf{H}_{0,0}^{(c)}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}_l\mathbf{K}'_{yz}). \quad (33)$$

Furthermore, the off-diagonal blocks are now rewritten as

$$\begin{aligned} \mathbf{H}_{0,1}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}'_l\mathbf{K}'_{yz}) \\ = \mathbf{T}_{0,1}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}'_l\mathbf{K}'_{yz}) + \mathbf{H}_{0,1}^{(c)}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}'_l\mathbf{K}'_{yz}), \end{aligned} \quad (34)$$

where Eqs. (29)–(31) still hold for the terms stemming from the kinetic energy operator.

The contributions due to the confinement operator cannot be given in an explicit, analytical form, and must be evaluated numerically by using an appropriate unitary transformation. To this purpose we define the unitary transformation matrix from  $x$  to  $K_x$  as

$$[\mathbf{U}_{K_x, x}] = \frac{1}{\sqrt{N_x}} \begin{bmatrix} e^{iK_{x1}x_1}\mathbf{I} & e^{iK_{x1}x_2}\mathbf{I} & \dots & e^{iK_{x1}x_{N_x}}\mathbf{I} \\ e^{iK_{x2}x_1}\mathbf{I} & e^{iK_{x2}x_2}\mathbf{I} & \dots & e^{iK_{x2}x_{N_x}}\mathbf{I} \\ \dots & \dots & \ddots & \vdots \\ e^{iK_{xN_x}x_1}\mathbf{I} & e^{iK_{xN_x}x_2}\mathbf{I} & \dots & e^{iK_{xN_x}x_{N_x}}\mathbf{I} \end{bmatrix}, \quad (35)$$

where  $\mathbf{I}$  is an identity matrix with rank of  $2N_GN_{k_z}/N_d$  and then write the confinement operator in the hybrid basis as

$$[\mathbf{H}_{x\mathbf{K}_{yz}}^{(c)}] = [\mathbf{U}_{K_x, x}]^\dagger [\mathbf{H}_{\mathbf{K}}^{(c)}] [\mathbf{U}_{K_x, x}], \quad (36)$$

where  $[\mathbf{H}_{\mathbf{K}}^{(c)}]$  is given by Eq. (20).

An inspection of the confining operator shows that  $\mathbf{H}^{(c)}(x_i\mathbf{K}_{yz}, x_j\mathbf{K}'_{yz})$  (where  $x_i$  is an individual discretization point) is periodic of  $a_0$  along  $x$ . Moreover, the operator is nonlocal, but the nonlocality is short range and we verified that  $\mathbf{H}^{(c)}(x_i\mathbf{K}_{yz}, x_j\mathbf{K}'_{yz})$  can be adequately described by considering only  $x_j$  such that  $|x_i - x_j|$  extends up to  $a_0$ . Consequently, if we evaluate  $\mathbf{H}^{(c)}(x_i\mathbf{K}_{yz}, x_j\mathbf{K}'_{yz})$  with Eq. (36) for  $x_i = 0, d, 2d \dots (a_0 - d)$  and  $x_j = 0, d, 2d \dots (2a_0 - d)$ , then  $\mathbf{H}_{0,0}^{(c)}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}'_l\mathbf{K}'_{yz})$  can be identified with the terms of  $\mathbf{H}^{(c)}(x_i\mathbf{K}_{yz}, x_j\mathbf{K}'_{yz})$  for  $x_i, x_j = 0, d, 2d \dots (a_0 - d)$ , whereas  $\mathbf{H}_{0,1}^{(c)}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}'_l\mathbf{K}'_{yz})$  corresponds

to the terms of  $\mathbf{H}^{(c)}(x_i\mathbf{K}_{yz}, x_j\mathbf{K}'_{yz})$  for  $x_i = 0, d, 2d \dots (a_0 - d)$ ,  $x_j = a_0, a_0 + d, \dots (2a_0 - d)$ .

We reiterate that  $\mathbf{H}_{0,0}^{(c)}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}'_l\mathbf{K}'_{yz})$ ,  $\mathbf{H}_{0,1}^{(c)}(\mathbf{x}_l\mathbf{K}_{yz}, \mathbf{x}'_l\mathbf{K}'_{yz})$  are the same in every section section  $l$  of the device (with  $l = 1, 2, \dots N_{cx}$ ).

#### IV. CALCULATIONS OF GREEN'S FUNCTIONS, CHARGE, AND CURRENT

The Hamiltonian matrices in Eqs. (24) and (32) have a block tridiagonal form, which reduces dramatically the computational burden to obtain the retarded,  $[\mathbf{G}_{x\mathbf{K}_{yz}}]$ , and the lesser-than,  $[\mathbf{G}_{x\mathbf{K}_{yz}}^<]$ , Green's function of the system.

##### A. Green's functions

In principle  $[\mathbf{G}_{x\mathbf{K}_{yz}}(E)]$ ,  $[\mathbf{G}_{x\mathbf{K}_{yz}}^<(E)]$  at a given energy  $E$  can be obtained by solving the kinetic equations,

$$[\mathbf{G}_{x\mathbf{K}_{yz}}(E)] = [(E + i0^+)\mathbf{I} - [\mathbf{H}_{x\mathbf{K}_{yz}}] - [\Sigma(E)]]^{-1}, \quad (37)$$

$$[\mathbf{G}_{x\mathbf{K}_{yz}}^<(E)] = [\mathbf{G}_{x\mathbf{K}_{yz}}(E)][\Sigma^<(E)][\mathbf{G}_{x\mathbf{K}_{yz}}(E)]^\dagger, \quad (38)$$

where  $[\Sigma] = [\Sigma_L] + [\Sigma_R] + [\Sigma_{\text{ph}}]$  and  $[\Sigma^<] = [\Sigma_L^<] + [\Sigma_R^<] + [\Sigma_{\text{ph}}^<]$  are, respectively, the retarded and the lesser-than self-energies accounting for the connection of the device to the external leads (i.e., left lead  $L$ , and right lead  $R$ ), and for the effects of a possible electron interaction with photons or phonons [33]. All matrices in Eqs. (37) and (38) have the same rank as  $[\mathbf{H}_{x\mathbf{K}_{yz}}]$ , which makes the complete solution computationally intractable.

Thanks to the block tridiagonal form of the  $[\mathbf{H}_{x\mathbf{K}_{yz}}]$  in the hybrid basis, however, we can limit the calculation of the Green's functions matrices to the blocks on the main diagonal and of the first diagonal above and below the main diagonal, which are necessary to express the carrier density and the spatial distribution of the current. These blocks can be computed with the recursive Green's function algorithms based on the Dyson equation [33], with a computational burden that is set by the rank of the diagonal and off-diagonal blocks in Eq. (24) or (32), as discussed in Ref. [40].

