Quantum Transport Simulations for Si:P δ-layer Tunnel Junctions

Juan P. Mendez
Cognitive & Emerging Computing
Sandia National Laboratories
Albuquerque, USA
jpmende@sandia.gov

Denis Mamaluy
Cognitive & Emerging Computing
Sandia National Laboratories
Albuquerque, USA
mamaluy@sandia.gov

Xujiao Gao
Electrical Models & Simulation
Sandia National Laboratories
Albuquerque, USA
xngao@sandia.gov

Shashank Misra
Multiscale Fab. Sci. & Tech. Dev.
Sandia National Laboratories
Albuquerque, USA
smisra@sandia.gov

***Abstract*—We present an efficient self-consistent implementation of the Non-Equilibrium Green Function formalism, based on the Contact Block Reduction method for fast numerical efficiency, and the predictor-corrector approach, together with the Anderson mixing scheme, for the self-consistent solution of the Poisson and Schrödinger equations. Then, we apply this quantum transport framework to investigate 2D horizontal Si:P δ-layer Tunnel Junctions. We find that the potential barrier height varies with the tunnel gap width and the applied bias and that the sign of a single charge impurity in the tunnel gap plays an important role in the electrical current.**

***Index Terms*—quantum transport, Si:P δ-layer tunnel junctions, contact block reduction, NEGF**

I. Introduction

The electronic structure and conductive properties of Si:P δ-layer systems have been a subject of intense experimental [1]–[3] and computational [4], [5] studies. In particular, tunnel junctions (TJ) in semiconductor δ-layer systems have raised a lot of interest due to their high potential to become one of the important building blocks of quantum [6] and classical (beyond-Moore) computing applications.

Previous theoretical approaches to investigate semiconductor δ-layer systems [4], [5] were based on the employment of periodic boundary conditions along the propagation direction, which become inapplicable in the case of TJs and any other more advanced (gated) devices. Additionally, in our previous work [7], [8], we have shown the need to extract the system's conductive properties from the quantum-mechanical flux and thus study highly conductive δ-layer systems from the first principles, without the use of semi-classical approximations for the current. The quantum-mechanical flux can be readily extracted using a Non-Equilibrium Green's Function (NEGF) treatment [8].

However, the NEGF simulations of δ-layer TJ systems possess two main challenges: 1) the convergence issues due to the existing sharp doping density in these systems; 2) the computational cost of the simulations for these systems. For the former one, we have successfully implemented a scheme, based on a predictor-corrector approach and the Anderson method [9], which considerably improve the convergence of the self-consistent solution of the non-linear Poisson equation and Schrödinger equation. For the latter one, we have used the Contact Block Reduction (CBR) method which reduces considerably the computational cost of the quantum transport calculations [10]–[13].

The aim of this work is to introduce this quantum transport framework and its application to δ-layer TJ systems. In Section II we present the efficient self-consistent quantum transport formalism [9]–[13]. In Section III we apply this framework to investigate Si:P δ-layer TJs. Finally, in Section IV we summarize the main findings.

II. Quantum Transport Framework

Our open-system quantum transport framework [9]–[13] relies on a self-consistent solution of Poisson-open system Schrödinger equation in the effective mass approximation and the Non-Equilibrium Green's Function (NEGF) formalism.

Within the NEGF formalism and taking into account only elastic scattering, the current $J_{\lambda\lambda'}$ from lead λ to λ' is computed from the Landauer formula

$$J_{\lambda\lambda'} = \frac{2e}{h}\int T_{\lambda\lambda'}(E)(f_\lambda(E) - f_{\lambda'}(E))dE, \qquad (1)$$

where e is the electron charge, h is the Planck's constant, E is the energy, $f_\lambda(E) = f(E - E_F - qV_\lambda)$ is the Fermi-Dirac

This work is funded under Laboratory Directed Research and Development Grand Challenge (LDRD GC) program, Project No. 213017, 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.

