



OPEN

Conductivity and size quantization effects in semiconductor δ -layer systems

Juan P. Mendez & Denis Mamaluy

We present an open-system quantum-mechanical 3D real-space study of the conduction band structure and conductive properties of two semiconductor systems, interesting for their beyond-Moore and quantum computing applications: phosphorus δ -layers and P δ -layer tunnel junctions in silicon. In order to evaluate size quantization effects on the conductivity, we consider two principal cases: nanoscale finite-width structures, used in transistors, and infinitely-wide structures, electrical properties of which are typically known experimentally. For devices widths $W < 10$ nm, quantization effects are strong and it is shown that the number of propagating modes determines not only the conductivity, but the distinctive spatial distribution of the current-carrying electron states. For $W > 10$ nm, the quantization effects practically vanish and the conductivity tends to the infinitely-wide device values. For tunnel junctions, two distinct conductivity regimes are predicted due to the strong conduction band quantization.

Highly conductive δ -layer systems, i.e. thin, high-density layers of dopants in semiconductors are actively used as a platform for exploration of the future quantum and classical computing when patterned in plane with atomic precision^{1–3}. Such structures, with the dopant densities above to the solid solubility limit⁴, have been shown to possess very high current densities^{3,5} and thus have a strong potential for quantum computing applications^{6,7} and advanced microelectronic devices^{8,9}. However, at the scale important for these applications¹⁰, i.e. devices with sub-20 nm physical gate/channel lengths and/or sub-20 nm widths, that could compete with the future CMOS, the conductive properties of such systems are expected to exhibit a strong influence of size quantization effects. At the same time, experimental assessment of the conductivity of δ -layer systems is typically performed using Hall measurements on samples of macroscopic dimensions ($> 1 \mu\text{m}$)^{5,11–13}.

The electronic structure and conductive properties of Si:P δ -layer systems have been a subject of previous studies based on either effective mass^{14–17}, tight-binding^{18–22}, density functional theory^{23–25} formalisms or semiclassical Boltzmann theory²⁶. Recently it has been demonstrated in^{15–17} that to accurately extract the conductive properties of highly-conductive, highly-confined systems, an open-system quantum-mechanical analysis is necessary. Such open-system treatment, that can be conducted for instance using the Non-Equilibrium Green's Function (NEGF) formalism^{27,28}, allows to compute the current and conductivity directly from the quantum-mechanical flux, thus avoiding semi-classical approximations, which are intrinsic to the traditional charge self-consistent closed-system or periodic boundary conditions band-structure calculation methods. It has been also demonstrated in¹⁶ that an open-system treatment in highly P doped δ -layer in Si: i) permits to reveal the quantization in space and energy of the free electrons around the δ -layer; ii) permits to explain the existence of the shallow 3Γ sub-band, which has been observed experimentally²⁹, without any fitting parameters; iii) predicts significant quantum-mechanical dependence of the current on the δ -layer sheet thickness for a fixed dopant sheet density; and iv) provides a very good agreement with the experimental electrical measurements^{5,11–13}.

An accurate computational description of electron tunneling in semiconductor δ -layer tunnel junctions (such the one shown in Fig. 1) is additionally required because the tunneling rate at a δ -layer junction is affected not only by the gap length and the conductivity of the δ -layers, but also by quantization of the conduction electrons in energy and space¹⁶. In this work we employ an efficient computational open-system quantum-mechanical treatment to explore the conductive band structure and the conductive properties of phosphorus δ -layer systems in silicon (Si:P δ -layer) for device widths, from nano-scale (< 20 nm) to macro-scale ($> 1 \mu\text{m}$) dimensions, and to analyze the influence of size quantization effects on the conductive properties for sub-12 nm device widths. All simulations are carried at the cryogenic temperature of 4K, in which we can neglect inelastic scattering events^{5,30}. The analysis of the influence of different kinds of non-idealities on the tunneling current is presented in³¹. The open-system treatment is based on an application of Keldysh formalism²⁷, known as NEGF²⁸, and the effective

Sandia National Laboratories, Albuquerque, NM 87123, USA. email: jpmende@sandia.gov; mamaluy@sandia.gov

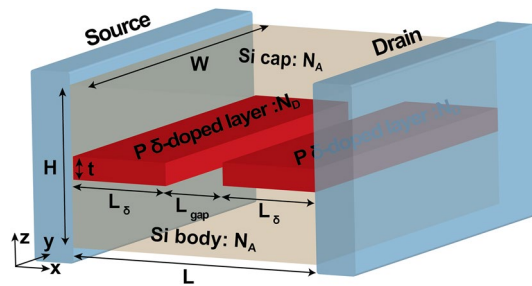


Figure 1. Geometry of the Si:P δ -layer tunnel junction. Our device consists of a semi-infinite source and drain, in contact with the engineering channel. The channel is composed of a lightly doped Si body and Si cap and a very thin, highly P doped-layer with an intrinsic gap of length L_{gap} .

mass theory. Generally for Si systems, the use of the effective mass approximation has been shown to be in a good agreement with tight-binding models for nanowire diameters down to 3 nm^{32,33}; thus, the fidelity of this approximation start to decline for very narrow systems (< 3 nm). Here we have employed an efficient implementation of NEGF, referred to as the Contact Block Reduction (CBR) method^{34–38}, which allows accurate computation of all quantum-mechanical quantities of interest (local density of states, transmission probability, current) and scales linearly $O(N)$ with the system size. We find that at the scale most interesting for applications, i.e. for device widths $W < 10$ nm, quantization effects strongly affect both the conductivity and the spatial distribution of the current-carrying electron states, which, similarly to the charge distribution in the hydrogen atom, is determined by a “quantum number”, i.e. the number of propagating modes. This strong spacial quantization of the current-carrying states can be utilized in novel electronic δ -layer devices. Conversely, for $W > 10$ nm, the quantization effects practically vanish and the conductivity tend to the values of infinitely-wide devices. Finally, two distinct conductivity regimes are predicted due to the strong conduction band quantization for tunnel junctions.

Results and discussion

Effects of the device width on the conductive properties. The conductivity of infinitely-wide Si:P δ -layers has been first studied in^{15–17} using the open-system quantum-mechanical approach for infinite-width systems (see “From finite-width to infinite-width systems”). Here we apply the open-system treatment to more realistic/utilitarian devices of finite width and/or with an intrinsic or lightly-doped tunnel gap as shown in Fig. 1. The device consists of a semi-infinite source and drain (represented by the NEGF open boundary conditions), in contact with a channel of length L , which is composed of a lightly doped Si body and Si cap and a very thin, highly P doped-layer with an intrinsic gap of length L_{gap} , as shown in Fig. 1. In this work we consider tunnel junction devices with an intrinsic gap of length L_{gap} , ranging from 0 nm to 12 nm, and width W , ranging from 2 nm to infinity, paying a particular attention to nano-scale dimensions ($W, L_{gap} < 12$ nm) which are most interesting for quantum and beyond-CMOS computing applications. The length of the channel is assumed to be $L = 30$ nm + L_{gap} to avoid the boundary effects, between the source and drain contacts, and the height of the device is chosen to be $H = 12$ nm. We also assume δ -layer thicknesses ranging from monoatomic layer, $t = 0.2$ nm, to few atomic layers, $t = 1.0$ nm, a doping density of the δ -layer of $N_D = 1.0 \times 10^{14}$ cm⁻² and an acceptor doping densities in the Si cap and Si body of $N_A = 5.0 \times 10^{17}$ cm⁻³. Note that the doping density in the δ -layer is given in cm⁻² (i.e. sheet doping density) to be consistent with experiments’ nomenclature: $N_D^{(2D)} = t \times N_D^{(3D)}$, where t is the δ -layer thickness, $N_D^{(3D)}$ is the doping density in cm⁻³ and $N_D^{(2D)}$ in cm⁻².

Figure 2 shows the computed conductivity in function of the device width W and for different gap lengths L_{gap} . The dashed lines represent the conductivity values for infinitely wide devices, $W \rightarrow \infty$. Details of the computational treatment for infinitely-wide systems are presented in “From finite-width to infinite-width systems”. The simulations suggest that quantization due to the finite size of the device width starts to appear for widths below to 7–10 nm. This size quantization is reflected as non-monotonic increase of the conductivity. Interestingly, there is a peak in the conductivity at $W = 5$ nm and a dip at $W = 3$ nm. Conversely, as W increases, the conductivity tends to the values for infinitely wide devices that therefore can be seen as lower bound limits. Additionally, the size quantization effect on the conductivity of δ -layer tunnel structures is most notable for large tunnel gaps $L_{gap} > 7$ nm. Indeed, in Fig. 2, one can see that the deviation from the conductivity values for infinitely wide devices is more significant for large tunnel gaps. As we will discussed later in “Quantization effects in δ -layer tunnel junctions”, this is due to the consequence of the quantization of the low-energy conduction band in δ -layer systems and the wave-functions decoupling between the left and right δ -layers in large tunnel gaps.

The oscillations of the conductivity for narrow device widths W arise due to a small number of the propagating modes in a “waveguide”, created by the finite-size width of the δ -layer. The corresponding dependence of the current on the device width is shown in Fig. 3a for a gapless δ -layer structure $L_{gap} = 0$. The existence of the conduction steps due to each new propagating mode is well known experimentally since 1980’s³⁹. Here we show, however, that in highly-confined, highly-conductive δ -layer systems, the quantum number m , representing the number of propagating modes, determines not just the total current, but also the spatial distribution of the corresponding current-carrying electrons. The total number of propagating modes m is determined by the number of peaks in the density of states (DOS) below the Fermi level or the number of steps in the electronic transmission function below the Fermi level.