As for the calculation of the self-energies of the leads, we note that in the contacts the external potential  $V_{\text{ext}}$  is zero, hence Eq. (25) shows that all diagonal blocks are the same  $\mathbf{H}_{l,l} \approx \mathbf{H}_{0,0}$  and the system is periodic in the  $x$  direction. Consequently, the surface Green's function  $\mathbf{G}_{0,0}$  of a semi-infinite chain of unit cells can be calculated by means of the Sancho-Rubio algorithm [41], or the eigenvalue method [42,43], that is by dealing only with the diagonal and off-diagonal blocks of Eq. (24) or (32). Then the self-energy of the left lead, for example, can be obtained as  $[\Sigma_L] = \mathbf{H}_{0,1}^\dagger \mathbf{G}_{0,0} \mathbf{H}_{0,1}$  and  $[\Sigma_L^<] = -([\Sigma_L] - [\Sigma_L]^\dagger)f_L$ , with  $f_L$  being the Fermi-Dirac occupation function of the left reservoir. An entirely similar expression holds for the self-energies  $[\Sigma_R]$ ,  $[\Sigma_R^<]$  of the right lead.

The above equations illustrate that in our formalism the inelastic scattering with phonons or photons can be included quite naturally by adding the appropriate expressions for the retarded and lesser-than self-energies  $[\Sigma_{\text{ph}}]$ ,  $[\Sigma_{\text{ph}}^<]$  [33,34]. The



results reported in this paper, however, are restricted to the ballistic transport regime.

### B. Charge, current, and self-consistent calculations

By treating the electron-electron interaction according to the Hartree approximation, we here couple the charge computed with the NEGF solver and the electrostatic potential by means of the 3D Poisson equation,

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -e[p(\mathbf{r}) - n(\mathbf{r}) + N_D(\mathbf{r}) - N_A(\mathbf{r})], \quad (39)$$

where  $e$  is the positive electron charge,  $\phi(\mathbf{r})$  is the electrostatic potential,  $\epsilon(\mathbf{r})$  is the material-dependent permittivity, and  $n(\mathbf{r})$ ,  $p(\mathbf{r})$ ,  $N_A(\mathbf{r})$ ,  $N_D(\mathbf{r})$  are the electron, hole, acceptor, and donor concentration, respectively.

In order to solve Eq. (39) it is necessary to express the carrier concentration in real space. To this purpose the electron concentration is first computed from the diagonal elements of the real-space, lesser-than Green's function,  $[\mathbf{G}_r^<]$ , which can be obtained by means of the unitary transformation  $[\mathbf{G}_r^<(E)] = [\mathbf{U}_{\mathbf{K}_{yz}, \mathbf{r}_{yz}}]^\dagger [\mathbf{G}_{x\mathbf{K}_{yz}}^<(E)] [\mathbf{U}_{\mathbf{K}_{yz}, \mathbf{r}_{yz}}]$ , and has the same spatial discretization  $d$  as the Hamiltonian matrix. More precisely, the electron concentration reads

$$n(\mathbf{r}) = \frac{-ig_s}{d^3} \int_{E_0(x_i)}^{\infty} \frac{dE}{2\pi} \mathbf{G}^<(\mathbf{r}, \mathbf{r}; E), \quad (40)$$

where  $g_s$  is the spin degeneracy,  $E_0(x_i)$  is the neutrality energy level at the abscissa  $x_i$ , here assumed to be in the center of the energy gap. A similar equation holds for hole concentration,

$$p(\mathbf{r}) = \frac{ig_s}{d^3} \int_{-\infty}^{E_0(x_i)} \frac{dE}{2\pi} \mathbf{G}^>(\mathbf{r}, \mathbf{r}; E), \quad (41)$$

where  $[\mathbf{G}_r^>(E)]$  is the real-space, greater-than Green's function defined as  $[\mathbf{G}_r^>(E)] = [\mathbf{G}_r^<(E)] + [\mathbf{G}_r(E)] - [\mathbf{G}_r(E)]^\dagger$ .

Then, because the electrostatic potential  $\phi(\mathbf{r})$  has fairly slow variations on the scale of the lattice constant  $a_0$ , the electron and hole concentrations are interpolated on a coarser mesh with a discretization  $d_c$  larger than  $d$  (typically  $d_c \simeq a_0/2$ ).

The spatial distribution of the current along the transport direction is expressed in the hybrid basis as

$$I_{x_l \rightarrow x_{l+1}} = \frac{g_s e}{\hbar} \int dE \operatorname{tr}\{[\mathbf{H}_{l,l+1} \mathbf{G}_{l+1,l}^< - \mathbf{G}_{l,l+1}^< \mathbf{H}_{l+1,l}]\}, \quad (42)$$

where  $\operatorname{tr}\{\dots\}$  denotes the trace of a matrix, and  $\mathbf{H}_{l,l+1}$ ,  $\mathbf{G}_{l+1,l}^<$  are the blocks of  $[\mathbf{H}_{x\mathbf{K}_{yz}}]$  and  $[\mathbf{G}_{x\mathbf{K}_{yz}}^<]$  on the first diagonal above and below the main diagonal. Finally, in the ballistic limit the current  $I_{LR}$  between the left and right contact is given by the Landauer formula

$$I_{LR} = \frac{g_s e}{\hbar} \int dE \operatorname{tr}\{[\Gamma_L][\mathbf{G}_{x\mathbf{K}_{yz}}][\Gamma_R][\mathbf{G}_{x\mathbf{K}_{yz}}]^\dagger\} \times [f_L(E) - f_R(E)], \quad (43)$$

where  $[\Gamma_{L(R)}] = i([\Sigma_{L(R)}] - [\Sigma_{L(R)}]^\dagger)$  and  $f_{L(R)}$  is the Fermi-Dirac occupation function at the  $L(R)$  contact.

A self-consistent calculation of the carrier densities and electrostatic potential is obtained by iteratively calculating the retarded and lesser-than Green's functions, and solving the Poisson equation Eq. (39) until convergence is reached. Equation (39) is numerically solved by using the well-known nonlinear formulation of the Poisson equation, which allows one to define a Jacobian matrix and thus employ the Newton-Raphson method.