978-1-6654-0685-7/21/$31.00 ©2021 IEEE

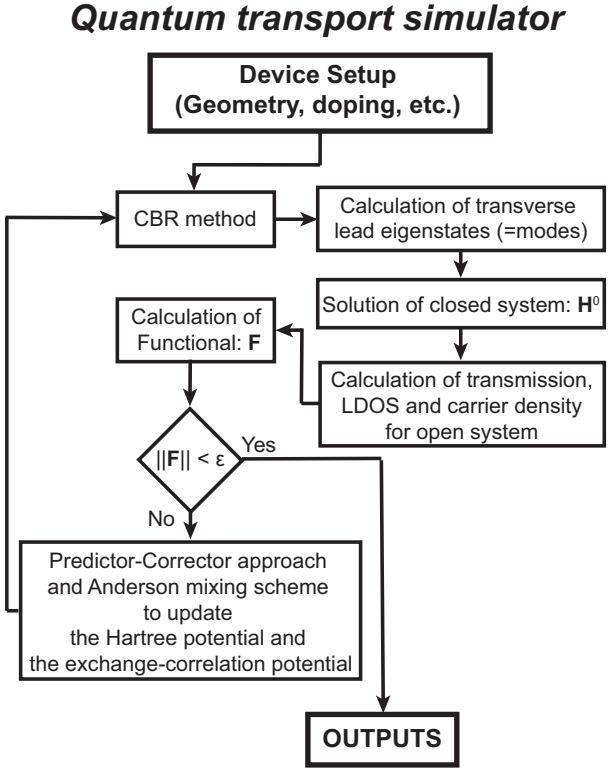


Fig. 1. Flow chart of the self-consistent quantum transport method [9]–[13].

function within the leads, V_λ is the applied voltage to the lead, E_F is the Fermi level and $T_{\lambda\lambda'}$ is the electronic transmission from lead λ to λ'. The transmission function is giving by

$$T_{\lambda\lambda'}(E) = \mathrm{Tr}(\mathbf{\Gamma}_\lambda \boldsymbol{G}_D \mathbf{\Gamma}_{\lambda'} \boldsymbol{G}_D^\dagger), \tag{2}$$

where $\mathbf{\Gamma}_\lambda = i(\mathbf{\Sigma}_\lambda - \mathbf{\Sigma}_\lambda^\dagger)$ are the coupling $(N_D \times N_D)$-matrices between the device and the leads, where N_D is the total grid-points of the discretized device domain, $\mathbf{\Sigma}_\lambda$ are the self-energy matrices of the leads and $\boldsymbol{G}_D$ and $\boldsymbol{G}_D^\dagger$ are the retarded and advanced Green's functions $(N_D \times N_D)$-matrices of the coupled device with the leads (open-system device), respectively. The retarded Green's function matrix can be computed using the Dyson equation

$$\boldsymbol{G}_D = [\boldsymbol{I} - \boldsymbol{G}_D^0 \mathbf{\Sigma}]^{-1} \boldsymbol{G}_D^0, \tag{3}$$

with

$$\mathbf{\Sigma} = \sum_{\lambda=\lambda_1}^{\lambda_L} \mathbf{\Sigma}_\lambda, \tag{4}$$

and

$$\boldsymbol{G}_D^0 = [\boldsymbol{E}^+ - \boldsymbol{H}_D^0]^{-1} = \sum_\alpha \frac{|\boldsymbol{\phi}_\alpha\rangle\langle\boldsymbol{\phi}_\alpha|}{E^+ - E_\alpha}, \tag{5}$$

where E_α and $|\boldsymbol{\phi}\rangle$ are the eigenvalues and eigenvectors of $\boldsymbol{H}_D^0 \boldsymbol{\phi}_\alpha = E_\alpha \boldsymbol{\phi}_\alpha$ and $E^\pm = E \pm i\epsilon, \epsilon \to 0+$.

