AIP Journal of Applied Physics

Modeling direct band-to-band tunneling: From bulk to quantum-confined semiconductor devices

H. Carrillo-Nuñez, A. Ziegler, M. Luisier, and A. Schenk

Citation: Journal of Applied Physics **117**, 234501 (2015); doi: 10.1063/1.4922427
View online: http://dx.doi.org/10.1063/1.4922427
View Table of Contents: http://scitation.aip.org/content/aip/journal/jap/117/23?ver=pdfcov
Published by the AIP Publishing

JOURNAL OF APPLIED PHYSICS **117**, 234501 (2015)

Modeling direct band-to-band tunneling: From bulk to quantum-confined semiconductor devices

H. Carrillo-Nuñez, A. Ziegler, M. Luisier, and A. Schenk
Integrated Systems Laboratory ETH Zürich, Gloriastrasse 35, 8092 Zürich, Switzerland

(Received 21 April 2015; accepted 1 June 2015; published online 15 June 2015)

A rigorous framework to study direct band-to-band tunneling (BTBT) in homo- and hetero-junction semiconductor nanodevices is introduced. An interaction Hamiltonian coupling conduction and valence bands (CVBs) is derived using a multiband envelope method. A general form of the BTBT probability is then obtained from the linear response to the "CVBs interaction" that drives the system out of equilibrium. Simple expressions in terms of the one-electron spectral function are developed to compute the BTBT current in two- and three-dimensional semiconductor structures. Additionally, a two-band envelope equation based on the Flietner model of imaginary dispersion is proposed for the same purpose. In order to characterize their accuracy and differences, both approaches are compared with full-band, atomistic quantum transport simulations of Ge, InAs, and InAs-Si Esaki diodes. As another numerical application, the BTBT current in InAs-Si nanowire tunnel field-effect transistors is computed. It is found that both approaches agree with high accuracy. The first one is considerably easier to conceive and could be implemented straightforwardly in existing quantum transport tools based on the effective mass approximation to account for BTBT in nanodevices. © *2015 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4922427]

I. INTRODUCTION

Technology computer aided design (TCAD) has been widely recognized as a valuable support of experimental research. TCAD tools built on classical and semi-classical models such as the drift-diffusion approach[1,2] or Boltzmann transport equation are, however, no longer accurate enough to predict the performance of nanodevices. At the nanometer scale, advanced tools are required that describe the wave nature of electrons and cover quantum effects like energy quantization and tunneling.

Tunnel field-effect transitors (TFETs), considered as promising candidates for energy-efficient transistors,[3] exploit the valence-band-to-conduction-band tunneling mechanism known as band-to-band tunneling (BTBT). In the drift-diffusion TCAD tool Sentaurus-Device (S-Device) from Synopsys,[4] a non-local model of BTBT based on Kane's two-band imaginary dispersion[5] can be used which neglects any quantum confinement effects on the semiconductor band structures.

Formalisms like $\mathbf{k}\cdot\mathbf{p}$,[6] tight-binding[7,8] (TB), and density functional theory[9,10] (DFT) are capable of computing the proper material band structure at different levels of accuracy. While $\mathbf{k}\cdot\mathbf{p}$ is usually restricted to the description of the Γ point, DFT tends to underestimate the band gap of most semiconductors. The atomistic TB approach offers a trade-off between accurate description of the band structure and lower computational cost for device transport simulations. However, the simulation of larger devices with atomistic quantum transport simulators such as the TCAD tool called OMEN, that uses a $sp^3d^5s^*$ TB representation of the band structure of various semiconductors,[8] becomes intractable and simpler models, for instance, the effective mass approximation (EMA), become inevitable.

In the literature, different approaches, such as Kane's two-band model,[4,5,11] Burt's envelope function method,[12,13] and one-band EMA models,[14–16] have been proposed to study BTBT in homo- and hetero-junction devices. Recently, a WKB approximation based on Flietner's two-band model[17] has been adapted for InAs-Si Esaki diodes and extremely narrow nanowire TFETs.[18] Despite the aforementioned progress on this line, the implementation of quantum confinement in TCAD tools based on analytical models is still a challenge, and only few articles report on BTBT in nano-TFETs under non-uniform fields.[13,18]

In this work, we introduce a new general expression for BTBT in semiconductors that can account for quantum confinement effects in nanodevices. It is obtained from the Multi-Band Envelope Function (MBEF) model proposed in Ref. 19 which leads to a two-band Hamiltonian within the EMA. Transforming it to its second quantization form, one can find an "interaction" Hamiltonian that couples conduction and valence bands (CVBs). Then, the expression for the BTBT current is obtained by solving the quantum Liouville equation to the lowest order in terms of the CVBs interaction. In Appendix C, it is shown how the transmission, or equivalently, the BTBT rate is derived from the solution of the two-band envelope equation for nanostructures by implementing Flietner's model. Since most TCAD tools relying on either the Non-Equilibrium Green's Function (NEGF) or the Wave Function (WF) formalism use the one-band EMA model, the implementation of the CVBs approach proposed here should be straightforward.

In Sec. II, we briefly summarize the MBEF model for quantum transport in tunnel devices. By combining the two-band second quantization Hamiltonian of the system with the solution of the Liouville equation in Sec. III, an expression for the BTBT probability is found in terms of the

one-electron and one-hole spectral functions. Results for quantum-confined devices are also derived taking into account the polarization of the momentum matrix elements, as explained in Appendix A. In the limit of the uniform field approximation, the WKB method is applied in Appendix B to analytically compare it with the previous results. In Sec. IV, the BTBT transmission obtained with the one-band EMA and Flietner models are compared with full-band, atomistic quantum transport (OMEN) simulations of Ge, InAs, and InAs-Si Esaki diodes. InAs-Si nanowire TFETs are then analyzed to reveal the differences between the models in low-dimensional systems. Finally, conclusions of this work are given in Sec. V, whereas the Flietner-model based envelope equations are presented in Appendix C.

II. MULTIBAND ENVELOPE EQUATION

A multiband quantum transport model for electrons, with rest mass m_0, in a crystal lattice can be derived within the Bloch theory and formulated in terms of the cell-averaged envelope functions $\chi(\mathbf{r})$ that are obtained in a general way from the set of coupled equations[19]

$$E\chi_n(\mathbf{r}) = E_n(-i\hbar\nabla)\chi_n(\mathbf{r}) + U(\mathbf{r})\chi_n(\mathbf{r}) - \frac{i\hbar}{m_0}\nabla U(\mathbf{r})\sum_{n'\neq n}\frac{\mathbf{p}_{nn'}}{\Delta E_{nn'}}\chi_{n'}(\mathbf{r}), \quad (1)$$

for which the following has been assumed: the potential $U(\mathbf{r})$ is almost constant within a single unit cell; each band denoted by the index n has a minimum/maximum at some point $\mathbf{k}_0$ with localized wave functions. At the Γ-point ($\mathbf{k}_0 = 0$), we may define the energy gap between the bands n and n' as $\Delta E_{nn'} = E_n(0) - E_{n'}(0)$ and momentum matrix elements

$$\mathbf{p}_{nn'} = -i\hbar\int_{\Omega_0} \mathrm{d}^3\mathbf{r}\, u^*_{n0}(\mathbf{r})\,\nabla\, u_{n'0}(\mathbf{r}), \quad (2)$$

where $u_{n\mathbf{k}}(\mathbf{r})$ are the periodic functions normalized over the unit cell (Ω_0). Equation (1) can be further simplified, for instance, if an isotropic parabolic EMA is assumed for the operator of the kinetic energy

$$E_n(-i\hbar\nabla) \approx E_{n0} - \frac{\hbar^2}{2m_n}\nabla^2, \quad (3)$$

with the electron effective mass m_n for band n at the Γ-point.

III. BAND-TO-BAND TUNNELING MODEL

BTBT can be accurately described by assuming that electron transitions take place mostly between the lowest conduction and one of the highest valence bands, i.e., light and heavy hole bands. Thus, we can neglect the contribution of remote bands, such as the valence split-off band, and write the electron (hole) envelope equation from Eq. (1) as

$$H_{\mathrm{c(v)}}\chi_{\mathrm{c(v)}} \mp \frac{i\hbar}{m_0}\nabla U\cdot\mathbf{p}_{\mathrm{cv}}E_{\mathrm{g}}^{-1}\chi_{\mathrm{v(c)}} = E\chi_{\mathrm{c(v)}}, \quad (4)$$

where $H_{\mathrm{c(v)}}$ is the single band effective mass Hamiltonian for the conduction (valence) band with $E_{\mathrm{g}} = E_{\mathrm{c}} - E_{\mathrm{v}}$. The coordinate vector $\mathbf{r}$ has been omitted for brevity. In Eq. (4), the electron and hole envelope equations are coupled through the external potential. The momentum matrix element is computed according to Kane:[5] $p_{\mathrm{cv}} = m_0\sqrt{E_{\mathrm{g}}/4m_r}$ with $m_r = m_{\mathrm{c}}m_{\mathrm{v}}/(m_{\mathrm{c}} + m_{\mathrm{v}})$ being the reduced effective mass at the Γ-point.