### V. IMPLEMENTATION AND COMPUTATIONAL COMPLEXITY

In order to set up our simulation framework we first have to identify the necessary reciprocal lattice vectors  $\mathbf{G}$  and the shape of the reduced zone of the bulk semiconductor. The maximum magnitude of the  $\mathbf{G}$  vectors is set by the spatial discretization mesh  $d = a_0/N_d$  according to Eq. (10). In this work, we verified that the minimum  $N_d$  necessary to obtain accurate band-structure calculations for silicon and germanium is  $N_d = 8$ , and we accordingly used a cubic expansion volume counting a number of  $\mathbf{G}$  vectors  $N_G = 128$  for all simulations.

We think it is useful to point out a few summarizing remarks about the reduced zone of the bulk semiconductor that we recommend to use for either band-structure or transport calculations. In fact for the purposes of this work it is convenient to tailor the reduced zone of the bulk crystal differently from the 3D first Brillouin zone.

As far as transport calculations are concerned, the reduced zone defined as  $-\pi/a_0 \leq k_x < \pi/a_0$ ,  $-\pi/a_0 \leq k_y < \pi/a_0$ , and  $-\pi/a_0 \leq k_z < \pi/a_0$  is appropriate for all electron gas dimensionalities because it guarantees that, for any  $\mathbf{K}_{yz} = [(k_y, k_z) + (G_y, G_z)]$  in Eqs. (24), (32), and (A5), the corresponding  $x$  components  $K_x = k_x + G_x$  cover with no voids the entire extended  $K_x$  range  $[-(N_d/2)(2\pi/a_0), (N_d/2)(2\pi/a_0)]$  defined by Eq. (2).

For a coherent transport such a reduced zone ensures also that, for a 3DEG with an external potential constant along the  $(y, z)$  plane transverse to the transport direction, a separated transport problem (with a corresponding transmission coefficient) can be written for any  $(k_y, k_z)$  with  $-\pi/a_0 \leq k_y < \pi/a_0$  and  $-\pi/a_0 \leq k_z < \pi/a_0$ . Similarly, for a 2DEG with an external potential constant along the  $y$  direction, a separated transport problem can be written for any  $k_y$  with  $-\pi/a_0 \leq k_y < \pi/a_0$ . No separation is instead possible for a 1DEG. The size of the transport problem is thus set by the rank of the blocks of the matrices in Eqs. (24), (32), and (A5). It is understood, however, that even for a 3DEG and a 2DEG the self-consistent calculation of charge density and electrostatic potential indirectly couples the otherwise separated transport problems.

For the band-structure calculation of a 2DEG the bulk crystal reduced zone defined above for transport calculations leads also to a separated eigenvalue problem for any  $(k_x, k_y)$ , consequently the reduced zone of the 2DEG can be defined as  $-\pi/a_0 \leq k_x < \pi/a_0$ ,  $-\pi/a_0 \leq k_y < \pi/a_0$ . For the band structure of a 1DEG, instead, the reduced zone defined for transport does not enable the separation of the eigenvalue problems for different  $k_x$ , which can be obtained by defining the bulk crystal reduced zone as  $-\pi/a_0 \leq k_y, k_z < \pi/a_0$  and  $-\pi/a_0 \leq k_x < \pi/a_0$ , that in turn results in a reduced zone for the 1DEG given by  $-\pi/a_0 \leq k_x < \pi/a_0$ .

We conclude this section with a comment about the rank of the blocks of the device Hamiltonian matrices in Eqs. (24), (32), and (A5). The overall number of wave vectors in the extended  $\mathbf{K}$  space defined in Sec. II is  $N_T = 4N_G N_{cx} N_{cy} N_{cz}$ , and it is equal to the number of points in the discrete real space lattice. In fact  $N_T$  can be written as  $N_G N_{\text{red}}$ , where  $N_{\text{red}} = 4N_{cx} N_{cy} N_{cz}$  is the number of wave vectors in the reduced zone of the bulk crystal, that in turn can be inferred

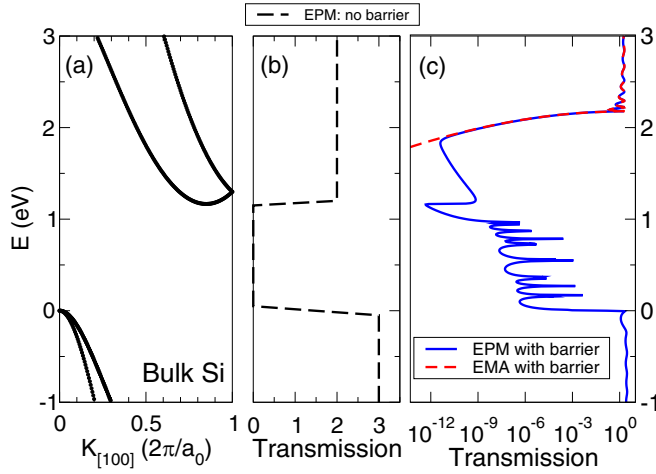


FIG. 3. (a) Bulk Si band structure along the  $\Delta$  direction, where the reference energy is taken at the top of the valence band. (b) Transmission probability through the band structure with no energy barrier. (c) Transmission probability in the presence of a rectangular energy barrier (with a barrier height  $E_B = 1$  eV and a width  $W_B = 5.43$  nm) and computed either with the EP model (blue solid line) or with the EMA Hamiltonian (red dashed line). Transport direction is along the  $[100]$  direction.

by the volume  $4(2\pi/a_0)^3$  of the reduced zone and the density of states  $(2\pi)^3/(L_x L_y L_z)$  in  $\mathbf{K}$  space. For a 3D electron gas each block in Eq. (24) corresponds to a fixed  $(k_y, k_z)$  in the reduced zone (with  $N_{k_y} = N_{c_y}$ ,  $N_{k_z} = 2N_{c_z}$  being the number of, respectively,  $k_y$ ,  $k_z$  values in the reduced zone), and to an  $a_0$  long device section along  $x$  (namely  $N_{c_x}$  times shorter than  $L_x = N_{c_x} a_0$ ). Consequently the size of the blocks in Eq. (24) is given by  $N_T/[N_{c_x} N_{c_y} (2N_{c_z})] = 2N_G$ . For a 2D electron gas each block in Eq. (32) corresponds to a fixed  $k_y$  and an  $a_0$  long device section, so that the size of the blocks is  $N_T/(N_{c_x} N_{c_y}) = 4N_G N_{c_z}$ . Finally for a 1D gas, because no separation is possible for transverse  $\mathbf{k}$  components, the size of the blocks in Eq. (A5) is simply given by  $N_T/N_{c_x} = 4N_G N_{c_y} N_{c_z}$ .

