Quantum Mechanical Solver for Confined Heterostructure Tunnel Field-Effect Transistors

UTD AUTHOR(S): William G. H. Vandenberghe

©2014 AIP Publishing LLC
Quantum mechanical solver for confined heterostructure tunnel field-effect transistors

Devin Verreck,1,2,a) Maarten Van de Put,1,3 Bart Soree,1,3 Anne S. Verhulst,1 Wim Magnus,1,3 William G. Vandenberghe,4 Nadine Collaert,1 Aaron Thean,1 and Guido Groeseneken1,2
1imec, Kapeldreef 75, 3001 Leuven, Belgium
2Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium
3Department of Physics, Universiteit Antwerpen, 2020 Antwerpen, Belgium
4Department of Materials Science and Engineering, University of Texas at Dallas, Richardson, Texas 75080, USA

(Received 28 November 2013; accepted 22 January 2014; published online 4 February 2014)

Heterostructure tunnel field-effect transistors (HTFET) are promising candidates for low-power applications in future technology nodes, as they are predicted to offer high on-currents, combined with a sub-60 mV/dec subthreshold swing. However, the effects of important quantum mechanical phenomena like size confinement at the heterojunction are not well understood, due to the theoretical and computational difficulties in modeling realistic heterostructures. We therefore present a ballistic quantum transport formalism, combining a novel envelope function approach for semiconductor heterostructures with the multiband quantum transmitting boundary method, which we extend to 2D potentials. We demonstrate an implementation of a 2-band version of the formalism and apply it to study confinement in realistic heterostructure diodes and p-n-i-n HTFETs. For the diodes, both transmission probabilities and current densities are found to decrease with stronger confinement. For the p-n-i-n HTFETs, the improved gate control is found to counteract the deterioration due to confinement. © 2014 AIP Publishing LLC.

I. INTRODUCTION

As scaling of semiconductor devices ventures into the nanometer realm, quantum mechanical (QM) effects are becoming increasingly important. These effects can be detrimental to device performance, causing, e.g., leakage currents through the gate oxide of a metal-oxide-semiconductor field-effect transistor (MOSFET).1 On the other hand, they give rise to novel transistor concepts that exploit the non-classical nature of the quantum world to achieve performance that remains otherwise unattainable. A prime example of this category is the tunnel field-effect transistor (TFET), which employs band-to-band tunneling (BTBT) to break the 60 mV/dec subthreshold swing limit at room temperature of the conventional MOSFET.2,3 The absence of a subthreshold limit allows to scale the supply voltage drastically, with the accompanying decrease in power consumption. Particularly promising is a TFET configuration which features a heterostructure of III-V materials at the tunnel junction, called a HTFET.4–6

However, in realizing an HTFET that outperforms a MOSFET, significant challenges remain not only for quantum transport but also for theoretical modeling. Current commercial device simulators are based on semi-classical (SC) models that apply approximations like the Kane formula to study BTBT, limiting their validity to the uniform field regime.7,8 They also fail to capture important QM effects like quantum confinement.9 SC simulators are therefore poorly suited for the optimization of TFETs and even less so for HTFETs.

Research has therefore been done to develop a fully QM description of TFET operation. However, the single-band effective mass approximation (EMA), often used in QM modeling,10,11 does not allow for direct bandgap BTBT. A multiband approach is therefore necessary. Existing QM multiband simulators are often based on the non-equilibrium Green’s function formalism (NEGF), combined with either a tight-binding12,13 or a kp14,15 basis set. The former, however, is only feasible either for very small structures or on large computing clusters due to its high computational demands. Current implementations of the latter, on the other hand, are more efficient, using techniques like the coupled-mode space approach in combination with the recursive Green’s function algorithm. Current kp-based implementations, however, do not take into account basis transformations between different materials in a heterostructure.