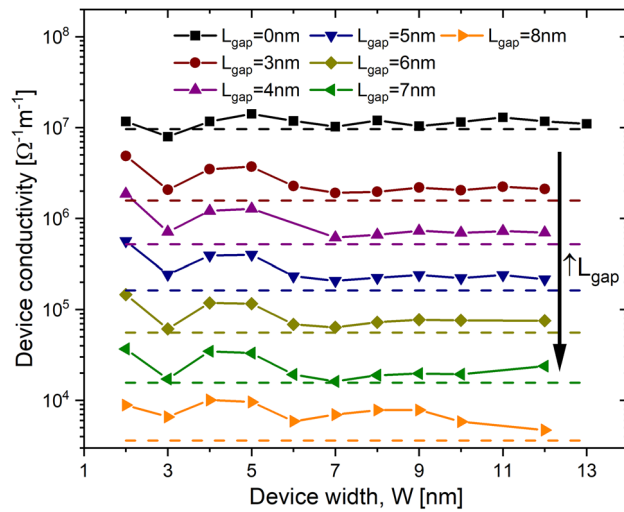


Figure 2. Conductivity for Si: P δ -layer devices. The conductivity is shown in function of the device width W for different tunnel gap lengths L_{gap} . The corresponding conductivity values for infinitely-wide devices are indicated in dashed lines. $N_D = 1.0 \times 10^{14} \text{ cm}^{-2}$, $N_A = 5.0 \times 10^{17} \text{ cm}^{-3}$, $t = 1.0 \text{ nm}$ and for an applied voltage of 1 mV.

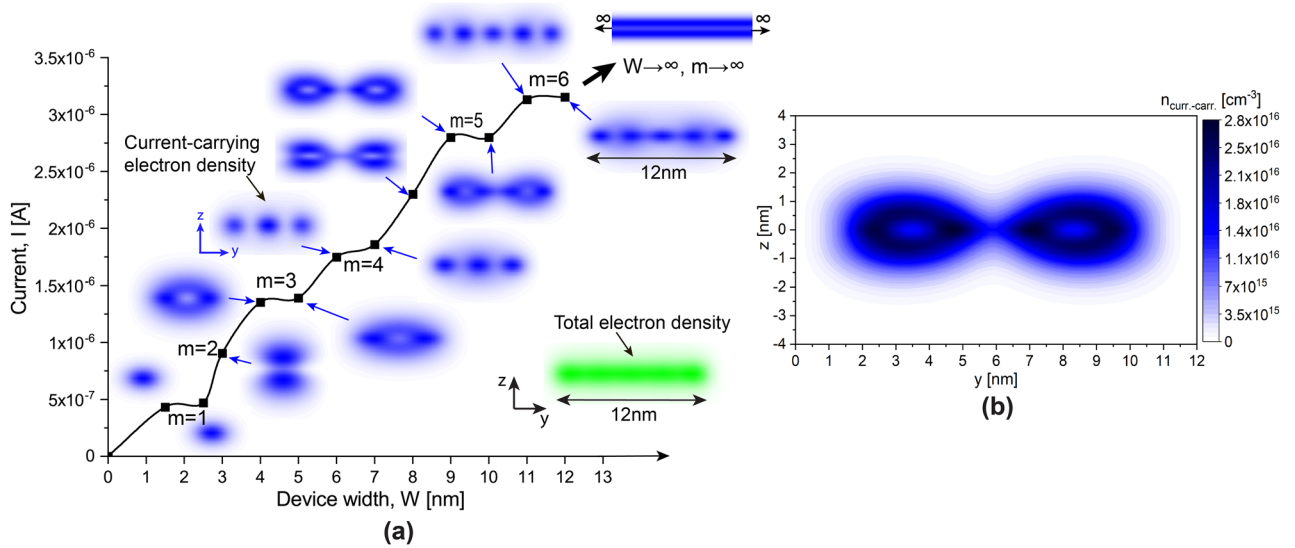


Figure 3. Propagation modes for Si: P δ -layer systems. (a) Current I vs device width W for δ -layer systems with $L_{gap} = 0 \text{ nm}$: the insets in blue color show the spatial distributions of current-carrying modes across a y - z plane, indicating the corresponding number of propagating modes m ; Inset in green color shows the total electron density that includes all (not just current-carrying) occupied electron states for a device width of $W = 12 \text{ nm}$ (Note that $n_{total} \sim 10^{20} \text{ cm}^{-3} \gg n_{curr.-carr.} \sim 10^{16} \text{ cm}^{-3}$). (b) Detail of the spatial distribution of current-carrying states $n_{curr.-carr.}(y, z)$ for a device width of $W = 9 \text{ nm}$. For all calculations, $N_D = 1.0 \times 10^{14} \text{ cm}^{-2}$, $N_A = 5.0 \times 10^{17} \text{ cm}^{-3}$, $t = 0.2 \text{ nm}$ and an applied voltage of 1 mV.

The spatial distribution of the current-carrying electron states, $n_{curr.-carr.}(y, z)$, can be obtained by performing the energy integration of the local density of electron states (LDOS) weighted by the corresponding current spectrum $i_e(E)$ as: $n_{curr.-carr.}(y, z) = \int LDOS(y, z, E) i_e(E) dE / \int i_e(E) dE$. The current spectrum and the local density of states can be obtained by expression in Eqs. (1) and (7), respectively. The spatial current-carrying electrons for the different modes is shown in Fig. 3a as insets in blue color, as well as the corresponding number of propagating modes. Additionally, the total electron density is also included in the figure as an inset in green color, demonstrating only weak spatial quantization along the y -direction. However, the specific portion of electrons with energies close to the Fermi level, i.e. the current-carrying states, do exhibit a strong spatial quantization. Indeed, for $m = 1$ the propagating mode reaches the maximum concentration at the center of the structure, the mode that corresponds to $m = 2$ is “excited” into the further penetration along the confinement direction (z -axis), leaving the center relatively depopulated (in terms of the current-carrying states), the mode $m = 3$ is again “pushed out” of the center along both z - and y -axis. One can further note that the modes $m = 1$, and $m = 4, 6$, etc. tend to

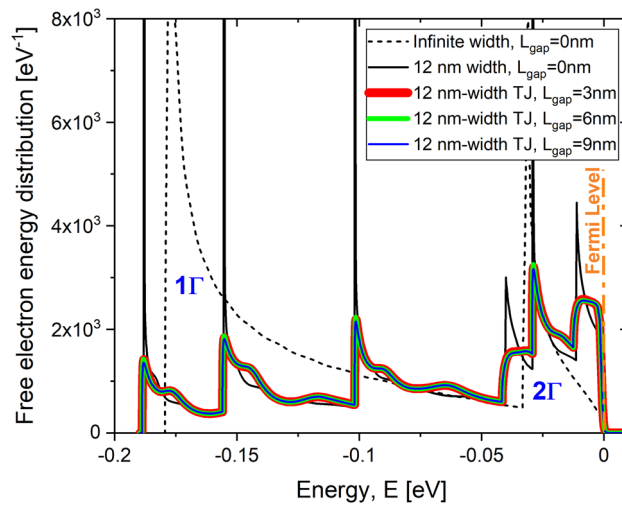


Figure 4. Free electron energy distribution for Si:P δ -layer systems. It shows a comparison of the free electron energy distribution for infinite and 12 nm-width δ -layer devices. The free electron energy distribution for the infinite-width δ -layer is normalized to a width of the device of $W = 12$ nm for comparison purpose. $N_D = 1.0 \times 10^{14} \text{ cm}^{-2}$, $N_A = 5.0 \times 10^{17} \text{ cm}^{-3}$ and $t = 1.0$ nm.

form a regular “phase” distribution of the current-carrying states (i.e. the states distributed closer to the center of the δ -layer along z -axis), while the modes $m = 2$, and $m = 3, 5$, etc. form “anti-phase” distributions (i.e. the states distributed further from the center of the δ -layer along z -axis) that have the maximum current being carried in the different regions of space, separated by a few nanometers. When $W \rightarrow \infty$, the number of propagation modes in y -direction becomes infinite $m \rightarrow \infty$ as expected. Fig. 3b shows the density of current-carrying electron states for the anti-phase case of δ -layer doping $N_D = 10^{14} \text{ cm}^{-2}$, thickness $t = 0.2$ nm (to approximate a monoatomic δ -layer) and device width $W = 9$ nm, which corresponds to $m = 5$. As final remark, the number of propagating modes m in δ -layer structures is mainly determined by three factors: (1) the δ -layer doping level N_D , (2) the the δ -layer doping thickness t and (3) the device width W .

Peculiarities of finite-width vs infinitely-wide devices. Next it is to get an insight into the transition from infinite-width to finite-width devices, before the quantum effects arise on the conductivity properties, i.e. the transition marked with a black arrow in the right upper part of Fig. 3. In the following, therefore, we will compare the free electron energy distribution, the transmission function and the current spectrum for devices of 12 nm-width, in which the size quantum effects on the conductivity are minimum as was seen in Fig. 2, and the corresponding infinitely wide counterparts. All results of infinitely wide devices are scaled to an effective device width of 12 nm for comparison purposes.

Figure 4 shows the free electron energy distribution of δ -layer structures of 12 nm-width (in continuous lines), together with the result of infinitely wide δ -layer structure with $L_{\text{gap}} = 0$ nm (in dashed line), for different gap lengths L_{gap} . The free electron energy distribution corresponds to the total density of states multiplied by the the Fermi-Dirac distribution function, which determines the probability of a state to be occupied. First we notice that in contrast to the infinite-width δ -layer systems, the finite size along the y -direction manifests itself in splitting of the occupied 1Γ sub-band, that is located around -0.18 eV, and 2Γ sub-band, that is located around -0.033 eV, into additional modes. Interestingly, for tunnel junctions ($L_{\text{gap}} > 0$), two of 2Γ modes are immediately dampened as manifested in the reduction of the corresponding peaks compared to the gapless case as shown in Fig. 4. In addition, as it is evident from Fig. 4, the free electron distribution in tunnel junctions is almost independent of the tunnel gap length L_{gap} .