The quantum transport problem could be solved by a numerical approach such as the NEGF method to extract all the desired physical quantities, e.g., transmission and current density. Alternatively, we can invoke the non-equilibrium density matrix formalism which, combined with the perturbation theory, allows us to solve the quantum Liouville equation. For that purpose, the two-band Hamiltonian in Eq. (4) has to be expressed in its second quantization form

$$H_{\text{2–band}} = \sum_{\lambda}(E_{\mathrm{c}\lambda}c^{\dagger}_{\lambda}c_{\lambda} + E_{\mathrm{v}\lambda}b^{\dagger}_{\lambda}b_{\lambda}) + \sum_{\lambda\lambda'}(\Lambda_{\lambda\lambda'}b^{\dagger}_{\lambda'}c_{\lambda} + \mathrm{h.c}), \quad (5)$$

where $E_{\mathrm{c(v)}\lambda}$ is the eigenvalue corresponding to the eigenfunction $\chi_{\mathrm{c(v)}\lambda}$. The operators c_{λ} ($b^{\dagger}_{\lambda}$) annihilate (create) a conduction (valence) electron. The CVBs interaction can now be defined as

$$H_{\mathrm{cv}} = \sum_{\lambda\lambda'}\Lambda_{\lambda\lambda'}b^{\dagger}_{\lambda'}c_{\lambda} + \mathrm{h.c.}, \quad (6)$$

with $\Lambda_{\lambda\lambda'}$ being the CVBs coupling

$$\Lambda_{\lambda\lambda'} = -\frac{i\hbar}{m_0}\int \mathrm{d}^3\mathbf{r}\,\nabla U(\mathbf{r})\cdot\mathbf{p}_{\mathrm{cv}}E_{\mathrm{g}}^{-1}\chi^*_{\mathrm{c}\lambda}(\mathbf{r})\chi_{\mathrm{v}\lambda'}(\mathbf{r}) = \int \mathrm{d}^3\mathbf{r}\, M_{\mathrm{cv}}(\mathbf{r})\chi^*_{\mathrm{c}\lambda}(\mathbf{r})\chi_{\mathrm{v}\lambda'}(\mathbf{r}). \quad (7)$$

The index λ contains all the relevant quantum numbers. Within the time-dependent perturbation theory, at $t \leq 0$, we can consider electrons in the conduction and valence bands as two uncoupled sub-systems locally in equilibrium in the presence of the external potential $U(\mathbf{r})$. The density matrix for the total system in equilibrium is then given by the direct product of the conduction and valence density matrices leading to

$$\varrho_0 = Z^{-1}\exp\left(-\frac{H_{\mathrm{el}} - \mu_{\mathrm{c}}N_{\mathrm{c}} - \mu_{\mathrm{v}}N_{\mathrm{v}}}{k_{\mathrm{B}}T}\right), \quad (8)$$

where Z is the partition function ensuring that $\mathrm{Tr}(\varrho_0) = 1$, H_{el} the electron Hamiltonian in second quantization incorporating the two first terms on the right-hand side of Eq. (5), whereas the conduction and valence band number operators are $N_{\mathrm{c}} = \sum_{\lambda} c^{\dagger}_{\lambda}c_{\lambda}$ and $N_{\mathrm{v}} = \sum_{\lambda} b^{\dagger}_{\lambda}b_{\lambda}$, respectively. The chemical potentials μ_{c} and μ_{v} are obtained as usual by imposing the condition of charge neutrality at the device contacts.

As $t > 0$, the perturbation H_{cv} is turned on and the system is driven out of equilibrium. This induces electron transitions from the valence band to conduction band and

therefore BTBT current. Within the interaction picture, the steady state current can be calculated in the linear response regime after solving the Liouville equation up to first order in $\tilde{H}_{\mathrm{cv}}(t)$ according to[20]

$$I = \frac{2e}{\hbar}\lim_{t\to\infty}\int_0^t \mathrm{d}t'\,\mathrm{Tr}\Big(\big[\big[N_{\mathrm{v}},\tilde{H}_{\mathrm{cv}}(t)\big],\tilde{H}_{\mathrm{cv}}(t')\big]\varrho_0\Big), \tag{9}$$

where $\tilde{H}_{\mathrm{cv}}(t)$ is written in the interaction representation as

$$\tilde{H}_{\mathrm{cv}}(t) = \sum_{\lambda\lambda'}\Lambda_{\lambda\lambda'}b^{\dagger}_{\lambda'}c_{\lambda}\exp\left(\frac{i}{\hbar}(E_{\mathrm{v}\lambda}-E_{\mathrm{c}\lambda'})t\right) + \mathrm{h.c.}. \tag{10}$$

The commutators in Eq. (9) yield

$$\big[\big[N_{\mathrm{v}},\tilde{H}_{\mathrm{cv}}(t)\big],\tilde{H}_{\mathrm{cv}}(t')\big] = \sum_{\lambda\lambda'}\Lambda_{\lambda\lambda'}\mathcal{F}_{\lambda'\lambda}\,\mathrm{e}^{\frac{i}{\hbar}(E_{\mathrm{v}\lambda}-E_{\mathrm{c}\lambda'})t} + \mathrm{h.c.}, \tag{11}$$

where

$$\mathcal{F}_{\lambda'\lambda} = \sum_{l}\Lambda^{\dagger}_{\lambda' l}b^{\dagger}_{\lambda}b_{l}\,\mathrm{e}^{-\frac{i}{\hbar}(E_{\mathrm{v}l}-E_{\mathrm{c}\lambda'})t'} - \Lambda^{\dagger}_{l\lambda}c^{\dagger}_{l}c_{\lambda'}\,\mathrm{e}^{-\frac{i}{\hbar}(E_{\mathrm{v}\lambda}-E_{\mathrm{c}l})t'}.$$

After evaluating the trace $\mathrm{Tr}(\cdots\varrho_0)$, the current can be re-written in the form

$$I = \frac{4e}{\hbar}\sum_{\lambda\lambda'}|\Lambda_{\lambda\lambda'}|^2\big(f_{\mathrm{v}}(E_{\mathrm{v}\lambda})-f_{\mathrm{c}}(E_{\mathrm{c}\lambda'})\big) \times\lim_{t\to\infty}\int_0^t \mathrm{d}t'\cos\left(\frac{1}{\hbar}(E_{\mathrm{v}\lambda}-E_{\mathrm{c}\lambda'})(t'-t)\right), \tag{12}$$

where $f_{\mathrm{v}}(E_{\mathrm{v}\lambda}) = \mathrm{Tr}(b^{\dagger}_{\lambda}b_{\lambda}\varrho_0)$ and $f_{\mathrm{c}}(E_{\mathrm{c}\lambda'}) = \mathrm{Tr}(c^{\dagger}_{\lambda'}c_{\lambda'}\varrho_0)$ are the Fermi-Dirac distribution functions for the valence and conduction bands, respectively. By exploiting the identity $\int_0^{\infty}\mathrm{d}t\cos xt \approx \pi\delta(x)$, Eq. (12) is reduced to

$$I = \frac{4\pi e}{\hbar}\sum_{\lambda\lambda'}|\Lambda_{\lambda\lambda'}|^2\big(f_{\mathrm{v}}(E_{\mathrm{v}\lambda})-f_{\mathrm{c}}(E_{\mathrm{c}\lambda'})\big)\delta(E_{\mathrm{v}\lambda}-E_{\mathrm{c}\lambda'}). \tag{13}$$

The properties of the delta-function allow us to finally express the BTBT current in a more familiar way

$$I = \frac{4\pi e}{\hbar}\sum_{\lambda\lambda'}|\Lambda_{\lambda\lambda'}|^2\int \mathrm{d}E\,\delta(E-E_{\mathrm{v}\lambda})\delta(E-E_{\mathrm{c}\lambda'})\big(f_{\mathrm{v}}(E)-f_{\mathrm{c}}(E)\big), \tag{14}$$

which is equivalent to the Landauer formula with the BTBT transmission probability

$$T(E) = (2\pi)^2\sum_{\lambda\lambda'}|\Lambda_{\lambda\lambda'}|^2\delta(E-E_{\mathrm{v}\lambda})\delta(E-E_{\mathrm{c}\lambda'}). \tag{15}$$

A. BTBT in bulk-like semiconductors

In case of a bulk-like semiconductor with the tunneling junction perpendicular to the transport direction (x-direction) and assuming periodic boundary conditions in the transverse direction, the expression for the BTBT transmission probability using Eqs. (7) and (15) can be written as

$$T_{\mathrm{bulk}}(E) = \sum_{k_x,k'_x}\sum_{\mathbf{k}_{\perp}}|\Lambda_{k_x,k'_x,\mathbf{k}_{\perp}}|^2\delta(E-E_{\mathrm{v}k'_x,\mathbf{k}_{\perp}})\delta(E-E_{\mathrm{c}k_x,\mathbf{k}_{\perp}}), \tag{16}$$

with the CVBs coupling given by

$$\Lambda_{k_x,k'_x,\mathbf{k}_{\perp}} = 2\pi\int \mathrm{d}x\,M_{\mathrm{cv}}(x)\chi^{*}_{\mathrm{c}k_x,\mathbf{k}_{\perp}}(x)\chi_{\mathrm{v}k'_x,\mathbf{k}_{\perp}}(x). \tag{17}$$