We have therefore developed a fully QM solver based on the multiband envelope function formalism developed by Bur16 and adapted for heterostructures by Van de Put.17 This formalism is combined for the first time with the multiband quantum transmitting boundary method (QTB), which we extend to 2D potentials allowing the developed solver to simulate confined HTFETs with realistic device dimensions and 2D potentials.

II. FORMALISM

A. Heterostructure envelope function formalism

In the envelope function formalism, the one-electron wavefunction \( \psi \) is expanded in a complete set of orthonormal, lattice-periodic basis functions \( U_n \). In this paper, we

a)devin.verreck@imec.be
take these basis functions to be the zone-center eigenfunctions of the bulk Hamiltonian in material layer $k$

$$\psi(r) = \sum_{n}^{N} F_n^{k}(r) U_n^{k}(r). \quad (1)$$

$F_n^{k}$ are the envelope functions referenced to layer $k$, containing only components that vary slowly on the scale of a lattice unit cell, with Fourier components limited to the first Brillouin zone. This ensures the uniqueness of the expansion. $N$ is the number of bands considered and is in principle infinite, but in practice limited by the available computational resources. The time independent one-electron Schrödinger equation for a semiconductor heterostructure reads

$$\left[-\frac{\hbar^2}{2m_e} \nabla^2 + V_e(r) + V_v(r)\right] \psi(r) = E \psi(r), \quad (2)$$

where $m_e$ is the free electron mass, $E$ is the total energy, $V_e(r)$ is the external potential, and $V_v(r)$ is the heterogeneous crystal potential

$$V_v(r) = \sum_{l}^{L} \theta_l(r) V_v^{l}(r), \quad (3)$$

where $V_v^{l}(r)$ is the bulk potential in material layer $l$ and $L$ is the number of layers present in the heterostructure. $\theta_l$ is a logical step function, which is only non-zero for $r$ in $\Omega_l$, the volume of layer $l$

$$\theta_l(r) = 0/1 \quad \text{if} \quad r \in / \not\Omega_l. \quad (4)$$

In this paper, we will only consider devices in which the potential and the material vary only in the transport direction $x$ and the confined direction $z$ (see Fig. 1 for the definition of the axes). The $y$-direction is assumed to be translationally invariant

$$F_n^{k}(r) = e^{i k y} F_n^{k}(x, z). \quad (5)$$

Inserting Eqs. (1), (3), and (5) into Eq. (2), and using the completeness and orthonormality of the basis functions, a system of $N$ coupled differential equations is obtained for the entire structure

$$-\frac{\hbar^2}{2m_e} \frac{\partial^2 F_n^{k}(x, z)}{\partial x^2} + k_y^2 \frac{\hbar^2}{2m_e} F_n^{k}(x, z) - \frac{\hbar^2}{2m_e} \frac{\partial^2 F_n^{k}(x, z)}{\partial z^2}
- \frac{i \hbar}{m_e} \sum_{m} \left\{ \langle p_{nm}^{k}(x, z) | \frac{\partial F_n^{k}(x, z)}{\partial x} \right\} + \frac{k_y \hbar}{m_e} \sum_{m} \left\{ \langle p_{nm}^{k}(x, z) | F_n^{k}(x, z) \right\}
- \frac{i \hbar}{m_e} \sum_{m} \left\{ \langle p_{nm}^{k}(x, z) | \frac{\partial F_n^{k}(x, z)}{\partial z} \right\} + \sum_{m} H_{nm}^{k}(x, z) F_m^{k}(x, z)
+ V_e(x, z) F_n^{k}(x, z) = EF_n^{k}(x, z), \quad (6)$$

where $p_{nm}^{k}$ represents the $k \cdot p$ interband momentum matrix elements, which couple band $n$ to band $m$, and $H_{nm}^{k}$ are the Hamiltonian matrix elements. The index $m$ is summed over all bands. Following the heterostructure formalism introduced by Van de Put, the envelope functions have been referenced to material layer $k$, and a transformation to this layer has been carried out for the Hamiltonian matrix elements

