A review of selected topics in physics based modeling for tunnel field-effect transistors

You may also be interested in:

Silicon and germanium nanowire electronics: physics of conventional and unconventional transistors
Walter M Weber and Thomas Mikolajick

Atomistic theory of transport
Alessandro Pecchia and Aldo Di Carlo

An improved empirical approach to introduce quantization effects in the transport direction in
multi-subband Monte Carlo simulations
P Palestri, L Lucci, S Dei Tos et al.

Microscopic theory of nanostructured semiconductor devices: beyond the envelope-function
approximation
Aldo Di Carlo

High-mobility Si and Ge structures
Friedrich Schäffler

2D Materials Advances: From Large Scale Synthesis and Controlled Heterostructures to Improved
Characterization Techniques, Defects and Applications
Zhong Lin, Amber McCreary, Natalie Briggs et al.

Novel nanoscale device architecture to reduce leakage currents in logic circuits
Deepanjan Datta, Samiran Ganguly, S Dasgupta et al.

Impact of technology scaling on analog and RF performance of SOI–TFET
P Kumari, S Dash and G P Mishra

Topical Review

A review of selected topics in physics based modeling for tunnel field-effect transistors

David Esseni[1], Marco Pala[2], Pierpaolo Palestri[1], Cem Alper[3] and Tommaso Rollo[1]

[1] DPIA—University of Udine, Via delle Scienze 206, I-33100 Udine, Italy
[2] Centre de Nanosciences et de Nanotechnologies, Centre National de la Recherche Scientifique, Université Paris-Sud, Université Paris-Saclay, F-91405 Orsay, France
[3] Nanoelectronic Devices Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland

E-mail: david.esseni@uniud.it

CrossMark

Abstract

The research field on tunnel-FETs (TFETs) has been rapidly developing in the last ten years, driven by the quest for a new electronic switch operating at a supply voltage well below 1 V and thus delivering substantial improvements in the energy efficiency of integrated circuits. This paper reviews several aspects related to physics based modeling in TFETs, and shows how the description of these transistors implies a remarkable innovation and poses new challenges compared to conventional MOSFETs. A hierarchy of numerical models exist for TFETs covering a wide range of predictive capabilities and computational complexities. We start by reviewing seminal contributions on direct and indirect band-to-band tunneling (BTBT) modeling in semiconductors, from which most TCAD models have been actually derived. Then we move to the features and limitations of TCAD models themselves and to the discussion of what we define *non-self-consistent quantum models*, where BTBT is computed with rigorous quantum-mechanical models starting from frozen potential profiles and closed-boundary Schrödinger equation problems. We will then address models that solve the open-boundary Schrödinger equation problem, based either on the non-equilibrium Green's function NEGF or on the quantum-transmitting-boundary formalism, and show how the computational burden of these models may vary in a wide range depending on the Hamiltonian employed in the calculations. A specific section is devoted to TFETs based on 2D crystals and van der Waals hetero-structures. The main goal of this paper is to provide the reader with an introduction to the most important physics based models for TFETs, and with a possible guidance to the wide and rapidly developing literature in this exciting research field.

Keywords: energy efficient switches, tunnel FETs, modeling and simulations, quantum transport, semi-classical transport, van-der-Waals transistors

Abbreviations and acronyms

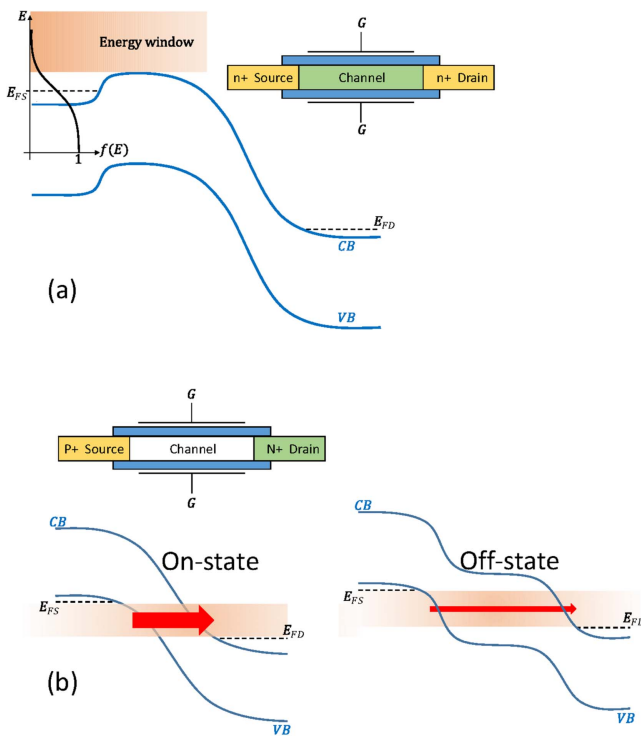|	
BJT	Bipolar junction transistor
BTBT	Band-to-band tunneling
CB	Conduction band
CMOS	Complementary metal-oxide-semiconductor
CMS	Coupled mode-space
CNT	Carbon nanotube
DFT	Density functional theory
DG	Double gate
DOS	Density of states

Figure 1. Sketch of the essential physics behind the electron injection from the source to the channel region of either an n-type MOSFET or an n-type tunnel FET. (a) MOSFET: where electrons are injected for energies above the top-of-the-barrier and according to the Fermi–Dirac occupation function in the source region; (b) tunnel FET: where electrons of the source are injected from the valence band, whose maximum sets an upper limit to the electron energy. A qualitative band-diagram is shown for both the OFF and the ON state of the tunnel FET.

EHBTFET	Electron–hole-bilayer TFETs
EMA	Effective mass approximation
FET	Field effect transistor
GAA NW	Gate-all-around NW
HH	Heavy holes
JDOS	Joint DOS
LDOS	Local DOS
LH	Light holes
MOSFET	Metal-oxide-semiconductor FET
NEGF	Non-equilibrium Green's function
NW	Nanowire
PAT	Phonon-assisted tunneling
PMI	Physical model interfaces
QCTP	Quantum-corrected tunneling path
QTB	Quantum transmitting boundary
SG	Single gate
SOI	Silicon-on-insulator
SRH	Shockley–Read–Hall
SS	Sub-threshold swing
TAT	Trap-assisted tunneling

TCAD	Technology computer aided design
TFET	Tunnel FET
TMD	Transition-metal dichalcogenides
UTB	Ultra-thin body
vdW	van der Waals
WKB	Wentzel Kramers Brillouin

1. Background and introduction

The progress in the computing and information technologies over the past four decades has been the enabler of a countless number of applications. In a foreseeable future, nanoelectronics will deliver self-powered, energy autonomous families of sensing, computing and communicating devices for many scenarios in the frameworks of the internet-of-things and internet-of-humans. As of today, however, power consumption is the main hindrance to the progress of the computing technologies: indeed integrated circuits simply do not have the energy budget necessary for the full exploitation of their potential performance. This utilization wall led to the so called *dark silicon age* where, at any point in time, significant fractions of the gates available on a chip are idle or significantly underclocked [1]. The origin of the utilization wall is that, in the CMOS technology generations after about the beginning of years 2000, it has been impossible to scale the power supply voltage, V_{DD}, as prescribed by the Dennardian scaling [2, 3], so that V_{DD} reached a plateau value of about 1 V. The dark silicon age officially marks the transition from the Dennardian scaling, where progress in CMOS technologies was measured in terms of improvements in transistor speed and number, to a new era where progress will be mainly measured in terms of energy efficiency.

A number of measures and design techniques against the power crisis have been devised at circuit and system level [4, 5], which in CMOS technology nodes from 130 to 90 nm allowed to improve the delay while reducing V_{DD} down to about 1 V. After the 90 nm technology node, however, it has been impossible to further scale V_{DD} and reduce the delay, so that designers had to find new avenues to convert geometrical scaling into performance enhancements. The most important of such avenues was parallelism, so that after 2005 the number of microprocessor cores started to double at each new technology cycle [6, 7]. The parallelism is not a solution to the utilization wall limit though, because its effectiveness ceases when each hardware unit approaches the minimum energy per operation [5, 6], and in fact the number of cores in microprocessors has already started to saturate.

The scaling of the power supply is probably the most effective measure to improve energy efficiency. The minimum V_{DD} for digital circuits has been discussed by several authors [8], and different authors driven by different perspectives found that V_{DD} could in principle be as small as a few $K_B T/e$ (i.e. about 0.1 V at room temperature), where K_B, T and e are respectively the Boltzmann constant, absolute

temperature and electron charge. It should be recalled that reducing V_{DD} by a factor of ten results in a $100\times$ save in dynamic power. A serious challenge to this aggressive V_{DD} scaling is the requirement to maintain the ratio $[I_{ON}/I_{OFF}]$ between on current I_{ON} and off current I_{OFF} to values of about 10^6: in fact $[I_{ON}/I_{OFF}]$ is ultimately set by the sub-threshold swing, *SS*, of the transistors. Consequently, several novel devices have been recently investigated [9, 10], to overcome the fundamental 60 mV/dec limit of the *SS* in CMOS transistors at room temperature [11].

CMOS transistors based on BTBT, usually referred to as TFETs, are the device concept most systematically developed and investigated in the last five to ten years [12]. The basic idea behind this device is to overcome the 60 mV/dec limit set by thermionic electron emission in CMOS transistors by injecting electrons in the channel (for an n-type device) from the valence band of the source, rather than from the conduction band (CB) as in conventional MOSFETs.

In fact in a nanoscale n-type MOSFET, illustrated in figure 1(a), electrons are injected from the source via thermionic emission and for energies above the top of the CB profile (neglecting the tunneling component below the top of the barrier, which should not be confused with a BTBT process). Consequently, when $\mathbf{V}_G$ is reduced and the top of the barrier is raised, I_{DS} is suppressed following the decay at large energy of the source occupation function. For energies larger than the source Fermi level, E_{FS}, the occupation decays as a Boltmann function, which leads to the well known thermally activated SS.

In the n-type TFET sketched in figure 1(b), instead, electrons are injected via BTBT from the valence band of the source region to the CB of the channel region. Hence the top of the valence band produces an energy filtering mechanism which suppresses the thermal tail of source electrons, provided that the source is not heavily degenerate, that is provided that E_{FS} is not several $K_B T$ below the top of the valence band. Such a transport mechanism is very different from a thermionic emission, and can in principle produce a subthreshold swing smaller than 60 mV/dec (at room temperature), and fairly independent of temperature.

The research field of TFETs has witnessed a rapid expansion in the last five to ten years and a lot of new material systems and device concepts have been proposed, with a number of papers devoted to performance benchmarking between TFETs and TFETs based circuits against mainstream CMOS transistors and circuits. While it is beyond the scope of this work to review the vast field of BTBT FETs, our aim is to review some of the methodologies developed and employed for the numerical modeling of TFETs. As can be seen, we do not even intend to cover all relevant modeling aspects, in fact we will not address compact models, and we will leave out also the methodologies employed for a circuit level benchmarking of TFETs against conventional CMOS transistors.

The present review paper is organized as follows. Section 2 describes a number of models for BTBT ranging from bulk materials subject to a uniform electric field, to nano-structured devices where quantum confinement results
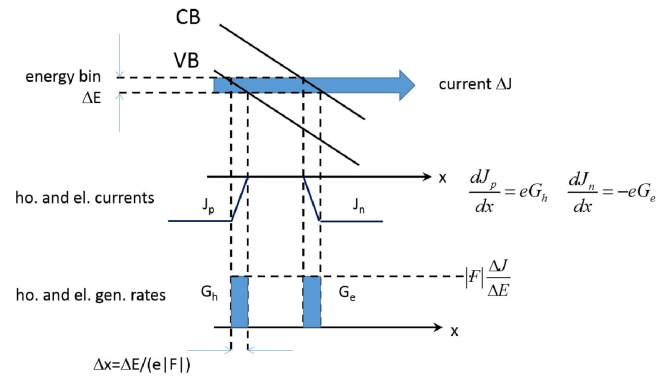

Figure 2. Sketch explaining the conversion between tunneling currents per unit energy ΔJ, the electron (J_n) and hole (J_h) currents inside the bands, and the BTBT generation rates, G_e and G_h. Case with uniform electric field F.

in significant changes of the band-structure compared to bulk materials. This broad section starts by reviewing seminal models in bulk materials, and then addresses models for complete device simulations including TCAD oriented models and more computationally intensive, full quantum transport models. Sections 3 and 4 describe several simulation case studies and comparisons to experiments concerning TFETs fabricated respectively with group IV or with III–V semiconductors and III–V based hetero-structures. Section 5 focuses on the recent but very vital field of TFETs based on two-dimensional (2D) semiconductors and, in particular, on van der Waals hetero-structures. In section 6 we finally propose a few concluding remarks.

2. Modeling of band-to-band-tunneling in Tunnel-FETs

In this section we review the main models developed for BTBT, focusing on their applications to TFETs. Models for BTBT have been developed since the 50s and applied to a large variety of devices such as reverse-biased diodes [13, 14], conventional MOSFETs (where BTBT takes place at the drain under high drain biases) [15, 16], and non-volatile memories [17–19]. The recent developments in TFETs have driven a renewed interest in this topic and many models have been revisited or originally proposed, in particular dealing with BTBT in the presence of quantum confinement, which was not so important in diodes and non-volatile-memories and thus not so much investigated.

Before entering the details of the individual models, it is worth recalling that BTBT is a quantum-mechanical mechanism corresponding to a flux of electrons from the valence to the CB. In principle a BTBT model should provide an energy resolved current spectrum. On the other hand, in a semi-classical modeling framework BTBT can be viewed as a carrier generation mechanism, where a hole is generated in the valence band at the beginning of the tunneling path, and an electron is generated in the CB at the end of the tunneling path. The two points of view are however equivalent, as illustrated pictorially in figure 2: in fact the current in the

forbidden gap corresponds to electron and hole currents flowing respectively at the left and right side of the tunneling path. These currents become null at the classical turning points, which results in a non null derivative of the currents over space, hence in a generation rate in the current continuity equation. As a result, the current density ΔJ in the energy bin ΔE corresponds to a generation rate of holes (left side of the tunneling path) and electrons (right side) given by:

$$G_{\mathrm{T}}(E) = |F|\frac{\Delta J(E)}{\Delta E}, \qquad (1)$$

where F is the electric field.

We now start reviewing the main models and modeling approaches for BTBT and first of all identify different categories. Since the BTBT current is due to tunneling inside the forbidden gap, we can identify models describing *direct tunneling* between valence and CB extrema both located at the Γ point, as well as *indirect*, *phonon-assisted* BTBT models that describe tunneling between valence and CB extrema located at different points of the Brillouin zone. In silicon, for example, indirect BTBT occurs between the maximum of the valence band in Γ and the minima of the CB in Δ. We can also classify the models based on the dimensionality of the involved carrier gases: models for 3D carrier gases are adequate for diodes and bulk-like or thick body MOSFETs; on the other hand, TFETs with ultra-thin-body architectures as well as devices such as the electron–hole-bilayer-TFETs, EHBTFETs, based on tunneling between quantized electron and hole gases require more advanced models. Those models are mainly based on the solution of the Schrödinger equation with open boundary conditions, such as the NEGF method. Since the computational burden of such approaches is very large, in particular when dealing with phonon assisted tunneling (PAT) and in large devices, alternative models have been developed which are based on the post-processing of the subband minima and wave-functions obtained by solving the Schrödinger equation with closed boundary conditions: in the following we will refer to such models as *non-self-consistent* quantum models.

2.1. Models for direct BTBT in bulk semiconductors

Kane proposed his BTBT theory in his seminal paper in 1959 [13]. The paper made use of perturbation theory to write the Schrödinger equation as a sum of interband and intraband coupling elements. Here we provide a different, simpler proof of his results based on the Landauer conduction formula and WKB approximation. We start with a so-called *local model*, that is a model that assumes a constant electric field along the tunneling path and an uniform structure.

For the derivation we start by writing the general expression of the current density using the Landauer's conduction formula [20]:

$$J = \frac{e}{\pi\hbar A}\int_{-\infty}^{\infty}\sum_{\mathbf{k}_\perp}T(E,\mathbf{k}_\perp)[f_v - f_c]\mathrm{d}E$$
$$= \frac{e}{4\pi^3\hbar}\int_{-\infty}^{\infty}\int_A T(E,\mathbf{k}_\perp)[f_v - f_c]\mathrm{d}\mathbf{k}_\perp\mathrm{d}E, \qquad (2)$$

where e is the positive electron charge, $\hbar$ is the reduced Plank constant, A is the area of the device in the direction normal to tunneling, $\mathbf{k}_\perp$ is the transverse wave-vector, $T(E,\mathbf{k}_\perp)$ is the transmission probability and f_v, f_c are the local occupation functions. For simplicity, hereafter we set $f_v = 1$ and $f_c = 0$, even if this implies that equation (2) gives non-zero current when zero bias is applied at the tunneling junction [21]; this problem is solved for example in [14]. The f_c and f_v will appear explicitly in the models in the section 2.2, but here we remain consistent with [13] and set $f_v = 1$ and $f_c = 0$. We now convert the tunneling current per unit energy in a generation rate according to equation (1), hence following the physical picture illustrated in figure 2. In particular, we evaluate the integral in equation (2) by moving to polar coordinates; we further assume that the transmission depends on the energy E and only on the magnitude $k_\perp$ of the wave-vector $\mathbf{k}_\perp$ and thus obtain

$$G_{\mathrm{T}}(E) = \frac{e|F|}{4\pi^3\hbar}2\pi\int_0^{+\infty}\mathrm{T}(E,k_\perp)k_\perp\mathrm{d}k_\perp, \qquad (3)$$

where the tunneling probability in equation (3) can be expressed by using the WKB approximation [22]:

$$T(E,k_\perp) = \frac{\pi^2}{9}\exp\left(-2\int_{x_i}^{x_f}\mathrm{Im}(k_x)\mathrm{d}x\right) \qquad (4)$$

and the $\pi^2/9$ factor is discussed in [23]. In order to evaluate the integral in equation (4) we need an E versus $\mathbf{k}$ relation valid in the energy gap, that is connecting the valence band to the CB. Moreover, since direct BTBT is an elastic process, total transverse momentum and energy need to be conserved. We employ here the Kane's two-band dispersion relation:

$$E^{\mathrm{valence}}(k) = \frac{E_G}{2} + \frac{\hbar^2 k^2}{2m_0} - \frac{1}{2}\sqrt{E_G^2 + \frac{E_G\hbar^2 k^2}{m_r}}, \qquad (5a)$$

$$E^{\mathrm{conduction}}(k) = \frac{E_G}{2} + \frac{\hbar^2 k^2}{2m_0} + \frac{1}{2}\sqrt{E_G^2 + \frac{E_G\hbar^2 k^2}{m_r}}, \qquad (5b)$$

$$E^{\mathrm{gap}}(\kappa) = \frac{E_G}{2} - \frac{\hbar^2\kappa^2}{2m_0} \pm \frac{1}{2}\sqrt{E_G^2 - \frac{E_G\hbar^2\kappa^2}{m_r}}, \qquad (5c)$$

where m_0 is the rest electron mass, $m_r^{-1} = (m_c^{-1} + m_v^{-1})^{-1}$, $k = |(k_x, \mathbf{k}_\perp)|$ is the magnitude of the wave vector. Inside the gap, since k_x is imaginary (and thus $k_x^2 = -\mathrm{Im}(k_x)^2$), it is easier to express the energy as a function of $\kappa^2 = -k^2 = [\mathrm{Im}(k_x)^2 - k_\perp^2]$. The $\pm$ sign in equation (5) indicates that inside the gap we have two possible energy values for given κ.