Figure 5 shows the transmission function of infinite-width devices (in dashed lines) and 12 nm-width devices (in continuous lines) for different tunnel gap lengths L_{gap} . The transmission function provides the sum of the probabilities for each mode of a carrier at certain given energy E to carry current from the source to drain. As can be seen, the transmission function is fairly similar for both systems, infinite and finite width, especially for wider tunnel gaps. The transmission function is reduced exponentially for the low-energy modes with the length of the tunnel gap. Additionally, the transmission function strongly dependent on the gap length: the energy window in which the carrier can be transmitted is quickly reduced with the increase of the tunnel gap length.

Figure 6 shows the current spectrum, $i_e(E)$: in dashed lines, for infinite-width devices and, in continuous lines, for 12 nm-width devices. It is evident that the current spectrum of infinite-width devices differs from the current spectrum of finite-width counterparts. The finite size along y -axis significantly affects the current spectrum limiting the current-carrying states in all cases, including the gapless case, to only a narrow energy window around the Fermi level, with the size of the energy window being proportional to the applied voltage. The mechanism of such current spectra limitation for finite-width devices can be understood referring to the derivation in “From finite-width to infinite-width systems”: for finite-width devices, the sum in Eq. (41) contains only a limited number of quantized values of k_y , therefore the contribution of the corresponding propagation modes with low

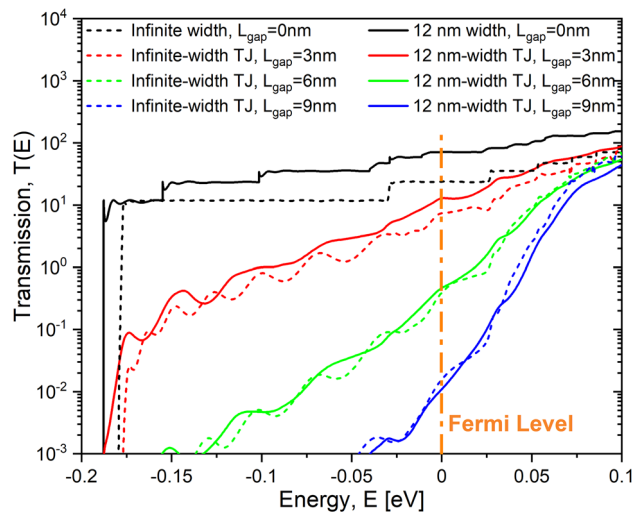


Figure 5. Electronic transmission for Si:P δ -layer systems. It shows a comparison of the transmission function (logarithmic scale) for an applied voltage of 1.0 mV for infinite and 12 nm-width δ -layer devices. The transmission function for the infinite-width δ -layer is normalized to a width of the device of $W = 12$ nm for comparison purpose. $N_D = 1.0 \times 10^{14} \text{ cm}^{-2}$, $N_A = 5.0 \times 10^{17} \text{ cm}^{-3}$ and $t = 1.0$ nm.

energies E_m is dampened, while only modes with the high energy (near Fermi level) can affect the current. The difference in the current spectra between finitely and infinitely-wide devices is diminished for very large tunnel junctions, as Fig. 6 indicates for $L_{\text{gap}} > 9$ nm. Indeed, for sufficiently large tunnel gaps, the tunneling current spectrum is exponentially suppressed for all low-energy modes, since the effective tunnel barrier height is larger for them than for higher-energy modes. Consequently, only the modes in the immediate vicinity of the Fermi level contribute to the total current for both finite and infinite-width tunnel junctions.

Despite the seemingly significant differences between finite-width systems and the corresponding infinitely wide counterparts, evidenced in the free electron energy distribution (see Fig. 4) and in their respective current spectra (see Fig. 6), the computed total currents, obtained by integrating the current spectra, only differ by about 20%. We thus conclude that while the density of states and the current spectra significantly differ, the sheet conductance (i.e. conductivity) calculations are accurate enough for finite-size systems at least down to 12 nm-widths. At the same time, for tunnel junctions with sufficiently large gaps, both the total current and the current spectra are very similar between infinite-width and finite-width systems, since in both cases the current is carried only by the high-energy electrons in the vicinity of system's Fermi level.

Quantization effects in δ -layer tunnel junctions. The tunneling current vs the tunnel gap length L_{gap} is shown in Fig. 7 for three different voltages $V = 1$ mV, 10 mV and 100 mV. In the range of gap lengths $L_{\text{gap}} = 0, \dots, 7$ nm, the I vs L_{gap} trend is practically exponential for all three voltages, i.e. $\ln I \sim \ln I_{L_{\text{gap}}=0} - L_{\text{gap}}/B_{\text{voltage}}$, where $I_{L_{\text{gap}}=0}$ is the current when $L_{\text{gap}} = 0$, L_{gap} is the tunnel gap length and B_{voltage} is a proportional constant related with the barrier height. Indeed, in this range of gap values we get $B_{1-10 \text{ mV}} \approx 1.9$ nm and $B_{100 \text{ mV}} = 2.9$ nm, which can be translated into the effective barrier heights of an equivalent rectangular barrier ($L_{\text{gap}} \times h_{\text{voltage}}$), where $h_{\text{voltage}} = d_{\text{val}} \hbar^2 / (8B_{\text{voltage}}^2 m_t)$, \hbar is the reduced Planck's constant, d_{val} is the valley degeneracy, and m_t is the transverse effective electron mass. Substituting the corresponding B_{voltage} values, we obtain $h_{1-10 \text{ mV}} \approx 80$ meV and $h_{100 \text{ mV}} \approx 30$ meV, which agrees well with the effective barrier height shown in Fig. 8 for a tunnel gap of $L_{\text{gap}} = 7$ nm, i.e. the barrier height from the Fermi level to the maximum value of the effective 1D electrostatic potential, computed using our charge self-consistent scheme. However, a deviation from the exponential trend can be seen for large tunnel gaps, $L_{\text{gap}} > 7$ nm, and it is the consequence of the quantization of the low-energy conduction band¹⁶ and the resulting mismatch between occupied and unoccupied quasi-discrete electron states in the left and right δ -layers, respectively, at low bias. Figure 8 shows the LDOS along x-direction for a tunnel junction of $L_{\text{gap}} = 7$ nm, i.e. the available states which can be occupied by the free electrons in space-energy dimension; in the low-temperature regime, only the states below the Fermi level are occupied. As one might observe from Fig. 8, the low-energy LDOS are quasi-quantized, highlighted with dashed lines in the figure, however, for high energies, the states are not quantized anymore, the LDOS are practically continuous in space-energy, as can be seen above the Fermi level for high energies. When the occupied quasi-discrete states from the left side overlap most with the unoccupied quasi-discrete states (corresponding to the ones above the Fermi level) from the right side, it results in a considerably increase of the tunneling current as shown in Fig. 7 for $L_{\text{gap}} = 10$ nm; on the contrary, if the overlap is reduced, as it happens for $L_{\text{gap}} = 11$ nm, there is a reduction of the current as the result of the mismatch. For low biases ($\lesssim 10$ mV), the mismatch can only exist for sufficiently large tunnel gaps, $L_{\text{gap}} > 7$ nm, because the existing coupling of the left and right δ -layer wave-functions for narrow tunnel gaps ($L_{\text{gap}} < 7$ nm) effectively equalizes the electron states on both sides, thus reducing this mismatch. Conversely, the quantization effect of the conduction band on the

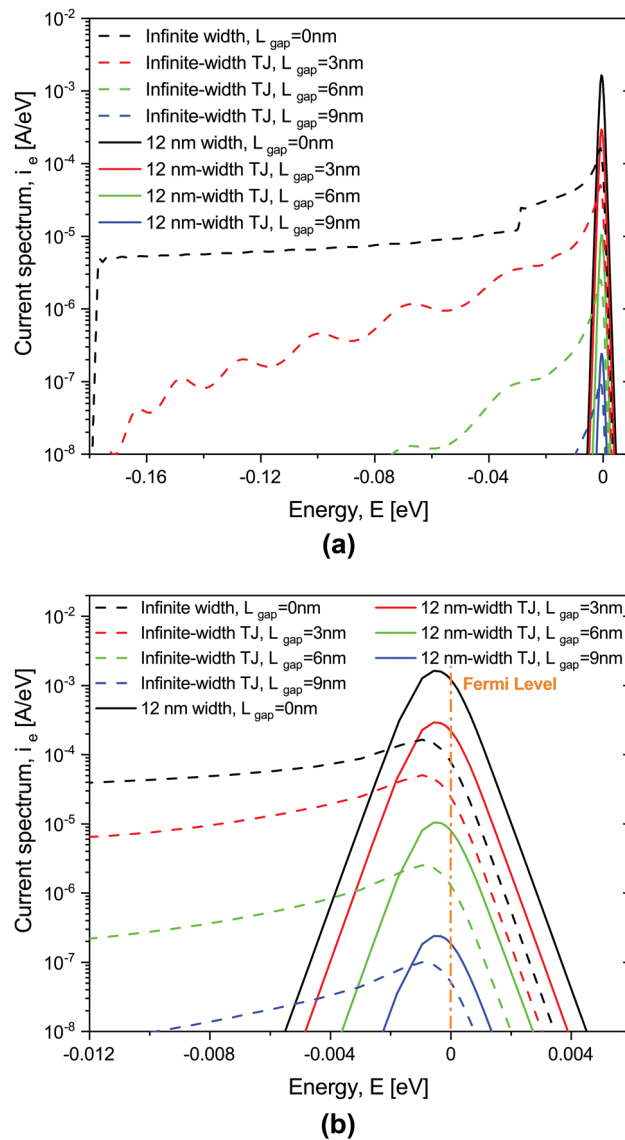


Figure 6. Current spectrum for Si:P δ -layer systems. (a) Comparison of the current spectrum i_e (logarithmic scale) for an applied voltage of 1.0 mV between infinite and 12 nm-width δ -layer devices. (b) Close-up of the current spectrum within the Fermi level. The currents for the infinite-width δ -layer systems are normalized to a width of $W = 12$ nm for comparison purpose. $N_D = 1.0 \times 10^{14} \text{ cm}^{-2}$, $N_A = 5.0 \times 10^{17} \text{ cm}^{-3}$ and $t = 1.0$ nm.

conductivity vanishes for high applied voltages ($\gtrsim 100$ mV) as shown in Fig. 7 for an bias of 100 mV. When a high bias is applied, e.g. to the right side (or drain) of the tunnel junction, it effectively makes the unoccupied high-energy continuous states of the right side available for the tunneling from the left side, thus negating the effects of the quantization of the occupied left side states on the current, as can be seen in Fig. 8b.