$$H_{nm}^{k}(x, z) = \sum_{l} \theta_l(x, z) \sum_{i,j} \langle \tilde{S}_{l-i}^{l-j} | n \rangle E_i^{l} \tilde{S}_{m}^{l-j}, \quad (7)$$

where $E_i^{l}$ are the band edge eigenenergies in layer $l$ and $\tilde{S}_{l-i}^{l-j}$ is a unitary transformation of the basis functions from layer $l$ to layer $k$ based on the interband momentum matrix elements. The matrix elements of this transformation are defined as

$$\tilde{S}_{l-i}^{l-j} = \langle U_i^{k} | \tilde{S}_{l-i}^{l-j} | U_j^{k} \rangle. \quad (8)$$

As a result of the presence of the different material layers, the Hamiltonian matrix elements are no longer diagonal and have become position dependent. The position dependence is introduced as well for the interband momentum matrix elements $p_{nm}^{k}$ if the number of bands considered $N$ is finite. This is the case in any practical simulation. The interband momentum matrix elements $p_{ij}^{l}$ of layer $l$ are then transformed to layer $k$ as follows:

$$p_{nm}^{k}(x, z) = \sum_{l} \theta_l(x, z) \sum_{i,j} \langle \tilde{S}_{l-i}^{l-j} | n \rangle p_{ij}^{l} \tilde{S}_{m}^{l-j}. \quad (9)$$

Because of the transformation of both the Hamiltonian matrix elements and the interband momentum matrix elements, the system in Eq. (6) is valid for the entire layered structure. In the remainder of the paper, the layer index $k$ will be suppressed for notational convenience.

With the heterostructure envelope system constructed, the next step is to determine boundary conditions at the contacts for the system in Eq. (6). To this end, we extend the multiband QTBM to 2D potentials.

B. The contact eigenvalue problem

The basic assumption underlying QTBM is that in the source and drain contacts, the external potential is constant in the transport direction. The envelope functions will therefore be plane waves in this direction. Since we are interested
in simulating confined structures with a gate contact, the potential may vary in the z-direction. The plane waves in the x-direction are therefore modulated by a confinement function \( \chi(z) \). Correspondingly, the envelope functions at the contacts take on the following form for a given energy \( E \) and \( k_{x} \):

\[
F_{n}(x, z) = \sum_{z} c_{x} e^{i k_{x} x} \chi_{m}(z),
\]

where each term corresponds to one of the subband modes \( z \) that arise due to the confinement in the z-direction, \( k_{x} \) is the wavenumber of the plane wave in the x-direction and \( c_{x} \) is a complex coefficient. Inserting Eq. (10) into Eq. (6), the following system is obtained:

\[
k_{x, y}^{2} \frac{\hbar^{2}}{2m_{e}} \chi(z) + k_{y}^{2} \frac{\hbar^{2}}{2m_{e}} \chi(z) - \frac{\hbar^{2}}{2m_{e}} \frac{d^{2} \chi_{m}(z)}{dz^{2}}
+ k_{x} \frac{\hbar}{m_{e}} \sum_{m} \left\{ P_{nm}(x_{c}, z) \right\} _{x} \chi_{m}(z)
+ k_{y} \frac{\hbar}{m_{e}} \sum_{m} \left\{ P_{nm}(x_{c}, z) \right\} _{y} \chi_{m}(z)
- \frac{i \hbar}{m_{e}} \sum_{m} \left\{ P_{nm}(x_{c}, z) \right\} _{z} \frac{d \chi_{m}(z)}{dz}
+ \sum_{m} H_{nm}(x_{c}, z) \chi_{m}(z) + V_{z}(z) \chi_{m}(z) = E \chi_{m}(z),
\]

with \( x_{c} \) denoting the x-location of the contact. Eq. (11) can be rewritten as an eigenvalue problem at each of the contacts

\[
\mathbf{H}(k_{x, y}, k_{x, y}) \chi(z) = E \chi(z),
\]

with \( \chi(z) \) as a vector containing the confinement functions for each band

\[
\chi(z) = \begin{bmatrix}
\chi_{1a}(z) \\
\chi_{2a}(z) \\
\vdots \\
\chi_{Na}(z)
\end{bmatrix}.
\]