We now set $x = 0$ at $E = 0$, that is at the point where the top of the valence band crosses the constant energy line that we use as reference. Assuming the electric field F is uniform and along x direction, the energy is given by $E^{\mathrm{gap}} = e|F|x$. Neglecting the $\hbar^2\kappa^2/(2m_0)$ term in equation (5c), the imaginary part of k_x inside the bandgap can be written as:

$$\mathrm{Im}(k_x) = \sqrt{\frac{m_r}{E_G\hbar^2}}\sqrt{E_G^2 + E_G\frac{\hbar^2 k_\perp^2}{m_r} - 4\left(e|F|x - \frac{E_G}{2}\right)^2}. \qquad (6)$$

After determining the classical turning points of the valence and CBs, which deviate from $x_i = 0$ and $x_f = eE_G/|F|$ due to non-null parallel momentum $\mathbf{k}_\perp$, the transmission coefficient can be obtained by evaluating the integral over x in equation (4), that leads to

$$T(k_\perp) = \frac{\pi^2}{9} \exp\left(-\pi \frac{\sqrt{m_r}E_G^{\frac{3}{2}}}{2\hbar e|F|}\right) \exp\left(-\frac{\pi\hbar k_\perp^2}{2e|F|}\sqrt{\frac{E_G}{m_r}}\right). \quad (7)$$

By substituting equation (7) in equation (3) we find that the integral over $k_\perp$ can be evaluated analytically and finally reach the well-known formula [13]:

$$G_T = \frac{e^2|F|^2\sqrt{m_r}}{18\pi\hbar^2\sqrt{E_G}} \exp\left(-\pi\frac{\sqrt{m_r}E_G^{\frac{3}{2}}}{2\hbar e|F|}\right). \quad (8)$$

The local models, such as the one presented above, tend to grossly overestimate the current when the electric field rapidly changes over small distances. In order to overcome this issue within the semiclassical modeling framework, the so-called *non-local* models were introduced [24, 25]. One main goal of these models was to account for the creation of carriers at either side of the tunneling barrier (see also figure 2), in contrast to the local models which unphysically assign electron and hole generation at the same position [26].

As an illustrative example, a non-local direct BTBT model based on Kane's two band dispersion can be obtained using an analytical expression for the imaginary wave-vector κ inside the forbidden gap derived from equation (5) as:

$$\kappa(x) = \frac{1}{\hbar}\sqrt{m_r E_G (1 - \alpha^2(x))}$$
$$\alpha(x) = -\frac{m_0}{2m_r} + 2\sqrt{\frac{m_0}{2m_r}\left(\frac{E - E_V(x)}{E_G} - \frac{1}{2}\right) + \frac{m_0^2}{16m_r^2} + \frac{1}{4}} \quad (9)$$

where the energy E is inside the gap, namely we have $E_V(x) \leqslant E \leqslant E_V(x) + E_G$.

Making use of equation (4) and assuming that only small $k_\perp$ values result in significant tunneling probability, we expand $\text{Im}(k_x)$ for small $[k_\perp/\kappa]$ as $\text{Im}(k_x) = \sqrt{\kappa^2 + k_\perp^2} \simeq \kappa + \frac{k_\perp^2}{2\kappa}$ and find:

$$T = \frac{\pi^2}{9}\exp\left(-2\int_{x_i}^{x_f}\text{Im}(k_x)dx\right)$$
$$= \frac{\pi^2}{9}\exp\left(-2\int_{x_i}^{x_f}\kappa dx\right)\exp\left(-k_\perp^2\int_{x_i}^{x_f}\frac{dx}{\kappa}\right), \quad (10)$$

where we implicitly assume that the physical system is uniform along the y and z direction. In principle the integration extremes x_i and x_f should depend on $k_\perp$. A further simplification is thus introduced by evaluating x_i and x_f as the classical turning points corresponding to $k_\perp = 0$. Inserting the expression for T into equation (3) and evaluating the integral over $k_\perp$, we have:

$$G_T(E) = \left|\frac{dE_V}{dx}\right|_{x_i}\frac{e}{36\hbar}\left(\int_{x_i}^{x_f}\frac{dx}{\kappa}\right)^{-1}$$
$$\times\left[1 - \exp\left(-k_m^2\int_{x_i}^{x_f}\frac{dx}{\kappa}\right)\right]\times\exp\left(-2\int_{x_i}^{x_f}\kappa dx\right)(f_c - f_v), \quad (11)$$

where we have now reintroduced the local occupation functions at the beginning and at the end of the tunneling path. The G_T from equation (11) corresponds to a generation of holes at position x_i and a generation of electrons at position x_f. The electric field F used in equation (1) is here replaced by the derivative of the valence band energy at the beginning of the tunneling path. By considering different energies E in the range $[E_{\min}, E_{\text{MAX}}]$ between minimum value of $E_C(x)$ and the maximum value of $E_V(x)$, equation (11) provides the generation rate of electrons and holes along the whole structure. In this respect, the parameter k_m represents the maximum value of $k_\perp$ and it is given by:

$$k_m^2 = \min\left(\frac{2m_v(E_{\text{MAX}} - E)}{\hbar^2}, \frac{2m_c(E - E_{\min})}{\hbar^2}\right). \quad (12)$$

Equation (11) is precisely the formula used in the *dynamic-path non-local model* implemented in some commercial simulators [25].

A similar expression has been used in [27], where the integral of the wave-vector along the tunneling path is performed by solving the equation of motion of the imaginary energy dispersion in the gap using the Monte Carlo method.

We will return to the discussion of non-local tunneling models in section 2.4.

2.2. Direct BTBT in quantized systems: non-self-consistent quantum models

In the presence of quantization, different modeling approaches can be followed depending on the different alignment between tunneling and quantization directions. In fact, in a planar device, tunneling can be in-line with the quantization direction, as in the EHBTFET [28], or it can be transverse to the quantization direction. In this paragraph we give more details for the case with tunneling aligned with quantization because the picture changes more dramatically compared to the transverse quantization case, for which expressions similar to equation (11) have been proposed, where only the prefactor is different due to the different dimensionality of $\mathbf{k}_\perp$ [29, 30]. We will return to this point in section 2.4.

The models we are reviewing here consider BTBT as a post-processing calculation after the electrostatics has been determined by the self-consistent solution of the closed boundary Schrödinger and Poisson equations (quantum mechanical approach), or the semi-classical continuity equations. This is unlike the NEGF approach or similar quantum transport formalisms (see section 2.5), where transport and electrostatics are inherently coupled and the Schrödinger equation is solved with open boundary conditions along the transport direction.
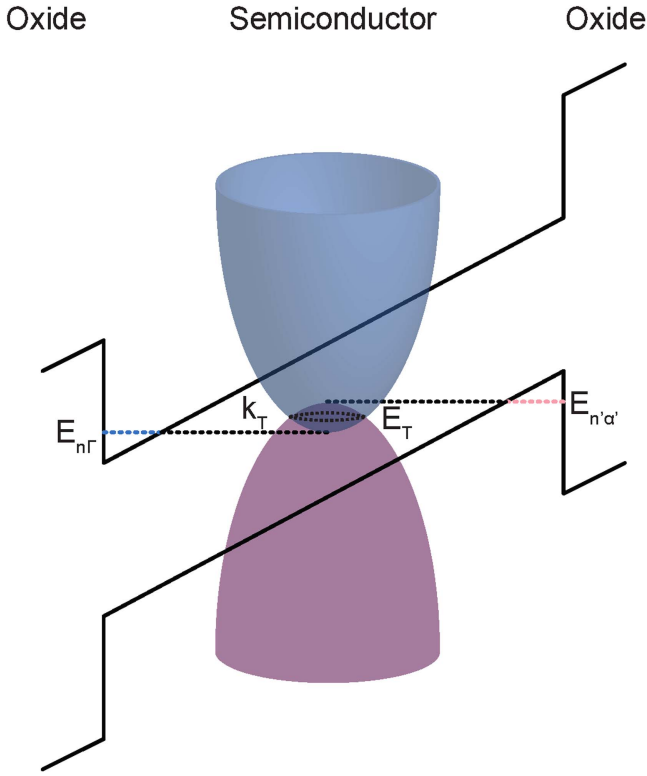
Oxide　　　　　Semiconductor　　　　　Oxide

Figure 3. Sketch of BTBT between two quantized carriers gases. The conservation of both the total energy and the carrier momentum perpendicular to the tunneling direction determines the tunneling energy $E_T = E_{n\Gamma} + \hbar^2 k_\perp^2/(2m_e) = E_{n'\alpha'} - \hbar^2 k_\perp^2/(2m_h)$.

For tunneling along the quantization direction, the integral over the total energy in equation (2) is replaced by a discrete sum since the energy spectrum of the carrier gas consists of a set of discrete subband levels. Using the Fermi's golden rule approach, one can write the following general expression:

$$J = \frac{4\pi e}{\hbar} \sum_n \sum_{\alpha'=\{HH,LH\},n'} |M_{\alpha',n'}^n|^2 \mathrm{JDOS}(E_T)$$
$$\times \Theta(E_{n'\alpha'} - E_{n\Gamma})[f_c(E_T) - f_v(E_T)], \qquad (13)$$

where the summation runs over all possible electron–hole subband pairs (n = electrons, n' = hole). Note that simultaneous conservation of total energy and transverse momentum (since direct BTBT is an elastic process) results in a single tunneling energy E_T for each pair of subbands (see figure 3). The joint density of states (JDOS) for the case of tunneling between two 2D carrier gases is given by

$$\mathrm{JDOS}_{2D}(E_T) = \frac{\overline{m}}{4\pi\hbar^2} \Theta(E_{n'\alpha'} - E_{n\Gamma}) \quad \text{with}$$

$$E_T = E_{n\Gamma} + \frac{\hbar^2 k_\perp^2}{2m_e} = E_{n'\alpha'} - \frac{\hbar^2 k_\perp^2}{2m_h}. \qquad (14)$$

Here $\overline{m}$ is defined as $\overline{m} = 2m_e m_h/(m_e + m_h)$ and we see how E_T is determined by energy and parallel momentum conservation.

There are two main approaches to calculate the interband coupling coefficient $M_{\alpha',n'}^n$ in the presence of quantization.

One is based on Bardeen's transfer Hamiltonian approach [31], where the matrix element for BTBT between two 2D carrier gases can be evaluated as [32]:

$$|M_{\alpha',n'}^n|^2 = \hbar^2 \frac{E_G^\Gamma}{2\overline{m}} \delta_{k_\perp,k'_\perp} |\psi_n \xi_{\alpha',n'}|^2 C_0(\theta), \qquad (15)$$

where E_G^Γ is the energy gap at the Γ point, $\delta_{k_\perp,k'_\perp}$ is the Kronecker symbol, ψ_n and $\xi_{\alpha',n'}$ are the envelope wave-functions for the two different carrier gases and the product $|\psi_n \xi_{\alpha',n'}|$ is somewhat arbitrarily taken at the point where it is maximum. An important difference compared to the bulk case is the inclusion of the term $C_0(\theta)$, where the angle θ represents the *direction* of the hole state. In [32], for example, the angle θ is estimated as

$$\cos^2\theta = \frac{E_h^z}{E_h^\perp + E_h^z}, \qquad (16)$$

where E_h^z and $E_h^\perp$ are effective kinetic energies along the quantization and transverse directions, respectively. Expressions for $C_0(\theta)$ have been given in [33, 34]. This term needs to be introduced in the presence of quantization since it accounts for the dependence of the coupling element on the direction of the electric field. In fact, ψ_n and $\xi_{\alpha',n'}$ are the envelope wave-functions, whereas the coupling between the electron and hole states is ruled by the underlying Bloch functions. Due to the symmetry properties of the atomic-like orbital functions, this dependence on the angle θ does not appear in the bulk case since for each envelope function we perform an average over all directions of the underlying Bloch functions. In the 2D gas, instead, the average takes place only over the plane normal to quantization, which introduces a dependence on θ.

A second approach to evaluate $M_{\alpha',n'}^n$ in equation (13) is similar to the perturbation method employed by Kane, where the presence of the electric field couples states in the conduction and valence band. This methodology has been presented in [35], and then recently employed to 2D–2D tunneling in [36]. The matrix element is given by:

$$M_{\alpha',n'}^n = e\sqrt{\frac{\hbar^2}{4\overline{m}E_G^\Gamma}} \delta_{k_\perp,k_\perp'} \int \psi_n^*(z) \sqrt{C_0(\theta)} |F(z)| \xi_{\alpha',n'}(z)\mathrm{d}z, \qquad (17)$$

where $F(z)$ is the electric field, that is allowed to vary only along the quantization direction z.

Figure 4 compares the currents of a Ge quantum-well diode as obtained with either equation (17) or equation (15) for the interband coupling. Results are similar, however equation (17) can be extended quite easily to describe physical systems where the electric field is non-uniform in more than one dimension (i.e. planar TFETs [36]), or quantization is present in more than one dimension (i.e. nanowire TFETs [37]). Reference [37] proposes a model for direct tunneling in bulk, 2D and 1D carrier gasses. It also presents a derivation of the non-local model for direct BTBT in bulk structures similar to our derivation in section 2.1.

An important point to note for tunneling in quantized gases is the large asymmetry between the effective masses of
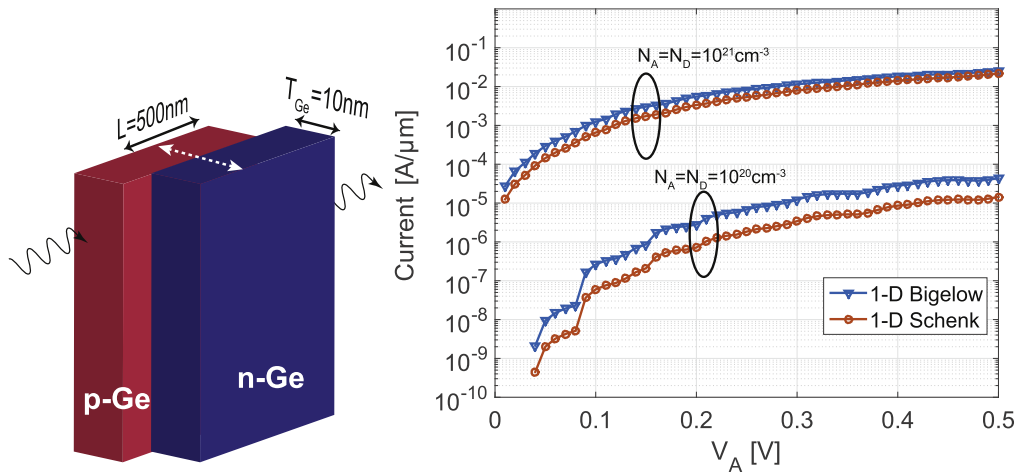
Figure 4. Comparison between the current in a Ge quantum-well diode (sketched in the left) as obtained with the equation (17) (*Schenk model*) or with equation (15) (*Bigelow model*).

the real and the imaginary dispersion for quantized holes. While quantization typically favors heavy hole (HH) sub-bands to contribute to the current, the calculation of the imaginary energy dispersion in the gap shows that the effective mass for the imaginary band is actually much closer to the light hole (LH) mass in the bulk crystal [38]. This implies that the interband tunneling matrix element in the presence of quantization is close to the one in the corresponding bulk material even if LHs do not come into play [39].

Band structure effects, such as the *anti-crossing* described above and the influence of quantization on the energy dispersion in the gap, are naturally accounted for in the **k·p** approach. We will discuss the simulation approach based on the NEGF with a **k·p** Hamiltonian in section 2.5. It is however worth mentioning that the **k·p** can be employed also in *non-self-consistent* quantum models. A relevant example is given in [40, 41], where the **k·p** method and the corresponding envelope wave-functions have been used with a quantum-transmitting-boundary formalism to study the transmission probability in hetero-junction III–V TFETs. The methodology in [40, 41] can be labeled as *non-self-consistent* because the potential profile is taken from TCAD simulations and then the BTBT current is calculated, but the possible influence on the potential profile of the charge produced by BTBT is not accounted for.

As an example of results that can be obtained with the models described in this section, we show in figure 5 the *I–V* characteristic of an EHBTFET with $In_{0.53}Ga_{0.47}As$ channel. In the EHBTFET we have two independent gates that create an electron and a hole inversion layer at the top and bottom of a thin film semiconductor [28], so that BTBT occurs between such two inversion layers. We assume that the potential profile is essentially uniform along the channel direction x, so that we can develop our calculations in a single vertical slice along z. The simulation methodology employed for the results in figure 5 starts by solving Schrödinger equation for electrons and holes for closed boundary conditions self-consistently with the Poisson equation. When computing the charge, electrons and holes are assumed to be at equilibrium
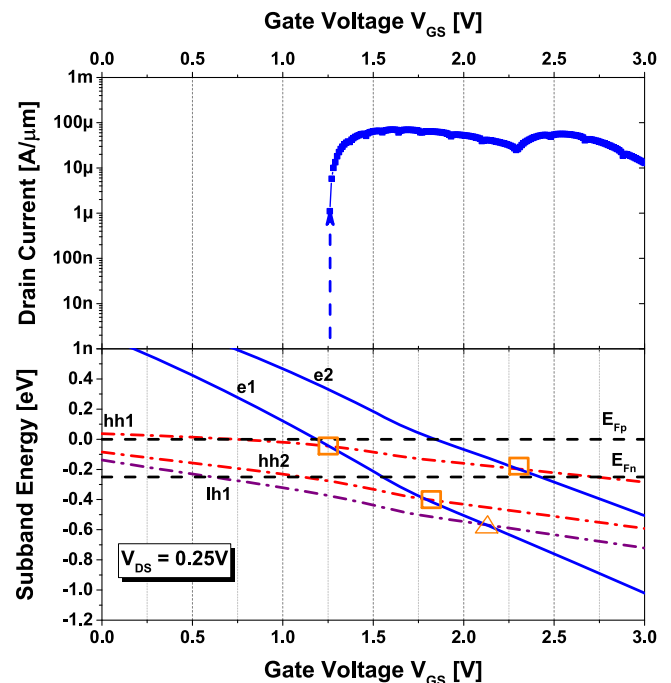

Figure 5. Drain current (top) and subband energies (bottom) versus the *n*-gate voltage for an $In_{0.53}Ga_{0.47}As$ EHBTFET with equivalent oxide thickness of 1 nm and channel thickness of 10 nm. The *p*-gate voltage is set to −1 V and V_{DS} is 0.25 V. In the bottom plot, the energy reference is the hole Fermi level E_{Fp} that is set by the source contact. The electron Fermi level is set by the drain voltage and it is $E_{Fn} = -eV_{DS}$.

with respectively the drain and source contacts: the hole Fermi-level E_{Fp} is set by the source contact, while electron Fermi-level E_{Fn} is set by the drain contact and is equal to $E_{Fp} - eV_{DS}$. The BTBT current is then computed as a post-processing using equations (13), (17), where the hole envelope functions are modified to account for the anti-crossing between LH and HH bands described above. The bottom of figure 5 shows the subband energy versus gate bias. We see that the current bumps in the top graph correspond to specific subband alignments which occur when the gate voltage

increases and moves the electron subbands to lower energy; the hole subbands are also slightly affected by the gate voltage due to coupling between the top and the back interface. In particular, we observe a first bump in the current when e1 (lowest electron subband) crosses the hh1 (highest HH subband). The second bump takes place when e2 aligns with hh1. On the other hand, the alignments between e1 and hh2 and between e1 and lh1 do not produce any current bump because they occur at energy values outside the window between E_{Fn} and E_{Fp}, so that the term $[f_c - f_v]$ is practically null.

2.3. Non-self-consistent quantum models for phonon-assisted BTBT

Phonon-assisted tunneling occurs in indirect gap semiconductors whose CB minima do not lie at the Γ point: Si and Ge are prominent examples. Keldysh first proposed an expression describing the phonon-assisted BTBT by making use of perturbation theory [42]. Later on, one of the first non-local BTBT models was proposed by Tanaka for phonon-assisted and direct tunneling [24]. In this model, the interband coupling due to electron–phonon interaction in the deformable ion model is used for interband coupling elements of the Wannier formula. The wavefunctions are obtained patching the plane wave solutions (in classically allowed regions) with the exponentially decaying components (in classically forbidden regions) by using the WKB approximation. Compared to the case of direct tunneling, in PAT the energy dispersion in the gap can be obtained from the imaginary branch of the energy relation of the two individual bands, i.e. the phonon connects two branches that can be unequivocally identified as belonging to the conduction or to the valence band.