Note that due to momentum conservation, conduction and valence electrons must have the same transverse mode $\mathbf{k}_{\perp}$. For each energy E and $\mathbf{k}_{\perp}$, the overlap of the hole and electron envelope functions results in a function with a sharp peak located at some point x_0. Expanding the overlap function around x_0, in Eq. (17), the BTBT transmission probability can be expressed as

$$T_{\mathrm{bulk}}(E) \approx \sum_{\mathbf{k}_{\perp}}\mathcal{S}^2_{\mathbf{k}_{\perp}}(x_0,E)\left|\int \mathrm{d}x\,M_{\mathrm{cv}}(x)\mathrm{e}^{-\frac{(x-x_0)^2}{\sigma^2}}\right|^2 = \sum_{\mathbf{k}_{\perp}}\left|\int \mathrm{d}x\,M_{\mathrm{cv}}(x)\,\mathcal{S}_{\mathbf{k}_{\perp}}(x,E)\right|^2, \tag{18}$$

where the function $\mathcal{S}_{\mathbf{k}_{\perp}}(x,E)$ contains the overlap of the electron $A_{\mathrm{c}\mathbf{k}_{\perp}}$ and hole $A_{\mathrm{v}\mathbf{k}_{\perp}}$ spectral functions

$$\mathcal{S}_{\mathbf{k}_{\perp}}(x,E) = \sqrt{A_{\mathrm{c}\mathbf{k}_{\perp}}(x,x;E)A_{\mathrm{v}\mathbf{k}_{\perp}}(x,x;E)}, \tag{19}$$

with $A_{\mathrm{c(v)}\mathbf{k}_{\perp}} = 2\pi\sum_{k_x}\chi^{*}_{\mathrm{c(v)}k_x,\mathbf{k}_{\perp}}\delta(E-E_{\mathrm{c(v)}k_x,\mathbf{k}_{\perp}})\chi_{\mathrm{c(v)}k_x,\mathbf{k}_{\perp}}$. The coefficient σ depends on the material parameters and the electric field evaluated at x_0, as demonstrated in Appendix B, where the analytical WKB approximation is compared with previous results. It is found that our model can exactly reproduce the expression for the electron generation rate as presented in Ref. 16 and it also agrees very well with the non-local dynamic BTBT model[4] of the S-Device TCAD tool.

B. BTBT in low-dimensional semiconductors

In low-dimensional devices, such as thin-body double gate, i.e., 2D electron gas (2DEG), or nanowire (1DEG) TFETs, the BTBT transmission probability can be calculated straightforwardly from

$$T_{2DEG}(E) = \sum_{k_z}\sum_{\nu\nu'}\left|\int \mathrm{d}x\,\tilde{M}^{2D}_{\nu\nu'}(x)\,\mathcal{S}_{k_z,\nu\nu'}(x,E)\right|^2, \tag{20}$$

$$T_{1DEG}(E) = \sum_{\nu\nu'}\left|\int \mathrm{d}x\,\tilde{M}^{1D}_{\nu\nu'}(x)\,\mathcal{S}_{\nu\nu'}(x,E)\right|^2, \tag{21}$$

where the functions $\mathcal{S}_{k_y,\nu\nu'}(x,E)$ and $\mathcal{S}_{\nu\nu'}(x,E)$ are

$$\mathcal{S}_{k_y,\nu\nu'}(x,E) = \sqrt{A_{\mathrm{c}k_z,\nu}(x,x;E)A_{\mathrm{v}k_z,\nu'}(x,x;E)}, \tag{22}$$

$$\mathcal{S}_{\nu\nu'}(x,E) = \sqrt{A_{\mathrm{c}\nu}(x,x;E)A_{\mathrm{v}\nu'}(x,x;E)}, \tag{23}$$

and the CVBs coupling strengths have been re-defined as

$$\tilde{M}^{2D}_{\nu\nu'}(x) = \int \mathrm{d}y\,M_{\mathrm{cv}}(x,y)F^{*}_{\mathrm{c}\nu}(y;x)F_{\mathrm{v}\nu'}(y;x), \tag{24}$$

$$\tilde{M}^{1D}_{\nu\nu'}(x) = \int \mathrm{d}^2\mathbf{r}_\perp M_{\mathrm{cv}}(x,\mathbf{r}_\perp) F^*_{\mathrm{c}\nu}(\mathbf{r}_\perp;x) F_{\mathrm{v}\nu'}(\mathbf{r}_\perp;x), \quad (25)$$

where $\mathbf{r}_\perp = (y,z)$. $F_{\mathrm{c(v)}\nu}$ is the electron (hole) wave function in the direction of confinement for the state ν. Special attention has to be payed to the CBVs coupling strength in low-dimensional structures since the electric field polarizes the momentum matrix elements.[21] Hence, the latter depend on the directionality of the total electron wave vector, and the CBVs coupling strength should be generally expressed as

$$M_{\mathrm{cv}} = \frac{\hbar p_{\mathrm{cv}}}{3m_0 E_{\mathrm{g}}} \sqrt{\sum_{j=1}^{n} C_j(\theta,\phi)\mathcal{E}_j^2}. \quad (26)$$

Here, j denotes the x, y ($n=1, 2$) and z ($n=3$) directions, and $\mathcal{E}_j$ is the j-component of the electric field. In case of the momentum matrix elements for coupling conduction and light hole band (see Appendix A), the functions $C_j(\theta,\phi)$ are[21]

$$C_x(\theta,\phi) = \frac{1}{2} + \frac{3}{2}\cos^2\theta, \quad (27)$$

$$C_y(\theta,\phi) = \frac{1}{2}\left(\cos^2\theta\sin^2\phi + \cos^2\phi\right) + 2\sin^2\theta\sin^2\phi, \quad (28)$$

$$C_z(\theta,\phi) = \frac{1}{2}\left(\cos^2\theta\cos^2\phi + \sin^2\phi\right) + 2\sin^2\theta\cos^2\phi. \quad (29)$$

The polar angle θ is usually related to the component of the wave vector parallel to the transport direction (k_x), whereas the azimuthal angle ϕ is linked with the transverse components of the wave vector (k_y, k_z). As shown in Appendix A, in case of quantum well structures, the functions $C_j(\theta,\phi)$ can be averaged over ϕ. However, for nanowires, it was found that the magnitude of the total wave vector is ϕ-independent. Then, setting $\phi = \pi/4$ for nanowires and taking the average over ϕ for QWs, the functions in Eqs. (27)–(29) for both types of nanostructure can be written in the same way

$$C_x(\theta) = \frac{1}{2} + \frac{3}{2}\cos^2\theta, \quad (30)$$

$$C_y(\theta) = C_z(\theta) = \frac{5}{4} - \frac{3}{4}\cos^2\theta, \quad (31)$$

with

$$\cos^2\theta = \begin{cases} \dfrac{W_\nu}{E} & \text{if } |E| \geq |W_\nu| \\ 1 & \text{otherwise,} \end{cases} \quad (32)$$

$$\cos^2\theta = \begin{cases} 1 - \dfrac{W_\nu}{E} & \text{if } |E| \geq |W_\nu| \\ 0 & \text{otherwise,} \end{cases} \quad (33)$$

for 2DEG and 1DEG systems, respectively. Note that in TFETs, the sub-band energies W_ν may be position-dependent. Finally, for low-dimensional TFETs, Eq. (26) takes the following explicit form:

$$M_{\mathrm{cv}} = \frac{\hbar p_{\mathrm{cv}}}{3m_0 E_{\mathrm{g}}} \sqrt{C_x(\theta)\mathcal{E}_x^2 + C_y(\theta)\mathcal{E}_y^2 + C_z(\theta)\mathcal{E}_z^2}. \quad (34)$$

The latter expression has been used in the calculation of the BTBT current of nanowire TFETs. The results are shown at the end of Sec. IV.

IV. ONE-BAND MODEL VS. FLIETNER MODEL

In this section, we compare the one-band EMA model with the Flietner model. In Fig. 1, the bulk band structures at Γ of InAs and Ge including their imaginary branches are plotted. The Flietner model, given by Eq. (C1) in Appendix C, can accurately reproduce not only the imaginary branch as obtained from the atomistic tool OMEN but also the conduction and valence real-band branches if the non-parabolicity corrections are used. The agreement is equally good for both materials. The one-band EMA model fits very well both the full-band structure and the Flietner imaginary-band dispersion for energies closer to the band edges. We also plotted the imaginary dispersion for Ge if a non-parabolicity correction of the conduction band is applied to the imaginary branch.