## VI. SIMULATION RESULTS

### A. Transmission through an analytical energy barrier

We start this section of simulation results showing a simple example consisting in the transmission through a rectangular barrier, in fact in this case a comparison with an analytical calculation based on the EMA is possible. The energy barrier is superimposed to the atomic scale pseudopotentials of the underlying silicon crystal and it has a width  $W_B = 10 a_0 = 5.43$  nm along the transport direction  $x$  (taken as  $[100]$ ), and an energy barrier height  $E_B = 1$  eV; such a potential energy profile creates a forbidden region for electrons and a quantum well for holes along  $x$ .

Figure 3 illustrates the band structure of bulk silicon close to the energy gap and the corresponding transmission at  $(k_y, k_z) = (0, 0)$  calculated with the formalism presented in Sec. III, and for a system with or without the above mentioned energy barrier. In particular, Fig. 3(b) shows the transmission for the case with no energy barrier, where sharp steps are observed at the energies corresponding to the band edges,

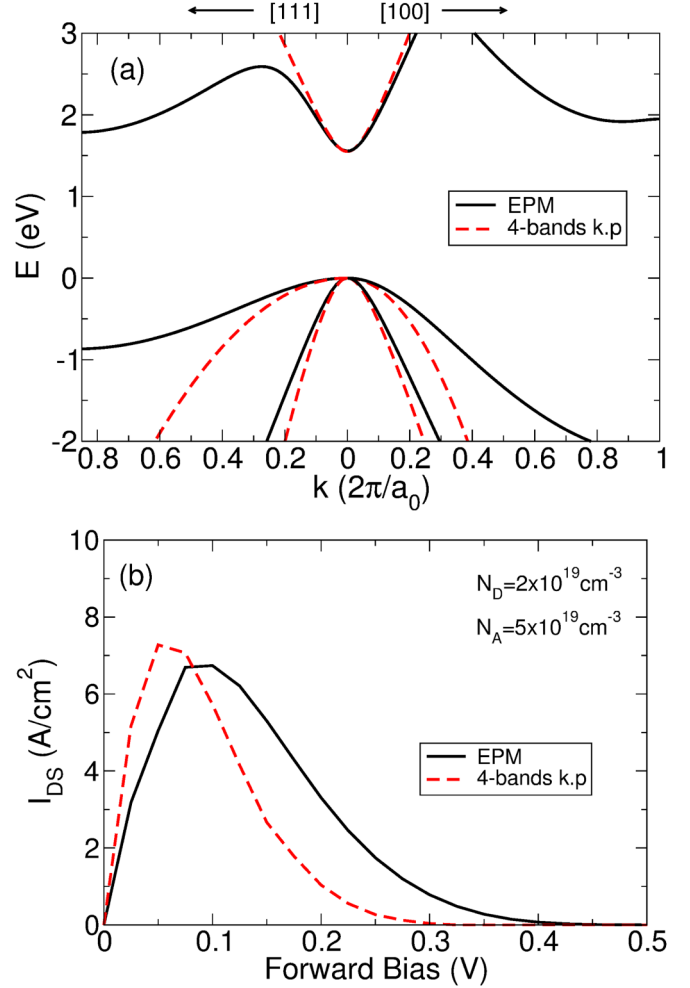


FIG. 4. (a) GaAs band structure in proximity of the  $\Gamma$  point obtained either with the EPM (solid black line) or with the four-band  $\mathbf{k}\cdot\mathbf{p}$  Hamiltonian (dashed red line). The parameters for EPM calculations are reported in Table I, while the  $\mathbf{k}\cdot\mathbf{p}$  model is based on the formulation in Ref. [47] and the parameters are  $\gamma_1^L = 6.98$ ,  $\gamma_2^L = 2.06$ ,  $\gamma_3^L = 2.93$ ,  $m_c = 0.067$ ,  $E_p = 20.8$  eV,  $E_G = 1.55$  eV, and  $\Delta_{so} = 0$  eV. (b) Simulated current versus forward bias for the GaAs Esaki diode obtained either with the EPM (solid black line) or with the four-band  $\mathbf{k}\cdot\mathbf{p}$  Hamiltonian (dashed red line). The doping concentration is  $N_D = 2 \times 10^{19} \text{ cm}^{-3}$  for the  $n$ -doped region and  $N_A = 5 \times 10^{19} \text{ cm}^{-3}$  for the  $p$ -doped region. Transport direction is along the  $[100]$  direction.

with two modes available in the conduction band and three modes in the valence band. Figure 3(c) illustrates the results in the presence of the energy barrier, and it can be seen that the transmission is much smaller than one for an energy up to about 1 eV above the bottom of the conduction band. Moreover, for energies close to the top of the energy barrier (i.e., around 2 eV) the decay of the transmission calculated with the EP method (solid line) is tracked very well by the analytical results obtained using a simple EMA model (dashed line) and using the silicon longitudinal effective mass  $0.91 m_0$ . When the energy falls deeper below the top of the energy barrier, however, a large discrepancy is observed between EMA and EP results, because in the EP model the transmission

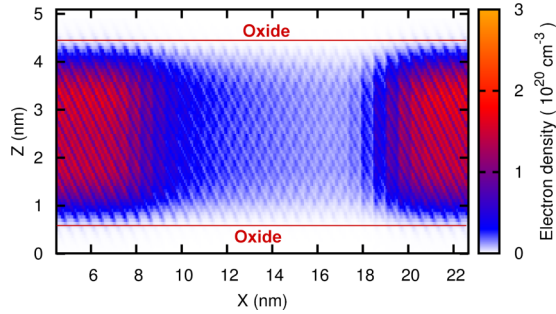


FIG. 5. Cross section of the simulated electron concentration at  $V_{GS} = 0.6$  V and  $V_{DS} = 0.6$  V for an  $n$ -type, silicon FET with gate length  $L_G \simeq 10$  nm and  $T_w = 7a_0 \simeq 3.8$  nm.

reaches a minimum and then increases for decreasing energy. For energies close to the conduction band edge, in particular, the transmission through energy barrier occurs mainly via the evanescent states of the valence band (whose top is at  $E = 1$  eV in the energy barrier region), which are not at all considered in the EMA model.

### B. GaAs Esaki diode

As a first device application for a 3DEG system we report simulation results for a GaAs Esaki diode consisting in a highly doped  $p$ - $n$  junction. This is a prototypical device demanding a quantum transport simulation approach because the forward bias current is determined by electron tunneling from the valence to the conduction band.