A recent model that is applicable to non-uniform potential profiles and to a carrier gas with different dimensionality has been proposed by Vandenberghe et al [43]. The approach makes use of the diagonal elements of the spectral functions that, for a 2D gas described by the parabolic effective mass approximation (EMA), can be written as

$$A_{c\alpha}(z, E) = \frac{g_\alpha m_{xy,\alpha}}{\hbar^2} \sum_{\mathbf{k}_\perp} \psi^*_{\mathbf{k}_\perp \alpha}(z) \psi_{\mathbf{k}_\perp \alpha}(z) \Theta(E - E_{\mathbf{k}_\perp \alpha}),$$

$$(18a)$$

$$A_{v\alpha}(z, E) = \frac{g_\alpha m_{xy,\alpha}}{\hbar^2} \sum_{\mathbf{k}_\perp} \psi^*_{\mathbf{k}_\perp \alpha}(z) \psi_{\mathbf{k}_\perp \alpha}(z) \Theta(E_{\mathbf{k}_\perp \alpha} - E),$$

$$(18b)$$

where z is the quantization direction and the structure is uniform along x and y directions, $\Theta(E)$ is the step function (corresponding to a step-like 2D DOS), g_α, $m_{xy,\alpha}$ are respectively the valley degeneracy and transverse effective mass for the subband α. Since wave-functions ψ can be assumed to be independent of $\mathbf{k}_\perp$, the spectral functions essentially reduce to summations over the 2D DOS functions of the different subbands weighted by the probability distribution of carriers.

The total phonon-assisted tunneling current is evaluated by first calculating the energy dependent transmission probability:

$$T_v^{\text{abs,em}}(E) = \frac{D_{\text{ph}}^2 \hbar^2}{2\rho E_{\text{ph}}} \sum_{\alpha,\alpha'} \int A_{v\alpha'}(z, E) A_{c\alpha}(z, E \pm E_{\text{ph}}) \mathrm{d}z,$$

$$(19)$$

where D_{ph} and E_{ph} are the deformation potential and the energy of the phonon, respectively, and ρ is the mass density of the material. The phonon-assisted BTBT current is finally calculated as the summation of phonon emission and absorbtion terms:

$$
\begin{aligned}
J_{\text{ph}-} = -\frac{e}{\pi\hbar} \int &(\{f_v(E)[1 - f_c(E - E_{\text{ph}})][n_{\text{B}}(E_{\text{ph}}) + 1] \\
&- f_c(E - E_{\text{ph}})[1 - f_v(E)]n_{\text{B}}(E_{\text{ph}})\} T_v^{\text{em}}(E) \\
&+ \{f_v(E)[1 - f_c(E + E_{\text{ph}})]n_{\text{B}}(E_{\text{ph}}) \\
&- f_c(E + E_{\text{ph}})[1 - f_v(E)][n_{\text{B}}(E_{\text{ph}}) + 1]\} \\
&\times T_v^{\text{abs}}(E)) \mathrm{d}E,
\end{aligned}
$$

$$(20)$$

where f_c and f_v are the occupation functions of the conduction and valence bands, respectively, and $n_{\text{B}}(E_{\text{ph}})$ is the Bose–Einstein distribution for the phonons. For device structures such as EHBTFETs, f_c and f_v can be taken as Fermi–Dirac functions with respectively electron and hole Fermi levels (whose difference is set to eV_{DS}), which is the assumption used also to obtain the results of figure 5 discussed at the end of section 2.2.

In [43] it is shown that, if the modeling framework described above is applied to the bulk crystal (using the corresponding spectral functions) and for a uniform electric field, then one obtains the same expression for the tunneling generation rate as in [42]. For a non uniform field, instead, it can be shown that, by introducing suitabe approximations for the envelope wave-functions, the formalism in equation (20) provides an expression for non-local tunneling in bulk systems similar to the dynamic-path, non-local-phonon-assisted model used in TCAD [25], and similar also to the model derived in [24].

2.4. Models for BTBT in commercial TCAD

Models for BTBT are present in many commercial TCAD tools. Besides the local models based on the expressions in [13, 42] (usually reformulated to assure a zero tunneling generation rate at equilibrium [14]), commercial TCAD tools have also recently implemented non-local models similar to equation (11) for direct tunneling and phonon-assisted tunneling [25], based on the approach developed in [24].

As discussed in section 2.1, these models have been derived for bulk structures, where the carriers are not confined. The effect of carrier quantization is manifold. First of all it increases the effective energy gap and changes the density of states (DOS) according to the dimensionality of the carrier gas. In addition, quantization changes the dispersion relationship inside the gap. Many efforts are being

devoted at present to account for these effects in TCAD models.

As for the increase of the effective energy gap, in [44] the energy profile of the CB close to the channel/dielectric interface is modified to mimic the presence of the electron subband splitting. More precisely the part of the profile $E_C(z)$ that is lower than the lowest electron subband E_{e0} is set to $E_C(z) = E_{e0}$. As an alternative approach, an effective gap obtained from the solution of the 1D Schrödinger equations for electrons and holes has been also used in [45, 46].

A similar methodology has been used in semi-classical Monte Carlo simulations [30] that implements non-local models for direct and phonon assisted BTBT as in [25]. First the 1D Schrödinger equation for electrons and holes is solved in each slice along the channel, then the conduction (E_C) and valence (E_V) band profiles are modified so as to suppress E_C values below the energy of the lowest electron subband E_{e0} and the E_V values above the highest hole subband E_{h0}. The non-local model with the effective gap correction was found to match quite well the full-quantum non-self-consistent model described in section 2.3 [43]. More recently, the authors of [47] proposed a methodology based on the rejection of those tunneling paths that involve states in the CB below E_{e0} or states in the valence band above E_{v0}. Thanks to the use of physical models interfaces, PMI, [25], this approach can be plugged directly in the TCAD, with no need for an external solver of the Schrödinger equation.

The effect of quantization on the energy dispersion inside the gap of III–V materials has been also analyzed using the $\mathbf{k} \cdot \mathbf{p}$ method [29]. A simple approach to modify the parameters of Kane's formula has been proposed, that requires only the knowledge of the effective gap (distance between E_{e0} and E_{h0}), because the effective mass m_r has an almost linear dependence on the effective gap. As a result the prefactor in Kane's formula scales as $(E_{e0} - E_{h0})^2$, whereas the exponential term remains unchanged. The expression of the prefactor is further modified based on the dimensionality of the carrier gases [29]. Similar expressions for low dimensionality gases have been reported also in [30].

The non-local models implemented in TCAD tools require the identification of a suitable *tunneling path*, defined as the line in real space over which performing, for example, the integral of equation (11) for the case of direct tunneling. In fact, when deriving equation (11) we assumed a purely 1D electric field profile, but in a realistic device simulation the electric field may follow quite complex patterns, so that the integration direction must be carefully defined and selected. The effect of different choices for the tunneling path has been analyzed in [26], going from simple *horizontal paths* to a more complicated *dynamic path* [25], where different directions of the path are identified in the different points of the discretization mesh based on the gradient of the valence band energy. This algorithm leads to more precise estimates of the BTBT current in complex device architectures, but has the disadvantage of not being available in AC simulations [25]. Another algorithm making use of Newton's law in the forbidden gap to estimate the tunneling path has been proposed

in [48], essentially similar to the Monte-Carlo procedure in [27].

Overall, the models implemented in commercial TCADs allow for an investigation of the design space for TFETs with a numerical efficiency that is not attainable with full-quantum tools (not even with the non-self-consistent models described in the previous section). With TCAD tools it is thus possible to analyze devices with a relatively large size, as most of the actual TFETs fabricated so far. Furthermore, models for trap-assisted-tunneling (TAT) are also available [25]. TAT has a strong effect on the subthreshold characteristic of many fabricated TFETs. Non-local models have been employed to analyze the subthreshold behavior of fabricated silicon [44, 49], as well as III–V TFETs [50, 51].

In some circumstances, however, the numerical efficiency of the TCAD tools comes at the cost of a limited accuracy and predictive capability. In fact, an inherently quantum effect is approximated by point-like generation of electrons and holes and the generation rate comes from approximated WKB integrals. Besides the approximations in the modeling of BTBT, the transport in TCAD tools is often described by using a drift-diffusion approach, whose results in terms of internal physical quantities are questionable in nanoscale transistors. For example, in nanoscale MOSFETs described with drift-diffusion, the carriers essentially move at the saturation velocity, whereas quasi-ballistic transport is expected to take place in nanoscale TFETs. Transport of the generated carriers affect the electrostatic of the device, that may influence the electric field in the BTBT region and thus the overall tunneling generation rate. In [26] a non-local BTBT model, similar to the models available in commercial TCAD simulators, has been implemented in a Monte-Carlo simulator. Comparison between different self-consistent schemes allowed authors to analyze the impact on the transistor electrostatics and current of the transport of generated carriers. In silicon TFETs this effect was found to be very limited, due to the low BTBT rates and low concentrations of generated carriers. However, in III–V hetero-junction TFETs where the BTBT rate is much higher (see section 4) this may not remain valid.

As a final remark, we notice that the equations for non-local BTBT (such as equation (11)) apply to the transitions from one specific branch of the valence band to one minimum of the CB. Consequently, in principle one should evaluate the BTBT rates considering all the possible combinations of valence and CBs, and each pair of bands should have specific values of the tunneling parameters entering the BTBT rate equations. We will return to this point in section 3, where we will analyze BTBT in strained silicon TFETs.

2.5. NEGF quantum models based on $\mathbf{k} \cdot \mathbf{p}$ or tight-binding (TB) Hamiltonians

The NEGF formalism is considered in many aspects the most general and rigorous method to simulate quantum transport in electronic devices and it is thus particularly suited to study TFETs. The NEGF method solves the Schrödinger equation with open boundary conditions, that is electron transport
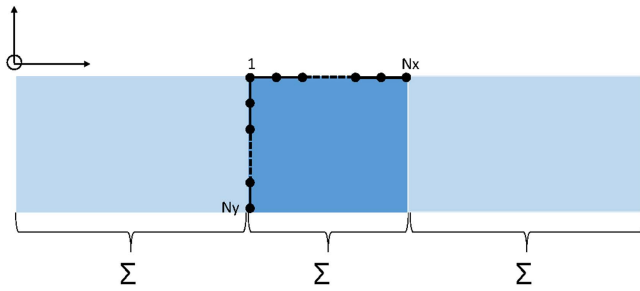
Figure 6. Schematic of the real-space representation of an electron device assumed for an NEGF based simulation approach, where the *x*-axis indicates the transport direction and the device cross-section is in the *y*–*z* plane. A possible discretization scheme is illustrated according on an uniform mesh with $N_x \times N_y$ nodes. The effect of the two semi-infinite leads on the device operation is described by means of the self-energies Σ_S and Σ_D. Phonon scattering in the device region is instead described by the self-energy Σ_{PH}.

through a quantum system connected to external electronic reservoirs and it can also consider the impact of different sources of scattering. Several reviews and textbooks describe the details of the theory [52–54], hence here we will only recall basic equations and their use in the simulation of TFETs.

2.5.1. Coherent transport regime. When dealing with quantum transport problems, the most natural choice is to express the Green's functions in the real-space representation [55]. In particular, either by using a local atomic basis (as in tight-biding method) [56–59], or by discretizing with finite difference (or other discretization methods) an EMA or a **k·p** Hamiltonian [60–62], all operators take the form of matrices, and the retarded Green's function matrix G is obtained by solving the problem

$$(EI - H - \Sigma_S - \Sigma_D)G = I, \qquad (21)$$

where I is the identity matrix, H is the Hamiltonian matrix describing the device and Σ_S, Σ_D are the so-called retarded self-energy matrices accounting for the injection/absorption of carriers from external leads, as illustrated in figure 6. The determination of the self-energies requires the knowledge of the Hamiltonian terms describing the coupling between the device and the contacts (or leads), as well as the Green's function of the leads themselves, which are described as semi-infinite periodic systems. The Green's function of the leads can be numerically evaluated with various approaches [63], among which we would like to mention the recursive Sancho–Rubio [64], and the eigenvalue method [65, 66]. From the knowledge of the Green's function and the contact self-energies, macroscopic quantities like DOS, carrier concentration and source-to-drain current I_{DS} can be readily evaluated. In the absence of inelastic interactions, the current is expressed by means of the Landauer–Büttiker formula [67, 68]

$$I_{DS} = \frac{e}{\pi \hbar} \int dE \; T(E)[f_S(E) - f_D(E)], \qquad (22)$$

where $f_{S,D}$ is the Fermi–Dirac function at source (S) and drain (D) and the transmission probability $T(E)$ is evaluated from

the retarded Green's function as

$$T(E) = \mathrm{Tr}\,[\Gamma_S(E)\, G(E)\, \Gamma_D(E) G^\dagger(E)], \qquad (23)$$

where $\mathrm{Tr}\,[\cdots]$ is the trace operation, and $\Gamma_{S,D} = i[\Sigma_{S,D} - \Sigma_{S,D}^\dagger]$ is the broadening function of the source and drain leads.

Equations (22) and (23) assume a coherent propagation of the electron wave function inside the device, hence they do not describe the phase-breaking interaction with phonons, that may be particularly important for indirect bandgap semiconductors or for PAT processes mediated by defects or interface states. Such a coherent quantum transport formalism has been used for instance to simulate the electric performance of carbon nanotube (CNT) [69, 70], graphene based TFETs [71] and InAs nanowire TFETs [57, 62]. Moreover, this approach can account for some sources of elastic scattering, such as surface roughness [72, 73], because these interactions do not break quantum-phase coherence and their effect can be included in the real-space electrostatic potential. We notice, however, that, in case a variability analysis is performed, this approach requires the generation of several different samples in order to accumulate a significant device statistics.

2.5.2. Possible forms for the matrix H. The form of the Hamiltonian matrix in equation (21) depends on how many spatial dimensions are actually included in the calculations (i.e. 1D, 2D or 3D simulations), and on the model employed to describe the electronic bandstructure. For example, in the simplest case of the EMA, the finite-difference discretization of the continuous operator gives rise to a tridiagonal matrix for a 1D system, a penta-diagonal matrix for a 2D system and epta-diagonal matrix for a 3D system, with a matrix rank equal to the number N of the nodes in the real-space mesh. It is very important to note that for 2D and 3D systems the full Hamiltonian can be written as a tri-diagonal block matrix with rank $N_x N_s$, where N_x is the number of cross-sections in the transport direction x and N_s is the number of real-space discretization points in each device cross-section, that is $N_s = N_y$ or $N_s = N_y N_z$ respectively for a 2D or a 3D device, as it is sketched in figure 6. Indeed, the tri-diagonal block structure of the Hamiltonian matrix allows us to calculate some blocks of the Green's function with the recursive algorithms to be mentioned in section 2.5.3. Simplified bandstructures based on the EMA for the conduction and the valence band, then linked by a phenomenological coupling term, have been used to simulate within the NEGF framework TFETs based on CNT [70], graphene [74] and 2D monolayer materials [75]. When dealing with Hamiltonians based on the **k·p** approximation, which naturally account for a realistic energy dispersion in the energy gap, the real-space Hamiltonian matrix can be obtained by following the standard quantum mechanics prescription for the wave-vector $\mathbf{k} \rightarrow (-i\nabla_\mathbf{r})$ and then discretizing the operator using, for instance, the finite-difference method. It turns out that a 3D device described with an N_b-band **k·p** results in an Hamiltonian matrix which also has a tri-diagonal block structure, with sub-matrices having rank of $N_b N_y N_z$. A tri-diagonal block

structure can be identified also in the H matrix corresponding to atomistic Hamiltonians based on the semi-empirical TB approach. In this case the rank of each sub-matrix is $N_s = N_a N_o$, where N_a is the number of atoms in a cross-section normal to the transport direction and N_o is the number of electronic orbitals used in the TB formulation.

2.5.3. Reduction of the computational burden.
As previously mentioned, in both **k·p** and TB models, the size of the matrices in equation (21) can rapidly increase with the lateral size of the system, thus making computationally prohibitive the determination of G via a direct inversion of the matrix $[EI - H - \Sigma_S - \Sigma_D]$. Fortunately, only a few elements of G are needed to obtain relevant physical quantities such as charge and current, so that it is possible to exploit recursive algorithms based on the Dyson equation to compute only specific elements of G [52], typically the main diagonal and the first sub-diagonal elements, which permits a dramatic reduction of the computational burden [52, 76]. These methods are called *recursive Green's function algorithms* and they have been widely and systematically used to simulate TFETs [60, 77].

A further reduction of the computational burden can be achieved by adopting the so-called coupled mode-space (CMS) representation, which consists in transforming the real-space Hamiltonian into a new basis composed of transverse modes defined at different positions along the transport direction. Such a basis is theoretically equivalent to the real-space basis, but has the advantage to give accurate, approximate results with a reduced set of basis functions chosen according to the energy range relevant for the problem at study. Therefore the CMS approach can be very convenient when dealing with TFETs, which exploit the energy-filtering mechanism and have a current spectrum peaked in a fairly narrow energy window. The CMS approach has been used to simulate TFETs based on CNT [56], InAs nanowire [60, 61], GaSb/InAs hetero-junction [78] and van der Waals hetero-junction TFETs [75].

2.5.4. Alternative approaches for coherent transport.
When the analysis is restricted to coherent transport, alternative methods to the NEGF exist and can be computationally advantageous. An important approach consists in solving the Schrödinger equation with open boundary conditions according to the quantum transmitting boundary (QTB) method [79]. In the QTB method the wave function in the leads is expressed as a linear combination of propagating and evanescent modes, whose complex coefficients are related to transmission and reflection amplitudes. The coefficients of such a linear combination have to be determined by imposing continuity conditions for the wave-function and its flux at the interface between the device and the leads. With such a procedure it is possible to specify the incoming and outgoing flux of carriers from the leads and solve the Schrödinger equation only in the device region. The method is equivalent to the scattering matrix method to compute transmission and reflection coefficients and, in the absence of inelastic

scattering, to the Green's function method [80]. Several examples of TFET simulation studies have been reported in the literature that use the QTB approach, and for different formulations of the Hamiltonian [81–83].

2.6. General considerations about phonon scattering in the NEGF formalism

The advantage of the NEGF method with respect to the QTB method is primarily in the possibility to include the effect of inelastic scattering mechanisms, such as the electron–phonon interaction [84]. This is obtained by introducing the concept of lesser(greater)-than Green's functions $G^{<(>)}$, which are associated to the electron(hole) statistics, whereas the retarded Green's fucntion G provides the information about the available electronic states and the carrier dynamics [54]. More precisely, in the steady-state regime the main diagonal of the lesser-than GF matrix (that is the $G^<(\mathbf{r}_i, \mathbf{r}_i)$ in a real-space representation), is used to compute the electron concentration $n(\mathbf{r}_i)$, the main diagonal of the greater-than GF matrix, $G^>(\mathbf{r}_i, \mathbf{r}_i)$, gives the hole concentration $p(\mathbf{r}_i)$ and the main diagonal of the retarded GF matrix, $G(\mathbf{r}_i, \mathbf{r}_i)$, is linked to the local density of states (LDOS) $\rho(\mathbf{r}_i)$.