In order to characterize their accuracy and differences when computing the BTBT current, both models are compared to atomistic quantum transport simulations of bulk-like $p-n$ diodes performed with OMEN. Within the one-band EMA model, the individual electron and hole spectral

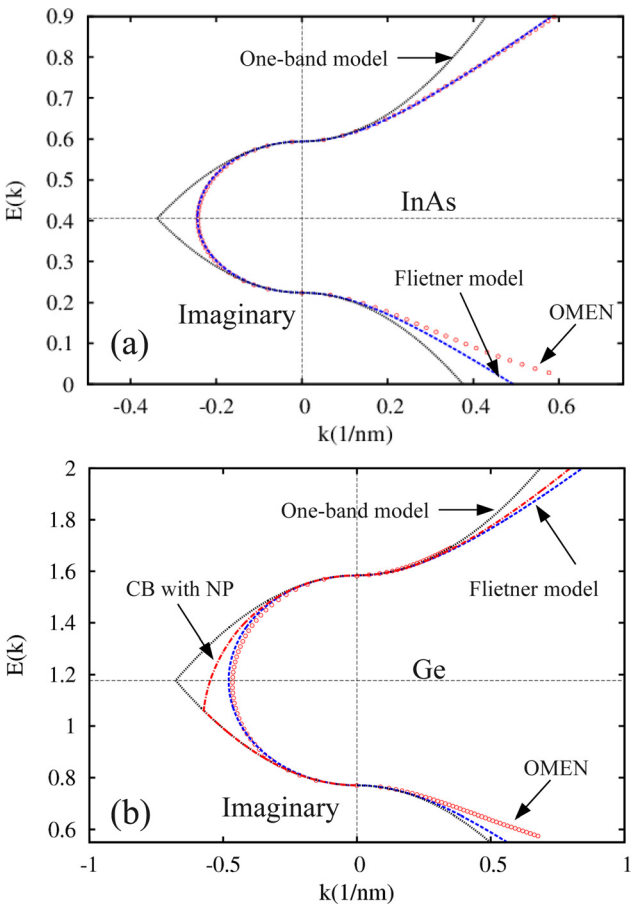

FIG. 1. Bulk imaginary dispersion of (a) InAs and (b) Ge at $\mathbf{k}_\perp = 0$. The one-band and Flietner models are compared to each other and to the full-band imaginary dispersion obtained with the atomistic tool OMEN. The material parameters are $m_{\mathrm{c}} = 0.023m_0$, $m_{\mathrm{v}} = 0.026m_0$, and $E_{\mathrm{g}} = 0.37$ eV for InAs and $m_{\mathrm{c}} = 0.041m_0$, $m_{\mathrm{v}} = 0.043m_0$, and $E_{\mathrm{g}} = 0.81$ eV for Ge.

functions are required. They can be computed from the solution of Eq. (4) without the interband coupling

$$H_{c(v)}\chi_{c(v)} = E\chi_{c(v),\mathbf{k}_\perp}, \tag{35}$$

with

$$H_{c(v)} = E_{c(v)} + E_{c(v),\mathbf{k}_\perp} \mp \frac{\hbar^2}{2m_{c(v)}}\frac{\partial^2}{\partial x^2} + U(x). \tag{36}$$

The BTBT transmission probability within the two-band Flietner model is obtained by solving the envelope Eq. (C6) in Appendix C. For the post-processing calculation of the BTBT, we have taken the self-consistent Poisson-Schrödinger solution from OMEN as an input for the electrostatic potential.

First, we consider a Ge $p-n$ diode. In this case, the polarity is inverted and the n (left) region is grounded while a negative bias is applied on the p (right) side of the diode. Fig. 2 shows the diagonal part of the electron and hole spectral functions for $\mathbf{k}_\perp = 0$. One can observe that the spectral function in Fig. 2(b) oscillates faster than the spectral functions in Fig. 2(a) due to the band nonparabolicity included in the real branch of the band structure in the Flietner model. In Fig. 2(a), both the electron and hole spectral functions decay exponentially inside the gap region. In contrast, a two-band model, as the Flietner model, features an electron-hole duality inside the gap region, e.g., electrons that are injected in the "source" go through the barrier and continue in direction to the "collector."

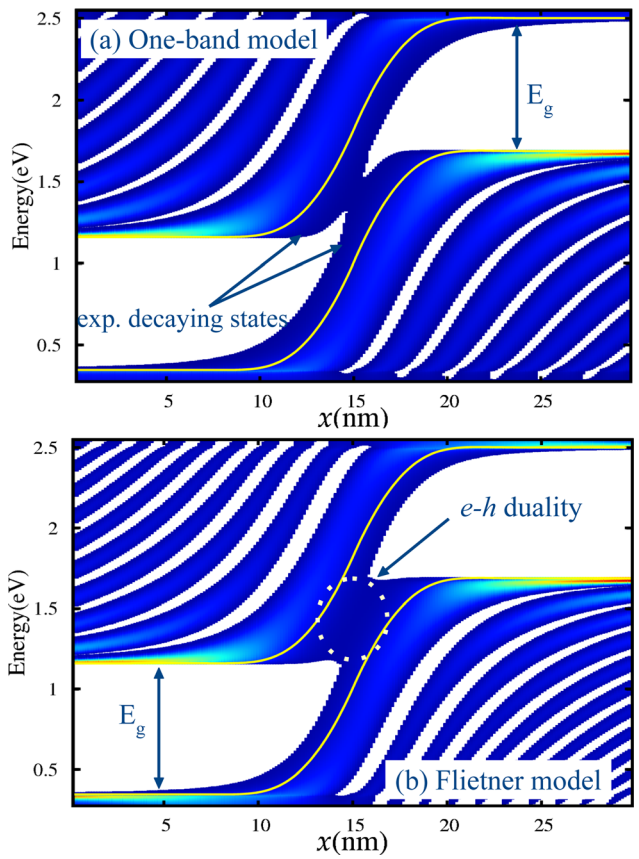

FIG. 2. Electron and hole spectral functions in a $p-n$ Ge diode at $T = 300$ K. (a) One-band EMA model: spectral functions decay exponentially within the tunneling barrier. (b) Flietner model: spectral functions oscillate faster and are connected within the tunneling barrier due to the electron-hole ($e-h$) duality. The material parameters are: $m_c = 0.041m_0$, $m_v = 0.043m_0$, and $E_g = 0.81$ eV. The doping concentrations in the p and n regions are $N_A = N_D = 5\times10^{19}\,\mathrm{cm}^{-3}$. The applied bias is $V = -0.5$ V.

The reverse $I-V$ characteristics of a Ge tunneling diode are plotted in Fig. 3. The BTBT current computed with the one-band EMA and Flietner models are compared to each other and to the full-band, atomistic simulations. It can be seen that atomistic results are accurately reproduced by means of the Flietner model. The BTBT current from the one-band EMA model is found to be shifted down by approximately a factor of 4 with respect to the atomistic simulations. This discrepancy is related to the band parabolicity which strongly affects the imaginary dispersion as observed in Fig. 1. Non-parabolicity corrections for the conduction band can be included in the one-band model via the Schrödinger equation[22]

$$H_c\,\chi_{c,\mathbf{k}_\perp} = \left(E + \alpha_{NP}(E - E_c - U(x))^2\right)\chi_{c,\mathbf{k}_\perp}, \tag{37}$$

where the non-parabolicity coefficient for Ge[23] is $\alpha_{NP} = 0.85\,\mathrm{eV}^{-1}$, and the parabolic EMA Hamiltonian H_c is given by Eq. (36). We have found that after incorporating this non-parabolicity in the one-band EMA model, the resulting BTBT current fits better the full-band simulations, especially in the high-bias regime, as demonstrated in Fig. 3. In the low-bias regime, the current is still slightly underestimated.

Fig. 4 shows the reverse $I-V$ characteristics of an InAs tunneling diode. We find that the full-band current is higher than the BTBT current computed with the Flietner model by a factor of 2.1 (not shown here), while the one-band model fairly reproduces the current obtained from atomistic simulations. The coupling terms of other bands, i.e., the heavy hole and split-off valence bands, tend to flatten the bands and decrease the effective energy gap $E_g + E_{\mathbf{k}_\perp}$ which is ignored by the Flietner model. Hence, the implementation of the EMA for the transverse direction should result in an underestimation of the BTBT current. This effect may be more pronounced in InAs since the spin-orbit coupling strength is known to be very strong in this material. A straightforward

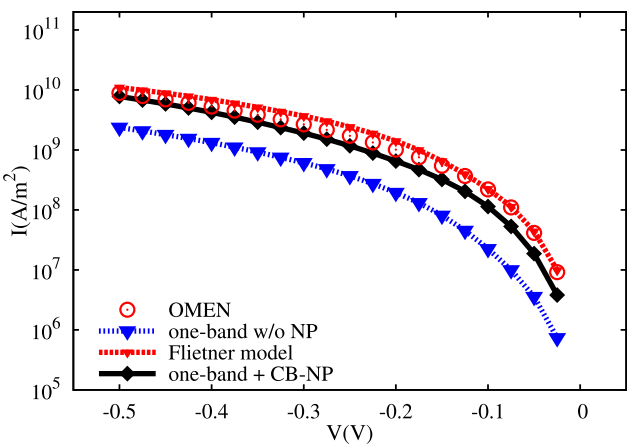

FIG. 3. Reverse $I-V$ characteristics of a Ge tunneling diode at $T = 300$ K. One-band EMA model and Flietner model are compared to the full-band current calculated with OMEN. Inclusion of non-parabolicity correction is necessary for the one-band EMA model to obtain a better agreement with OMEN. The doping concentrations in the p and n regions are $N_A = N_D = 5\times10^{19}\,\mathrm{cm}^{-3}$.