The Hamiltonian can be written as a quadratic expression in \( k_{x, y} \)

\[
\mathbf{H}(x_{c}, k_{x, y}, k_{y}) = \mathbf{H}^{(2)} k_{x, y}^{2} + \mathbf{H}^{(1)}(x_{c}, z) k_{x, y} + \mathbf{H}^{(0)}(k_{y, y}, z).
\]

Since the goal is to obtain the different \( k_{x, y} \) along with the corresponding \( \chi(z) \) for a given \( E \) and \( k_{y} \), we convert Eq. (14) to an eigenvalue problem for \( k_{x, y} \), for which \( E \) and \( k_{y} \) are inputs. Suppressing the dependencies of the Hamiltonian matrices for notational convenience, we obtain

\[
\begin{bmatrix}
0 \\
-\mathbf{H}^{(2)} \mathbf{H}^{(0)} - E \\
-\mathbf{H}^{(2)} \mathbf{H}^{(1)}
\end{bmatrix}
\begin{bmatrix}
\chi_{1a}(z) \\
\chi_{2a}(z) \\
\vdots \\
\chi_{Na}(z)
\end{bmatrix}
= k_{x, y}
\begin{bmatrix}
\chi_{1a}(z) \\
\chi_{2a}(z) \\
\vdots \\
\chi_{Na}(z)
\end{bmatrix}.
\]

This system has the same form as obtained by Liu et al.,

\[
F_{n}(x, z) = \sum_{\gamma} f_{\gamma} e^{i k_{x, y} x} \chi_{m}(z) + \sum_{\gamma} r_{\gamma} e^{i k_{x, y} x} \chi_{n}(z),
\]

with \( f_{\gamma} \) and \( r_{\gamma} \) denoting the coefficients of, respectively, the incoming and the reflected modes. At the drain side we have

\[
F_{n}(x, z) = \sum_{\delta} t_{\delta} e^{i k_{x, y} x} \chi_{n}(z) + \sum_{\delta} U_{\delta} e^{i k_{x, y} x} \chi_{n}(z),
\]

where \( t_{\delta} \) are the coefficients of the transmitted modes and \( U_{\delta} \) are the coefficients of the incoming modes. The direction of propagation of a subband mode can be determined from the sign of its associated probability current in the x-direction \( J_{p} \). If we assume \( k_{x, y} = q_{x} - k_{x} \) to be in general a complex number, then the probability current of one subband mode can be determined to be (see the Appendix)

\[
\left\{ J_{p} \right\} = \frac{e^{2i \alpha x} |c_{x}|^{2}}{m_{e}} \left[ \hbar q_{x} \chi_{a}(z) + \text{Re}[\chi_{a}(z) P \chi_{a}(z)] \right].
\]

The sorted modes are collected in a matrix equation for each of the contacts

\[
\begin{bmatrix}
\mathbf{F}_{1, z} \\
\mathbf{F}_{2, z} \\
\vdots \\
\mathbf{F}_{n, z}
\end{bmatrix} = \begin{bmatrix}
\mathbf{S}_{11} & \mathbf{S}_{12} \\
\mathbf{S}_{21} & \mathbf{S}_{22}
\end{bmatrix} \begin{bmatrix}
\mathbf{r} \\
\mathbf{u}
\end{bmatrix},
\]