The lesser(greater)-than self-energy describing the phonon scattering is evaluated as $\Sigma_{PH}^{<(>)} = D^{<(>)} G^{<(>)}$ [84], where $D^{<(>)}$ is the phonon lesser(greater)-than Green's function. The problem is here most often simplified by assuming that the phonon bath is at equilibrium, so that $D^{<(>)}$ can be written in terms of the Bose–Einstein distribution function $n_B(\omega) = [\exp(\hbar\omega/k_B T) - 1]^{-1}$, where $\hbar\omega$ is the phonon energy. In this case the lesser-than and the greater-than self-energies for the jth phonon branch read [84]

$$\Sigma_{PH,j}^<(E) = |M_j|^2 \{ G^<(E - \hbar\omega_j) n_B(\omega_j) \\ + G^<(E + \hbar\omega_j)[n_B(\omega_j) + 1] \}, \quad (24a)$$

$$\Sigma_{PH,j}^>(E) = |M_j|^2 \{ G^>(E + \hbar\omega_j) n_B(\omega_j) \\ + G^>(E - \hbar\omega_j)[n_B(\omega_j) + 1] \}, \quad (24b)$$

where M_j is the matrix element expressing the microscopic details of the electron–phonon interaction [85]. Equation (24) describe a dissipative phenomenon and in fact the self-energy at energy E depends on the Green's function at a different energy $E \pm \hbar\omega_j$, where the first term in equations (24) represents a phonon-absorption and the second term a phonon-emission process. A very similar formalism can be used also to consider approximately elastic phonons, such as intra-valley acoustic phonons in the long-wavelength approximation.

2.6.1. Self-consistent Born approximation.
Since the retarded phonon self-energy Σ_{PH} has to be expressed in terms of the greater-than and lesser-than self-energies according to the relation $\Sigma_{PH} - \Sigma_{PH}^\dagger = \Sigma_{PH}^> - \Sigma_{PH}^<$, the retarded and the lesser-than Green's function are non-linearly coupled and have to be calculated self-consistently. This is obtained by

solving self-consistently the kinetic equations

$$[EI - H - \Sigma_S - \Sigma_D - \Sigma_{PH}] G = I, \qquad (25)$$

$$[EI - H - \Sigma_S - \Sigma_D - \Sigma_{PH}] G^< = [\Sigma_S^< + \Sigma_D^< + \Sigma_{PH}^<]G^\dagger \qquad (26)$$

until convergence is reached. This method to include phonon scattering within the NEGF formalism is called self-consistent Born approximation, and it has the remarkable merit that it guarantees current conservation along the device also in the presence of dissipative phenomena [54]. In case of dissipative transport the current can no longer be expressed in terms of the retarded Green's function alone, and the lesser-than function is also necessary [84]. For example, the steady-state current at the lead $L = $ S, D may be written as

$$I_L = \frac{e}{\pi \hbar} \int dE \, \text{Tr} \left[\Sigma_L^<(G - G^\dagger) + i\Gamma_L G^<\right], \qquad (27)$$

where the broadening function is given by $\Gamma_L = i[\Sigma_L - (\Sigma_L)^\dagger]$.

The impact of phonon scattering on the transfer characteristics of TFETs within the NEGF formalism has been discussed with more details in [58, 86, 87].

2.6.2. Main difficulties and approximations. In principle the self-consistent Born approximation gives an excellent description of the electron–phonon interaction, as stated by the Migdal theorem [88], but in practice further simplifications are very often necessary to perform the calculations. The main difficulty in the iterative solution of equations (25) and (26) is the computational burden, because $\Sigma_{PH}, \Sigma_{PH}^<, \Sigma_{PH}^>$ are dense matrices, so that all the entries of $G, G^<, G^>$ are necessary in the calculations. This makes it practically impossible to resort to recursive schemes to calculate only specific elements of the Green's functions, which has a dramatic impact on the memory and CPU time requirements [76]. Because of this computational problem only the diagonal elements of the self-energies in equations (24) are most often retained in practical calculations. This implies that simulations usually account only for local interactions, but neglect all spatial coherence terms of the electron–phonon interaction. The simplified, local formulation of the phonon self-energies is an acceptable approximation when dealing with acoustic phonons since their self-energy can be considered as independent of the phonon wave-vector, but it becomes more delicate and questionable for optical phonons. The local approximation is not fully justified, in particular, when dealing with polar optical phonons, which are a dominant phonon scattering mode in III–V compounds. In fact polar optical phonon scattering is described by a squared matrix element $|M(\mathbf{q})|$ proportional to $|\mathbf{q}|^{-2}$, where $\mathbf{q}$ is the phonon wave-vector [85], hence it is an inherently non-local physical mechanism.

Indeed, the description of the electron phonon interaction beyond the local approximation is arguably one of the most challenging problems for the dissipative quantum transport simulations of TFETs based on the NEGF formalism.
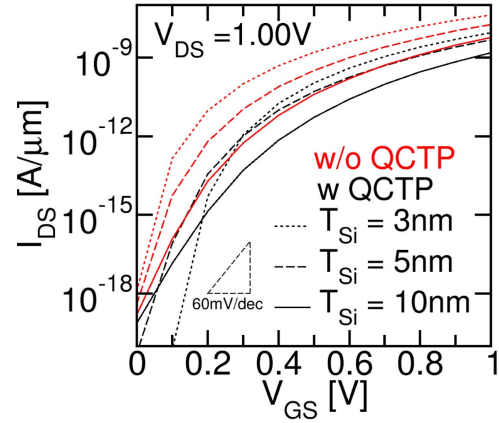

Figure 7. Simulated drain current versus gate voltage characteristics for double gate silicon TFETs with different film thickness. The equivalent oxide thickness is 0.7, 0.9 and 1.1 nm for the silicon film thickness $T_{Si} = 3$, 5 and 10 nm, respectively. More details about this simulation study may be found in [30].

3. Silicon and group IV semiconductor tunnel FETs

In this section we report selected modeling results for Si and Ge TFETs, providing also some considerations about the use of strained silicon. Specific results for group IV alloys such as SiGe and GeSn, are not covered by this review paper. Results for these alloys using TCAD-like models can be found for example in [89, 90], while a full-quantum model has been used in [91], neglecting phonon-assisted tunneling because GeSn becomes direct bandgap for suitable Sn concentrations and strain conditions [92].

Si and Ge are indirect gap semiconductors, so that modeling of phonon-assisted BTBT is needed which, as seen in section 2.6, is very computationally demanding in full-quantum simulations [58]. For this reason, the vast majority of modeling results, in particular for Si, have been obtained with TCAD models or with *non-self-consistent* quantum models, as those described in section 2.3. It is worth noting that in Si nanowires with a small cross-section the direct tunneling dominates over phonon-assisted tunneling [58].

3.1. Silicon TFETs

Due to the maturity of the silicon CMOS technology, at first the implementation of TFETs in Si has been investigated as a quite natural option. In fact most of the fabrication steps are shared with the ones used for conventional CMOS transistors and this allows for MOSFET-TFET co-integration [93].

A good electrostatic control is mandatory in order to enhance the electric field at the source-channel junction, increase the BTBT rate and obtain steep sub-threshold I_{DS} versus V_{GS} characteristic. For this reason TCAD simulations have investigated TFETs with thin SOI films and a double-gate biasing, as well as GAA NWs TFETs [94]. The use of thin silicon films on one side improves the electrostatic control (resulting in a shorter tunneling path), but on the other hand increases the effective gap (see section 2.4). To discuss this trade-off, we plot in figure 7 the trans-characteristic of double-gate Si TFETs as obtained with the multi-subband

Monte Carlo described in [30]. The simulator implements a non-local model for BTBT similar to the one in TCAD [25], but with a quantum-corrected tunneling path (QCTP) where the effective gap is modified in each point of the real space according to the solution of the 1D Schrödinger equation (see discussion in section 2.4).

We see in figure 7 that the effect of the QCTP is stronger at low V_{GS} and for thin films. However, the improved electrostatic control obtained with thin films is not offset by the increase of the effective gap, so that for $T_{Si} = 3\,\mathrm{nm}$ we observe the best SS and on-current: the minimum point SS is around $20\,\mathrm{mV/dec}$, however the on-current is only a few $\mathrm{nA}\,\mu\mathrm{m}^{-1}$. The low on-current of Si-TFETs found in simulations is consistent with available experimental data [9, 12, 93, 95]. In fact some experiments have reported currents approaching the $\mu\mathrm{A}\,\mu\mathrm{m}^{-1}$, but for very high biases. If extrapolated to supply voltages around $0.5\,\mathrm{V}$, the experimental on-current is still in the few $\mathrm{nA}\,\mu\mathrm{m}^{-1}$ range.

It must be noted that device structures as the ones analyzed in figure 7 (and in several experimental studies [93, 95]) are dominated by *point-tunneling*, that is not very efficient since the tunneling direction is normal to the vertical electric field induced and modulated by the gate. Many structures based instead on *line-tunneling* have been proposed recently [49, 96, 97], where the tunneling direction is aligned with the electric field induced by the gate. One example is the EHBTFET that we have discussed in section 2.2, while a second example is the core–shell nanowire proposed in [98]. However, also these architectures are expected to result in poor on-current if implemented in silicon. For the Si EHBTFET this has been shown both by experiments [99] and by simulations [100]. The simulations for the core–shell nanowire indicate an on-current of about $30\,\mathrm{nA}\,\mu\mathrm{m}^{-1}$ if one sets I_{off} to $1\,\mathrm{pA}\,\mu\mathrm{m}^{-1}$ in figure 2 of [98], and then projects the on-current to a supply voltage below $0.5\,\mathrm{V}$. Also the experimental data for line tunneling in [49] results in an on-current of a few tens $\mathrm{nA}\,\mu\mathrm{m}^{-1}$ if it is projected to a $0.5\,\mathrm{V}$ power supply. Core–shell TFETs using III–V materials, instead, look promising and relatively large on currents have been experimentally reported [101].

3.2. BTBT in strained silicon

Strain modifies the band structure and suitable strain configurations may enhance the BTBT rate. Indeed, experimental data for strained Si and SiGe nanowires have shown promising on-currents [102–105]. The effect of strain on the current of TFETs has been analyzed experimentally in [106], while TCAD simulations of Si TFETs with non-uniform strain profiles have been reported in [107].

In this section we illustrate a worked out example of how to use a commercial TCAD environment [25] to investigate possible strain induced BTBT enhancements in silicon TFETs. The BTBT rate for phonon assisted transitions is given by [25]:

$$G_{\mathrm{T}} = |\nabla E_v(0)| \cdot C_p \exp\left(-2\int_0^l \kappa_{\mathrm{im}}\,\mathrm{d}x\right)[f_v - f_c]$$

$$C_p = \frac{C_{\mathrm{path}}}{2^6\pi^2 E_g}\sqrt{\frac{m_c m_v}{h\sqrt{2m_r E_g}}}\left(\int_0^l \frac{\mathrm{d}x}{\kappa_{\mathrm{im}}}\right)^{-1}$$

$$\times\left[1 - \exp\left(-k_{v,\mathrm{max}}^2 \int_0^{l_c}\frac{\mathrm{d}x}{\kappa_v}\right)\right]$$

$$\times\left[1 - \exp\left(-k_{c,\mathrm{max}}^2 \int_{l_c}^l\frac{\mathrm{d}x}{\kappa_c}\right)\right]$$

$$\kappa_{\mathrm{im}} = \min(\kappa_c, \kappa_v), \quad \kappa_c = \frac{1}{\hbar}\sqrt{2m_c(E_C - E)},$$

$$\kappa_v = \frac{1}{\hbar}\sqrt{2m_v(E - E_V)}, \tag{28}$$

where 0 and l are the coordinate of the beginning and the end of the tunneling path, while l_c is the point where $\kappa_c = \kappa_v$. The terms $k_{v,\mathrm{max}}$ and $k_{c,\mathrm{max}}$ have the same meaning as the term k_m in equation (11) and C_{path} contains the phonon parameters (see discussion below). The reduced mass is $m_r = 2m_c m_v/(m_c + m_v)$. We see that the modeling *ingredients* for the BTBT rate are the effective masses for the valence and CBs, the energy gap and the phonon energy and deformation potential.

Strain changes both the energy gap, inducing splitting between the different valleys of the conduction and valence bands, and modifies the effective masses. With the tool *sband* included in [25] we computed the effective masses inside the band (real branch) and in the gap (imaginary branch). A good correspondence is found between (real) mass values in literature and *sband* for both unstrained and strained silicon. For the CB, the masses extracted from the imaginary branches of the energy relation in the gap are almost the same as the masses corresponding to the real energy branch. For the valence band, instead, the masses extracted in the energy gap are remarkably different compared to the values corresponding to real energy branches. More precisely, figure 8 shows that the effective mass of the imaginary branch corresponding to the HH valley has a small mass, similar to the one of the real branch of the LH valley, whereas the imaginary branch of the LH valley has a mass similar to the real branch of the HH valley. These imaginary valence bands can still be reproduced, in first approximation and for small energies, with a parabolic expression, as can be seen in figure 8. The effect of biaxial strain on the valence band parameters is shown in figure 9.

Equation (28) refers to a single type of tunneling transition, for example between LH or HH and one of the minima of the CB. To simulate more realistically the BTBT process in silicon, we used six different tunneling paths, from HH and LH to the three Δ valleys of the CB. All these paths produce different contributions to the generation rate, particularly when strain breaks the symmetry between the CB valleys. For each valley combination we used masses and energy gaps resulting from the *sband* analysis. In particular, for each path we used the effective masses (m_c and m_v) obtained from
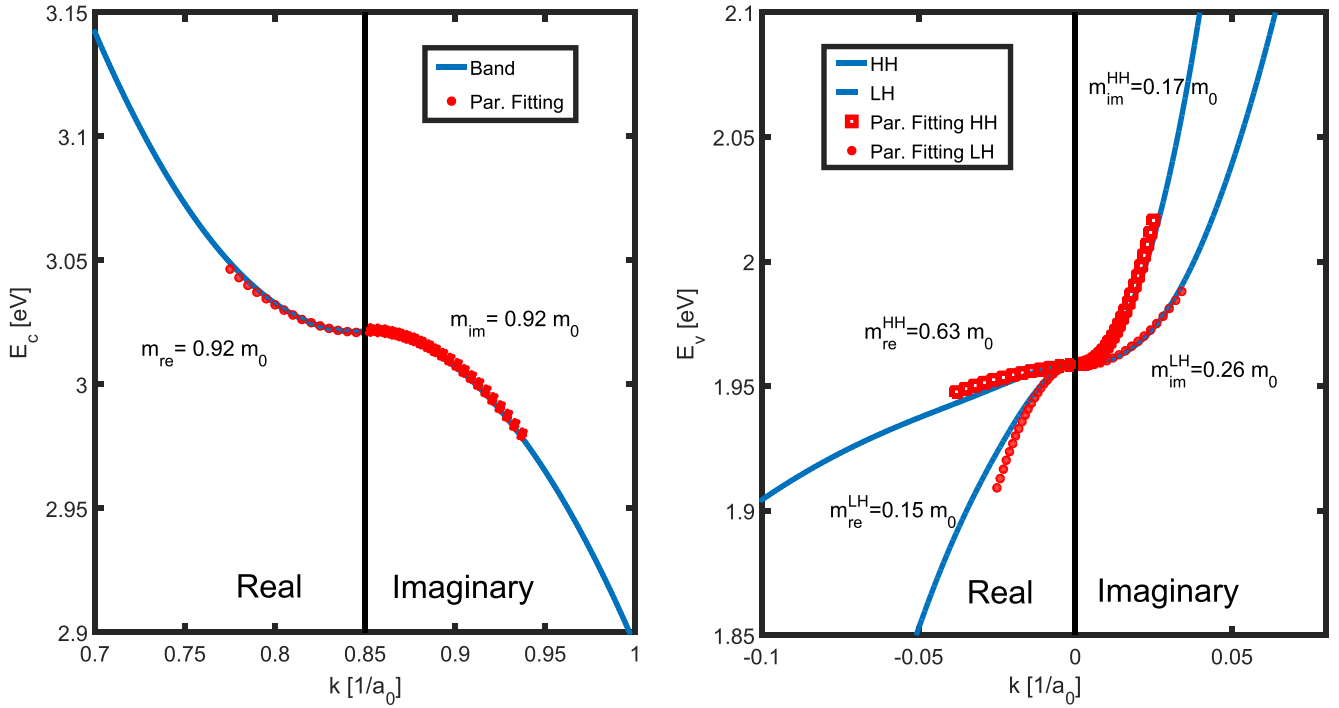
Figure 8. Real and imaginary energy dispersion for the Δ minimum of the conduction band along [100] (left) and for the valence band along [110](right) of unstrained silicon. A fitting with a parabolic energy relation is also shown (red circles).
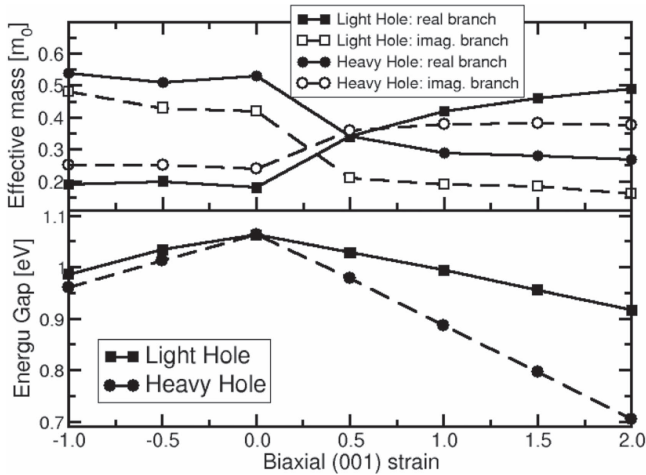

Figure 9. Hole effective masses and energy difference between the lowest CB and the LH and HH valleys in biaxially strained silicon in the (100) plane.

sband along the corresponding tunneling direction and corresponding to the imaginary energy relation in the gap. The phonon energy is assumed to be the same as in unstrained Si. The problem with equation (28) is that it does not take into account the fact that the CB minima are anisotropic. In fact, the parameter m_c in equation (28) is used to compute the transmission probability (and in this respect it should be the effective mass inside the gap), but also as a prefactor that accounts for the DOS involved in tunneling. To solve this issue, we set as m_c and m_v the masses inside the energy gap (that is for imaginary branches of the energy relation) from *sband*, and modified the C_{path} parameters by the ratio between

the DOS masses and the imaginary masses:

$$C_{\text{path}} = C_{\text{path0}} \left(\frac{m_{\text{dos},c} m_{\text{dos},v}}{m_c m_v} \right)^{3/2}. \tag{29}$$

Equation (29) implicitly assumes that the phonon deformation potential does not change with strain. The power of $3/2$ stems from the fact that, under constant lateral electric field, equation (28) reduces to Keldysh formula with prefactor proportional to $(m_c m_v)^{3/2} m_r^{-5/4}$. While the $m_r^{-5/4}$ is related to the dispersion relationship in the gap, the term $(m_c m_v)^{3/2}$ instead results from the DOS. The correction in equation (29) is thus exact under uniform lateral field and approximated in general cases. The value of C_{path0} has been calibrated to reproduce the experimental Si diodes in [9, 108], once the effective masses have been extrated from *sband* calculations.

To assess the impact of strain on BTBT, we have considered uniform structures with constant lateral electric field and different doping levels. Results are reported in figure 10, where we see that, after proper calibration of the C_{path0}, the approach consisting of six tunneling paths reproduces the experimental data with better accuracy compared to the default SDevice calibration. Biaxial strain of 2% is shown to enhance the BTBT rate by more than an order of magnitude.

The contributions of the different tunneling paths are reported in table 1 for one particular value of F_{MAX}. For unstrained silicon the tunneling is dominated by the transitions from HH (low tunneling mass) to Δ_y and Δ_z (low tunneling mass along the electric field direction x). In the presence of biaxial tensile stress, the energy of the Δ_z valley is reduced, and transitions between HH and Δ_z are significantly enhanced. The change in the gap and effective masses also enhances the LH to Δ_z transitions.
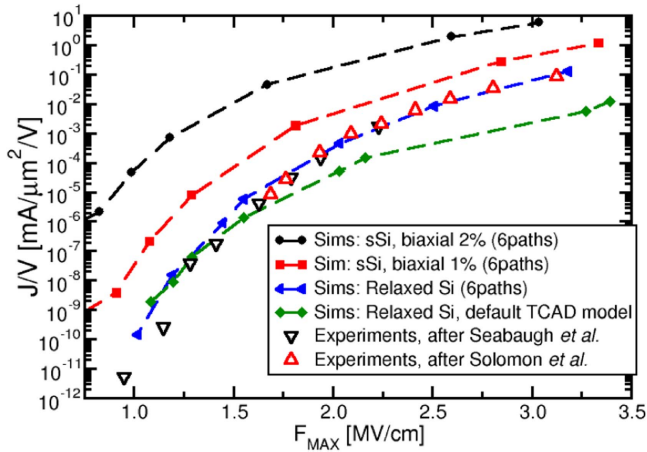
Figure 10. Generation rate versus electric field in uniform slabs of biaxially strained silicon. The figure also shows the calibration of the model for unstrained silicon against the experiments in [9, 108].