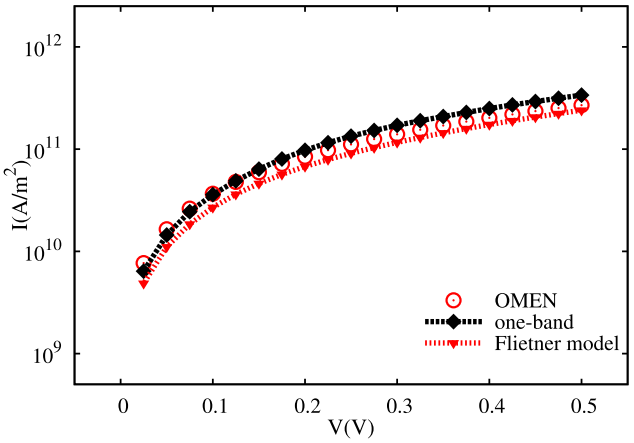

FIG. 4. Reverse $I - V$ characteristics of an InAs tunneling diode at $T = 300$ K computed with the one-band EMA and Flietner models in comparison to the full-band current calculated with OMEN. Non-parabolicity corrections for $E_{\mathbf{k}_\perp}$ included in the Fietner model improve the calculated current. The doping concentrations in the p and n regions are $N_A = N_D = 5 \times 10^{19}\,\mathrm{cm}^{-3}$.

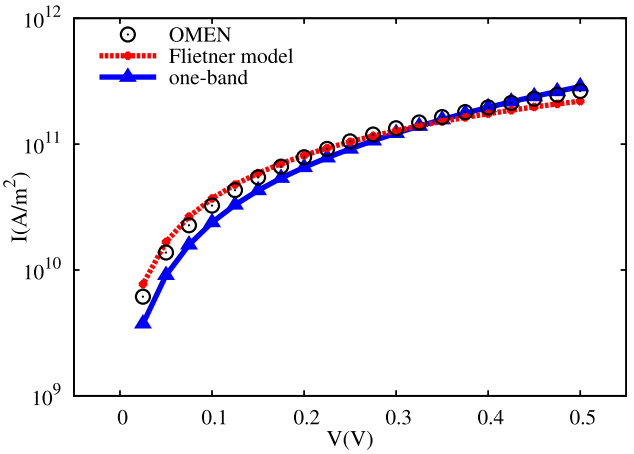

FIG. 5. Reverse $I - V$ characteristics of an InAs-Si hetero Esaki diode at $T = 300$ K. The one-band EMA and Flietner models are compared to the full-band current calculated with OMEN. The doping concentrations in the n (InAs) and p (Si) regions are $N_A = N_D = 5 \times 10^{19}\,\mathrm{cm}^{-3}$. Inside the Si region, the material parameters used in the Flietner model are: $m_v = 0.144 m_0$, $E_g = 1.13$ eV.

solution can be the inclusion of a non-parabolicity correction for the transverse energy.[11] Here, we have used a Flietner-model-like dispersion for $E_{\mathbf{k}_\perp}$

$$\frac{\hbar^2 \mathbf{k}_\perp^2}{2m_c} = E_{\mathbf{k}_\perp}\left(1 + \frac{E_{\mathbf{k}_\perp}}{E_g}\right)\left(1 + \gamma\frac{E_{\mathbf{k}_\perp}}{E_g}\right)^{-2}, \qquad (38)$$

with $\gamma = 1 - \sqrt{m_c/m_v}$. Including Eq. (38) in the calculation of the BTBT current, the $I - V$ curve matches very well the atomistic simulations, as seen in Fig. 4, also showing that the impact of a non-parabolic $E_{\mathbf{k}_\perp}$-dispersion may be important for a proper description of InAs tunneling devices by means of two-band models such as Kane's one.[5] The effect of non-parabolic corrections for $E_{\mathbf{k}_\perp}$ has been found to be irrelevant for Ge, although, it is included for the results showed above.

A. BTBT in hetero-structures

TFETs are considered as energy-efficient transistors, but they provide lower ON-currents than MOSFETs. III-V/Si hetero-structures have been proposed to address this issue and others.[24,25] For hetero-junctions between a direct and an indirect semiconductor, such as InAs and Si,[24] no analytical BTBT theory exists. Recently, in Ref. 18, it was found from full-band, atomistic simulations that BTBT is mainly direct in the InAs-Si material system. A BTBT model based on a WKB-Flietner model was introduced for hetero-structures showing high accuracy as long as appropriate material parameters and electrostatic potentials are used as an input. Inter-material tunneling processes can also be incorporated in the one-band EMA model by treating material parameters in Eqs. (18), (20), and (21) as position-dependent quantities.

In order to demonstrate the accuracy of our models for hetero-structures, a bulk InAs-Si Esaki diode has been first simulated with OMEN. Then, in a post-processing step, the electrostatic potential was used as an input for the calculation of the BTBT current with the analytical models. From the $I - V$ characteristics in Fig. 5, one can observe that the Flietner model plus Eq. (38) fits the atomistic simulations better than the one-band EMA model over the whole bias range considered here. Still, the analytical models result in an overall good agreement with the atomistic simulations, both qualitatively and quantitatively.

As a numerical exercise, the Flietner model and the one-band EMA model have also been compared for devices with size quantization such as the InAs-Si nanowire p-TFET proposed by IBM.[26] Fig. 6(a) shows the schematic of the cross section of the nanowire TFETs used in this work. The InAs source region is n^+–doped and the Si drain region p^+–doped. The gate is all-around the 20 nm long intrinsic Si region. The total device length is 80 nm, and there is no gate overlap on the InAs side. The doping concentrations are $N_D = 10^{19}\,\mathrm{cm}^{-3}$ and $N_A = 2 \times 10^{19}\,\mathrm{cm}^{-3}$. The oxide thickness is 1 nm with the permittivity $\varepsilon_{\mathrm{high}-\kappa} = 20$.

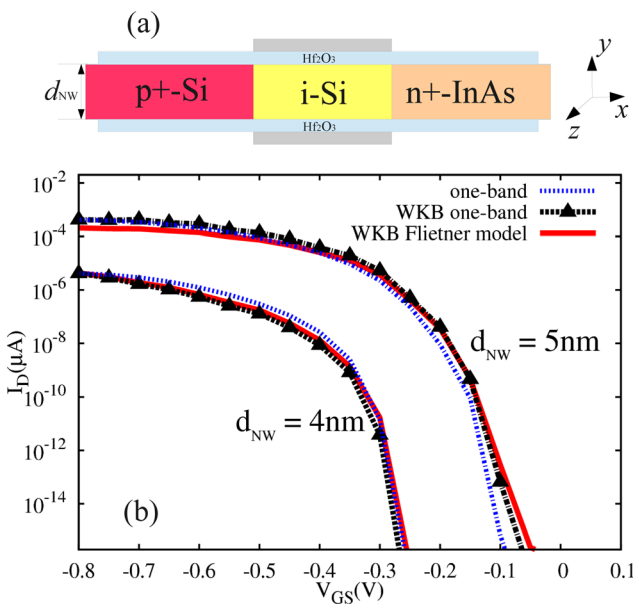

FIG. 6. (a) Schematic of the cross section and (b) room-temperature I_D–V_{GS} characteristics of InAs-Si nanowire p-TFETs with diameters of 4 nm and 5 nm. The one-band and Flietner models are compared. The doping concentrations in the source n^+–InAs and drain p^+–Si regions are $N_D = 10^{19}\,\mathrm{cm}^{-3}$ and $N_A = 2 \times 10^{19}\,\mathrm{cm}^{-3}$, respectively, and $V_{DS} = -0.5$ V.

The quantum transport problem for electrons and holes has been solved within the EMA using the mode-space NEGF version of OMEN,[27] assuming that the conduction and valence bands are uncoupled. Parameters of the bulk materials have been used for InAs and Si. Once the Poisson-Schrödinger solver converged, the electrostatic potential was used as input to compute the BTBT transmission probability and current by means of the one-band EMA model Eq. (21), the modified WKB one-band model Eq. (B13) and the WKB Flietner model Eq. (C5) for hetero-structures.[18] The expressions for the last two are briefly introduced in Appendixes B and C, respectively.

Fig. 6 shows the $I_D - V_{GS}$ characteristics of an InAs-Si nanowire p-TFET with two different diameters, $d_{NW} = 4$ nm and $d_{NW} = 5$ nm. One can see that all three approaches agree, although the agreement for the thicker wire is slightly worse. At high gate voltages, the one-band BTBT current is lower than its counterparts obtained with the WKB approximation. For gate voltage lower than -0.4 V, the currents obtained with both the one-band and WKB one-band models are about the same and slightly higher than the WKB-Flietner model. The difference in the $I_D - V_{GS}$ characteristics of both nanowires TFETs can be explained as follows. The tunneling paths in the TFET with thinner wire are mainly parallel to the transport direction, and a WKB approximation as presented here is sufficient to calculate the BTBT current with high accuracy. For TFETs with thicker nanowires, the WKB approximation in Eqs. (B13) and (C5) becomes less accurate especially at higher gate voltages. Although the BTBT transmission is inter-material, electrons can tunnel along paths that are not necessarily aligned with the transport direction. These tunneling paths cannot be modelled with Eqs. (B13) and (C5). Instead, a non-local path BTBT approach, such as Eqs. (B12) and (B14), should be implemented. Direct evaluation of the envelope Eq. (C4) should also provide an alternative to study quantum transport in 1DEG nanowires TFETs.

Fig. 7 shows the impact of the polarization of momentum matrix elements (MMEs) on $I_D - V_{GS}$ characteristics of InAs-Si nanowire p-TFETs with diameters of (a) 5 nm and (b) 4 nm (inset figure). The thicker is the nanowire, the weaker is the impact of the polarization of MMEs. Thus for TFETs with larger size, the effect of the polarization of MMEs is expected to become less pronounced, as observed from the comparison of Figs. 7(a) and 7(b). Without the polarization of MMEs, the current is overestimated by a factor of 1.6 and 2.2, respectively. Hence, the inclusion of the polarization of MMEs due to the electric field might be crucial for a properly quantitative prediction of the BTBT current in nano-structures TFETs.