Conclusions. In this work, we have used an efficient open-system quantum-mechanical treatment to explore the conductive band structure and the conductive properties in Si:P δ -layers for device widths ranging from nano-scale to macro-scale dimensions, and analyze the influence of size quantization effects on the conductive properties for sub-12 nm device widths. For device widths $W < 10$ nm, quantization effects strongly affect not only the conductivity, but also the spatial distribution of the current-carrying electron states. Conversely, for $W > 10$ nm, the quantization effects practically vanish and the conductivity tends to the values of infinitely-wide devices. Additionally, we have revealed and discussed the mechanism of two conductivity regimes in δ -layer tunnel junctions with $L_{gap} > 7$ nm: a low-voltage regime, where conduction band quantization effects play a very significant role, and a high-voltage regime, where the quantum effects on the current are negligible.

Finally, we point out that the strong spacial quantization of the current-carrying states could be utilized in novel electronic δ -layer switches, where the number of propagating modes and their match/mis-match could be controlled by external electric fields, thus strongly affecting the current. In regular δ -layer conductors the

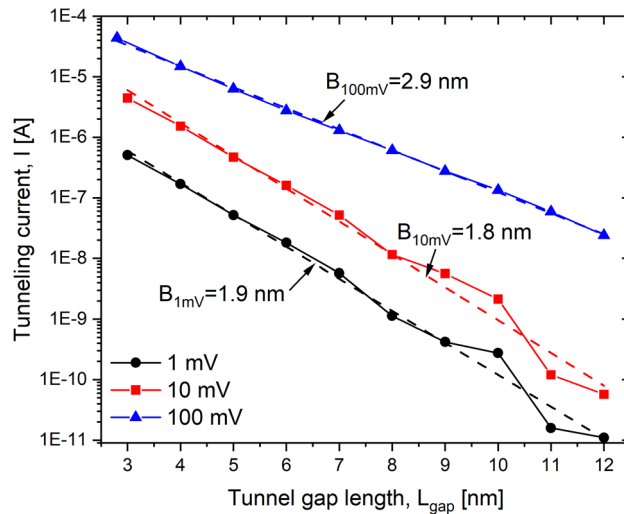


Figure 7. Characteristic tunneling current curves. The total current (logarithmic scale) I vs. tunnel gap length L_{gap} for three different applied voltages of $V = 1$ mV, 10 mV and 100 mV is shown. $N_D = 1.0 \times 10^{14} \text{ cm}^{-2}$ and $N_A = 5.0 \times 10^{17} \text{ cm}^{-3}$, $W = 12$ nm and $t = 1$ nm. Dashed lines represent least-square fits to the exponential trend.

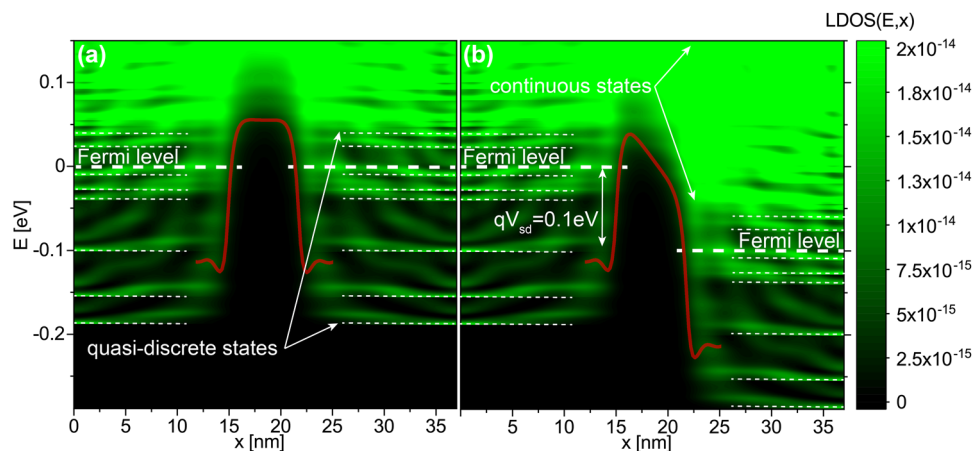


Figure 8. Local density of states for δ -layer tunnel junctions. The $LDOS(E, x)$ for a tunnel junction of $L_{\text{gap}} = 7$ nm is shown in (a) and (b) when a voltage of 1 mV and 100 mV is applied to the right side of the tunnel junction, respectively. Subfigure a indicates that the quantized states around the Fermi level might affect the conductivity for low applied biases due to a possible mismatch between the left and right quasi-discrete states, i.e., when the corresponding quasi-discrete peaks from the left and right sides are not aligned. Subfigure (b) shows that for high applied voltages the mismatch becomes impossible due to the availability of unoccupied continuum states on the right side for tunneling from the left side. In (a) and (b), the corresponding effective 1D potentials are also shown, calculated by integrating over the (y, z) -plane the actual charge self-consistent 3D potentials weighted with the electron density. $N_D = 1.0 \times 10^{14} \text{ cm}^{-2}$, $N_A = 5.0 \times 10^{17} \text{ cm}^{-3}$, $W = 12$ nm and $t = 1$ nm.

particular distribution of current-carrying states directly affects their penetration depth into Si body and cap, which typically has a large concentration of impurities (see e.g.³). Thus, the control of the number of propagating modes may give an additional degree of control over the rate of impurity scattering.

Methods

Open-system treatment. The open-system framework used in this work is based on an application of Keldysh formalism²⁷, known as Non-Equilibrium Green Function (NEGF)²⁸ (“Non-equilibrium green’s function formalism”), and the effective mass theory (“Device effective-mass Hamiltonian”). The NEGF formalism, described in “Non-equilibrium green’s function formalism”, defines the Green’s function matrix through the inverse $G(E) = [1 - H_0 - \Sigma(E)]^{-1}$, which allows to compute the electron density and current through a charge self-consistent scheme as described in “Charge self-consistency”. However, the computation of an inverse matrix

generally also scales as $O(N^3)$, with N being the size of the device Hamiltonian matrix H_0 (“**Device effective-mass Hamiltonian**”). For instance, for a device of dimensions $12 \text{ nm} \times 15 \text{ nm} \times 50 \text{ nm}$ with a uniform grid of 0.2 nm , $N > 10^6$ and $N^3 > 10^{18}$, making a direct computation of the inverse very challenging. Here we have employed an efficient implementation of NEGF, referred to as the Contact Block Reduction (CBR) method, which allows accurate computation of all quantum-mechanical quantities of interest (the local density of states, transmission probability, current) and scales linearly $O(N)$ with the system size. The CBR method in 3D real-space is summarized in “**Contact block reduction method**” and “**Incomplete set of CBR eigenstates**”.

Non-equilibrium Green’s function formalism. The current for two-contact device (D) from source (s) to drain (d) can be computed within the Landauer–Buttiker formalism through the transmission function

$$I_{sd} = \int i_e(E) dE = \frac{2e}{h} \int T_{sd}(E) (f_s(E) - f_d(E)) dE, \quad (1)$$

where $i_e(E)$ is the current spectrum, e is the electron charge, h is the Planck’s constant, E is the energy, $f_{s(d)}(E) = f(E + qV_{s(d)})$ is the Fermi–Dirac distribution function within source (drain), to which a voltage $V_{s(d)}$ is applied, and $T_{sd}(E)$ is the electronic transmission from source to drain. Within the Green’s function formalism²⁸, the transmission function is given by

$$T_{sd}(E) = \text{Tr}(\Gamma_s \mathbf{G}_D \Gamma_d \mathbf{G}_D^\dagger) \quad (2)$$

with

$$\Gamma_{s(d)} = i(\Sigma_{s(d)} - \Sigma_{s(d)}^\dagger) \quad (3)$$

where $\Gamma_{s(d)}$ are the coupling ($N_D \times N_D$)-matrices between the device and the source (drain), $\Sigma_{s(d)}$ are the self-energy ($N_D \times N_D$)-matrices, which describe the effects of the source (drain) on the electronic structure of the device by providing the appropriate open-system boundary conditions, \mathbf{G}_D and \mathbf{G}_D^\dagger are the retarded and advanced Green’s function ($N_D \times N_D$)-matrices of the coupled device with the source and drain (open-system device), respectively. N_D is the total number of grid points of the discretized device domain. The retarded Green’s function matrix is defined by

$$\mathbf{G}_D = [E^+ - \mathbf{H}_D^0 - \Sigma]^{-1} = [E^+ - \mathbf{H}_D^0 - \Sigma_s - \Sigma_d]^{-1}, \quad (4)$$

where $\mathbf{H}_D^0 + \Sigma_s + \Sigma_d$ is the non-Hermitian effective Hamiltonian ($N_D \times N_D$)-matrix of the open device. Eq. (4) can be also expressed in terms of the Green’s function of the decoupled device (i.e. the closed system) $\mathbf{G}_D^0 = [E^+ - \mathbf{H}_D^0]^{-1}$, where $E^+ = I(E + i\varepsilon)$, with $\varepsilon \rightarrow 0+$, via the Dyson equation:

$$\mathbf{G}_D = [I - \mathbf{G}_D^0 \Sigma]^{-1} \mathbf{G}_D^0. \quad (5)$$

The decoupled device Green function \mathbf{G}_D can be computed using its spectral representation:

$$\mathbf{G}_D^0 = [E^+ - \mathbf{H}_D^0]^{-1} = \sum_{\alpha} \frac{|\psi_{\alpha}\rangle\langle\psi_{\alpha}|}{E^+ - E_{\alpha}}, \quad (6)$$

and E_{α} and $|\psi_{\alpha}\rangle$ are the eigenvalues and eigenvectors of $\mathbf{H}_D^0 \psi_{\alpha} = E_{\alpha} \psi_{\alpha}$. The electron density $n(\mathbf{r}_i)$ is given by

$$\begin{aligned} n(\mathbf{r}_i) &= \int_{-\infty}^{\infty} LDOS(\mathbf{r}_i, E) dE \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} \langle \mathbf{r}_i | \mathbf{G}_D \Gamma_s \mathbf{G}_D^\dagger | \mathbf{r}_i \rangle f_s(E) dE \\ &\quad + \int_{-\infty}^{\infty} \frac{1}{2\pi} \langle \mathbf{r}_i | \mathbf{G}_D \Gamma_d \mathbf{G}_D^\dagger | \mathbf{r}_i \rangle f_d(E) dE. \end{aligned}$$

where $LDOS(\mathbf{r}_i, E)$ is the local density of states. Note that the dimension of \mathbf{G}_D and Γ_{λ} is ($N_D \times N_D$).

Contact block reduction method. We next review the Contact Block Reduction (CBR) method presented in Mamaluy et al.^{34–36}, which allows a very efficient calculation of the density matrix, transmission function, etc. of an arbitrarily shaped, multiterminal two- or three-dimensional open device within the NEGF formalism. In the following we apply the CBR method to the two-contact device shown in Fig. 9. The device consists of two semi-infinite contacts, source (s) and drain (d), which are in contact to the engineering channel, named as device (D).

We first discretize the domain of the device in N_D grid-points, and subdivide them into $N_C = N_{C_s} + N_{C_d}$ boundary grid-points and into $N_{D_i} = N_D - N_C$ interior grid-points. $N_{C_{s(d)}}$ corresponds to the boundary grid-points between the device and source (drain). Furthermore, we assume that the real-space Hamiltonian matrix that corresponds to this discretization only couples sites within some finite range with one another, typically first nearest-neighbors. The total grid-points in the device domain is described by the following set

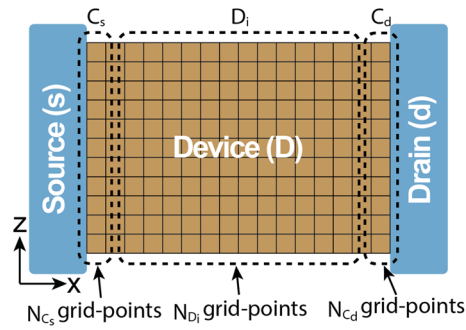


Figure 9. Schematic computational model for the CBR method. The two-contacts model is composed of two semi-infinite contacts, source (s) and drain (d), and a channel or device (D).

$$\begin{aligned} \Omega &= \{C \cup \{\text{interior of device}\}\} = \{C_s \cup C_d \cup D_i\} \\ &= \{\{1, \dots, N_{C_s}\} \cup \{N_{C_s} + 1, \dots, N_{C_s} + N_{C_d}\} \cup \{N_C + 1, \dots, N_D\}\} \end{aligned} \tag{7}$$

With this discretization of the device domain, the self-energy matrices $\Sigma_{s(d)}$, which represent the coupling of the source (drain) to the device, is given by

$$\Sigma_{s(d)}(i, j) = \begin{cases} [W_{Ds(d)} G_{s(d)}^0 W_{Ds(d)}^\dagger]_{ij}, & i, j \in C_{s(d)} \\ 0, & i, j \notin C_{s(d)} \end{cases} \tag{8}$$

where $W_{Ds(d)}$ are the Hamiltonian coupling matrices between the device D and contact $s(d)$ and $G_{s(d)}^0$ are the retarded Green's function matrix of decoupled source (drain). The total self-energy matrix ($N_D \times N_D$)- Σ can be expressed by the following block-diagonal matrix of the form

$$\Sigma = \Sigma_s + \Sigma_d = \begin{pmatrix} \Sigma_{C_s} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{C_d} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{9}$$

where

$$\Sigma_C = \begin{pmatrix} \Sigma_{C_s} & \mathbf{0} \\ \mathbf{0} & \Sigma_{C_d} \end{pmatrix}, \tag{10}$$

is of dimension of ($N_C \times N_C$). Analogously, the retarded Green's function matrix for the device can be also rewritten by the following submatrices

$$\mathbf{G}_D = \begin{pmatrix} \mathbf{G}_C & \mathbf{G}_{CD_i} \\ \mathbf{G}_{D_iC} & \mathbf{G}_{D_i} \end{pmatrix}, \tag{11}$$

where \mathbf{G}_C is of dimension of ($N_C \times N_C$), \mathbf{G}_{D_i} is of dimension of ($N_{D_i} \times N_{D_i}$) and \mathbf{G}_{CD_i} is of dimension of ($N_C \times N_{D_i}$).

Using the new representation for the above matrices, the Dyson equation can be then rewritten as

$$\mathbf{G}_D = \mathbf{A}^{-1} \mathbf{G}_D^0 = [\mathbf{I} - \mathbf{G}_D^0 \Sigma]^{-1} \mathbf{G}_D^0, \tag{12}$$

where the matrices have the following forms:

$$\mathbf{G}_D^0 = \begin{pmatrix} \mathbf{G}_C^0 & \mathbf{G}_{CD_i}^0 \\ \mathbf{G}_{D_iC}^0 & \mathbf{G}_{D_i}^0 \end{pmatrix} \tag{13}$$

and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_C & \mathbf{0} \\ \mathbf{R} & \mathbf{1} \end{pmatrix}. \tag{14}$$

The submatrices $\mathbf{A}_C = \mathbf{1} - \mathbf{G}_C^0 \Sigma_C$ and \mathbf{G}_C^0 are of dimension of ($N_C \times N_C$), the submatrices $\mathbf{R} = -\mathbf{G}_{D_iC}^0 \Sigma_C$ and $\mathbf{G}_{D_i}^0$ are of dimension of ($N_{D_i} \times N_C$), and $\mathbf{G}_{CD_i}^0$ is of dimension of ($N_C \times N_{D_i}$). Thus, the retarded Green's function of the open system is given by

$$\begin{aligned} \mathbf{G}_D &= \begin{pmatrix} \mathbf{G}_C & \mathbf{G}_{CD_i} \\ \mathbf{G}_{D_iC} & \mathbf{G}_{D_i} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_C^{-1} \mathbf{G}_C^0 & \mathbf{A}_C^{-1} \mathbf{G}_{CD_i}^0 \\ -\mathbf{R} \mathbf{A}_C^{-1} \mathbf{G}_C^0 + \mathbf{G}_{D_iC}^0 & -\mathbf{R} \mathbf{A}_C^{-1} \mathbf{G}_{CD_i}^0 + \mathbf{G}_{D_i}^0 \end{pmatrix} \end{aligned}$$

and the Green's function within the contact region by

$$G_C = A_C^{-1} G_C^0 = [I - G_C^0 \Sigma_C]^{-1} G_C^0. \tag{15}$$

The electron transmission from source s to drain d can be rewritten as

$$T_{sd}(E) = \text{Tr}(\Gamma_{C_s} G_C \Gamma_{C_d} G_C^\dagger) \tag{16}$$

where

$$\Gamma_{C_{s(d)}} = i(\Sigma_{C_{s(d)}} - \Sigma_{C_{s(d)}}^\dagger). \tag{17}$$

We note that the dimension of all submatrices involved in Eqs. (16) and (17) is of $(N_C \times N_C)$.

Similarly, the electron density can also be written as

$$n(\mathbf{r}_i) = \sum_{\alpha, \beta} \int_{-\infty}^{\infty} \left(\langle \mathbf{r}_i | \alpha \rangle \langle \beta | \mathbf{r}_i \rangle \Xi_{\alpha\beta}^s(E) f_s(E) + \langle \mathbf{r}_i | \alpha \rangle \langle \beta | \mathbf{r}_i \rangle \Xi_{\alpha\beta}^d(E) f_d(E) \right) dE$$

where the density matrix due to source(drain), $\Xi_{\alpha, \beta}^{s(d)}(E)$, is given in the closed-system basis (indexed by α, β)

$$\Xi_{\alpha, \beta}^{s(d)}(E) = \frac{1}{2\pi} \frac{\text{Tr} \left[[|\beta\rangle\langle\alpha|]_C \mathbf{B}_C^{-1} \Gamma_{C_{s(d)}} \mathbf{B}_C^{-1\dagger} \right]}{(E^+ - E_\alpha)(E^- - E_\beta)}, \tag{18}$$

with

$$E^\pm = E \pm i\varepsilon, \varepsilon \rightarrow 0+, \tag{19}$$

and

$$\mathbf{B}_C = \mathbf{1}_C - \Sigma_C G_C^0, \tag{20}$$

where $\mathbf{1}_C$ is the identity matrix of dimension of $(N_C \times N_C)$. Note that the dimension of Σ_C and G_C^0 is $(N_C \times N_C)$ as well.