3.3. Ge-based TFETs

Germanium has been deeply investigated as channel material for TFETs due to the lower band-gap and smaller effective masses compared to Si. In Ge the direct gap is only 0.14 eV larger than the indirect gap between the top of the valence band and the CB minimum at the L point. For this reason direct and phonon-assisted BTBT may both contribute to the overall tunneling current. In this respect, figure 11 shows that the experimental data for the tunneling currents in the Ge diodes collected in [109] can be reproduced quite well by the model for direct BTBT, whereas the contribution of phonon-assisted transitions is one order of magnitude lower.

The default calibration in TCAD tools is quite empirical and assumes that phonon-assisted tunneling is dominant in Ge [25]. This point is thoroughly discussed in [110], which proposes a more realistic calibration where Ge is instead dominated by direct tunneling. It can be verified that in uniform structures, the direct BTBT rate provided by the model in [110] is very close to the phonon-assisted BTBT rate obtained with the default calibration in TCAD, whereas the phonon assisted model from [110] provides a much lower BTBT rate.

To assess the effect of different calibrations on realistic devices, we simulated with TCAD the planar device with Ge source and Si channel fabricated in [111]. In figure 12 the experimental results of [111] are reported by black solid line and look promising, with an I_{ON} close to 1 μA μm^{-1} and a minimum SS significantly below 60 mV/dec at low bias. Simulations using the default SDevice calibration (dashed lines) slightly underestimate the current but are, somewhat surprisingly, in quite good agreement with the experiments. The dotted curves with symbols in figure 12 have been obtained with the calibration proposed in [110]: as can be seen they are significantly lower than the experiments and predict a poor device performance. In particular, as discussed above, the calibration in [110] considers a much smaller phonon assisted BTBT (circles) than the one assumed as default in TCAD. Direct BTBT (triangles) is high, but its

onset is at large gate voltage due to the Ge–Si hetero-junction that essentially inhibits direct tunneling from the Ge valence band to the Γ point in the CB of Si. In fact, direct BTBT becomes significant only when the gate voltage is large enough to allow for an essentially vertical tunneling within the Ge source. On the other hand, phonon assisted BTBT takes place between the Ge source and the Si channel.

The default calibration in SDevice and the calibration in [110] for BTBT in Ge result in orders of magnitude difference in on-current and also large differences in SS. Notice that the phonon model in [110] has been recently revised in [112]. The new set of parameters (that we used in figure 11) features a significantly higher phonon deformation potential (7.8×10^8 eV cm^{-1} instead of 0.8×10^8 eV cm^{-1} in [110]) and slightly different phonon energy (6 meV versus 8.6 meV). It is easy to show that these modifications result in a BTBT rate 200 times higher than using the set proposed in [110]. If we multiply the dotted curve with symbol in figure 12 (obtained with the set in [110]) by a factor of 200, the simulations approach the experiments (not shown), but are still more than one decade below them.

From the analysis above, we can conclude that there are still open issues in the calibration of TCAD BTBT models for Ge TFETs. In fact, the uniform structures used for model calibration are dominated by direct BTBT, whereas in realistic devices phonon-assisted tunneling dominates, and different sets of parameters have thus been proposed in the literature. As a further example of interplay between direct and phonon-assisted BTBT, we consider in figure 13 a Ge EHBTFET. We have employed the non self-consistent quantum model for direct and phonon-assisted BTBT described in sections 2.2 and 2.3 respectively. Due to quantization, the offset between the direct and the indirect gap becomes larger than in bulk Ge. In particular, we can see that the first bump in the current takes place when subband e1, originating from the L-valleys, gets aligned with the heavy-hole subband hh1. A second bump is due to alignment between e1 and hh2. We see direct tunneling only at much higher gate voltages, such that subband Γ1 aligns with hh1, but in the same gate voltage range we also have another bump due to phonon-assisted transitions between e2 and hh1.

The effect of strain in Ge TFETs has been analyzed in [113] by using quantum-corrected TCAD simulations to estimate the current and $\mathbf{k} \cdot \mathbf{p}$ calculations for the band parameters. It was found that tensile strain significantly enhances the BTBT rate, similarly to the results for strained Si discussed in section 3.2.

4. Tunnel FETs employing III–V semiconductors

III–V compounds, in particular arsenides and antimonides, have bandstructure properties that make them particularly suited to be employed in TFETs. They have direct band gap centered around the Γ point of the Brillouin zone with small band-gap energies and electron effective masses smaller than Si and Ge [114]. This means that BTBT in III–V semiconductors does not need to be phonon assisted and can

Table 1. Contribution of different tunneling paths to the currents in figure 10.

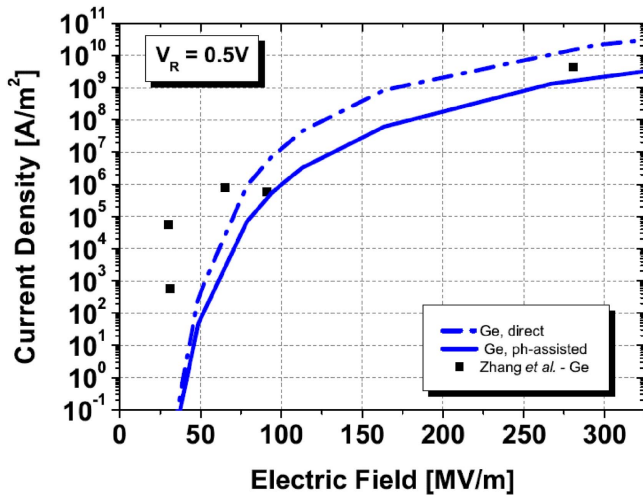| Paths	J_{unstr}/V @$F_{\text{MAX}} = 1.45$ MV cm^{-1} [mA μm^{-2} V^{-1}]	$J_{\text{tens},1\%(001)}/V$ @$F_{\text{MAX}} = 1.29$ MV cm^{-1} [mA μm^{-2} V^{-1}]
LH $\to \Delta_x$	5.5×10^{-13}	5.4×10^{-11}
HH $\to \Delta_x$	5.9×10^{-10}	5.4×10^{-11}
LH $\to \Delta_y$	9.3×10^{-9}	9.0×10^{-10}
HH $\to \Delta_y$	4.5×10^{-7}	3.9×10^{-9}
LH $\to \Delta_z$	9.3×10^{-9}	1.1×10^{-6}
HH $\to \Delta_z$	4.5×10^{-7}	2.9×10^{-6}

Figure 11. BTBT current in Ge diodes. The experimental data in [109] are compared with results of direct and phonon-assisted BTBT simulations. The model parameters for the direct BTBT model are the direct gap (0.8 eV) and the effective masses for electron and holes ($m_e = 0.043\, m_0$, $m_{hh} = 0.33\, m_0$, $m_{lh} = 0.043\, m_0$), whereas the phonon parameters for phonon-assisted BTBT are taken from [112]. Simulations account for non parabolicity of the Γ conduction band (with $\alpha = 0.85$ eV^{-1}) as well as for the *anti-crossing* of the LH and HH valleys, namely the LH mass is used in the gap for the HH valleys (see the discussion concerning strained silicon in section 3.2).
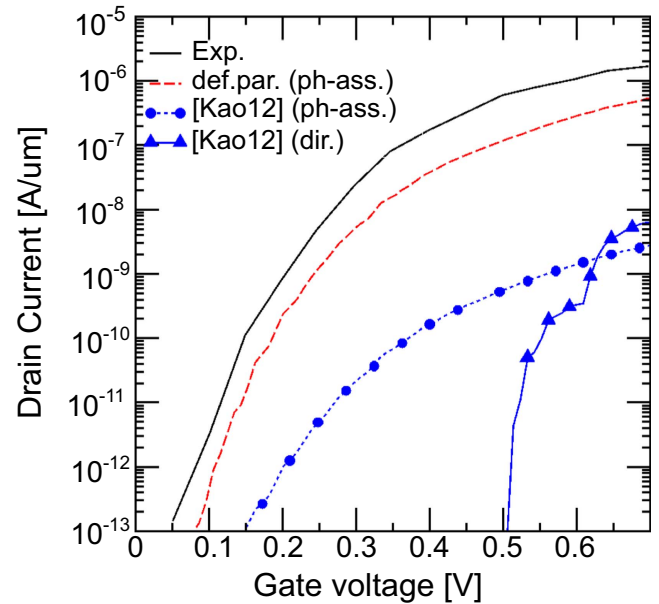

Figure 12. Comparison between TCAD simulations using different calibration strategies, that is either the default calibration of [25] or the calibration proposed in [110], and the experimental data for a TFET with germanium source reported in [111].

follow a direct path from valence to CB. Moreover, thanks to the small effective mass, the electron wave function experiences a deeper penetration in the channel and gives rise to a larger tunneling probability.

The main disadvantage of these materials compared to Si or SiGe is that the technology is not as mature as it is for group IV semiconductors and, for instance, contact resistances and interface defects at the semiconductor-oxide interface are typically worse in III–V based FETs compared to their Si or SiGe counterpart. Several experimental works reported good on-state currents related to the high BTBT [115–118], but also SS higher than the thermionic limit of 60 mV/dec [12], which has been attributed to both inefficient electrostatic control and to the high density of defects [119, 120]. It is a recent and welcome news the report of a vertical InAs/GaAsSb/GaSb TFETs with a subthreshold swing below 60 mV/dec and very competitive on current at a V_{DS} of 0.3 V [121].

As a matter of fact many of the III-V TFETs reported in experiments are relatively large devices and it is difficult to describe them by using full quantum modeling, so that a semi-classical modeling approach has been often used as, for example, the multi-subband Monte Carlo simulations with quantum corrected tunneling path illustrated in figure 14. Likewise, an analysis based on a commercial TCAD has been recently used to interpret the electrical characteristics of large diameter, InAs nanowire TFETs [51].

4.1. Simulation studies based on NEGF simulations

Due to their small energy gap and electron effective mass, InAs based TFETs have been extensively simulated with the NEGF approach by using either an atomistic, TB Hamiltonian or a **k·p** approach.

The TB method was used in [57] to analyze InAs p–i–n single-gate, dual-gate, ultrathin-body and gate-all-around nanowire (GAA NW) devices with a gate length of 20 nm, showing that a reduced subthreshold swing can only be achieved if the electrostatic potential is efficiently controlled
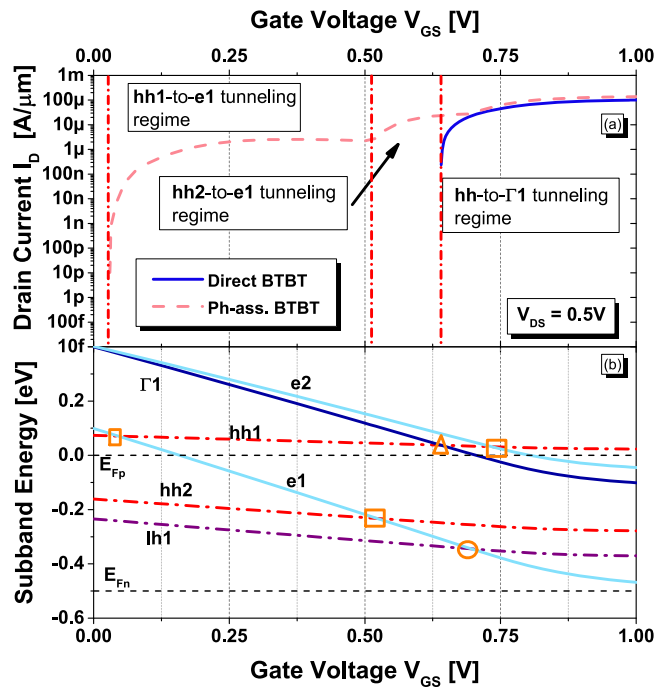
Figure 13. Top: simulated drain current versus gate voltage characteristic for a Ge EHBTFET. Bottom: electron and hole subbands versus gate voltage. The overlap and underlap regions are all 50 nm long. The equivalent oxide thickness is 0.53 nm (i.e. 3 nm of HfO$_2$) and the Ge film thickness is 10 nm.
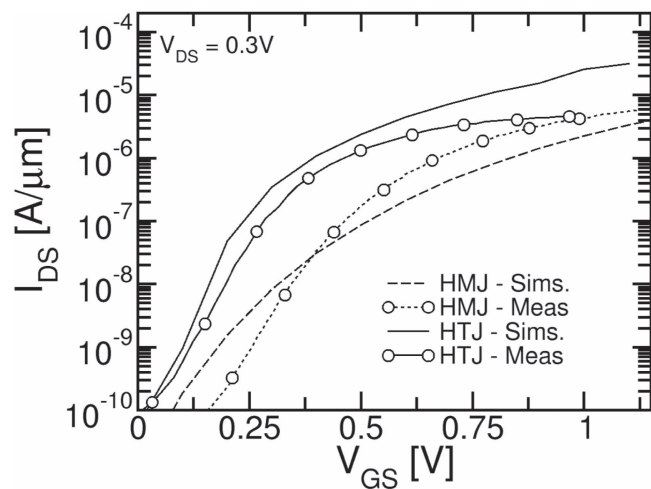

Figure 14. Comparison between MSMC simulations with quantum-corrected-tunneling-path discussed in section 2.4 [30], and the experimental data for III–V TFETs in [115]. HMJ stands for homo-junction TFET; HTJ stands for hetero-junction TFET.

by the gate. Similarly, in [58] it is shown that in direct-gap semiconductors, such as InAs, the BTBT is often dominated by a single branch of the energy relation in the gap, and therefore a fairly good estimate of the tunneling rate can be obtained with the WKB approximation.

In order to reduce the computational effort and simulate larger structures within the NEGF method, the **k·p** approximation has been successfully employed to simulate transport at the Γ point in homojunction InAs TFETs [40, 60, 61, 122, 123], as well as in hetero-junction TFETs

such as GaSb/InAs based transistors [124]. In [124] it was shown that ideal InAs NW TFETs with gate length of 17 nm require a lateral width of 7 nm or less to achieve a SS smaller than 60 mV/dec due to the loss of electrostatic integrity at small gate lengths. In such narrow FETs the energy dispersion is strongly altered by quantum confinement compared to the energy dispersion in the corresponding bulk materials. An example is shown in figure 15 reporting the highest valence subband and the lowest conduction subband profiles along the channel, as well as the current spectra for a p–i–n hetero-junction nanowire FET having GaSb in the source and InAs in the channel and drain regions. As can be seen, for a lateral diameter of $D_W = 7$ nm or smaller this material system is no longer broken-gap.

4.2. Strain modeling and engineering

The requirement for narrow multi-gate transistors necessary to preserve a strong control of the gate terminal on the channel region results in a significant carrier confinement that in turn degrades the on current delivered by TFETs. Consequently, several material and device design options have been investigated to improve the on current and preserve a steep subthreshold swing.

As a prominent example mechanical strain has been proposed, in analogy to its use as mobility booster in MOS-FETs, as an efficient way to improve the on current of InAs TFETs [77]. In [61], the impact of different strain conditions on the bandstructure and I_{DS} versus V_{GS} characteristics of an InAs nanowire TFET was investigated by using NEGF simulations based on an 8-band **k·p** Hamiltonian [125], and with a strain interaction matrix whose deformation potentials were taken from [114]. It was shown that both compressive uniaxial and tensile biaxial stress can be used to modify the imaginary branch of the InAs nanowire and reduce the effective gap between the highest valence and the lowest conduction subband. The tensile biaxial strain in the device cross-section was found to be particularly effective in order to reduce the energy gap and enhance BTBT. The introduction of strain, however, implies also a deterioration of the SS due to the change of the valence subband profile in the source region [124], so that specific design options at the source region may be needed to preserve a steep subthreshold characteristic in the presence of strain [126]. The impact of strain on the bandstructure and the turn-on characteristics of III–V based TFETs is still under investigation at the time of writing [127–129].

4.3. Hetero-junction modeling and engineering

III–V compounds have also the remarkable merit to form hetero-junctions with interesting properties. In particular, the GaSb/InAs hetero-junction is approximately lattice matched and forms a broken bandgap material system, that is a system where the valence band edge of GaSb is higher than the CB edge of InAs [130]. Such a material system has attracted a lot of interest in the quest for large on currents in tunnel FETs [62, 131].
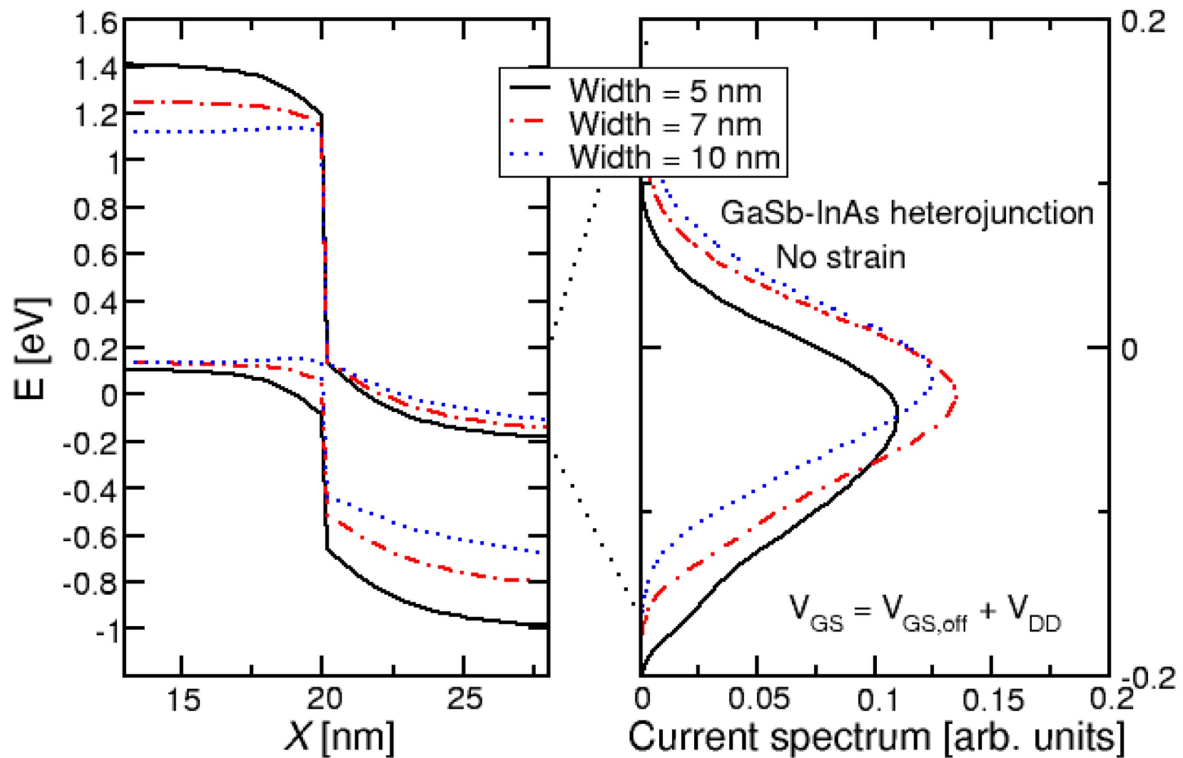
Figure 15. Subband profile along the device length (left), and current spectrum as a function of energy (right) for a GaSb–InAs hetero-junction TFET with a squared cross-section with a side $D_W = 5$, 7 and 10 nm; the source Fermi level $E_{F,S}$ is taken as zero. $\mathbf{V}_{GS} = [V_{GS,off} + V_{DD}]$ and $V_{DS} = 0.3$ V, where $V_{GS,off}$ is the gate voltage corresponding to $I_{OFF} = 5$ nA μm^{-1}. The gate length is $L_G = 17$ nm.