Figure 4 shows a comparative analysis for simulation results obtained either with the EPM Hamiltonian of this work or with a four-band  $\mathbf{k}\cdot\mathbf{p}$  Hamiltonian developed for the  $\Gamma$  point. The four-band  $\mathbf{k}\cdot\mathbf{p}$  model is a simplified version of the eight-band  $\mathbf{k}\cdot\mathbf{p}$  Hamiltonian discussed and employed in some of our previous publications [15,44], where we have removed the spin-orbit interaction in order to have a fair comparison with the EPM Hamiltonian used in this work, which does not include spin-orbit related terms. Figure 4(a) reports the bulk GaAs band structure in proximity of the  $\Gamma$  point and shows that the agreement between the  $\mathbf{k}\cdot\mathbf{p}$  and the EPM band structure is good close to the valence and conduction band edge.

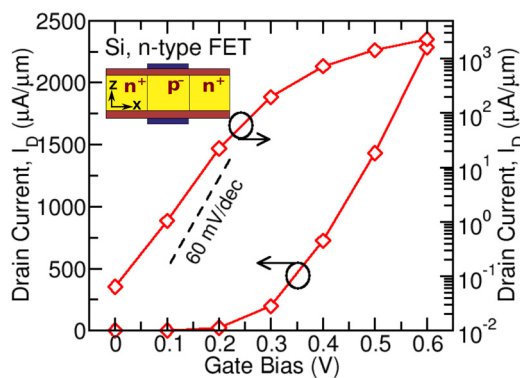


FIG. 6. Simulated  $I_{DS}$  versus  $V_{GS}$  characteristic at  $V_{DS} = 0.6$  V for an  $n$ -type, silicon FET with gate length  $L_G \simeq 10$  nm and  $T_w = 7a_0 \simeq 3.8$  nm.

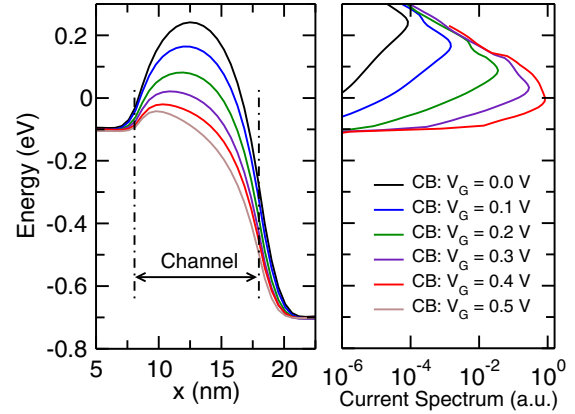


FIG. 7. Conduction sub-band profile (left) and corresponding current density plot (right) for the  $n$ -type silicon FET of Fig. 6 and at  $V_{DS}$  is 0.6 V.

As for the GaAs Esaki diode, we here performed self-consistent simulations of a diode having a donor and an acceptor-type dopant concentration of, respectively,  $N_D = 2 \times 10^{19} \text{ cm}^{-3}$  and  $N_A = 5 \times 10^{19} \text{ cm}^{-3}$ , and assumed that the external potential varies only along the longitudinal direction ( $x$  axis). Figure 4(b) reports the current versus voltage characteristic under forward bias, which exhibits the negative differential resistance behavior distinctive of an Esaki diode. The  $I$ - $V$  curves obtained with the two different Hamiltonians are in fairly good agreement especially in terms of peak current, suggesting that the current is dominated by direct tunneling at the  $\Gamma$  point and for an energy range where the band structure obtained with the two models is in good agreement.

### C. Silicon ultra-thin-body FET

We here report simulations for an ultra-thin-body silicon  $n$ -FET having a silicon thickness  $T_w = 7a_0 \simeq 3.8$  nm and a gate length  $L_G \simeq 10$  nm. The confinement along the  $z$  direction is described using an energy barrier region having band edge discontinuities  $V_{cb} = 3$  eV and  $V_{vb} = -3$  eV (respectively, for the conduction and valence band), and an overall thickness in the  $z$  direction of  $2a_0$ . In fact we verified that such a barrier thickness is sufficient to ensure an appropriate suppression of the electron wave function in the oxide region, namely an appropriate confinement of the electron wave function and electron concentration in the semiconductor region. This is exemplified in Fig. 5 reporting the self-consistent electron density at  $V_{GS} = 0.6$  V and  $V_{DS} = 0.6$  V, where the reader can also recognize the atomic scale features of the carrier concentration due to the periodic parts of the crystal Bloch functions. The same parameters for the energy barrier region have been used also for the simulations of thin-body TFETs in Sec. VID.

Figure 6 reports the results of self-consistent simulations for the  $I_{DS}$  versus  $V_{GS}$  curve at  $V_{DS} = 0.6$  V. As it can be seen the current characteristics at  $V_{DS} = 0.6$  V are well behaved and the quite large  $I_{DS}$  values are due to the fact that neither scattering nor series resistance effects are included. Figure 7 illustrates the corresponding profile of the lowest conduction subband along the channel and the current spectral density

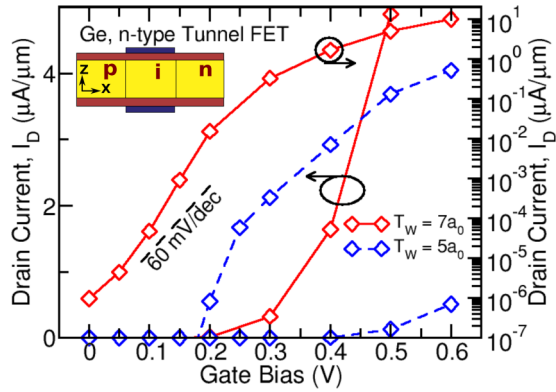


FIG. 8. Simulated drain current versus gate voltage characteristics for an  $n$ -type, germanium tunnel FET at  $V_{DS} = 0.4$  V with  $T_w$  of either  $7a_0 \simeq 3.95$  nm or  $5a_0 \simeq 2.82$  nm. Gate length is  $L_G \simeq 17$  nm.

$J(E)$ . As it can be seen the peak of  $J(E)$  follows closely the top of the barrier energy because the source-drain tunneling component is small in silicon at  $L_G \simeq 10$  nm [6].