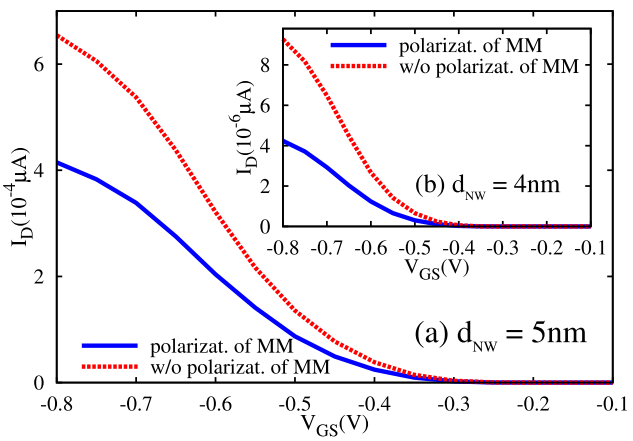

FIG. 7. Room-temperature $I_D - V_{GS}$ characteristics with and without the polarization of momentum matrix (MM) elements of InAs-Si nanowire p-TFETs with diameters of (a) 5 nm and (b) 4 nm (inset figure). The doping concentrations in the source n^+–InAs and drain p^+–Si regions are $N_D = 10^{19}\,cm^{-3}$ and $N_A = 2 \times 10^{19}\,cm^{-3}$, respectively, and $V_{DS} = -0.5$ V.

V. CONCLUSIONS

Based on a multi-band envelop function method, we have analytically derived an expression for the BTBT transmission probability in semiconductor devices within the EMA (labeled as "one-band EMA model"). The interaction H_{cv} between CVBs has been tackled within perturbation theory. Then, the non-equilibrium matrix density formalism was invoked to solve the quantum Liouville equation in the limit of linear response, i.e., up to first order in H_{cv}. Comparisons to the two-band Flietner model and to full-band, atomistic simulations were given for Ge and InAs p–n tunnel diodes. The calculation of the BTBT current was carried out in a post-processing step that used the converged potential from atomistic simulations as input. In case of the Ge $p - n$ diode, we have found that non-parabolicity corrections for the conduction band are necessary in the one-band model to obtain a satisfactory agreement with the atomistic simulations. The Flietner model can accurately reproduce the atomistic simulations in case of the Ge $p - n$ diode. The opposite was found for the InAs $p - n$ diode. A Flietner-model-like dispersion must be applied here to the transverse energy modes $E_{\mathbf{k}_\perp}$ in order to capture their non-parabolic shape. After the inclusion of non-parabolicity corrections for $E_{\mathbf{k}_\perp}$, the BTBT current computed with the Flietner model also reproduced the atomistic simulation with high accuracy. Hence, the band non-parabolicity of $E_{\mathbf{k}_\perp}$ could have a relevant impact on the proper simulation of InAs-based tunnel devices.

The one-band EMA model has also been applied to hetero-structures and devices with size quantization as nanowire TFETs. The model was tested on a InAs-Si Esaki diode and compared with both the Flietner model and the atomistic simulations yielding good agreement. Furthermore, InAs-Si nanowire TFETs were analyzed to gain insight into the accuracy and the differences between the two analytical models when used for low-dimensional systems. For the one-band EMA model, the polarization of the momentum matrix elements due to the electric field was included, otherwise the BTBT current is overestimated, e.g., by a factor slightly higher than 2 for the thinner nanowire considered in this work with diameter equals 4 nm, suggesting that the inclusion of the polarization of the momentum matrix element might be determining for the performance prediction of quantum-confined semiconductor devices. The quantum transport problem for electrons and holes was solved within the EMA using a mode-space NEGF simulator which neglects the CVBs. The self-consistent electrostatic potential from the converged Poisson-Schrödinger solution was taken

as input for the computation of the BTBT current. From the comparison with a modified WKB approximation, it turned out that the one-band model not only results in the same I_D–V_{GS} curve but is also expected to be more accurate in the case of thicker wires since it covers all the different tunnel paths and not only those parallel to the transport direction as in the modified WKB approximation.

From the aforementioned findings, we may finally conclude that the rigorous one-band EMA model for BTBT introduced in this work could be applied to larger homo- and hetero-structures with reduced computational burden, compared to an atomistic tool like OMEN, while keeping a high accuracy. Moreover, being based on the EMA, it could be implemented straightforwardly in existing quantum transport tools to account for BTBT in both 2DEG and 1DEG nanodevices.

ACKNOWLEDGMENTS

Funding from the European Community's Seventh Framework Programme under Grant Agreement No. 619509 (Project E^{2}SWITCH) is acknowledged.

APPENDIX A: MOMENTUM MATRIX ELEMENTS IN NANOSTRUCTURES

Assuming the total wave vector **k** parallel to the growth direction of zinc-blende-type semiconductor structures, e.g., the z-direction, the symmetrized p-like valence band wave functions at the Γ-point are represented by[6]

$$|1,\quad 1\rangle = \frac{1}{\sqrt{2}}(|X\rangle + i|Y\rangle), \tag{A1}$$

$$|1,\quad 0\rangle = |Z\rangle, \tag{A2}$$

$$|1,-1\rangle = \frac{1}{\sqrt{2}}(|X\rangle - i|Y\rangle). \tag{A3}$$

Similar to atoms, p states correspond to the orbital angular momentum $l=1$ triply degenerated with the magnetic quantum numbers $m_l=-1$, 0, and 1. In the presence of spin-orbit interaction, the total angular momentum $\mathbf{J}=\mathbf{1}+\mathbf{1/2}$ takes the values: 3/2 and 1/2 with $m_J=\pm 3/2$, $\pm 1/2$ and $m_J=\pm 1/2$, respectively. The states with $J=3/2$ are used to describe the heavy hole (HH) and light hole (LH) bands, whereas for the split-off band states with $J=1/2$ are employed. Since for the momentum matrix elements the HH band and LH bands are the only relevant bands, the states with $J=1/2$ will be disregarded in the following.

The states with $J=3/2$ are expressed as linear combination of the p-like and spin eigenfunctions[6] leading to

$$\begin{aligned}|3/2,\pm 3/2\rangle &= |1,\pm 1\rangle|\uparrow\rangle(|\downarrow\rangle)\\ &= \frac{1}{\sqrt{2}}(|X\rangle \pm i|Y\rangle)|\uparrow\rangle(|\downarrow\rangle),\end{aligned} \tag{A4}$$

$$\begin{aligned}|3/2,\pm 1/2\rangle &= \frac{1}{\sqrt{3}}\left[|1,\pm 1\rangle|\downarrow\rangle(|\uparrow\rangle) + \sqrt{2}|1,0\rangle|\uparrow\rangle(|\downarrow\rangle)\right]\\ &= \frac{1}{\sqrt{6}}[(|X\rangle \pm i|Y\rangle)|\downarrow\rangle(|\uparrow\rangle) + 2|Z\rangle|\uparrow\rangle(|\downarrow\rangle)],\end{aligned} \tag{A5}$$

where $|3/2,\pm 3/2\rangle$ and $|3/2,\pm 1/2\rangle$ are the eigenfunctions for the doubly degenerated HH and LH bands, respectively. In general, the electron wave vector **k** is not necessarily aligned with the symmetry axis z. Hence, the $|3/2,\pm 3/2\rangle$ and $|3/2,\pm 1/2\rangle$ eigenfunctions must be transformed by a spherical rotation of the basis functions $|X\rangle$, $|Y\rangle$, $|Z\rangle$.[28] Then, Eqs. (A4) and (A5) can be written in a more general form as[21]

$$\begin{aligned}|3/2,\pm 3/2\rangle = \frac{1}{\sqrt{2}}\{&(\cos\theta\cos\phi \mp i\sin\phi)|X\rangle\\ &+ (\cos\theta\sin\phi \pm i\cos\phi)|Y\rangle\\ &-\sin\theta|Z\rangle\}|\uparrow\rangle(|\downarrow\rangle),\end{aligned} \tag{A6}$$

for the HH band and

$$\begin{aligned}|3/2,\pm 1/2\rangle = \frac{1}{\sqrt{6}}\{&(\cos\theta\cos\phi \mp i\sin\phi)|X\rangle\\ &+ (\cos\theta\sin\phi \pm i\cos\phi)|Y\rangle\\ &-\sin\theta|Z\rangle\}|\downarrow\rangle(|\uparrow\rangle) + 2(\sin\theta\cos\phi|X\rangle\\ &+ \sin\theta\sin\phi|Y\rangle + \cos\theta|Z\rangle)|\uparrow\rangle(|\downarrow\rangle),\end{aligned} \tag{A7}$$

for the LH band. The angles θ and ϕ are the polar and azimuthal angles of the **k** vector measured from the symmetry axes z and x, respectively. Finally, for the conduction band, the eigenfunctions are given by $|iS\rangle|\downarrow\rangle(|\uparrow\rangle)$.