Incomplete set of CBR eigenstates. The major feature of the CBR method is the ability to use a greatly reduced, incomplete set of specially defined eigenstates to represent the true open-system solution. As was shown in Mamaluy et al.³⁴, it can be accomplished by imposing a special kind boundary condition to the decouple device. The idea is to be able to rewrite the Green's function matrix of the open system $G_D(E)$, Eq. (4), as

$$G_D(E) = \left(E^+ - H_D^0 - \Sigma_s - \Sigma_d \right)^{-1} = \left(E^+ - H_D^N - \Sigma_s^N(E) - \Sigma_d^N(E) \right)^{-1} \tag{21}$$

where H_D^N is independent of the energy, and $\Sigma_s^N(E)$ and $\Sigma_d^N(E)$ tend toward zero for values of E that lies not far from the band edge. This enables us to solve the Dyson equation with an incomplete basis.

We start assuming that all nonzero coupling matrix elements of $W_{s(d)D}$ is equal to a real constant value, $W_{s(d)}$. This leads to the following expression for the self-energy matrix $\Sigma_{s(d)}$ within each source(drain) contact, $s(d)$, see e.g. Datta²⁸

$$\Sigma_{s(d)} = -W_{s(d)} \sum_m^{M_T} \chi_{s(d)}^m \exp(ik_{s(d)}^m a_{s(d)}) (\chi_{s(d)}^m)^\dagger \tag{22}$$

where $a_{s(d)} = a$ is the constant parameter of the lattice, M_T is the total number of propagating modes, and $\chi_{s(d)}^m$ are the modes of the Schrödinger equation $H_{s(d)}^0 \chi_{s(d)}^m = \varepsilon_{s(d)}^m \chi_{s(d)}^m$. The wave vectors $k_{s(d)}^m$ are functions of energy E as

$$E = \varepsilon_{s(d)}^m + 2W_{s(d)}(1 - \cos(k_{s(d)}^m a)) \tag{23}$$

The exponential term in Eq. (22) can be approximated in power of $k_{s(d)}^m$ as

$$-W_{s(d)} \exp(ik_{s(d)}^m(E)a) = -W_{s(d)} - iW_{s(d)} a k_{s(d)}^m(E) + \dots = -W_{s(d)} + F_{s(d)}(k_{s(d)}^m(E)), \tag{24}$$

which leads to

$$\Sigma_{s(d)}(E) = -W_{s(d)}\delta_{ij} + \sum_m \langle i | \chi_{s(d)}^m \rangle F(k_{s(d)}^m(E)) \langle \chi_{s(d)}^m | j \rangle, i, j \in C_{s(d)}. \tag{25}$$

Thus, the self-energy matrix in each contact, s and d , consists of a first term, which is an independent-energy diagonal matrix, and a second term, that vanishes not too far from the low energy band edge. The total self-energy matrix can be then rewritten as

$$\Sigma_s(E) + \Sigma_d(E) = -K + \Sigma^N(E), \tag{26}$$

where

$$K = \begin{cases} W_{s(d)}\delta_{ij}, & i, j \in C_{s(d)} \\ 0, & i, j \notin C_{s(d)}, \end{cases} \tag{27}$$

and the retarded Green's function matrix as

$$\begin{aligned} G_D(E) &= [E^+ - H_D^0 - \Sigma_s(E) - \Sigma_d(E)]^{-1} \\ &= [E^+ - H_D^0 + K - \Sigma^N(E)]^{-1} \\ &= [E^+ - H_D^N - \Sigma^N(E)]^{-1}. \end{aligned} \tag{28}$$

The self-energy $\Sigma^N(E)$ is small for E close to the band edge, and the Hamiltonian matrix of the device is defined by $H_D^N = H_D^0 - K$. It can be shown that this Hamiltonian matrix corresponds to generalized Neumann boundary conditions for its eigenfunctions³⁶. Thus, the eigenfunctions of H_D^N are approximate solutions of the open-boundary problem in the low-energy limit. As a consequence it suffices to include an incomplete set in the spectral representation of H_D^N in the calculation of the open-system quantities of interest within some limited energy range.

Device effective-mass Hamiltonian. The effective mass Schrödinger equation for Γ -valley electrons in Si is given by

$$H_D^0(\mathbf{r}_i)\psi_\alpha(\mathbf{r}_i) = E_\alpha\psi_\alpha(\mathbf{r}_i), \tag{29}$$

where H_D^0 is the effective-mass Hamiltonian operator

$$\begin{aligned} H_D^0(\mathbf{r}_i) &= -\frac{\hbar^2}{2m_t} \frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m_t} \frac{\partial^2}{\partial y^2} - \frac{\hbar^2}{2m_l} \frac{\partial^2}{\partial z^2} \\ &+ \phi^H(\mathbf{r}_i) + \phi^{XC}(\mathbf{r}_i), \end{aligned} \tag{30}$$

and $\phi^H(\mathbf{r}_i)$ is the Hartree potential and $\phi^{XC}(\mathbf{r}_i)$ is the exchange-correlation potential.

We discretize the 3D domain in N_x equidistant grid-points along the x-axis, in N_y equidistant grid-points along the y-axis and in N_z equidistant grid-points along the z-axis. For all directions, the separation between grid-points is chosen to be $\Delta x = \Delta y = \Delta z = 0.2$ nm. A grid-point is defined by the following triple indices (i, j, k) , with $i = 1, \dots, N_x$, $j = 1, \dots, N_y$ and $k = 1, \dots, N_z$, and the global index $n = k + N_z(N_x(j - 1) + i - 1)$. Therefore, the effective mass Hamiltonian matrix for the decoupled device can be expressed with seven non-zero diagonals as follows

$$H_D^0 = \begin{pmatrix} d_1 & l_1 & & b_1 & & & g_1 \\ l_1 & d_2 & l_2 & & b_2 & & \ddots \\ & l_2 & \ddots & \ddots & & & g_n \\ & & \ddots & & b_n & & \ddots \\ b_1 & & & \ddots & \ddots & & \ddots \\ b_2 & & & \ddots & d_n & l_n & \\ & & & \ddots & l_n & \ddots & \ddots \\ g_1 & & & b_n & & \ddots & \\ & \ddots & & & & \ddots & \\ & & g_n & & & & \ddots \end{pmatrix}, \tag{31}$$

where

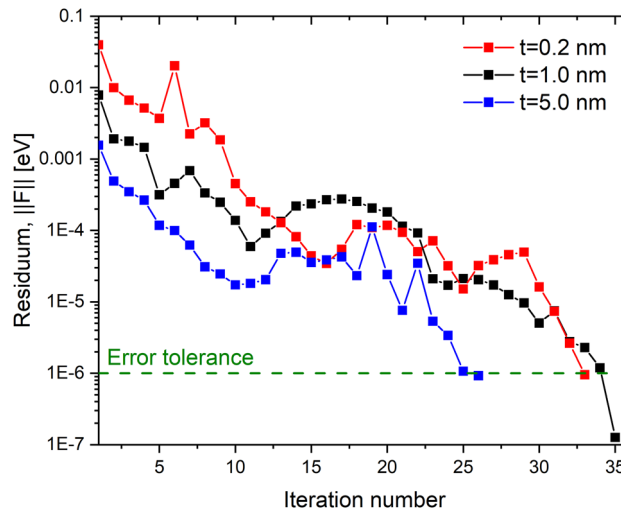


Figure 10. Convergence evaluation of the method. The convergence rate of the charge self-consistent open-system Schrodinger-Poisson equations for different δ -layer thicknesses (t) is shown.

$$d_n = \frac{\hbar^2}{m_1^{(i,j,k)} \Delta z^2} + \frac{\hbar^2}{m_t^{(i,j,k)} \Delta y^2} + \frac{\hbar^2}{m_t^{(i,j,k)} \Delta x^2}, \text{ with } n = 1, \dots, N_x N_y N_z, \tag{32}$$

$$l_n = \left(\frac{1}{m_1^{(i,j,k)}} + \frac{1}{m_1^{(i,j,k-1)}} \right) \frac{-\hbar^2}{4\Delta z^2}, \text{ with } n = 1, \dots, N_x N_y N_z - 1, \tag{33}$$

$$b_n = \left(\frac{1}{m_t^{(i,j,k)}} + \frac{1}{m_t^{(i-1,j,k)}} \right) \frac{-\hbar^2}{4\Delta x^2}, \text{ with } n = 1, \dots, (N_x N_y - 1) N_z, \tag{34}$$

and

$$g_n = \left(\frac{1}{m_t^{(i,j,k)}} + \frac{1}{m_t^{(i,j-1,k)}} \right) \frac{-\hbar^2}{4\Delta y^2}, \text{ with } n = 1, \dots, (N_y - 1) N_x N_z, \tag{35}$$

The number of zero diagonals between the diagonals l and b is $(N_z - 2)$ and between the diagonals b and g is $(N_x - 1) N_z - 1$. The exchange-correlation potential $\phi^{XC}(\mathbf{r}_i)$ that accounts for electron-electron interaction is computed using the terms given in Perdew and Zunger⁴⁰.

Charge self-consistency. We solve charge self-consistently the open-system effective mass Schrödinger equation and the non-linear Poisson equation^{37,38,41}. We employ a single-band effective mass approximation with a valley degeneracy of $d_{val} = 6$. For the charge self-consistent solution of the non-linear Poisson equation we use a combination of the predictor-corrector approach and Anderson mixing scheme^{37,38}.