Heterojunction TFETs have been simulated by using the NEGF approach with both a TB [83, 115, 132, 133], and a **k·p** Hamiltonian [62, 124, 126, 131, 134]. The use of the GaSb/InAs hetero-junction usually provides a larger I_{ON} compared to InAs homojunction TFETs, but such a potential improvement is partly frustrated by quantum confinement effects that tend to turn the supposedly broken-bandgap GaSb/InAs hetero-junction into an actually staggered band alignment system, as illustrated in figure 15. These effects reduce the advantages of the GaSb/InAs hetero-junction compared to an InAs homo-junction TFET, and also compared to the ideal broken-bandgap, 1D system considered in [135]. According to our results, the GaSb/InAs hetero-junction TFET is still unable to reach the very demanding ITRS specifications for low power applications, which prescribe an I_{ON}/I_{OFF} ratio larger than 10^7 and $I_{OFF} = 10$ pA μm^{-1}. Therefore several additional design options have been proposed to increase the on current, including the grading of the Al molar fraction in an AlGaSb source [131], or the insertion and engineering of quantum wells in the source region [78].

Recently an alterative device design based on the concept of polarization-engineered tunnel diodes has been proposed [136], which exploits III-nitrides hetero-junctions [82]. By means of TB NEGF simulations it was shown that interband tunneling can be significantly large in GaN/InN/GaN hetero-junctions, leading to an on current close to 100 μA μm^{-1} and to a very competitive subthreshold swing.

4.4. Effects of non-ideal interfaces, defects and traps

Non-ideal properties of interfaces are considered the most relevant problem affecting the off-state behavior of TFETs and impeding the experimental realization of devices with a subthreshold swing well below 60 mV/dec. The impact of traps has been first included in NEGF simulations with a phenomenological description in [137], where interface traps were modeled as 0D electrically active states modifying simultaneously the electron transport and the device electrostatics. This was obtained by superimposing cubic potential wells to the CB with different volume size and potential height in order to tune the energy trap levels [138]. It was found that only a few traps can significantly deteriorate the SS of narrow InAs NW FETs with a 5×5 nm^2 cross-section, because traps can act as stepping stones for BTBT and thus enhance the corresponding current, with a tunneling mechanism that can be phonon assisted and thus inelastic. The phonon assisted tunneling also brings along a sensitivity to the temperature of the otherwise temperature independent I_{DS} versus $\mathbf{V}_{GS}$ characteristics [138].

Several papers using a semi-classical approach to describe trap assisted tunneling by using a modified Shockley–Read–Hall formula found that traps can produce a sizable degradation of the SS, and investigated the minimum trap densities necessary for a steep slope behavior [51, 120, 139–141]. In very recent references, for instance, it has been argued that the trap density should be decreased by 100 times with respect to the state of the art values in order to obtain IIIV TFETs with an SS below

60 mV/dec at room temperature [142]. The key role played by traps and defects for the interpretation of the experimental IV characteristics of InAs nanowire TFETs has been clearly underlined in [51]. Analitycal formulas using a detailed balance analysis have been also used to include the effects of traps in the simulation of GaSb/InAs TFETs [143].

Interface traps can be also an important source of device variability in TFETs [144, 145], together with surface roughness [72, 73], dopant fluctuations in the source [146, 147] and work-function variation associated to the granularity of the metal gate [141, 148]. Traps typically influence the variability of I_{OFF} and subthreshold swing much more than the I_{ON} variability [144], because in the on state the direct tunneling is dominant with respect to trap-assisted tunneling.

5. Tunnel FETs based on 2D crystals and van der Waals hetero-structures

As already mentioned in sections 3 and 4, the design of TFETs is very demanding in terms of electrostatic integrity and requires very small semiconductor film thicknesses (for planar MOSFETs) or small effective diameters (for FinFETs or nanowire MOSFETs) [141, 149], which in turn result in significant bandgap widening effects and consequent degradation of the on current [124]. The subthreshold region is furthermore vulnerable to the effects of interface defects, that are among the most serious hurdles to achieve small sub-V_T swings [138, 144]. In this respect, monolayers of graphene or transition-metal dichalcogenides are very attractive because they have a sub-nanometer thickness and are in principle free of dangling bonds at the surface, in virtue of their native 2D nature. Several arrangements for BTBT transistors based on 2D crystals have been proposed [150], ranging from the lateral transistor sketched in figure 16(a) and having an architecture similar to conventional TFETs based on 3D semiconductors [151–153], to several possible embodiments based on vertical van der Waals hetero-structures between 2D crystals or between 3D and 2D crystals [154, 155], an example of which is shown in figure 16(b). The weak bonding in the out-of-plane direction is expected to ease the fabrication of vertical hetero-structures with limited strain even in presence of a significant lattice mismatch [150].

In the family of vdW-TFETs one can further distinguish resonant tunneling transistors aiming at a gate controlled negative differential resistance element for tera-hertz applications [156–158], from density of states switches that, similar in concept to the EHBTFET discussed in section 2.3, target instead a very abrupt turn-on characteristic [154, 159].

5.1. Physical mechanisms and modeling methodologies for van der Waals TFETs

From a modeling perspective it is interesting to notice that both resonant tunneling FETs and DOS switches based on van der Waals hetero-structures have been originally conceived and described using the Bardeen's transfer
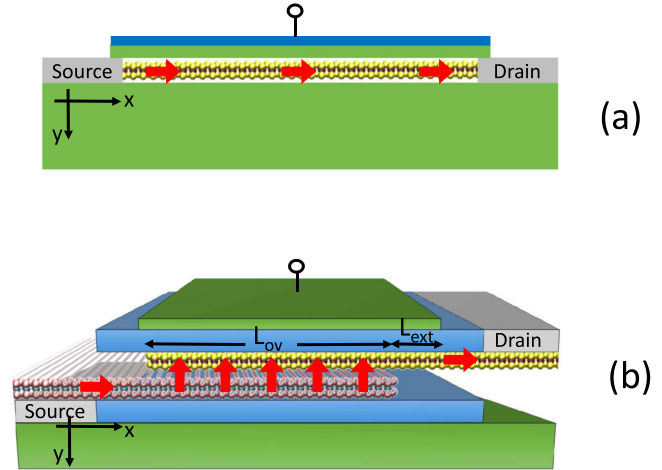
Figure 16. Sketch of possible architectures for tunnel FETs based on 2D materials. (a) Homojunction, lateral TFET where the current flux occurs in the plane of the 2D crystal; (b) van der Waals hetero-junction transistor (vdW-TFETs) where the bottom and top 2D layer act respectively as the source and the drain of the device. In the overlap area between the 2D layers a BTBT current flows in the direction normal to the plane of the 2D layers. L_{ov} and L_{ext} indicate respectively the length of the overlap region between the 2D materials and the extension of the gate electrode beyond the overlap region.

Hamiltonian approach [31, 160, 161]. Based on such formalism, the current per unit area can be written as [154]:

$$I = g_v \frac{4\pi e}{\hbar} \sum_{\mathbf{k}_T, \mathbf{k}_B} |M(\mathbf{k}_T, \mathbf{k}_B)|^2 \delta(E_B(\mathbf{k}_B) - E_T(\mathbf{k}_T))(f_B - f_T),$$

(30)

where g_v is the valley degeneracy, $\mathbf{k}_B$, $\mathbf{k}_T$ are the wave-vectors respectively in the bottom (source) and top (drain) 2D material, $E_B(\mathbf{k}_B)$ and $E_T(\mathbf{k}_T)$ denote the corresponding energies, while f_B and f_T are the Fermi occupation functions in the bottom and top layer that depend on the respective Fermi levels E_{FB} and E_{FT}. This expression for the current density assumes that the majority carriers of the two 2D materials are at thermodynamic equilibrium with their Fermi levels, where the difference between the Fermi levels, $E_{FB} - E_{FT} = eV_{DS}$, is simply set by the drain to source voltage V_{DS}.

The reader may notice the similarity between equations (30) and (13) introduced for 2D to 2D tunneling (e.g. EHBTFETs). However in equation (30) the wave-vectors $\mathbf{k}_B$, $\mathbf{k}_T$ belong to two different materials and, furthermore, the equation does not enforce the conservation of the in plane wave-vector (as discussed in more details below), whereas the JDOS in equation (13) entails the conservation of both energy and in plane wave-vector in the BTBT process.

The matrix element $M(\mathbf{k}_T, \mathbf{k}_B)$ expresses the coupling between the two layers and entirely governs the current, but the Dirac function $\delta(E_B(\mathbf{k}_B) - E_T(\mathbf{k}_T))$ is also of utmost importance in equation (30), because it enforces energy conservation by prescribing that the sum over $\mathbf{k}_B$, $\mathbf{k}_T$ in equation (30) be restricted to the states of the two layers that have the same energy. This implies that the current vanishes if the top of the valence band in the bottom layer is lower than

the bottom of the CB in the top layer. Under these circumstances, in fact, no valence band electrons exist in the bottom layer that can transfer to the CB of the top layer via an elastic tunneling process.

As can be seen the current expression in the Bardeen's transfer Hamiltonian approach describes quite effectively the nature of DOS switch of a vdW-TFET. In fact, by modulating the band edge alignment in the top and bottom layer by means of an external gate bias, one can open or close the tunneling window and thus drive the device respectively in the on- or in the off-state.

It is usually assumed that in actual van der Waals hetero-structures several physical mechanisms can occur in the interlayer region that relax the conservation of the in plane wave-vector $\mathbf{k}$ in the tunneling process [157]. This implies that in equation (30) we have significant contributions to the current also for different $\mathbf{k}_B$ and $\mathbf{k}_T$ values, unlike in the physical picture described in section 2.2 and formally expressed by equation (13). Physical mechanisms that can assist the tunneling and relax momentum conservation may be charged impurities [162], short-range disorder [163], or Moiré patterns that have been observed at the graphene-hBN interface [164, 165]. When the tunneling is assisted by a perturbation potential $U_{sc}(r, z)$ in the interlayer region, the matrix element in equation (30) can be expressed as [154]:

$$M(\mathbf{k}_T, \mathbf{k}_B) = \int_A \mathrm{d}r \int \mathrm{d}z \, \psi_{T,\mathbf{k}_T}^\dagger(r, z) \, U_{sc}(r, z) \, \psi_{B,\mathbf{k}_B}(r, z),$$
(31)

where $\mathbf{r}$ and z are the in plane and vertical spatial coordinates, while $\psi_{B,\mathbf{k}_B}$ and $\psi_{T,\mathbf{k}_T}$ are the electron wave-function respectively in the bottom and top 2D layers.

The wave-functions $\psi_{T,\mathbf{k}}(r, z)$ and $\psi_{B,\mathbf{k}}(r, z)$ in equation (31) are the Bloch functions of the two 2D layers and they entail a decay in the out of plane direction that has not been made explicit in the expression of the matrix element. The calculation of such a decay is the most important and arguably most difficult problem in the application of Bardeen's transfer Hamiltonian approach to vdW-TFETs, and it is a necessary step if one wants to use equations (30), (31) as a quantitative and predictive model, as opposed to a conceptual tool for an insightful but essentially qualitative analysis. As a matter of fact, however, the decay of the wave-function in the interlayer or in the van der Waals gap has been frequently taken as an adjustable parameter to reproduce experimental data [156, 157].

The problem of the coupling between the wave-functions in the bottom and top layer remains a central and somewhat thorny problem even in different transport formalisms, such as the NEGF modeling based on the empirical TB method. In fact the parameters of a TB Hamiltonian are typically calibrated to reproduce the electronic band-structure, as determined via *ab initio* density functional theory (DFT) calculations or inferred from experiments. Such a calibration is effective for the in-plane coupling energies of a given 2D crystal, whereas it is far more difficult to determine the vertical coupling between the atomic sites of different materials stacked to form a van der

Waals hetero-structure. An interesting alternative to the empirical TB approach is to transform the wave-functions obtained by DFT calculations, that are typically obtained using a plane-wave expansion [166], into maximally localized Wannier functions centered on the ions, which can be accomplished by using the *wannier90* package [167]. This effectively produces a TB Hamiltonian matrix with no need of an empirical calibration of the parameters [168]. Such a methodology is becoming quite popular, and it is particularly effective for 2D crystals and vdW-TFETs[169, 170].

Also *ab initio* calculations for van der Waals hetero-structures, however, encounter some conceptual and practical difficulties, starting right from the possible lattice mismatch of the 2D crystals and the resulting problem in the definition of the super-cell to be used for the analysis of the physical system. As an example, if we consider a van der Waals hetero-structure consisting of WTe_2 and MoS_2 monolayers, then the lattice parameter for the two unstrained crystals is about 3.55 Å and 3.19 Å, respectively, so that the super-cell necessary to include an integer number of unstrained unit cells of the individual materials is very large. It is thus a common practice to assume a strain in the two layers so that a matching of the lattice parameters can be obtained. For the case of WTe_2 and MoS_2 monolayers, so as to conclude the example, authors have introduced a compressive strain on WTe_2 and tensile strain on MoS_2 layer to obtain a commensurate lattice parameter of 3.411 Å in the two materials [171–173]. As can be seen for the MoS_2 this implies a strain larger than 6%, which alters significantly the bandstructure of an isolated MoS_2 monolayer compared to the unstrained case. Thus, quite paradoxically, although it has been claimed that the weak van der Waals bonding in the vertical direction may ease the fabrication of high quality hetero-structures even in the presence of a significant lattice mismatch, in most DFT studies of van der Waals hetero-structures a significant strain has been introduced in order to achieve a lattice matching condition.

A cutting-edge methodology in the modeling of van der Waals hetero-structures and vdW-TFETs is the employment of *ab initio* methods to calculate directly the transmission across the physical structures [174], which is a capability that has become available even in some popular, open-source programs such as Quantum Espresso and Siesta [166, 175].

As of today, some aspects remain very hard to be adequately described in the numerical simulation of van der Waals hetero-structures and vdW-TFETs. As an example, it should be recalled that in most fabrication techniques it is very challenging to obtain an accurate rotational alignment of the two 2D materials. The rotational alignment of the two 2D crystals is crucial for the working principle of resonant tunneling vdW-TFETs [156, 157], whereas it is expected to be less critical for DOS switches targeting a small SS [154]. In any case it is difficult to account for rotational misalignment in transport simulations and even in bandstructure calculations, in fact such a misalignment poses new difficulties in the definition of the super-cell of the physical system, besides the already mentioned problems due to a possible
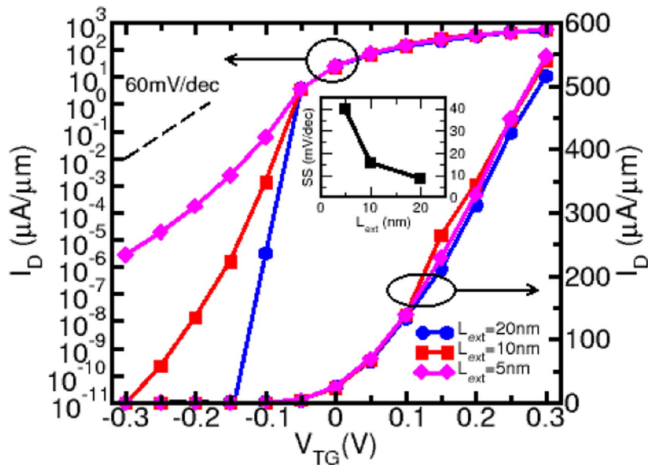
Figure 17. I_{DS} per unit device width versus top gate voltage at $V_{DS} = 0.3$ V for different lengths L_{ext} of the top gate extension indicated in figure 16(b).

lattice mismatch. Furthermore, similarly to the case of TFETs consisting of conventional 3D semiconductors, even for vdW-TFETs it is difficult to include in a self-consistent transport and device simulation the tails of the DOS in the energy gap, whose effect on the minimum SS of vdW-TFETs can be important.

5.2. A few operation and design considerations for van der Waals TFETs

While the original concept of a van der Waals TFET has been conceived as a device where the current should be fairly constant across the overlap area between the two 2D layers, numerical simulations have shown that the current injection from the bottom to the top layer can be strongly non uniform in the overlap area [170, 172, 173], and very sensitive to the length L_{ext} of the extension between the gate and the top 2D layer (see figure 16(b)). In this latter regard, figure 17 reports the I_{DS} versus top gate voltage, V_{TG}, characteristic for a van der Waals TFET consisting of a WTe$_2$ bottom layer (being the source region), and an MoS$_2$ top layer (being the drain); the simulation approach has been described in [172, 173]. As can be seen the sub-threshold region is strongly influenced by L_{ext} and, in particular, for $L_{ext} = 20$ nm the device can attain a subthreshold swing as low as 11 mV/dec (average value for I_{DS} between 10 pA μm^{-1} and 1 μA μm^{-1}), whereas for smaller L_{ext} the swing tends to degrade. This behavior has been explained by a close analysis of the LDOS in the off-state of the device. In particular, it was found that electrons in the valence band of WTe$_2$ have energy values that, in the overlap region, correspond to states deep in the energy band gap of the MoS$_2$ top layer. Consequently, the most favorable tunneling path from the bottom to the top layer was found to be at the right edge of the bottom layer and the tunneling distance increases by enlarging L_{ext} [172, 173]. As a result, the current in the off-state is exponentially suppressed by increasing L_{ext}, which explains the large sensitivity of I_{DS} to L_{ext} in figure 17, and the degradation of SS for small L_{ext} values.

The edge transport discussed above is an undesired effect in van der Waals TFETs, and it can lead to an I_{DS} that does not scale with the length of the overlap region [172, 173]. While a number of design aspects have been recently analyzed by means of numerical simulations, experiments for van der Waals TFETs are still at a quite embryonal stage. However new experimental data for Esaki diodes or transistors consisting of hetero-structure van der Waals systems is appearing with a steady pace [176–178], and a transistor based of an MoS$_2$–Ge hetero-structure has recently shown sub-60 mV/dec subthreshold swing over several orders of magnitude of I_{DS} variation [155].

6. Concluding remarks

This paper has reviewed several aspects related to physics based modeling in TFETs, whose description requires remarkable innovations compared to conventional MOSFETs and, moreover, poses challenges that are at the cutting-edge of our present ability to develop quantitative and predictive models for nanoscale transistors.

The BTBT itself is of course a genuinely quantum mechanical effect governed by the imaginary energy dispersion, namely the branches of the energy dispersion corresponding to imaginary components of the electron wave-vector and connecting the valence to the CB. This is a distinct feature compared to most transistors (e.g. MOSFETs and bipolar junction transistors), where transport essentially occurs inside the conduction and valence bands, so that semi-classical transport models are often still adequate. Not only in TFETs the imaginary energy relation plays a central role, but in the working principle of these DOS switches the subthreshold swing may be ultimately limited by how steeply the DOS decays in the energy gap, so that the tails of the bands become an important physical ingredient that should be accounted for in numerical models.

In indirect bandgap semiconductors BTBT can be dominated by a phonon-assisted mechanism, which makes it particularly challenging the use of full quantum transport approaches, such as NEGF, because they must include dissipative phenomena rather than being restricted to coherent transport. Furthermore, many of the proposed TFET architectures make use of hetero-junctions: III–V semiconductor systems with broken-bandgap alignment, for example, are very promising for the enhancement of the tunneling current. At the same time, the requirement of a strong gate control and good electrostatic integrity is driving towards and aggressive geometrical scaling with narrow (or thin) device cross-sections. This technological trend implies a growing influence of quantum confinement effects, that alter significantly the energy dispersion in nanoscale FETs compared to the corresponding bulk materials, and affect also the band-alignments of hetero-junctions. A sound modeling of hetero-junctions and quantum confining effects is very challenging and critically important. Strain engineering has also been explored for TFETs, which adds one more variable (and one more challenge) to the description of the band-structure, in

addition to quantum confinement and hetero-junctions. The role of hetero-junctions is even more crucial in TFETs based on 2D semiconductors, such as van der Waals hetero-structure TFETs.