where \( \mathbf{S} \) and \( \mathbf{D} \) refer to, respectively, the source and drain contact. The x-axis has been discretized into \( n_{x} \) grid points, whereas \( \mathbf{F}_{1, z} \) and \( \mathbf{F}_{n, z} \) correspond to the envelope function vector at the x-coordinate of, respectively, the source and the drain contact. E.g., at the source

\[
\mathbf{F}_{1, z} = \begin{bmatrix}
\mathbf{F}_{1, 1} \\
\mathbf{F}_{1, 2} \\
\mathbf{F}_{1, 3} \\
\vdots \\
\mathbf{F}_{1, n_{x}}
\end{bmatrix}.
\]
where the z-axis has been discretized into \( n_z \) points. \( I, r, t \), and \( U \) contain the coefficients for, respectively, the incoming modes at the source side, the reflected modes, the transmitted modes, and the incoming modes at the drain side. The \( S \) and \( D \)-matrices are constructed from the \( S_k(z) \) and \( k_{s,z} \)-values that were obtained from the eigenvalue problem in Eq. (15) at each contact. E.g., for the source contact

\[
S_{11} = \left[ S_1^1(z), S_2^2(z), \ldots, S_{N_{NR}}^1(z) \right], \\
S_{12} = \left[ S_{N_{NR}+1}^1(z), S_{N_{NR}+2}^2(z), \ldots, S_{2N_{NR}}^1(z) \right], \\
S_{21} = \left[ e^{-i\Delta x_k S^1_1(z)}, e^{-i\Delta x_k S^1_2(z)}, \ldots, e^{-i\Delta x_k S^1_{N_{NR}}(z)} \right], \\
S_{22} = \left[ e^{-i\Delta x_k S^2_1(z)}, e^{-i\Delta x_k S^2_2(z)}, \ldots, e^{-i\Delta x_k S^2_{N_{NR}}(z)} \right],
\]

where \( N_{NR}(N_{DL}) \) is the number of right (left) propagating states, and \( \Delta x \) is the discretization stepsize in the x-direction. The \( S \)-matrices for the drain contact are completely analogous. We can see that \( F_{1,z} (F_{n,z}) \) is \( F_{1,z} (F_{n,-1,z}) \) propagated as a plane wave over a distance \( \Delta x \). The QTBM boundary conditions are finally obtained by eliminating \( r \) and \( t \) from, respectively, Eqs. (19) and (20)

\[
F_{1,z} - S_{12} S_{22}^I F_{2,z} = S_{11} I - S_{12} S_{22}^I S_{21} I, \\
-D_{21} D_{11}^I F_{n,-1,z} + F_{n,z} = D_{23} U - D_{21} D_{11}^I D_{13} U.
\]

These equations form the boundary conditions to the system of envelope function equations in Eq. (6).

D. Calculating the electric current

In order to obtain a prediction for the real electron current, we need to calculate the transmission probabilities for the different subband modes. The transmission probability for a single mode, say \( \beta \), can be obtained by injecting only that mode (see Fig. 1). We only have to consider modes injected at the source side, as modes injected from the drain side have the same transmission probability as a result of time reversal symmetry. This means that in Eq. (16) for the incoming modes \( I_\beta = 1 \) and \( I_{\beta',\beta} = 0 \). We set the corresponding elements in the vector \( I \) in the boundary condition Eq. (26) and set \( U = 0 \) in Eq. (27). Next, we solve the envelope function system Eq. (6) with its boundary conditions Eqs. (26) and (27). This corresponds to injecting a probability current (see Eq. (18))

\[
\{J^R_p\}_s = \frac{\hbar q_e |I_\beta|^2}{m_e} |S_\beta(z)|^2 + \frac{|F_\beta|^2}{m_e} \text{Re}\left[ S_\beta(z) P S_\beta(z) \right].
\]