#### D. Germanium ultra-thin-body tunnel FET

To illustrate the flexibility of the new transport methodology, we show in Fig. 8 the self-consistently calculated  $I_{DS}$  versus  $V_{GS}$  characteristics at  $V_{DS} = 0.4$  V for an  $n$ -type, germanium tunnel FET with two different semiconductor film thicknesses  $T_w$ . The TFET exhibits a subthreshold swing (SS) below 60 mV/dec and, as expected, the minimum SS improves with reducing  $T_w$  [6]. Figure 9 reports the subband profile and the corresponding  $J(E)$ , where the tunneling from the source valence band to the conduction band in the channel is clearly observed. We here emphasize that BTBT in an indirect bandgap semiconductor, such as germanium, cannot be described with a  $\mathbf{k}\cdot\mathbf{p}$  method restricted to the  $\Gamma$  point, and it is a distinct feature of a full-band transport model, such as the pseudopotential NEGF method developed in this work.

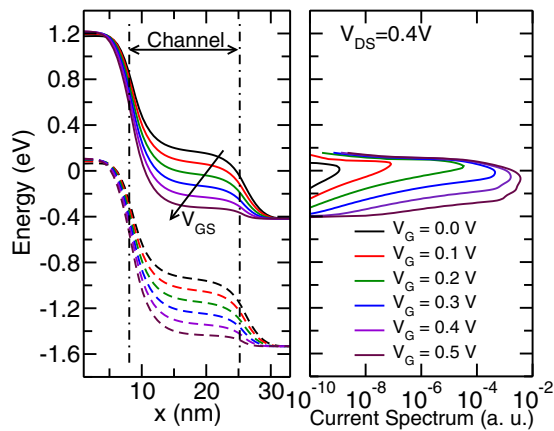


FIG. 9. Conduction and valence sub-band profile (left) and corresponding current density plot (right) for the germanium tunnel FETs of Fig. 8 with  $T_w = 7a_0$ .

## VII. CONCLUSIONS

We have presented a full-band quantum transport modeling approach based on the NEGF formalism and an empirical pseudopotential Hamiltonian. The model relies on a hybrid real-space/plane-wave basis that improves the computational efficiency compared to a full plane-wave approach, and has been described for different electron gas dimensionalities, which makes it in principle suitable for a large variety of devices.

We have implemented the proposed approach in a self-consistent simulation framework, and reported complete device simulations for a tunneling diode, and for ultra-thin-body transistors (both FETs and tunnel FETs) with geometrical features representative of forthcoming CMOS technologies.

The paper delivers novel contributions in several aspects (already mentioned in Sec. I), which include the hybrid real-space/plane-wave basis itself, the treatment of quantum confinement by means of an on-purpose developed, nonlocal confining operator, a discussion of the separability of either the band structure or transport problems leading to definitions of the reduced zone alternative to the conventional first Brillouin zone, and finally the size of the systems addressed by our self-consistent simulations, that is substantially larger than reported in previous papers using a similar Hamiltonian for quantum transport calculations.

While all the results in this work have been obtained using a local EP formulation and assuming a coherent transport regime, a model extension to nonlocal pseudopotentials seems possible and, moreover, the NEGF-based formalism is in principle capable of including also scattering with phonons or photons. Both these topics are outside the scope of the present work, but they are interesting directions for future developments.

The computational complexity of the method remains a challenge, in terms of both memory occupation and CPU time. In particular, the size of the problem increases for a 2D and especially for a 1D electron gas, because the quantum confinement precludes the separation of the problem in transverse modes identified by some  $\mathbf{k}$  components normal to the transport direction.

We believe that the formalism developed in this work may find useful applications also for transport methods based on density functional theory Hamiltonians, which typically rely on a full plane-wave basis. Further investigations are necessary to establish how useful and feasible the transport method of this work can be for an accurate analysis of nanoscale electronic and photonic devices.

## ACKNOWLEDGMENT

Authors are thankful to Oves Badami for fruitful discussions and for his help with some of the simulations reported in the paper.

## APPENDIX: CONFINEMENT AND TRANSPORT FOR A 1D ELECTRON GAS (1DEG)

The quantization and transport for a quasi-1D electron gas (1DEG), namely a gas where quantum confinement occurs

along both directions normal to the transport direction  $x$ , can be described through an extension of the approach described for a 2D electron gas in Secs. II B and III B. The step function  $\Theta(y, z)$  used to describe the quantum confinement in a 1DEG can be written in  $\mathbf{K}$  space as

$$\Theta(\mathbf{K}, \mathbf{K}') = \delta_{K_x, K'_x} \Theta_{yzF}(\mathbf{K}_{yz} - \mathbf{K}'_{yz}), \quad (\text{A1})$$

where  $\Theta_{yzF}(\mathbf{K}_{yz} - \mathbf{K}'_{yz})$  is the 2D discrete Fourier transform in the  $(y, z)$  plane.

### 1. Band-structure calculations

For band-structure calculations it is convenient to define the reduced zone of the bulk crystal with inequalities  $-\pi/a_0 \leq k_y, k_z < 2\pi/a_0$ , and  $-\pi/a_0 \leq k_x < \pi/a_0$  which ensure that, for any two different wave vectors  $\mathbf{k} \neq \mathbf{k}'$  in the reduced zone, none of the corresponding wave vectors in the extended  $\mathbf{K}$  space have the same  $x$  component; namely for any two reciprocal lattice vectors  $\mathbf{G}, \mathbf{G}'$  we have  $(k_x + G_x) \neq (k'_x + G'_x)$ .

Such a definition of the reduced zone implies that the unitary step function  $\Theta(n\mathbf{k}, n'\mathbf{k}')$  in the Bloch function basis is non-null only for Bloch functions  $\Phi_{n\mathbf{k}_x\mathbf{k}_{yz}}, \Phi_{n'\mathbf{k}'_x\mathbf{k}'_{yz}}$  corresponding to the same  $k_x$ . Consequently the transformation of the step function from plane waves to the Bloch function basis is governed by a separated unitary matrix  $\mathbf{U}_{\mathbf{K}, \Phi}^{(k_x)}$  for  $-\pi/a_0 \leq k_x < \pi/a_0$ . The unitary matrix  $\mathbf{U}_{\mathbf{K}, \Phi}^{(k_x)}$  is defined similarly to Eq. (16), but the rank of the matrix for a 1DEG is  $N_G N_{k_y} N_{k_z}$ , with  $N_{k_y}, N_{k_z}$  being the number of, respectively,  $k_y, k_z$  values in the reduced zone of the bulk crystal.