Firstly, the Schrödinger equation is solved in a specially defined (see generalized Neumann BC in “[Incomplete set of CBR eigenstates](#)”) closed-system basis taking into account the Hartree potential $\phi^H(\mathbf{r}_i)$ and the exchange and correlation potential $\phi^{XC}(\mathbf{r}_i)$. It is found that out of more than 1,000,000 eigenstates only about 700 ($< 0.1\%$) of lowest-energy states is needed (see “[Incomplete set of CBR eigenstates](#)”), which is generally determined by the material properties (e.g. doping level), but not the device size. Then, the LDOS of the open system, $\rho(\mathbf{r}_i, E)$, and the electron density, $n(\mathbf{r}_i)$, are computed using the CBR method. The potential and the carrier density are then used to calculate the residuum F of the Poisson equation

$$\|F[\phi^H(\mathbf{r}_i)]\| = \|A\phi^H(\mathbf{r}_i) - (n(\mathbf{r}_i) - N_D(\mathbf{r}_i) + N_A(\mathbf{r}_i))\|, \tag{36}$$

where A is the matrix derived from the discretization of the Poisson equation and N_D and N_A are the total donor and acceptor doping densities arrays, respectively. If the residuum is larger than a predetermined threshold ε , the Hartree potential is updated using the predictor-corrector method, together with the Anderson mixing scheme. Using the updated Hartree potential and the corresponding carrier density, the exchange-correlation is computed again for the next step, and an iteration of Schrodinger-Poisson is repeated until the convergence is reached with $\|F[\phi^H(\mathbf{r}_i)]\| < \varepsilon = 10^{-6}$ eV. A typical convergence rate for Si:P δ -layer structures of different thicknesses is shown in Fig. 10. It illustrates that the proposed charge self-consistent scheme has a robust convergence for different structures, providing the residuum error reduction of 5 orders of magnitude within 40 iterations. In all

simulation in this work the standard values of electron effective masses $m_l = 0.98 \times m_e$, $m_t = 0.19 \times m_e$ and the dielectric constant $\epsilon_{Si} = 11.7$ of Silicon were employed.

From finite-width to infinite-width systems. In the scenario that the width of a conducting structure along the y -direction is assumed to be infinite (see Fig. 9) with a flat electrostatic potential, we can write the solutions of the Schrödinger equation as the product of plane-waves along y -axis and the solutions of the 2D Schrödinger equation in the X - Z plane of the device.

$$\psi_\alpha(\mathbf{r}_i) = \frac{\exp(ik_y y)}{\sqrt{L_y}} \psi_m(x, z), \quad (37)$$

which is normalized to a length of L_y , and

$$E_\alpha = E_m + \frac{\hbar^2}{2m_t} k_y^2. \quad (38)$$

The electron density can be expressed by summing over all (generally - open-system) states α as

$$n(\mathbf{r}_i) = 2 \sum_\alpha |\psi_\alpha(\mathbf{r}_i)|^2 f(E_\alpha), \quad (39)$$

where $k_B = 8.617 \times 10^{-5} \text{ eV K}^{-1}$ is the Boltzmann constant, T is the temperature of the system and $f(E)$ is the Fermi-Dirac distribution function, which provides the probability that a state is occupied or unoccupied. The Fermi-Dirac distribution function is given by

$$f(E) = \frac{1}{1 + \exp[E/k_B T]}, \quad (40)$$

Using Eqs. (37), (38) and (39) we get

$$\begin{aligned} n(\mathbf{r}_i) &= 2 \sum_\alpha |\psi_\alpha(x, z)|^2 f(E_\alpha) \\ &= 2 \sum_m |\psi_m(x, z)|^2 \left(\frac{1}{L_y} \sum_{k_y} f\left(E_m + \frac{\hbar^2}{2m_t} k_y^2\right) \right) \\ &= 2 \sum_m |\psi_m(x, z)|^2 f^{2D}(E_m), \end{aligned} \quad (41)$$

where we denote the term in parenthesis as the effective 2D Fermi-Dirac distribution function $f^{2D}(E)$. Replacing the summation over k_y with an integral as $\frac{1}{L_y} \sum_{k_y} \rightarrow \int_{-\infty}^{\infty} \frac{dk_y}{2\pi}$, we can get the distribution function in an integral form as follows

$$\begin{aligned} f^{2D}(E) &= \frac{1}{L_y} \sum_{k_y} f\left(E + \frac{\hbar^2}{2m_t} k_y^2\right) \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} f\left(E + \frac{\hbar^2}{2m_t} k_y^2\right) dk_y \\ &= \int_0^{\infty} \frac{1}{\pi} f\left(E + \frac{\hbar^2}{2m_t} k_y^2\right) dk_y \\ &= \int_0^{\infty} \frac{1}{\pi} \frac{1}{1 + \exp\left[\frac{E}{k_B T} + \frac{\hbar^2 k_y^2}{2m_t k_B T}\right]} dk_y. \end{aligned} \quad (42)$$

Finally, performing a variable change $l = \frac{\hbar^2 k_y^2}{2m_t k_B T}$ we obtain

$$\begin{aligned} f^{2D}(E) &= \left(\frac{m_t k_B T}{2\pi \hbar^2}\right)^{1/2} \frac{1}{\sqrt{\pi}} \int_0^{\infty} \frac{1}{1 + \exp\left(\frac{E}{k_B T} + l\right)} \frac{1}{\sqrt{l}} dl \\ &= \left(\frac{m_t k_B T}{2\pi \hbar^2}\right)^{1/2} f_{-1/2}\left(-\frac{E}{k_B T}\right), \end{aligned} \quad (43)$$

where $f_{-1/2}(x)$ is the Fermi-Dirac Integral of the order of $-1/2$. It can be expressed, for instance, through the polylogarithm special function as $f_{-1/2}(x) = -\text{Li}_{1/2}(-\exp(x))^{42}$.

Figure 11a shows the Fermi-Dirac distribution function, Eq. (40), and the effective 2D Fermi-Dirac distribution function, Eq. (43), in equilibrium conditions and a temperature of 4.0 K. In both cases the occupied states exist only below the Fermi level, but the occupation rate for the effective 2D distributions increases as $\sim \sqrt{-E}$ for $E \rightarrow -\infty$. We next examine the behaviour of the current integral given in the Eq. (1) at low temperatures. In general, the difference in the distribution functions $f_s(E) - f_d(E) = f(E) - f(E + qV_{bias})$ determines the energy range of non-zero contributions to the total current. The difference between Fermi-Dirac

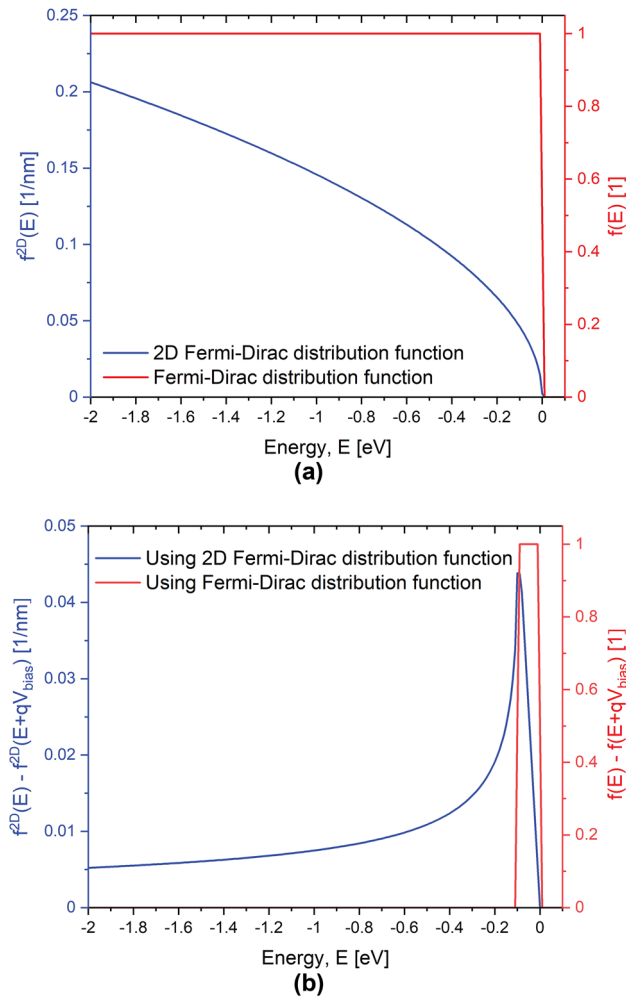


Figure 11. Fermi-Dirac distribution functions. **(a)** Comparison between the Fermi-Dirac distribution function, Eq. (40), and the effective 2D Fermi-Dirac distribution function, Eq. (43), for equilibrium conditions and $T = 4.0$ K. **(b)** Comparison between the Fermi-Dirac function difference between source and drain, $f_s(E) - f_d(E) = f(E) - f(E + qV_{bias})$, and the corresponding effective 2D Fermi-Dirac distribution function, $f_s^{2D}(E) - f_d^{2D}(E) = f^{2D}(E) - f^{2D}(E + qV_{bias})$, for an applied voltage of $V_{bias} = 0.1$ V and $T = 4.0$ K.

distributions is non-zero only in a small range of energies in the vicinity of the Fermi level as the result in red line in Fig. 11b indicates. Furthermore, the energy range for non-zero contribution is proportional to the applied voltage $V_{bias} = 0.1$ V. On the contrary, the difference in the effective 2D distribution functions is non-zero for all energies below the Fermi level as the plot in blue color in Fig. 11b illustrates. This is a reflection of the asymptotic behaviour of $f^{2D}(E)$ at $E \rightarrow -\infty$; for the difference in the distributions one gets $f_s^{2D}(E) - f_d^{2D}(E) = f^{2D}(E) - f^{2D}(E + qV_{bias}) \sim qV_{bias}/\sqrt{-E}$.

We thus conclude that, within the Landauer–Buttiker/NEGF formalism, finite size structures in the transverse directions have a principally different current spectrum from the structures that are infinite along a transverse direction (e.g. when $W \rightarrow \infty$ in δ -layer structures shown in Fig. 1).

Data availability

The data that support the plots within this paper are available from the corresponding authors upon reasonable request.