The outgoing probability current (in the x-direction) is obtained by decomposing the calculated envelope functions at the drain contact into the outgoing modes

\[
t = (D_{21})^{-1} F_{n,z},
\]

and using the same current formula as in Eq. (28), now summed over all outgoing modes

\[
\{J^\text{OUT}_p\}_s = \sum_\delta \left\{ \frac{\hbar q_e |I_\delta|^2}{m_e} |S_\delta(z)|^2 + \frac{|F_\delta|^2}{m_e} \text{Re}\left[ S_\delta(z) P S_\delta(z) \right] \right\}.
\]

The transmission probability for the injected mode \( \beta \) is then the outgoing probability current divided by the probability current that was injected

\[
T_\beta(E, k_y) = \frac{\{J^\text{OUT}_p\}_s}{\{J^R_p\}_s}.
\]

The total electron current density formula is obtained from the well-known Tsu-Esaki formula for ballistic transport

\[
J_{3D} = \frac{2q_e}{\hbar} \int \int T(E, k_y, k_z) (f_5(E) - f_3(E)) \frac{dk_y dk_z}{2\pi 2\pi} dE,
\]

where \( q_e \) is the elementary electron charge and \( f_5(E) \) and \( f_3(E) \) represent the Fermi-Dirac distributions, respectively, in the source and drain contact. For a structure with confinement in the z-direction, the translational invariance in that direction is lost. The integral over \( k_z \) is therefore replaced with a discrete sum over the subband transmission probabilities

\[
J_{3D} = \frac{2q_e}{\hbar} \int \sum_\gamma T_\gamma(E, k_y) (f_5(E) - f_3(E)) \frac{dk_y}{2\pi} dE,
\]

where \( J \) is now in units \( \text{A/m} \). This means the transmission probability has to be calculated for each energy, \( k_y \) and subband mode. The entire procedure has been summarized in a flowchart in Fig. 2.

III. APPLICATION BASED ON 2-BAND IMPLEMENTATION

We implemented the formalism outlined in Sec. II with the aim of simulating heterostructure diodes and, in a next step, HTFETs with realistic dimensions, with the specific aim of investigating the effects of size-induced confinement. To limit the computational burden, the implementation was carried out for the 2-band case. For symmetry reasons, there is no coupling between conduction and valence band in the \( y \)-direction or in the confined \( z \)-direction in the 2-band model, so, correspondingly, tunneling only takes place in the \( x \)-direction. Field-induced quantum confinement (FIQC) underneath the gate dielectric is therefore not well described in the 2-band model. These limitations will be somewhat alleviated by a proper choice of device configuration (see Sec. III B). The 2-band approximation with lack of coupling in \( y \) and \( z \)-direction also results in upwards curvature of the valence band in the directions orthogonal to transport. To reincorporate the effect of higher bands on the band curvature in the \( y \) and \( z \)-direction, the EMA is used. We introduced
a heavy hole mass for the valence band and an electron effective mass for the conduction band. The $k\cdot p$ interband momentum matrix parameters along with other material parameters were obtained from literature. The electrostatic potential was extracted from Synopsys Sentaurus Device (SDevice), which solves a nonlinear Poisson equation together with electron and hole continuity equations. The potential is calculated in SDevice taking into account the non-parabolicity at the $\Gamma$-point and the presence of possible satellite valleys. Doping induced bandgap narrowing has been neglected. Discretization in the QM solver was carried out using a first order central finite difference scheme on a rectangular grid.

A. Heterostructure diodes

In a first step, heterostructure diodes with varying body thickness were simulated. Two band alignments were investigated: straddled (InGaAs/InP) and staggered (GaAsSb/InGaAs) (see Fig. 3), both lattice matched combinations. Dirichlet boundary conditions are assumed in the $z$-direction, which means the wavefunction goes to zero at the horizontal edges. Fig. 4 shows the configuration of the diodes simulated, along with the configuration details in Table I.