In this case the passages from Eq. (17) to Eq. (23) remain valid for a 1DEG, so that we can write the overall Hamiltonian in  $\mathbf{K}$  space as

$$\mathbf{H}_{k_x}(\mathbf{K}, \mathbf{K}') = T(\mathbf{k} + \mathbf{G})\delta_{\mathbf{k}_{yz}, \mathbf{k}'_{yz}} \delta_{\mathbf{G}, \mathbf{G}'} + V_L(\mathbf{G} - \mathbf{G}')\delta_{\mathbf{k}_{yz}, \mathbf{k}'_{yz}} + \mathbf{H}^{(c)}(\mathbf{K}, \mathbf{K}'), \quad (\text{A2})$$

where  $\mathbf{K} = [(k_x, \mathbf{k}_{yz}) + \mathbf{G}]$ ,  $\mathbf{K}' = [(k_x, \mathbf{k}'_{yz}) + \mathbf{G}']$ , where  $\mathbf{H}^{(c)}(\mathbf{K}, \mathbf{K}')$  is given by

$$[\mathbf{H}_{\mathbf{K}}^{(c)}] = [\Theta_{\mathbf{K}}][\mathbf{D}_{c, \mathbf{K}}], \quad (\text{A3})$$

with

$$[\mathbf{D}_{c, \mathbf{K}}] = [\mathbf{U}_{\mathbf{K}, \Phi}^{(k_x)}][\mathbf{D}_{c, \Phi}][\mathbf{U}_{\mathbf{K}, \Phi}^{(k_x)}]^\dagger. \quad (\text{A4})$$

It is understood that  $\Theta(\mathbf{K}, \mathbf{K}')$  is now given by Eq. (A1) and that all matrices in Eq. (A2) have a rank  $N_G N_{k_y} N_{k_z}$ , with  $N_{k_y}, N_{k_z}$  being the number of  $k_y, k_z$  in the reduced zone of the bulk crystal. The band structure of the 1DEG gas is obtained varying  $k_x$  in the range  $-\pi/a_0 \leq k_x < \pi/a_0$ .

### 2. Transport

As for the transport in a 1DEG, let us consider a system confined along  $y, z$  and with a length  $L_x = N_{cx}a_0$ , where  $N_{cx}, N_{cy}, N_{cz}$  are the number of unit cells, respectively, in the transport direction  $x$  and in the confinement directions  $y, z$ .

We start by recalling that, as already mentioned after Eqs. (24) and (32), for transport it is necessary that the reduced zone of the bulk crystal is defined such that, for each transverse wave vector  $\mathbf{K}_{yz} = [(k_y, k_z) + (G_y, G_z)]$  in the extended  $\mathbf{K}$  space, the corresponding  $x$  components  $K_x = k_x + G_x$  cover with no voids the entire extended  $K_x$  range  $[-(N_d/2)(2\pi/a_0), (N_d/2)(2\pi/a_0)]$ . This condition is not fulfilled by the reduced zone for band-structure calculations introduced in the previous subsection and corresponding to  $-\pi/a_0 \leq k_x < \pi/a_0$ . Consequently, for transport calculations we have to use a reduced zone defined by inequalities  $-\pi/a_0 \leq k_x < 2\pi/a_0, -\pi/a_0 \leq k_y < \pi/a_0$  and  $-\pi/a_0 \leq k_z < 2\pi/a_0$ , which is thus suitable for transport calculations for all electron gas dimensionalities. This definition implies that the unitary step function  $\Theta(n\mathbf{k}, n'\mathbf{k}')$  in the Bloch function basis is non-null only for Bloch functions  $\Phi_{n\mathbf{k}_x\mathbf{k}_{yz}}, \Phi_{n'\mathbf{k}'_x\mathbf{k}'_{yz}}$  corresponding to couples of  $k_x$  and  $k'_x$  such that  $|k_x - k'_x| = 2\pi/a_0$ . Consequently the transformation of the step function in Eq. (A1) from a plane-wave to Bloch-function basis is governed by a separated unitary matrix,  $\mathbf{U}_{\mathbf{K}, \Phi}^{(k_x, k'_x)}$ , identified by the couple  $(k_x, k'_x)$  with  $-\pi/a_0 \leq k_x, k'_x < 2\pi/a_0$ . The unitary matrix  $\mathbf{U}_{\mathbf{K}, \Phi}^{(k_x, k'_x)}$  is defined similarly to Eq. (16), but the rank of the matrix for a 1DEG is  $2N_G N_{k_y} N_{k_z}$ .

In the hybrid basis the Hamiltonian with closed boundary conditions along  $x$  can still be written,

$$[\mathbf{H}_{x\mathbf{K}_{yz}}] = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{0,1} & 0 & 0 & \cdots & 0 \\ \mathbf{H}_{0,1}^\dagger & \mathbf{H}_{2,2} & \mathbf{H}_{0,1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{H}_{0,1}^\dagger & \mathbf{H}_{N_{cx}, N_{cx}} \end{bmatrix}, \quad (\text{A5})$$

where the  $\mathbf{K}_{yz}$  are the  $(y, z)$  components of the wave vectors  $\mathbf{K} = [(k_x, k_y, k_z) + \mathbf{G}]$  involved in the solution of the  $N_{k_x} N_{k_y} N_{k_z}$  eigenvalue problems given by Eq. (5) for  $\mathbf{k} = (k_x, k_y, k_z)$  and with  $-\pi/a_0 \leq k_x, k_z < 2\pi/a_0, -\pi/a_0 \leq k_y < \pi/a_0$ . Each block  $\mathbf{H}_{l,l}, \mathbf{H}_{0,1}$  has now a  $2N_G N_{k_y} N_{k_z}$  rank, with  $N_{k_y} = N_{cy}, N_{k_z} = 2N_{cz}$  being the number of, respectively,  $k_y, k_z$  values in the reduced zone of the bulk crystal (see the end of Sec. V for more details). As it can be seen Eq. (A5) is similar to Eq. (32), however, the Hamiltonian in Eq. (A5) couples all the  $k_y, k_z$  in the plane normal to the transport direction and, differently from the  $[\mathbf{H}_{x\mathbf{K}_{yz}}]$  in Eq. (32), no separation of the problem is possible for any transverse wave-vector component.

Equations (33)–(36) remain formally valid for a 1DEG system, however, all the blocks have now a  $2N_G N_{k_y} N_{k_z}$  rank, which increases substantially the computational burden compared to a 2DEG system.