Moreover, despite all efforts to enhance and engineer BTBT in TFETs by using optimal material and device systems, the I_{DS} in TFETs is frequently dominated by tunneling mechanisms assisted by defects and interface states, in particular at small current levels where the off-current should be suppressed to the very low values prescribed by the ITRS roadmap for low power applications, namely to a few tens of pA μm^{-1}. Inclusion of trap-assisted tunneling and generation-recombination mechanisms has been proved to be very important in order to reproduce experimental results with physics based simulations and, at the same time, these physical mechanisms are a serious hindrance to the implementation of TFETs with a subthreshold swing smaller than 60 mV/dec in a fairly large range of drain current.

It should be also noticed, moreover, that suppressing the off-current of a real world transistor down to a few tens of pA μm^{-1} demands a tight control and effective suppression of all leakage current mechanisms, which may pose a lower limit to the minimum current that can be attained in practice. While gate leakage, as well as defects assisted generation-recombination mechanisms are at work even in conventional CMOS transistors, and consequently they are frequently neglected in the analysis of TFETs based on the optimistic assumption that a carefully optimized CMOS technology can deal with these issues, the inclusion of all leakage mechanisms becomes crucial in the comparison between simulations and experimental data.

Because of the complexity of the physical mechanisms at work in an actual TFET, a hierarchy of models has been developed with a quite wide range of accuracy and predictive capabilities, different computational burden and, consequently, different geometrical dimensions of the transistors that can be practically analyzed and simulated with such models.

This paper has reviewed the seminal contributions on direct and indirect BTBT modeling in semiconductors, from which most of the models implemented in TCAD simulators have been actually derived; then we reviewed the main features and limitations of the TCAD models themselves. We have introduced and analyzed what we defined non-self-consistent quantum models, that is models that calculate BTBT rates using quite sophisticated formulations, but as a post-processing of the band-structure and electrostatic potential profile obtained by using either TCAD tools or in-house developed Schrödinger–Poisson solvers that do not include BTBT. The influence of the carriers generated by BTBT and of their transport towards the electrodes is not accounted for in these models. This feature differentiates such models from full quantum models, based on NEGF or on the quantum-transmitting-boundary formalism, that instead solve the open-boundary Schrödinger equation problem, with an accuracy and a computational burden that vary in a wide range depending on the Hamiltonian employed in the simulations. A specific section has been also devoted to tunnel

FETs based on 2D crystals and van der Waals hetero-structures, because they are very timely and innovative device structures, which also pose some specific challenges from a modeling and simulation perspective.

Our review on TFETs modeling and simulations does not have an ambition of completeness and in the introduction we already mentioned that, for instance, we have addressed neither compact models for TFETs nor the methodologies employed for a circuit level evaluation of TFETs. The research field on tunnel FETs has been expanding and evolving quite hectically in the last ten years, so that the main goal of this paper is to provide the reader with an introductory description of the most important physics based models for TFETs, as well as a possible guidance in the already wide and moreover rapidly enlarging literature. It is easy to foresee that in the near future new developments and substantial progress will be reported for TFETs modeling because this is, at the time of writing, a very vital research field in the electron device and applied materials worldwide arena.

Acknowledgments

References

[1] Taylor M B 2012 Is dark silicon useful? Harnessing the four horsemen of the coming dark silicon apocalypse *2012 49th ACM/EDAC/IEEE, Design Automation Conf. (DAC)* pp 1131–6

[2] Dennard R H, Gaensslen F H, Kuhn L, Yu N Y, Ridout V L, Bassous E and LeBlanc A 1974 Design of ion-implanted MOSFETs with very small physical dimensions *IEEE J. Solid State Circuits* **SC-9** 256–68

[3] Baccarani G, Wordeman M R and Dennard R H 1984 Generalized scaling theory and its application to a 1/4 micrometer mosfet design *IEEE Trans. Electron Devices* **ED-31** 452–62

[4] Chandrakasan A P, Sheng S and Brodersen R W 1992 Low-power CMOS digital design *IEEE J. Solid State Circuits* **27** 473–84

[5] Chandrakasan A P, Daly D C, Finchelstein D F, Kwong J, Ramadass Y K, Sinangil M E, Sze V and Verma N 2010 Technologies for ultradynamic voltage scaling *IEEE Proc.* **98** 191–214

[6] Calhoun B H and Chandrakasan A 2004 Characterising and modelling minimum energy operation for subthreshold circuits *Proc. IEEE Int. Symp. on Low-Power Electronics and Design* pp 90–5

[7] Horowitz M, Alon E, Patiland D, Naffziger S, Kumar R and Bernstein K 2005 Scaling, power, and the future of CMOS *IEEE IEDM Technical Digest* pp 15–8

[8] Theis T N and Solomon P M 2010 In quest of the next switch: prospects for greatly reduced power dissipation in a successor to the silicon field-effect transistor *IEEE Proc.* **98** 2005–14

[9] Seabaugh A and Zhang Q 2010 Low-voltage tunnel transistors for beyond CMOS logic *Proc. IEEE* **98** 2095–110

[10] Ionescu A and Riel H 2011 Tunnel field-effect transistors as energyefficient electronic switches *Nature* **479** 329–37

[11] Taur Y and Ning T 1998 *Fundamentals of Modern VLSI Devices* (Edimburgh: Cambridge University Press)

[12] Lu H and Seabaugh A 2014 Tunnel field-effect transistors: state-of-the-art *IEEE J. Electron Devices Soc.* **2** 44–9

[13] Kane E 1959 Zener tunneling in semiconductors *J. Phys. Chem. Solids* **12** 181–8

[14] Hurkx G A M, Klaassen D B M and Knuvers M P G 1992 A new recombination model for device simulation including tunneling *IEEE Trans. Electron Devices* **39** 331–8

[15] Chan T Y, Chen J, Ko P K and Hu C 1987 The impact of gate-induced drain leakage current on MOSFET scaling *1987 Int. Electron Devices Meeting* vol 33, pp 718–21

[16] Endoh T, Shirota R, Momodomi M and Masuoka F 1990 An accurate model of subbreakdown due to band-to-band tunneling and some applications *IEEE Trans. Electron Devices* **37** 290–6

[17] Takayanagi M and Iwabuchi S 1991 Theory of band-to-band tunneling under nonuniform electric fields for subbreakdown leakage currents *IEEE Trans. Electron Devices* **38** 1425–31

[18] Habas P, Lugbauer A and Selberherr S 1992 Two-dimensional numerical modeling of interband tunneling accounting for nonuniform electric field *NUPAD IV. Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits* pp 135–40

[19] Ingrosso G, Selmi L and Sangiorgi E 2002 Monte Carlo simulation of program and erase charge distributions in NROM(TM) devices *32nd European Solid-State Device Research Conf.* pp 187–90

[20] Knoch J and Appenzeller J 2008 Tunneling phenomena in carbon nanotube field-effect transistors *Phys. Status Solidi* a **205** 679–94

[21] De Michielis L, Dağtekin N, Biswas A, Lattanzio L, Selmi L, Luisier M, Riel H and Ionescu A M 2013 An innovative band-to-band tunneling analytical model and implications in compact modeling of tunneling-based devices *Appl. Phys. Lett.* **103** 123509

[22] Schiff L I 1968 *Quantum Mechanics* 3rd edn (New York: McGraw-Hill) pp 268–79

[23] Kane E 1961 Theory of tunneling *J. Appl. Phys.* **32** 83–91

[24] Tanaka S 1994 A unified theory of direct and indirect interband tunneling under a nonuniform electric field *Solid-State Electron.* **37** 1543–52

[25] Synopsys Inc. 2014 *Sentaurus Device User Guide, Version 2014.09* (Mountain View, CA: Synopsys Inc.)

[26] Michielis L D, Iellina M, Palestri P, Ionescu A M and Selmi L 2012 Effect of the choice of the tunnelling path on semiclassical numerical simulations of tfet devices *Solid-State Electron.* **71** 7–12

[27] Fischetti M, O'Regan T, Sudarshan N, Sachs C, Jin S, Jiseok K and Zhang Y 2007 Theoretical study of some physical aspects of electronic transport in n-MOSFETs at the 10-nm gate-lengt *IEEE Trans. Electron Devices* **54** 2116–36

[28] Lattanzio L, Dagtekin N, De Michielis L and Ionescu A M 2012 On the static and dynamic behavior of the germanium electron–hole bilayer tunnel FET *IEEE Trans. Electron Devices* **59** 2932–8

[29] Pan A and Chui C O 2014 Modeling direct interband tunneling: II. Lower-dimensional structures *J. Appl. Phys.* **116** 054509

[30] Revelant A, Palestri P, Osgnach P and Selmi L 2013 Calibrated multi-subband Monte Carlo modeling of tunnel-fets in silicon and iii–v channel materials *Solid-State Electron.* **88** 54–60

[31] Bardeen J 1961 Tunnelling from a many-particle point of view *Phys. Rev. Lett.* **6** 57–9

[32] Bigelow J and Leburton J 1994 Self-consistent modeling of resonant interband tunneling in bipolar tunneling field-effect transistors *IEEE Trans. Electron Devices* **41** 125–31

[33] Yamanishi M and Suemune I 1984 Comment on polarization dependent momentum matrix elements in quantum well lasers *Japan. J. Appl. Phys.* **23** L35–6

[34] Asada M, Kameyama A and Suematsu Y 1984 Gain and intervalence band absorption in quantum-well lasers *IEEE J. Quantum Electron.* **20** 745–53

[35] Schenk A, Stahl M and Wünsche H-J 1989 Calculation of interband tunneling in inhomogeneous fields *Phys. Status Solidi* b **154** 815–26

[36] Alper C, Palestri P, Lattanzio L, Padilla J and Ionescu A 2015 Two dimensional quantum mechanical simulation of low dimensional tunneling devices *Solid-State Electron.* **113** 167–72

[37] Carrillo-Nuñez H, Ziegler A, Luisier M and Schenk A 2015 Modeling direct band-to-band tunneling: from bulk to quantum-confined semiconductor devices *J. Appl. Phys.* **117** 234501

[38] Alper C, Visciarelli M, Palestri P, Padilla J L, Gnudi A, Gnani E and Ionescu A M 2015 Modeling the imaginary branch in III-V tunneling devices: effective mass versus $k \cdot p$ *2015 Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)* (Piscataway, NJ: IEEE) pp 273–6

[39] Pan A and Chui C O 2014 Modeling direct interband tunneling: I. Bulk semiconductors *J. Appl. Phys.* **116** 054508

[40] Verreck D, de Put M V, Sore B, Verhulst A S, Magnus W, Vandenberghe W G, Collaert N, Thean A and Groeseneken G 2014 Quantum mechanical solver for confined heterostructure tunnel field-effect transistors *J. Appl. Phys.* **115** 053706

[41] Verreck D, Verhulst A S, Soree B, Collaert N, Mocuta A, Thean A and Groeseneken G 2016 Non-uniform strain in lattice-mismatched heterostructure tunnel field-effect transistors *ESSDERC* pp 412–5

[42] Keldysh L V 1958 Behaviour of non-metallic crystals in strong electric fields *Sov. Phys. J. Exp. Theor. Phys.* **6** 763–70

[43] Vandenberghe W, Sorée B, Magnus W and Fischetti M 2011 Generalized phonon-assisted Zener tunneling in indirect semiconductors with non-uniform electric fields: a rigorous approach *J. Appl. Phys.* **109** 124503

[44] Walke A, Verhulst A, Vandooren A, Verreck D, Simoen E, Rao V R, Groeseneken G, Collaert N and Thean A V-Y 2013 Part I: Impact of Field-induced quantum confinement on the subthreshold swing behavior of line TFETs *IEEE Trans. Electron Devices* **60** 4057–64

[45] Padilla J L, Gámiz F and Godoy A 2012 A simple approach to quantum confinement in tunneling field-effect transistors *IEEE Electron Device Lett.* **33** 1342–4

[46] Beneventi G B, Gnani E, Gnudi A, Reggiani S and Baccarani G 2015 Optimization of a pocketed dual-metal-gate TFET by means of TCAD simulations accounting for quantization-induced bandgap widening *IEEE Trans. Electron Devices* **62** 44–51

[47] Schenk A 2016 III–V-based hetero tunnel FETs: a simulation study with focus non-ideality effects *Joint Int. EUROSOI Workshop Int. Conf. Ultimate Integr. Silicon*

[48] Shen C, Yang L-T, Samudra G and Yeo Y C 2011 A new robust non-local algorithm for band-to-band tunneling simulation and its application to Tunnel-FET *Solid-State Electron.* **57** 23–30

[49] Walke A *et al* 2013 Part II: Investigation of subthreshold swing in line tunnel fets using bias stress measurements *IEEE Trans. Electron Devices* **60** 4065–72

[50] Moselund K E, Cutaia D, Schmid H, Riel H, Sant S and Schenk A 2016 Complementary III–V heterostructure tunnel FETs *ESSDERC* pp 403–7

[51] Sant S, Moselund K, Cutaia D, Schmid H, Borg M, Riel H and Schenk A 2016 Lateral InAs/Si p-Type tunnel FETs integrated on Si: II. Simulation study of the impact of interface traps *IEEE Trans. Electron Devices* **63** 4240–7

[52] Ferry D and Goodnick S 1997 *Transport in Nanostructures* (Cambridge: Cambridge University Press)

[53] Datta S 1998 *Electronic Transport in Mesoscopic Systems* (Cambridge: Cambridge University Press)

[54] Datta S 2005 *Quantum Transport—Atom to Transistor* (Cambridge: Cambridge University Press)

[55] Datta S 2000 Nanoscale device modeling: the Green's function method *Superlattices Microstruct.* **28** 253–78

[56] Koswatta S O, Lundstrom M S and Nikonov D E 2009 Performance comparison between p–i–n tunneling transistors and conventional MOSFETs *IEEE Trans. Electron Devices* **56** 456–65

[57] Luisier M and Klimeck G 2009 Atomistic full-band design study of InAs band-to-band tunneling field-effect transistors *IEEE Electron Device Lett.* **30** 602–4

[58] Luisier M and Klimeck G 2010 Simulation of nanowire tunneling transistors: From the Wentzel–Kramers–Brillouin approximation to full-band phonon-assisted tunneling *J. Appl. Phys.* **107** 084507

[59] Luisier M 2011 Performance comparison of GaSb, strained-Si, and InGaAs double-gate ultrathin-body n-FETs *IEEE Electron Device Lett.* **32** 1686–8

[60] Shin M 2009 Full-quantum simulation of hole transport and band-to-band tunneling in nanowires using the k · p method *J. Appl. Phys.* **106** 054505

[61] Conzatti F, Pala M, Esseni D, Bano E and Selmi L 2012 Strain-Induced performance improvements in InAs nanowire tunnel FETs *IEEE Trans. Electron Devices* **59** 2085–92

[62] Baravelli E, Gnani E, Grassi R, Gnudi A, Reggiani S and Baccarani G 2014 Optimization of n- and p-type TFETs integrated on the same InAs/Al$_x$Ga$_{1-x}$Sb technology platform *IEEE Trans. Electron Devices* **61** 178–85

[63] Velev J and Butler W 2004 On the equivalence of different techniques for evaluating the Green function for a semi-infinite system using a localized basis *J. Phys.: Condens. Matter* **16** R637

[64] Sancho M P L, Sancho J M L and Rubio J 1984 Quick iterative scheme for the calculation of transfer matrices: application to Mo (100) *J. Phys. F: Met. Phys.* **14** 1205

[65] Lee D H and Joannopoulos J D 1981 Simple scheme for surface-band calculations: I *Phys. Rev.* B **23** 4988–96

[66] Lee D H and Joannopoulos J D 1981 Simple scheme for surface-band calculations: II. The Green's function *Phys. Rev.* B **23** 4997–5004

[67] Landauer R 1957 Spatial variation of currents and fields due to localized scatterers in metallic conduction *IBM J. Res. Dev.* **1** 233

[68] Büttiker M 1986 Four-terminal phase-coherent conductance *Phys. Rev. Lett.* **57** 1761

[69] Koswatta S O, Nikonov D E and Lundstrom M S 2005 Computational study of carbon nanotube p–i–n tunnel FETs *IEEE Int. Electron Devices Meeting, 2005. IEDM Technical Digest* pp 518–21

[70] Poli S, Reggiani S, Gnudi A, Gnani E and Baccarani G 2008 Computational study of the ultimate scaling limits of CNT tunneling devices *IEEE Trans. Electron Devices* **55** 313–21

[71] Fiori G and Iannaccone G 2009 Ultralow-voltage bilayer graphene tunnel FET *IEEE Electron Device Lett.* **30** 1096–8

[72] Conzatti F, Pala M and Esseni D 2012 Surface-roughness-induced variability in nanowire InAs tunnel FETs *IEEE Electron Device Lett.* **33** 806–8

[73] Carrillo-Nuñez H, Rhyner R, Luisier M and Schenk A 2016 Effect of surface roughness and phonon scattering on extremely narrow InAs–Si Nanowire TFETs *2016 46th European Solid-State Device Research Conf. (ESSDERC)* pp 188–91

[74] Grassi R, Gnudi A, Gnani E, Reggiani S and Baccarani G 2009 An investigation of performance limits of conventional and tunneling graphene-based transistors *J. Comput. Electron.* **8** 441

[75] Cao J, Cresti A, Esseni D and Pala M 2016 Quantum simulation of a heterojunction vertical tunnel FET based on 2D transition metal dichalcogenides *Solid-State Electron.* **116** 1–7

[76] Anantram M P, Lundstrom M S and Nikonov D E 2008 Modeling of nanoscale devices *Proc. IEEE* **96** 1511–50

[77] Conzatti F, Pala M, Esseni D, Bano E and Selmi L 2011 A simulation study of strain induced performance enhancements in InAs nanowire Tunnel-FETs *IEEE IEDM Technical Digest* pp 5.2.1–5.2.4

[78] Pala M G and Brocard S 2015 Exploiting hetero-junctions to improve the performance of III–V nanowire tunnel-FETs *IEEE J. Electron Devices Soc.* **3** 115–21

[79] Lent C S and Kirkner D J 1990 The quantum transmitting boundary method *J. Appl. Phys.* **67** 6353–9

[80] Fisher D S and Lee P A 1981 Relation between conductivity and transmission matrix *Phys. Rev.* B **23** 6851–4

[81] Ilatikhameneh H, Novakovic B, Tan Y, Salmani-Jelodar M, Kubis T, Povolotskyi M, Rahman R and Klimeck G 2014 Atomistic simulation of steep subthreshold slope Bi-layer MoS$_2$ transistors *2014 IEEE, Silicon Nanoelectronics Workshop (SNW)* pp 1–2

[82] Li W *et al* 2015 Polarization-engineered III-nitride heterojunction tunnel field-effect transistors *IEEE J. Exploratory Solid-State Comput. Devices Circuits* **1** 28–34

[83] Long P, Huang J Z, Povolotskyi M, Klimeck G and Rodwell M J W 2016 High-current tunneling FETs with (1 1̄0) orientation and a channel heterojunction *IEEE Electron Device Lett.* **37** 345–8

[84] Mahan G 1990 *Many-Particle Physics* (New York: Plenum)

[85] Lundstrom M S 1990 *Fundamentals of Carrier Transport* (New York: Addison-Wesley)

[86] Koswatta S, Lundstrom M and Nikonov D 2008 Influence of phonon scattering on the performance of p–i–n band-to-band tunneling transistors *Appl. Phys. Lett.* **92** 043125

[87] Pala M G, Grillet C, Cao J, Logoteta D, Cresti A and Esseni D 2016 Impact of inelastic phonon scattering in the OFF state of tunnel-field-effect transistors *J. Comput. Electron.* **15** 1240–7

[88] Fetter A and Walecka J 2003 *Quantum Theory of Many-Particle Systems* (New York: Dover) (https://doi.org/10.1063/1.3071096)