The square momentum matrix elements, e.g., between the conduction band and LH band, can now be evaluated from

$$|P_j|^2 = |\langle iS\downarrow\,|p_j|\,3/2,\,1/2\rangle|^2 + |\langle iS\downarrow\,|p_j|\,3/2,-1/2\rangle|^2, \tag{A8}$$

where j denotes the x, y or z directions. Then, using Eqs. (A7) and (A8), expressions for the square momentum matrix elements become $|P_j|^2 = C_j(\theta,\phi)p_{cv}^2/3$, where the functions $C_j(\theta,\phi)$ are

$$C_x(\theta,\phi) = \frac{1}{2}\left(\cos^2\theta\cos^2\phi + \sin^2\phi\right) + 2\sin^2\theta\cos^2\phi, \tag{A9}$$

$$C_y(\theta,\phi) = \frac{1}{2}\left(\cos^2\theta\sin^2\phi + \cos^2\phi\right) + 2\sin^2\theta\sin^2\phi, \tag{A10}$$

$$C_z(\theta,\phi) = \frac{1}{2} + \frac{3}{2}\cos^2\theta, \tag{A11}$$

and $p_{cv}^2 = |\langle iS\,|\,p_j\,|\,j\rangle|^2$. For a bulk crystal, by averaging each of the above elements over both angles θ and ϕ, they approach the single value $p_{cv}^2/3$.

In the case of a quantum well structure parallel to the xy-plane with the quantization direction along the z-axis, the square momentum matrix elements can be averaged over the angle ϕ[21]

$$C_x(\theta) = C_y(\theta) = \frac{1}{4}(1+\cos^2\theta) + \sin^2\theta, \tag{A12}$$

$$C_z(\theta) = \frac{1}{2} + \frac{3}{2}\cos^2\theta. \tag{A13}$$

θ can be found from the relation $E_{z,\nu} = E\cos^2\theta$. $E_{z,\nu}$ is the electron subband energy from the solution of Schrödinger equation in the confinement direction and E is the total energy. In nanowires, assuming confinement in the xy-cross

section and the transport direction along the z-axis, the average over the angle ϕ is not straightforward. Invoking cylindrical symmetry for the **k**-vector, one can observe that its polar $k_{\perp,\nu}$-component is the constant cylinder radius for a given electron subband energy W_ν. Hence, the changes in the magnitude of **k** are only due to the variation of its k_z-component. As a consequence, physical quantities such as the BTBT current in nanowire TFETs should become ϕ-independent. Finally, Eqs. (A12) and (A13) can be also used to calculate the square momentum matrix elements in nanowires with θ computed according to $\cos^2\theta = 1 - W_\nu/E$.

APPENDIX B: WKB APPROXIMATION

Between the valence (x_v) and conduction (x_c) band turning points, the WKB envelope wave functions are defined as

$$\chi_{v(c)\mathbf{k}_\perp}(x;E) = \sqrt{\frac{m_{v(c)}}{2\hbar^2}}k_{v(c)}^{-1/2}(x;E)\mathrm{e}^{\mp\int_{x_{v(c)}}^{x}\mathrm{d}x\,k_{v(c)}(x;E)}, \quad \text{(B1)}$$

and the diagonal elements of the spectral functions are given by

$$A_{v(c)\mathbf{k}_\perp}(x,x;E) = |\chi_{v(c)}(x,\mathbf{k}_\perp;E)|^2. \quad \text{(B2)}$$

The product of the spectral functions in Eq. (18) can be obtained by expanding the imaginary wave vector in the exponential, i.e., $k_{v(c)} = \sqrt{k_{x,v(c)}^2 + \mathbf{k}_\perp^2} \approx k_{x,v(c)} + \frac{1}{2k_{x,v(c)}}\mathbf{k}_\perp^2$, leading to

$$\begin{aligned}&\sqrt{A_{c\mathbf{k}_\perp}(x,x;E)A_{v\mathbf{k}_\perp}(x,x;E)}\\ &\approx \sqrt{\frac{m_c m_v}{4\hbar^4}}\frac{\mathrm{e}^{-\frac{\mathbf{k}_\perp^2}{2}\left(\int_{x_v}^{x}\mathrm{d}x k_{x,v}^{-1}+\int_{x}^{x_c}\mathrm{d}x k_{x,c}^{-1}\right)}}{k_{x,c}^{1/2}(x)k_{x,v}^{1/2}(x)}\mathrm{e}^{-\left(\int_{x_v}^{x}\mathrm{d}x k_{x,v}+\int_{x}^{x_c}\mathrm{d}x k_{x,c}\right)}\\ &\approx f_{\mathbf{k}_\perp}(x)\mathrm{e}^{-\left(\int_{x_v}^{x}\mathrm{d}x k_{x,v}+\int_{x}^{x_c}\mathrm{d}x k_{x,c}\right)}.\end{aligned} \quad \text{(B3)}$$

The second exponential will decay rapidly with increasing distance from its maximum x_0. By expanding its argument up to second order it is found that

$$\begin{aligned}&\sqrt{A_{c\mathbf{k}_\perp}(x,x;E)A_{v\mathbf{k}_\perp}(x,x;E)}\\ &\quad\approx f_{\mathbf{k}_\perp}(x_0)\mathrm{e}^{-\left(\int_{x_v}^{x_0}\mathrm{d}x k_{x,v}+\int_{x_0}^{x_c}\mathrm{d}x k_{x,c}\right)}\mathrm{e}^{-\frac{(x-x_0)^2}{2}\frac{(m_c+m_v)}{\hbar^2}\frac{|\nabla U(x_0)|}{k_0}},\end{aligned} \quad \text{(B4)}$$

where all the prefactors have been evaluated at x_0 and $k_{x,c(v)}(x_0) = k_0 = \sqrt{2\mu E_g/\hbar^2}$. Finally, the BTBT transmission probability takes the form

$$\begin{aligned}T(E) = &\left|\int \mathrm{d}x\, M_{cv}(x)\mathrm{e}^{-\frac{(x-x_0)^2}{2}\frac{(m_c+m_v)}{\hbar^2}\frac{|\nabla U(x_0)|}{k_0}}\right|^2\\ &\times\frac{A_\perp}{4\pi^2}\int \mathrm{d}\mathbf{k}_\perp f_{\mathbf{k}_\perp}^2(x_0)\cdot T_{WKB}(E),\end{aligned} \quad \text{(B5)}$$

with $T_{WKB}(E) = \exp\{-2(\int_{x_v}^{x_0}\mathrm{d}x k_{x,v}(x) + \int_{x_0}^{x_c}\mathrm{d}x k_{x,c}(x))\}$ and $A_\perp$ being the area across the transport direction.

For a 1D uniform electric field, we can proceed in two different ways starting from Eq. (B5). First, in the limit $F \to 0$, we could neglect the argument in the exponential and find that $\int \mathrm{d}x M_{cv}(x) \approx \sqrt{\hbar^2 E_g/4\mu}$. The integration over the transverse modes and the argument in the exponential of T_{WKB} then leads to

$$T = \frac{A_\perp(m_c+m_v)F}{128\pi\hbar\sqrt{2\mu E_g}}\exp\left(-\frac{4}{3}\frac{\sqrt{2\mu}E_g^{3/2}}{\hbar F}\right), \quad \text{(B6)}$$

and the BTBT generation rate, defined as $G = TF/\pi\hbar A_\perp$,[29] is

$$G = \frac{(m_c+m_v)}{128\pi^2\hbar^2\sqrt{2\mu E_g}}F^2\exp\left(-\frac{4}{3}\frac{\sqrt{2\mu}E_g^{3/2}}{\hbar F}\right). \quad \text{(B7)}$$

The pre-exponential factor contains the electric field to the same power as found by Kane in Ref. 5. However, if the argument in the exponential is not neglected, then

$$\frac{F^2\hbar^2}{4\mu E_g}\left|\int \mathrm{d}x\,\mathrm{e}^{-\frac{(x-x_0)^2}{2}\frac{(m_c+m_v)}{\hbar^2}\frac{F}{k_0}}\right|^2 = \frac{F^2\hbar^2}{4\mu E_g}\frac{2\pi\hbar^2 k_0}{F(m_c+m_v)}. \quad \text{(B8)}$$

Replacing k_0 by its explicit value, and integrating again over the transverse modes, the BTBT transmission probability T and generation rate G become, respectively,

$$T = \frac{A_\perp F^2}{64E_g^2}\exp\left(-\frac{4}{3}\frac{\sqrt{2\mu}E_g^{3/2}}{\hbar F}\right), \quad \text{(B9)}$$

$$G = \frac{F^3}{64\pi\hbar E_g^2}\exp\left(-\frac{4}{3}\frac{\sqrt{2\mu}E_g^{3/2}}{\hbar F}\right). \quad \text{(B10)}$$

Equation (B10) is exactly the same expression as given by Schenk *et al.* in Ref. 16. The pre-exponential factor contains the electric field to the third power.