Figs. 5 and 6 illustrate the effects of confinement for the InGaAs/InP case. Fig. 5 shows that due to confinement, a discrete set of subband modes is formed in the contacts. As the body thickness decreases, the subband energy level separation increases, such that the bandgap for the traveling modes enlarges. The larger bandgap decreases their transmission probability, which can be seen in Fig. 6. For the 5 nm configuration, one mode has become dominant, with a negligible contribution of the second mode. The transmission probabilities of the first mode coincide with those of the second 10 nm mode. Since the energy of the $n$th subband level can be estimated using the EMA as

$$E_n = \frac{n^2\hbar^2\pi^2}{2mL_z^2},$$

with $L_z$ the body thickness, the first confined energy level of a 5 nm configuration coincides with the second level in a

![FIG. 2. Flowchart of the procedure to calculate the current.](image)

![FIG. 3. Alignment of conduction band $E_c$ and valence band $E_v$ at the heterojunction for the two studied material combinations.](image)

![FIG. 4. (a) Diode and (b) HTFET configuration under study. The configuration parameters are listed in Table I.](image)

<table>
<thead>
<tr>
<th>Table I. Details of simulated configurations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diode</td>
</tr>
<tr>
<td>Lsource [nm]</td>
</tr>
<tr>
<td>Lchannel [nm]</td>
</tr>
<tr>
<td>Ldrain [nm]</td>
</tr>
<tr>
<td>Lgate/Lgs [nm]</td>
</tr>
<tr>
<td>Doping S/D [cm$^{-3}$]</td>
</tr>
<tr>
<td>Doping pocket [cm$^{-3}$]</td>
</tr>
<tr>
<td>EOT [nm]</td>
</tr>
<tr>
<td>WF gate [eV]</td>
</tr>
</tbody>
</table>
10 nm configuration. In the remainder of this paper, transmission probabilities will be normalized to the amount of subbands available, which corresponds to the width of the diode $L_z$ (see Eq. (34)). In this way, we isolate the effect of confinement on the subbands itself, rather than the amount of subbands available.

Figs. 7 and 8 show the decrease in current density for both material configurations as the body thickness is scaled down to below 30 nm. Figs. 9 and 10 illustrate the effect of the increasing quantum confinement directly on the normalized transmission coefficients for a given $V_{ds}$ and $k_y$. Below 30 nm, the transmission drops significantly as the confinement in the $z$-direction becomes more important.

**B. Heterostructure TFET**

Next, we investigated dual gate GaAsSb/InGaAs p-n-i-n HTFETs, where the heterojunction is located at the tunnel junction (see Fig. 4 and Table I). We only simulated the material combination with the staggered band alignment, since this is the most promising for application in HTFETs.\(^6\) In a p-n-i-n configuration, a counterdoped pocket is added at the source-channel interface to generate a stronger electric field at the tunnel junction, causing shorter tunnel paths and an accompanying increase in on-current.\(^{22,26}\) The induced electric field favors the alignment of tunnel paths more parallel to the gate dielectric, justifying the absence of tunnel paths perpendicular to the gate in the 2-band model. This also makes the configuration less sensitive to FIQC. Research has shown that the optimal pocket thickness lies slightly above the depletion width of the junction, such that the gate does not lose control over the tunnel junction. We therefore chose a pocket thickness of 2 nm, as the depletion width at the GaAsSb/InGaAs junction for the given doping levels is around 1.9 nm. For the 5 nm configuration, a gate-source underlap of 1 nm was introduced to prevent depletion of the source region.\(^{26}\)

As Fig. 11 shows, the detrimental effects of size confinement in the GaAsSb/InGaAs p-n-i-n TFET are counteracted by the improved gate control in the thinner devices. The 20 nm configuration is clearly too thick, resulting in a deteriorated SS due to the weak gate control. For the 10 nm and 5 nm configurations, the gate control is stronger as a result of the closer proximity of the two gates.\(^{27}\) In the 10 nm configuration, the on-current that is reached 0.5 V beyond the onset voltage, is even higher than in the 20 nm case. The 5 nm
configuration, being strongly confined, sees a noticeable decrease in on-current, although the SS is slightly better than the 10 nm configuration. Inspection of the transmission probabilities (see Fig. 12) reveals that the impact of confinement is strongest in the 5 nm configuration, while the improved gate control counteracts the effects of confinement in the 10 nm configuration, retaining a similar maximum in transmission probability.