Received: 14 June 2022; Accepted: 8 September 2022

Published online: 30 September 2022

References

- Škřeň, T., Koester, S. A., Douhard, B., Fleischmann, C. & Fuhrer, A. Bipolar device fabrication using a scanning tunnelling microscope. *Nat. Electron.* **3**, 524. <https://doi.org/10.1038/s41928-020-0445-5> (2020).
- Bussmann, E. *et al.* Atomic-precision advanced manufacturing for Si quantum computing. *MRS Bull.* **46**, 607. <https://doi.org/10.1557/s43577-021-00139-8> (2021).
- Ward, D. *et al.* Atomic precision advanced manufacturing for digital electronics. *EDFAAO* **22**, 4 (2020).

4. Goh, K. E. J., Oberbeck, L., Simmons, M. Y., Hamilton, A. R. & Clark, R. G. Effect of encapsulation temperature on Si:P δ -doped layers. *Appl. Phys. Lett.* **85**, 4953. <https://doi.org/10.1063/1.1827940> (2004).
5. Goh, K. E. J., Oberbeck, L., Simmons, M. Y., Hamilton, A. R. & Butcher, M. J. Influence of doping density on electronic transport in degenerate Si:P δ -doped layers. *Phys. Rev. B* **73**, 035401. <https://doi.org/10.1103/PhysRevB.73.035401> (2006).
6. He, Y. *et al.* A two-qubit gate between phosphorus donor electrons in silicon. *Nature* **571**, 371. <https://doi.org/10.1038/s41586-019-1381-2> (2019).
7. Yang, C. *et al.* Operation of a silicon quantum processor unit cell above one kelvin. *Nature* **580**, 350. <https://doi.org/10.1038/s41586-020-2171-6> (2020).
8. Fuechsle, M. *et al.* A single-atom transistor. *Nat. Nanotech.* **7**, 242. <https://doi.org/10.1038/nnano.2012.21> (2012).
9. Škerenč, T. *et al.* CMOS platform for atomic-scale device fabrication. *Nanotechnology* **29**, 435302. <https://doi.org/10.1088/1361-6528/aad7ab> (2018).
10. Moore, S. K. Intel: Back on top by 2025?. *IEEE Spectrum* (2021), <https://spectrum.ieee.org/intel-says-its-manufacturing-tech-will-lead-the-world-by-2025>.
11. Goh, K. E. J. & Simmons, M. Y. Impact of Si growth rate on coherent electron transport in Si:P delta-doped devices. *Appl. Phys. Lett.* **95**, 142104. <https://doi.org/10.1063/1.3245313> (2009).
12. Reusch, T. C. G. *et al.* Morphology and electrical conduction of Si:P δ -doped layers on vicinal Si(001). *J. Appl. Phys.* **104**, 066104. <https://doi.org/10.1063/1.2977750> (2008).
13. McKibbin, S. R., Polley, C. M., Scappucci, G., Keizer, J. G. & Simmons, M. Y. Low resistivity, super-saturation phosphorus-in-silicon monolayer doping. *Appl. Phys. Lett.* **104**, 123502. <https://doi.org/10.1063/1.4869111> (2014).
14. Drumm, D. W., Hollenberg, L. C. L., Simmons, M. Y. & Friesen, M. Effective mass theory of monolayer δ doping in the high-density limit. *Phys. Rev. B* **85**, 155419. <https://doi.org/10.1103/PhysRevB.85.155419> (2012).
15. Mendez, J. P., Mamaluy, D., Gao, X., Anderson, E. M., Campbell, D. M., Ivie, J. A., Lu, T.-M., Schmucker, S. W., & Misra, S. Quantum Transport in Si:P δ -Layer Wires. in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)* (2020) pp. 181–184.
16. Mamaluy, D., Mendez, J. P., Gao, X. & Misra, S. Revealing quantum effects in highly conductive δ -layer systems. *Commun. Phys.* **4**, 205. <https://doi.org/10.1038/s42005-021-00705-1> (2021).
17. Mendez, J. P., Mamaluy, D., Gao, X. & Misra, S. Quantum Transport Simulations for Si:P δ -layer Tunnel Junctions. in *2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)* (2021).
18. Lee, S. *et al.* Electronic structure of realistically extended atomistically resolved disordered Si:P δ -doped layers. *Phys. Rev. B* **84**, 205309. <https://doi.org/10.1103/PhysRevB.84.205309> (2011).
19. Ryu, H. *et al.* “Atomistic modeling of metallic nanowires in silicon”. *Nanoscale* **5**, 8666. <https://doi.org/10.1039/C3NR01796F> (2013).
20. Smith, J. S., Cole, J. H. & Russo, S. P. Electronic properties of δ -doped Si:P and Ge:P layers in the high-density limit using a Thomas-Fermi method. *Phys. Rev. B* **89**, 035306. <https://doi.org/10.1103/PhysRevB.89.035306> (2014).
21. Smith, J. S. *et al.* Electronic transport in Si:P δ -doped wires. *Phys. Rev. B* **92**, 235420. <https://doi.org/10.1103/PhysRevB.92.235420> (2015).
22. Dusko, A., Koiller, B. & Lewenkopf, C. Disordered Si:P nanostructures as switches and wires for nanodevices. *Phys. Rev. B* **99**, 205422. <https://doi.org/10.1103/PhysRevB.99.205422> (2019).
23. Carter, D. J., Warschkow, O., Marks, N. A. & McKenzie, D. R. Electronic structure models of phosphorus δ -doped silicon. *Phys. Rev. B* **79**, 033204. <https://doi.org/10.1103/PhysRevB.79.033204> (2009).
24. Carter, D. J., Marks, N. A., Warschkow, O. & McKenzie, D. R. Phosphorus δ -doped silicon: mixed-atom pseudopotentials and dopant disorder effects. *Nanotechnology* **22**, 065701. <https://doi.org/10.1088/0957-4484/22/6/065701> (2011).
25. Drumm, D. W., Budi, A., Per, M. C., Russo, S. P. & Hollenberg, L. C. Ab initio calculation of valley splitting in monolayer δ -doped phosphorus in silicon. *Nanoscale Res. Lett.* **8**, 111 (2013).
26. Hwang, E. H. & Das Sarma, S. Electronic transport in two-dimensional Si: P δ -doped layers. *Phys. Rev. B* **87**, 125411. <https://doi.org/10.1103/PhysRevB.87.125411> (2013).
27. Keldysh, L. V. Diagram Technique for Nonequilibrium Processes. *Sov. Phys. J. Exp. Theor. Phys.* **20**, 1018 (1965).
28. Datta, S. *Electronic transport in mesoscopic systems* (Cambridge university press, 1997).
29. Holt, A. J. *et al.* Observation and origin of the Δ manifold in Si:P δ layers. *Phys. Rev. B* **101**, 121402. <https://doi.org/10.1103/PhysRevB.101.121402> (2020).
30. Mazzola, F., Polley, C. M., Miwa, J. A., Simmons, M. Y. & Wells, J. W. Disentangling phonon and impurity interactions in δ -doped Si(001). *Appl. Phys. Lett.* **104**, 173108. <https://doi.org/10.1063/1.4874651> (2014).
31. Mendez, J. P., Misra, S. & Mamaluy, D. Influence of imperfections on tunneling rate in δ -layer junctions (Manuscript submitted for publication) (2022).
32. Wang, J., Rahman, A., Ghosh, A., Klimeck, G. & Lundstrom, M. On the validity of the parabolic effective-mass approximation for the I-V calculation of silicon nanowire transistors. *IEEE Trans. Electron Devices* **52**, 1589. <https://doi.org/10.1109/TED.2005.850945> (2005).
33. Neophytou, N., Paul, A., Lundstrom, M. S. & Klimeck, G. Simulations of nanowire transistors: atomistic vs. effective mass models. *J. Comput. Electron.* **7**, 363. <https://doi.org/10.1007/s10825-008-0188-4> (2008).
34. Mamaluy, D., Sabathil, M. & Vogl, P. Efficient method for the calculation of ballistic quantum transport. *J. Appl. Phys.* **93**, 4628. <https://doi.org/10.1063/1.1560567> (2003).
35. Mamaluy, D., Mannargudi, A., Vasileska, D., Sabathil, M. & Vogl, P. Contact block reduction method and its application to a 10 nm MOSFET device. *Semicond. Sci. Technol.* **19**, S118. <https://doi.org/10.1088/0268-1242/19/4/042> (2004).
36. Mamaluy, D., Vasileska, D., Sabathil, M., Zibold, T. & Vogl, P. Contact block reduction method for ballistic transport and carrier densities of open nanostructures. *Phys. Rev. B* **71**, 245321. <https://doi.org/10.1103/PhysRevB.71.245321> (2005).
37. Khan, H. R., Mamaluy, D. & Vasileska, D. Quantum Transport Simulation of Experimentally Fabricated Nano-FinFET. *IEEE T. Electron Dev.* **54**, 784. <https://doi.org/10.1109/TED.2007.892353> (2007).
38. Gao, X. *et al.* Efficient self-consistent quantum transport simulator for quantum devices. *J. Appl. Phys.* **115**, 133707. <https://doi.org/10.1063/1.4870288> (2014).
39. van Wees, B. J. *et al.* Quantized conductance of point contacts in a two-dimensional electron gas. *Phys. Rev. Lett.* **60**, 848. <https://doi.org/10.1103/PhysRevLett.60.848> (1988).
40. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048. <https://doi.org/10.1103/PhysRevB.23.5048> (1981).
41. Mamaluy, D. & Gao, X. The fundamental downscaling limit of field effect transistors. *Appl. Phys. Lett.* **106**, 193503. <https://doi.org/10.1063/1.4919871> (2015).
42. <https://dlmf.nist.gov/25.12>

Acknowledgements

This work is funded under Laboratory Directed Research and development (LDRD) program, Project No. 227155, at Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell

International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. This article has been authored by employees of National Technology and Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employees own all right, title and interest in and to the article and are solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan <https://www.energy.gov/downloads/doe-public-access-plan>.

Author contributions

J.P.M. and D.M. performed equally the central calculations, analysis and discussion presented in this work. The manuscript was written by all the authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.P.M. or D.M.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022