- [1] *The International Technology Roadmap for Semiconductors (ITRS)* (IEEE, Piscataway, 2015).
- [2] J. Del Alamo, *Nature (London)* **479**, 317 (2011).
- [3] G. Doornbos and M. Passlack, *IEEE Electron. Device Lett.* **31**, 1110 (2010).
- [4] M. Passlack, R. Droopad, and G. Brammertz, *IEEE Trans. Electron Devices* **57**, 2944 (2010).
- [5] R. Kim, U. E. Avci, and I. A. Young, *IEEE Trans. Electron Devices* **62**, 713 (2015).
- [6] D. Esseni, M. Pala, and T. Rollo, *IEEE Trans. Electron Devices* **62**, 3084 (2015).
- [7] C. Grillet, D. Logoteta, A. Cresti, and M. G. Pala, *IEEE Trans. Electron Devices* **64**, 2425 (2017).
- [8] A. Seabaugh and Q. Zhang, *Proc. IEEE* **98**, 2095 (2010).
- [9] D. Esseni, M. Pala, P. Palestri, C. Alper, and T. Rollo, *Semicond. Sci. Technol.* **32**, 083005 (2017).
- [10] S. De Franceschi, L. Hutin, R. Maurand, L. Bourdet, H. Bohuslavskyi, A. Corna, D. Kotekar-Patil, S. Barraud, X. Jehl, M. S. Y.-M. Niquet, and M. Vinet, *Proceedings of the IEEE International Electron Devices Meeting, San Francisco, CA* (IEEE, Piscataway, 2016), pp. 13.4.1-4.
- [11] R. Venugopal, Z. Ren, S. Datta, and M. Lundstrom, *J. Appl. Phys.* **92**, 3730 (2002).
- [12] J. Wang, E. Polizzi, and M. Lundstrom, *J. Appl. Phys.* **96**, 2192 (2004).
- [13] S. Poli, M. G. Pala, T. Poiroux, S. Deleonibus, and G. Baccarani, *IEEE Trans. Electron Devices* **55**, 2968 (2008).
- [14] M. Shin, *J. Appl. Phys.* **106**, 054505 (2009).
- [15] F. Conzatti, M. Pala, D. Esseni, E. Bano, and L. Selmi, *IEEE Trans. Electron Devices* **59**, 2085 (2012).
- [16] E. Baravelli, E. Gnani, R. Grassi, A. Gnudi, S. Reggiani, and G. Baccarani, *IEEE Trans. Electron Devices* **61**, 178 (2014).
- [17] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, *Phys. Rev. B* **74**, 205323 (2006).
- [18] G. Klimeck, S. S. Ahmed, H. Bae, N. Kharche, R. Rahman, S. Clark, B. Haley, S. Lee, M. Naumov, H. Ryu, F. Saied, M. Prada, M. Korkusinski, and T. B. Boykin, *IEEE Trans. Electron Devices* **54**, 2079 (2007).
- [19] J. Taylor, H. Guo, and J. Wang, *Phys. Rev. B* **63**, 121104 (2001).
- [20] Y. Fujimoto and K. Hirose, *Phys. Rev. B* **67**, 195315 (2003).
- [21] T. Ono, Y. Egami, and K. Hirose, *Phys. Rev. B* **86**, 195406 (2012).
- [22] H. Takaki, N. Kobayashi, and K. Hirose, *J. Nanomater.* **2014**, 172169 (2014).
- [23] L.-W. Wang, A. Franceschetti, and A. Zunger, *Phys. Rev. Lett.* **78**, 2819 (1997).
- [24] L.-W. Wang and A. Zunger, *Phys. Rev. B* **59**, 15806 (1999).
- [25] F. Chirico, A. Di Carlo, and P. Lugli, *Phys. Rev. B* **64**, 045314 (2001).
- [26] D. Esseni and P. Palestri, *Phys. Rev. B* **72**, 165342 (2005).
- [27] J.-L. van der Steen, D. Esseni, P. Palestri, L. Selmi, and R. J. E. Huetting, *IEEE Trans. Electron Devices* **54**, 1843 (2007).
- [28] H. J. Choi and J. Ihm, *Phys. Rev. B* **59**, 2267 (1999).
- [29] X.-W. Jiang, S.-S. Li, J.-B. Xia, and L.-W. Wang, *J. Appl. Phys.* **109**, 054503 (2011).
- [30] A. Garcia-Lekue, M. Vergniory, X. Jiang, and L. Wang, *Prog. Surf. Sci.* **90**, 292 (2015).
- [31] J. Fang, W. G. Vandenberghe, B. Fu, and M. V. Fischetti, *J. Appl. Phys.* **119**, 035701 (2016).
- [32] J. Fang, S. Chen, W. G. Vandenberghe, B. Fu, and M. V. Fischetti, *IEEE Trans. Electron Devices* **64**, 2758 (2017).
- [33] G.D. Mahan, *Many-Particle Physics* (Plenum Press, New York, 1988).
- [34] S. Datta, *Quantum Transport—Atom to Transistor* (Cambridge University Press, Cambridge, 2005).
- [35] M. G. Pala, O. Badami, and D. Esseni, *Proceedings of the IEEE International Electron Devices Meeting, San Francisco, CA* (IEEE, Piscataway, 2017), pp. 35.1.1-4.
- [36] J. R. Chelikowsky and M. L. Cohen, *Phys. Rev. B* **10**, 5095 (1974).
- [37] M. L. Cohen and J. R. Chelikowsky, *Electron Structure and Optical Properties of Semiconductors*, Springer Series in Solid-State Sciences (Springer-Verlag, Berlin, Heidelberg, 1988).
- [38] J. R. Chelikowsky, N. Troullier, and Y. Saad, *Phys. Rev. Lett.* **72**, 1240 (1994).
- [39] V. Sverdlov, O. Baumgartner, H. Kosina, S. Selberherr, F. Schanovsky, and D. Esseni, in *Proceedings of the 13th International Workshop on Computational Electronics, Beijing* (IEEE, Piscataway, 2009), pp. 1–4.
- [40] M. P. Anantram, M. S. Lundstrom, and D. E. Nikonov, *Proc. IEEE* **96**, 1511 (2008).
- [41] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, *J. Phys. F: Metal Phys.* **14**, 1205 (1984).
- [42] D. H. Lee and J. D. Joannopoulos, *Phys. Rev. B* **23**, 4988 (1981).
- [43] D. H. Lee and J. D. Joannopoulos, *Phys. Rev. B* **23**, 4997 (1981).
- [44] M. Pala and D. Esseni, *IEEE Trans. Electron Devices* **60**, 2795 (2013).
- [45] L. R. Saravia and D. Brust, *Phys. Rev.* **176**, 915 (1968).
- [46] M. L. Cohen and T. K. Bergstresser, *Phys. Rev.* **141**, 789 (1966).
- [47] T. B. Bahder, *Phys. Rev. B* **41**, 11992 (1990).