IV. CONCLUSION

We have presented a fully quantum mechanical multiband formalism for the simulation of ballistic quantum transport, including BTBT, in confined direct-bandgap heterostructure devices. The formalism combines a novel heterostructure envelope function approach with the multiband QTBM. We have shown its use in the study of quantum confinement, both in heterostructure diodes and p-n-i-n HTFETs. As to the diodes, size confinement was shown to decrease drastically the transmission probability for the individual confined subband modes, resulting in an overall decrease in current density. However, for the p-n-i-n HTFETs, the improved gate control in smaller devices was found to counteract this effect to some degree, with the 10 nm wide device having a lower SS and higher on-current than the 20 nm configuration. This confirms that gate control and size-induced quantum confinement play an important opposing role in strongly scaled devices.

As the formalism can accommodate for an arbitrary number of bands, future work will focus on the implementation of more bands. This will allow the simulation of HTFET-structures, where the tunneling is perpendicular to the gate dielectric, known as line tunneling.

ACKNOWLEDGMENTS

D. Verreck acknowledges the support of a Ph.D. stipend from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This work was supported by imec’s Industrial Affiliation Program.
APPENDIX: PROBABILITY CURRENT OF AN INDIVIDUAL SUBBAND MODE

We start from the formula for the probability current of a set of envelope functions as given by Burt:\(^1\(^6\))

\[
\{J_p\}_x = \frac{1}{m_e} Re \left[ \sum_n F_n^* \left( -i\hbar \frac{\partial F_n}{\partial x} \right) + \sum_{n,m} F_n^* \{p_{nm}\}_x F_m \right].
\]  
(A1)

In vector notation

\[
\{J_p\}_x = \frac{1}{m_e} Re \left[ -i\hbar F^* \left( \frac{\partial F}{\partial x} \right) + F^* P F \right].
\]  
(A2)

with \(F\) a vector containing all the envelope functions and \(P\) a matrix containing the interband momentum matrix elements in the \(x\)-direction. If \(k_{c,s} = q_s - k_{c,s,i}\), then the envelope functions have the following form in the contacts:

\[
F_n(x,z) = \sum_s C_s \psi_{n,s}(z)e^{i(q_s x + k_{c,s} z)}. \quad (A3)
\]

Inserting this into Eq. (A1) for a single mode, we obtain

\[
\{J_p\}_x = \frac{e^{2x|x|}}{m_e} \left( \frac{d}{2} \left| \sum_n |\psi_{n,s}(z)|^2 (\hbar q_s - i\hbar k_{c,s}) \right|^2 + \sum_{n,m} \left| \psi_{n,s}(z) \right|^2 \right). \quad (A4)
\]

Taking the real part of the first term, we obtain the final equation for the probability current

\[
\{J_p\}_x = \frac{e^{2x|x|}}{m_e} \left[ \left( \frac{d}{2} \right) \left| \sum_n |\psi_{n,s}(z)|^2 \right|^2 + Re \left[ \sum_{n,m} \left| \psi_{n,s}(z) \right|^2 \{p_{nm}\}_x \psi_{m,s}(z) \right] \right]. \quad (A5)
\]

or in vector form

\[
\{J_p\}_x = \frac{e^{2x|x|}}{m_e} \left[ \left| \sum_n |\psi_{n,s}(z)|^2 \right|^2 + Re \left[ \psi_{x,s}(z) P \psi_{x,s}(z) \right] \right]. \quad (A6)
\]

---