In the case of non-uniform fields no analytical solution can be found for Eq. (B5), and the BTBT transmission probability must be numerically calculated by means of

$$T(E) = A_\perp C(x_0)\, T_{non-local}^{WKB}(E), \quad \text{(B11)}$$

where $C(x_0)$ and $T_{non-local}^{WKB}$ are, respectively,

$$\begin{aligned}C(x_0) = &\frac{1}{2\pi}\frac{m_c m_v}{4\hbar^4}k_0^{-2}\left|\int_{x_v}^{x_c}\mathrm{d}x\, M_{cv}(x)\exp\right.\\ &\left.\times\left\{-\frac{(x-x_0)^2}{2}\frac{(m_c+m_v)}{\hbar^2}\frac{|\nabla U(x_0)|}{k_0}\right\}\right|^2,\end{aligned} \quad \text{(B12)}$$

$$\begin{aligned}T_{non-local}^{WKB}(E) = &\left(\int_{x_v}^{x_c}\mathrm{d}x k_x^{-1}\right)^{-1}\left\{1-\exp\left(-k_m^2\int_{x_v}^{x_c}\mathrm{d}x k_x^{-1}\right)\right\}\\ &\times\exp\left\{-2\int_{x_v}^{x_c}\mathrm{d}x k_x\right\}.\end{aligned}$$

For brevity, the imaginary dispersion is written in an equivalent form as $k_x = \min(k_{x,v}, k_{x,c})$. The integration over the transverse modes $\mathbf{k}_\perp$ takes energy conservation into account. Hence, the maximum value that $\mathbf{k}_\perp$ can take is $k_m = \min(|\mathbf{k}_{v,m}|, |\mathbf{k}_{c,m}|)$. From Eq. (B13), one finds that the hole (electron) BTBT generation rate is given by

$$G_{\mathrm{v(c)}}\left(E_{\mathrm{v(c)}}\right) = \frac{|\nabla U(x)|}{\pi\hbar} C(x_0) T^{\mathrm{WKB}}_{\mathrm{non-local}}\left(E_{\mathrm{v(c)}}\right) \times \left(f_{\mathrm{v}}\left(E_{\mathrm{v(c)}}\right) - f_{\mathrm{c}}\left(E_{\mathrm{v(c)}}\right)\right). \quad \text{(B14)}$$

Note that Eq. (B14) agrees well with the *non-local path band-to-band tunneling model* formula[4] used by the commercial device simulator S-Device.

APPENDIX C: FLIETNER MODEL

The two-band Flietner model of the imaginary dispersion is defined as

$$\frac{\hbar^2 k^2}{2m_0} = \frac{E_{\mathrm{g}}(E - E_{\mathrm{c}})(E - E_{\mathrm{v}})}{\left[\sqrt{\frac{m_0}{m_{\mathrm{c}}}}(E - E_{\mathrm{v}}) - \sqrt{\frac{m_0}{m_{\mathrm{v}}}}(E - E_{\mathrm{c}})\right]^2}. \quad \text{(C1)}$$

Contrary to the Kane model, both the real conduction and valence bands in the vicinity of their extrema are correctly reproduced by the Flietner model

$$E \approx E_{\mathrm{c(v)}} \pm \frac{\hbar^2 k^2}{2m_{\mathrm{c(v)}}} + \cdots. \quad \text{(C2)}$$

Note that expression Eq. (C1) inherently involves non-parabolic corrections for the real conduction and valence bands. This is advantageous since by including an external potential $U(\mathbf{r})$ in Eq. (C2) one can set up the appropriate envelope equations for bulk and low-dimensional semiconductors that incorporate both real and imaginary branches of the total band structure. Starting from

$$E \approx E_{\mathrm{c(v)}} + U(\mathbf{r}) \pm \frac{\hbar^2 k^2}{2m_{\mathrm{c(v)}}} + \cdots, \quad \text{(C3)}$$

one can proceed as Flietner in Ref. 17 to find that electrons are subject to the following two-band envelope equation

$$-\frac{\hbar^2}{2m_0}\left(\frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial x^2}\right)\chi = \frac{E_{\mathrm{g}}\left(E - U_{\mathrm{c}}\right)\left(E - U_{\mathrm{v}}\right)}{\left[\sqrt{\frac{m_0}{m_{\mathrm{c}}}}(E - U_{\mathrm{v}}) - \sqrt{\frac{m_0}{m_{\mathrm{v}}}}(E - U_{\mathrm{c}})\right]^2}\chi. \quad \text{(C4)}$$

Here, $U_{\mathrm{c(v)}} = E_{\mathrm{c(v)}} + U$, and the coordinate vector $\mathbf{r}$ is omitted for brevity. Further simplifications could be applied depending on the method to solve Eq. (C4). In case of nanowire TFETs, the BTBT transmission probability can be computed from a variant of the WKB approximation proposed in Ref. 18

$$T_{\mathrm{WKB}}(E) = \frac{1}{A_{\mathrm{NW}}}\int_{A_{\mathrm{NW}}} \mathrm{d}^2\mathbf{r}_{\perp} \exp\left\{-2\int \kappa(x, \mathbf{r}_{\perp})\mathrm{d}x\right\}, \quad \text{(C5)}$$

where A_{NW} is the cross section area of the nanowire with $\mathbf{r}_{\perp} = (y, z)$ and the imaginary dispersion for $\kappa(x, \mathbf{r}_{\perp})$ is given by Eq. (C1). Note that in Eq. (C5), the transmission is averaged over the nanowire cross section. For bulk-like structures, the envelope equation

$$\left(E_{\mathbf{k}_{\perp}} - \frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial x^2}\right)\chi_{\mathbf{k}_{\perp}} = \frac{E_{\mathrm{g}}\left(E - U_{\mathrm{c}}\right)\left(E - U_{\mathrm{v}}\right)}{\left[\sqrt{\frac{m_0}{m_{\mathrm{c}}}}(E - U_{\mathrm{v}}) - \sqrt{\frac{m_0}{m_{\mathrm{v}}}}(E - U_{\mathrm{c}})\right]^2}\chi_{\mathbf{k}_{\perp}} \quad \text{(C6)}$$

must be directly solved for each given transverse energy $E_{\mathbf{k}_{\perp}} = \hbar^2\mathbf{k}_{\perp}^2/2m_0$. For Eq. (C6), the potential in $U_{\mathrm{c(v)}}$ only varies in the transport direction being completely one-dimensional.

[1] W. Fichtner, D. J. Rose, and R. E. Bank, IEEE Trans. Electron Dev. **30**, 1018 (1983).
[2] C. S. Rafferty, M. R. Pinto, and R. W. Dutton, IEEE Trans. Electron Dev. **32**, 2018 (1985).
[3] A. M. Ionescu and H. Riel, Nature **479**, 329 (2011).
[4] *Sentaurus-Device User Guide*, 9th ed. (Synopsys Inc., Mountain View, California, 2013).
[5] E. O. Kane, J. Phys. Chem. Solids **12**, 181 (1959).
[6] P. Yu and M. Cardona, *Fundamentals of Semiconductors: Physics and Materials Properties*, 3rd ed. (Springer, 2005), p. 68.
[7] Y. M. Niquet, A. Lherbier, N. H. Quang, M. V. Fernández-Serra, X. Blase, and C. Delerue, Phys. Rev. B **73**, 165319 (2006).
[8] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, Phys. Rev. B **74**, 205323 (2006).
[9] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[10] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[11] A. Pan and C. O. Chui, J. Appl. Phys. **116**, 054508 (2014).
[12] M. G. Burt, Semicond. Sci. Technol. **3**, 739 (1988).
[13] D. Verreck, M. V. de Put, B. Sorée, A. S. Verhulst, W. Magnus, W. G. Vandenberghe, N. Collaert, A. Thean, and G. Groeseneken, J. Appl. Phys. **115**, 053706 (2014).
[14] I. I. Ivanchik, Sov. Phys. Solid State **3**, 75 (1961).
[15] D. J. BenDaniel and C. B. Duke, Phys. Rev. **152**, 683 (1966).
[16] A. Schenk, M. Stahl, and H.-J. Wünsche, Phys. Status Solidi B **154**, 815 (1989).
[17] H. Flietner, Phys. Status Solidi B **54**, 201 (1972).
[18] H. Carrillo-Nuñez, M. Luisier, and A. Schenk, in *Proceeding of ESSDERC* (2014) pp. 118–121.
[19] O. Morandi and M. Modugno, Phys. Rev. B **71**, 235331 (2005).
[20] H. J. Wünsche and K. Henneberger, Phys. Status Solidi B **73**, 245 (1976).
[21] M. Yamanishi and I. Suemune, Jpn. J. Appl. Phys., Part 2 **23**, L35 (1984).
[22] H. Tanimoto, N. Yasuda, and K. Taniguchi, Jpn. J. Appl. Phys., Part 1 **27**, 563 (1988).
[23] M. V. Fischetti, IEEE Trans. Electron Dev. **38**, 634 (1991).
[24] M. T. Björk, H. Schmid, C. D. Bessire, K. E. Moselund, H. Ghoneim, S. Karg, E. Lörtscher, and H. Riel, Appl. Phys. Lett. **97**, 163501 (2010).
[25] A. S. Verhulst, W. G. Vandenberghe, K. Maex, S. D. Gendt, M. M. Heyns, and G. Groeseneken, IEEE Electron Dev. Lett. **29**, 1398 (2008).
[26] K. E. Moselund, H. Schmid, C. Bessire, M. T. Björk, H. Ghoneim, and H. Riel, IEEE Electron Dev. Lett. **33**, 1453 (2012).
[27] M. Luisier, A. Schenk, and W. Fichtner, J. Appl. Phys. **100**, 043713 (2006).
[28] E. O. Kane, J. Phys. Chem. Solids **1**, 249 (1956).
[29] E. O. Kane, J. Appl. Phys. **32**, 83 (1961).