[89] Revelant A, Palestri P, Osgnach P, Lizzit D and Selmi L 2013 On the optimization of SiGe and III–V compound hetero-junction tunnel FET devices *2013 Proc. European Solid-State Device Research Conf. (ESSDERC)* pp 49–52

[90] Sant S and Schenk A 2015 Band-offset engineering for GeSn-SiGeSn hetero tunnel FETs and the role of strain *IEEE J. Electron Devices Soc.* **3** 164–75

[91] Kotlyar R, Avci U E, Cea S, Rios R, Linton T D, Kuhn K J and Young I A 2013 Bandgap engineering of

group IV materials for complementary n and p tunneling field effect transistors *Appl. Phys. Lett.* **102** 113106

[92] Wirths S *et al* 2015 Lasing in direct-bandgap GeSn alloy grown on Si *Nat. Photon.* **9** 88–92

[93] Mayer F, Royer C L, Damlencourt J F, Romanjek K, Andrieu F, Tabone C, Previtali B and Deleonibus S 2008 Impact of SOI, Si$_{1-x}$Ge$_x$OI and GeOI substrates on CMOS compatible tunnel FET performance *2008 IEEE Int. Electron Devices Meeting* pp 1–5

[94] Boucart K and Ionescu A M 2007 Double-gate tunnel FET with high-κ; gate dielectric *IEEE Trans. Electron Devices* **54** 1725–33

[95] Wan J, Royer C L, Zaslavsky A and Cristoloveanu S 2011 Tunneling FETs on SOI: suppression of ambipolar leakage, low-frequency noise behavior, and modeling *Solid-State Electron.* **65–66** 226–33

[96] Vandenberghe W G, Verhulst A S, Groeseneken G, Soree B and Magnus W 2008 Analytical model for point and line tunneling in a tunnel field-effect transistor *2008 Int. Conf. on Simulation of Semiconductor Processes and Devices* pp 137–40

[97] Schmidt M, Schaefer A, Minamisawa R A, Buca D, Trellenkamp S, Hartmann J M, Zhao Q T and Mantl S 2014 Line and point tunneling in scaled Si/SiGe heterostructure TFETs *IEEE Electron Device Lett.* **35** 699–701

[98] Michielis L D, Lattanzio L, Palestri P, Selmi L and Ionescu A M 2011 Tunnel-FET architecture with improved performance due to enhanced gate modulation of the tunneling barrier *2011 69th Annual, Device Research Conference (DRC)* pp 111–2

[99] Revelant A, Villalon A, Wu Y, Zaslavsky A, Royer C L, Iwai H and Cristoloveanu S 2014 Electron-Hole bilayer TFET: experiments and comments *IEEE Trans. Electron Devices* **61** 2674–81

[100] Alper C, Palestri P, Padilla J L and Ionescu A M 2016 The electron–hole bilayer TFET: dimensionality effects and optimization *IEEE Trans. Electron Devices* **63** 2603–9

[101] Dey A W, Svensson J, Ek M, Lind E, Thelander C and Wernersson L-E 2013 Combining axial and radial nanowire heterostructures: radial esaki diodes and tunnel field-effect transistors *Nano Lett.* **13** 5919–24

[102] Villalon A *et al* 2012 Strained tunnel FETs with record ION: first demonstration of ETSOI TFETs with SiGe channel and RSD *2012 Symp. on VLSI Technology (VLSIT)* pp 49–50

[103] Knoll L, Richter S, Nichau A, Scharfer A, Bourdelle K K, Zhao Q T and Mantl S 2013 Gate-all-around Si nanowire array tunnelling FETs with high on-current of 75 μA/μm @ VDD = 1.1 V *2013 14th Int. Conf. on Ultimate Integration on Silicon (ULIS)* pp 97–100

[104] Villalon A *et al* 2014 First demonstration of strained SiGe nanowires TFETs with ION beyond 700 μA/μm *2014 Symp. on VLSI Technology (VLSI-Technology): Digest of Technical Papers* pp 1–2

[105] Luong G V, Narimani K, Tiedemann A T, Bernardy P, Trellenkamp S, Zhao Q T and Mantl S 2016 Complementary strained Si GAA nanowire TFET inverter with suppressed ambipolarity *IEEE Electron Device Lett.* **37** 950–3

[106] Guo P F, Yang L T, Yang Y, Fan L, Han G Q, Samudra G S and Yeo Y C 2009 Tunneling field-effect transistor: effect of strain and temperature on tunneling current *IEEE Electron Device Lett.* **30** 981–3

[107] Boucart K, Riess W and Ionescu A M 2009 Lateral strain profile as key technology booster for all-silicon tunnel FETs *IEEE Electron Device Lett.* **30** 656–8

[108] Solomon P M, Jopling J, Frank D J, Demic C, Dokumaci O, Ronsheim P and Haensch W E 2004 Universal tunneling behavior in technologically relevant P/N junction diodes *J. Appl. Phys.* **95** 5800–12

[109] Zhang Q, Sutar S, Kosel T and Seabaugh A 2009 Fully-depleted Ge interband tunnel transistor: modeling and junction formation *Solid-State Electron.* **53** 30–5

[110] Kao K-H, Verhulst A S, Vandenberghe W G, Soree B, Groeseneken G and De Meyer K 2012 Direct and indirect band-to-band tunneling in germanium-based TFETs *IEEE Trans. Electron Devices* **59** 292–301

[111] Kim S H, Kam H, Hu C and Liu T J K 2009 Germanium-source tunnel field effect transistors with record high ION/IOFF *2009 Symp. on VLSI Technology* pp 178–9

[112] Vandenberghe W G and Fischetti M V 2015 Deformation potentials for band-to-band tunneling in silicon and germanium from first principles *Appl. Phys. Lett.* **106** 013505

[113] Kao K-H, Verhulst A S, Van de Put M, Vandenberghe W G, Soree B, Magnus W and Meyer K De 2014 Tensile strained ge tunnel field-effect transistors: kp material modeling and numerical device simulation *J. Appl. Phys.* **115** 044505

[114] Vurgaftman I, Meyer J R and Ram-Mohan L R 2001 Band parameters for III–V compound semiconductors and their alloys *J. Appl. Phys.* **89** 5815–75

[115] Dewey G *et al* 2011 Fabrication, characterization, and physics of III–V heterojunction tunneling field effect transistors (H-TFET) for steep sub-threshold swing *IEEE IEDM Technical Digest* pp 33.6.1–33.6.4

[116] Dewey G, Chu-Kung B, Kotlyar R, Metz M, Mukherjee N and Radosavljevic M 2012 III–V field effect transistors for future ultra-low power applications *IEEE Symp. on VLSI Technology—Technical Digest* pp 45–6

[117] Zhou G *et al* 2012 Novel gate-recessed vertical InAs/GaSb TFETs with record high I_{ON} of 180 μA/μm at V_{DS} = 0.5 V *IEEE IEDM Technical Digest* pp 777–80

[118] Li R *et al* 2012 AlGaSb/InAs tunnel field-effect transistor with on-current of 78 μA/μm at 0.5 V *IEEE Electron Device Lett.* **33** 363–5

[119] Mookerjea S, Mohata D, Mayer T, Narayanan V and Datta S 2010 Temperature-dependent *I–V* Characteristics of a Vertical In$_{0.53}$Ga$_{0.47}$As tunnel FET *IEEE Electron Device Lett.* **31** 564–6

[120] Vandooren A, Walke A M, Verhulst A S, Rooyackers R, Collaert N and Thean A V Y 2014 Investigation of the subthreshold swing in vertical tunnel-FETs using H$_2$ and D$_2$ anneals *IEEE Trans. Electron Devices* **61** 359–64

[121] Memisevic E, Svensson J, Hellenbrand M, Lind E and Wernersson L-E 2016 Vertical InAs/GaAsSb/GaSb tunneling field-effect transistor on Si with S = 48 mV/decade and I_{on} = 10 μA/μm for I_{off} = 1 nA/μm at V_{DS} = 0.3 V *IEEE Int. Electron Devices Meeting (IEDM)* pp 500–3

[122] Baravelli E, Gnani E, Gnudi A, Reggiani S and Baccarani G 2014 Capacitance estimation for InAs tunnel FETs by means of full-quantum k.p simulation *2014 15th Int. Conf. on Ultimate Integration on Silicon (ULIS)* pp 17–20

[123] Huang J Z, Zhang L, Long P, Povolotskyi M and Klimeck G 2016 *Quantum Transport Simulation of III–V TFETs with Reduced-Order K.P Method* (Cham: Springer International Publishing) pp 151–80

[124] Brocard S, Pala M and Esseni D 2013 Design options for hetero-junction tunnel FETs with high on current and steep sub-threshold voltage slope *2013 IEEE Int. Electron Devices Meeting (IEDM)* pp 5.4.1–5.4.4

[125] Bahder T B 1990 Eight-band **k·p** model of strained zinc-blende crystals *Phys. Rev.* B **41** 11992–2001

[126] Verreck D, Verhulst A S, de Put M L V, Sorée B, Collaert N, Mocuta A, Thean A and Groeseneken G 2016 Uniform strain in heterostructure tunnel field-effect transistors *IEEE Electron Device Lett.* **37** 337–40

[127] Conzatti F, Pala M, Esseni D and Bano E 2013 Investigation of localized versus uniform strain as a performance booster in InAs tunnel-FETs *Solid-State Electron.* **88** 49–53

[128] Visciarelli M, Gnani E, Gnudi A, Reggiani S and Baccarani G 2016 Impact of strain on tunneling current and threshold voltage in III–V nanowire TFETs *IEEE Electron Device Lett.* **37** 560–3

[129] Huang J Z, Wang Y, Long P, Tan Y, Povolotskyi M and Klimeck G 2016 High-performance complementary III–V tunnel FETs with strain engineering arXiv:1605.00955

[130] Knoch J and Appenzeller J 2010 Modeling of high-performance p-type III–V heterojunction tunnel FETs *IEEE Electron Device Lett.* **31** 305–7

[131] Brocard S, Pala M and Esseni D 2014 Large on-current enhancement in hetero-junction tunnel-FETs via molar fraction grading *IEEE Electron Device Lett.* **35** 184–6

[132] Avci U E and Young I A 2013 Heterojunction TFET Scaling and resonant-TFET for steep subthreshold slope at sub-9 nm gate-length *2013 IEEE Int. Electron Devices Meeting* pp 4.3.1–4.3.4

[133] Long P, Wilson E, Huang J Z, Klimeck G, Rodwell M J W and Povolotskyi M 2016 Design and simulation of GaSb/InAs 2D transmission-enhanced tunneling FETs *IEEE Electron Device Lett.* **37** 107–10

[134] Verreck D, Van de Put M, Sorée B, Verhulst A S, Magnus W, Vandenberghe W G, Collaert N, Thean A and Groeseneken G 2014 Quantum mechanical solver for confined heterostructure tunnel field-effect transistors *J. Appl. Phys.* **115** 053706

[135] Koswatta S O, Koester S J and Haensch W 2010 On the possibility of obtaining MOSFET-Like performance and sub-60-mV/dec swing in 1-D Broken-gap tunnel transistors *IEEE Trans. Electron Devices* **57** 3222–30

[136] Krishnamoorthy S, Nath D N, Akyol F, Park P S, Esposto M and Rajan S 2010 Polarization-engineered GaN/InGaN/GaN tunnel diodes *Appl. Phys. Lett.* **97** 203502

[137] Pala M, Esseni D and Conzatti F 2012 Impact of interface traps on the IV curves of InAs tunnel-FETs and MOSFETs: a full quantum study *IEEE IEDM Technical Digest* pp 6.6.1–6.6.4

[138] Pala M and Esseni D 2013 Interface traps in InAs nanowire Tunnel-FETs and MOSFETs: I. Model description and single trap analysis in tunnel-FETs *IEEE Trans. Electron Devices* **60** 2795–801

[139] Teherani J T, Chern W, Agarwal S, Hoyt J L and Antoniadis D A 2015 A framework for generation and recombination in tunneling field-effect transistors *2015 4th Berkeley Symp. on Energy Efficient Electronic Systems (E3S)* pp 1–3

[140] Biswal S M, Baral B, De D and Sarkar A 2016 Study of effect of gate-length downscaling on the analog/rf performance and linearity investigation of inas-based nanowire tunnel FET *Superlattices Microstruct.* **91** 319–30

[141] Avci U E *et al* 2005 Study of TFET Non-ideality effects for determination of geometry and defect density requirements for Sub-60 mV/dec Ge TFET *IEEE IEDM Technical Digest* pp 891–4

[142] Sajjad R N, Chern W, Hoyt J L and Antoniadis D A 2016 Trap assisted tunneling and its effect on subthreshold swing of tunnel FETs *IEEE Trans. Electron Devices* **63** 4380–7

[143] Lind E, Memišević E, Dey A W and Wernersson L E 2015 III–V heterostructure nanowire tunnel FETs *IEEE J. Electron Devices Soc.* **3** 96–102

[144] Esseni D and Pala M 2013 Interface traps in inas nanowire tunnel FETs and MOSFETs: II. Comparative analysis and trap-induced variability *IEEE Trans. Electron Devices* **60** 2802–7

[145] Fan M L, Hu V P H, Chen Y N, Su P and Chuang C T 2013 Investigation of single-trap-induced random telegraph noise

for tunnel FET based devices, 8T SRAM cell, and sense amplifiers *2013 IEEE Int. Reliability Physics Symp. (IRPS)* pp CR.1.1–CR.1.5

[146] Damrongplasit N, Kim S H and Liu T J K 2013 Study of random dopant fluctuation induced variability in the raised-Ge-source TFET *IEEE Electron Device Lett.* **34** 184–6

[147] Leung G and Chui C O 2013 Stochastic variability in silicon Double-Gate lateral tunnel field-effect transistors *IEEE Trans. Electron Devices* **60** 84–91

[148] Fan M L, Hu V P H, Chen Y N, Su P and Chuang C T 2013 Analysis of Single-Trap-induced random telegraph noise and its interaction with work function variation for tunnel FET *IEEE Trans. Electron Devices* **60** 2038–44

[149] Esseni D, Pala M and Rollo T 2015 Essential physics of the off-state current in nanoscale mosfets and tunnel fets *IEEE Trans. Electron Devices* **62** 3084–91

[150] Jena D 2013 Tunneling transistors based on graphene and 2-D crystals *IEEE Proc.* **1585–602**

[151] Zhang Q, Iannaccone G and Fiori G 2014 Two-dimensional tunnel transistors based on Bi_2Se_3 thin film *IEEE Electron Device Lett.* **35** 129–31

[152] Das S, Prakash A, Salazar R and Appenzeller J 2014 Toward low-power electronics: tunneling phenomena in transition metal dichalcogenides *ACS Nano* **8** 1681–9

[153] Ilatikhameneh H, Tan Y, Novakovic B, Klimeck G, Rahman R and Appenzeller J 2015 Tunnel field-effect transistors in 2-D transition metal dichalcogenide materials *IEEE J. Exploratory Solid-State Comput. Devices* **1** 12–9

[154] Mingda L, Esseni D, Snider G, Jena D and Xing H G 2014 Single particle transport in two-dimensional heterojunction interlayer tunneling field effect transistor *J. Appl. Phys.* **115** 074508

[155] Sarkar D, Xie X, Liu W, Cao W, Kang J, Gong Y, Kraemer S, Ajayan P M and Banerjee K 2015 A subthermionic tunnel field-effect transistor with an atomically thin channel *Nature* **526** 91–5

[156] Feenstra R M, Jena D and Gu G 2012 Single-particle tunneling in doped graphene-insulator-graphene junctions *J. Appl. Phys.* **111** 043711

[157] Britnell L, Gorbachev R, Geim A, Ponomarenko L, Mishchenko A, Greenaway M, Fromhold T, Novoselov K and Eaves L 2013 Resonant tunnelling and negative differential conductance in graphene transistors *Nat. Commun.* **4** 1794

[158] Britnell L *et al* 2012 Electron tunneling through ultrathin boron nitride crystalline barriers *Nano Lett.* **12** 1707–10

[159] Li M O, Esseni D, Nahas J J, Jena D and Xing H G 2015 Two-dimensional heterojunction interlayer tunneling field effect transistors (thin-TFETs) *IEEE J. Electron Devices Soc.* **3** 200–7

[160] Harrison W A 1961 Tunneling from an independent-particle point of view *Phys. Rev.* **123** 85

[161] Duke C B 1969 *Tunneling in Solids* vol 1999 (New York: Academic)

[162] Yan J and Fuhrer M S 2011 Correlated charged impurity scattering in graphene *Phys. Rev. Lett.* **107** 206601

[163] Li Q, Hwang E, Rossi E and Sarma S D 2011 Theory of 2D transport in graphene for correlated disorder *Phys. Rev. Lett.* **107** 156601

[164] Xue J, Sanchez-Yamagishi J, Bulmash D, Jacquod P, Deshpande A, Watanabe K, Taniguchi T, Jarillo-Herrero P and LeRoy B J 2011 Scanning tunnelling microscopy and spectroscopy of ultra-flat graphene on hexagonal boron nitride *Nat. Mater.* **10** 282–5

[165] Decker R, Wang Y, Brar V W, Regan W, Tsai H-Z, Wu Q, Gannett W, Zettl A and Crommie M F 2011 Local electronic properties of graphene on a BN substrate via scanning tunneling microscopy *Nano Lett.* **11** 2291–5

[166] Giannozzi P *et al* 2009 QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials *J. Phys.: Condensed Matter* **21** 395502

[167] Marzari N and Vanderbilt D 1997 Maximally localized generalized Wannier functions for composite energy bands *Phys. Rev.* B **56** 12847–65

[168] Luisier M, Rhyner R, Szabo A and Pedersen A 2016 Atomistic simulation of nanodevices *Proc. SISPAD* pp 281–6

[169] Bruzzone S, Iannaccone G, Marzari N and Fiori G 2014 An open-source multiscale framework for the simulation of nanoscale devices *IEEE Trans. Electron Devices* **61** 48–53

[170] Szabo A, Koester S and Luisier M 2015 *Ab-Initio* simulation of van der Waals MoTe$_2$-SnS$_2$ heterotunneling FETs for low-power electronics *IEEE Electron Device Lett.* **36** 514–6

[171] Szabo A, Koester S and Luisier M 2014 Metal-dichalcogenide hetero-TFETs: Are they a viable option for low power electronics? *2014 72nd Annual, Device Research Conf. (DRC)* pp 19–20

[172] Cao J, Logoteta D, Özkaya S, Biel B, Cresti A, Pala M and Esseni D 2015 A computational study of van der Waals tunnel transistors: fundamental aspects and design challenges *2015 IEEE Int. Electron Devices Meeting (IEDM)* pp 12.5.1–12.5.4

[173] Cao J, Logoteta D, Özkaya S, Biel B, Cresti A, Pala M G and Esseni D 2016 Operation and design of van der Waals tunnel transistors: a 3-D quantum transport study *IEEE Trans. Electron Devices* **63** 4388–94

[174] Stokbro K, Engelund M and Blom A 2012 Atomic-scale model for the contact resistance of the nickel–graphene interface *Phys. Rev.* B **85** 165442

[175] Brandbyge M, Mozos J-L, Ordejn P, Taylor J and Stokbro K 2002 Density functional method for nonequilibrium electron transport *Phys. Rev.* B **65** 165401

[176] Roy T, Tosun M, Cao X, Fang H, Lien D-H, Zhao P, Chen Y-Z, Chueh Y-L, Guo J and Javey A 2012 Dual-gated MoS$_2$/ WSe$_2$ van der Waals tunnel diodes and transistors *Nano Lett.* **9** 2071–9

[177] Yan R *et al* 2015 Esaki diodes in van der Waals Heterojunctions with broken-gap energy band alignment *Nano Lett.* **15** 5791–8

[178] Li M O, Yan Li, R, Jena D and Xing H G 2016 Two-dimensional heterojunction interlayer tunnel FET (thin-TFET): from theory to applications *IEEE Int. Electron Devices Meeting (IEDM)* 19.2.1–19.2.4