The electron density $n(\boldsymbol{r}_i)$ is defined as

$$n(\boldsymbol{r}_i) = \sum_{\lambda=\lambda_1}^{\lambda_L} \int_{-\infty}^{\infty} \frac{1}{2\pi} \langle \boldsymbol{r}_i | \boldsymbol{G}_D \mathbf{\Gamma}_\lambda \boldsymbol{G}_D^\dagger | \boldsymbol{r}_i \rangle f_\lambda(E) dE. \tag{6}$$

Notice that all matrices involved in the above operations are of size $(N_D \times N_D)$. Thus, for instance, the inversion matrix cost in Eq. (6) is of $O(N_D^3)$, and the calculation cost of the eigenstates of $\boldsymbol{H}_D^0$ in Eq. (6) is of $O(N_D N_e^2)$, where N_e is the number of calculated eigenstates. To reduce the computational cost of these intensive calculations, we utilize the Contact Block Reduction (CBR) method [9]–[13].

The CBR is an efficient method to calculate the electronic transmission function of an arbitrarily shaped, multi-terminal open device. Within this method, the N_D grid-points are subdivided into N_C boundary grid-points with the leads and N_{D_i} interior grid-points of the device domain $(N_D = N_C + N_{D_i},\ N_{D_i} >> N_C)$. 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. With this domain discretization, the self-energy matrix $\mathbf{\Sigma}$, the open-system device Hamiltonian $\boldsymbol{H}_D$, and the Green's function matrix of the open-system device $\boldsymbol{G}_D$ can be rewritten with the following submatrices

$$\boldsymbol{H}_D = \begin{pmatrix} \boldsymbol{H}_C & \boldsymbol{H}_{CD_i} \\ \boldsymbol{H}_{D_iC} & \boldsymbol{H}_{D_i} \end{pmatrix}, \tag{7}$$

$$\boldsymbol{G}_D = \begin{pmatrix} \boldsymbol{G}_C & \boldsymbol{G}_{CD_i} \\ \boldsymbol{G}_{D_iC} & \boldsymbol{G}_{D_i} \end{pmatrix}, \tag{8}$$

and

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \sum_{\lambda=\lambda_1}^{\lambda=\lambda_L} \mathbf{\Sigma}_{C_\lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{9}$$

where the size of the sub-matrices $\boldsymbol{H}_C$, $\boldsymbol{G}_C$, and $\mathbf{\Sigma}_C$ is $(N_C \times N_C)$, the size of the sub-matrices $\boldsymbol{H}_{CD_i}$ and $\boldsymbol{G}_{CD_i}$ is $(N_C \times N_{D_i})$, and the size of the submatrices $\boldsymbol{H}_{D_i}$ and $\boldsymbol{G}_{D_i}$ is $(N_{D_i} \times N_{D_i})$.

After some algebra and using the submatrices in Eqs. (7)-(9), the electronic transmission from lead λ to λ' can be computed as

$$T_{\lambda\lambda'}(E) = \mathrm{Tr}(\mathbf{\Gamma}_{C_\lambda} \boldsymbol{G}_C \mathbf{\Gamma}_{C_\lambda} \boldsymbol{G}_C^\dagger), \tag{10}$$

where

$$\mathbf{\Gamma}_{C_\lambda} = i(\mathbf{\Sigma}_{C_\lambda} - \mathbf{\Sigma}_{C_\lambda}^\dagger), \tag{11}$$

and the Dyson equation as

$$\boldsymbol{G}_C = [\boldsymbol{I} - \boldsymbol{G}_C^0 \mathbf{\Sigma}_C]^{-1} \boldsymbol{G}_C^0. \tag{12}$$

Similarly, the electron density $n(\boldsymbol{r}_i)$ can be computed as

$$n(\boldsymbol{r}_i) = \sum_{\lambda=\lambda_1}^{\lambda_L} \sum_{\alpha,\beta} \int_{-\infty}^{\infty} \langle \boldsymbol{r}_i | \beta \rangle \langle \alpha | \boldsymbol{r}_i \rangle \Xi_{\alpha,\beta}^\lambda(E) f_\lambda(E) dE, \tag{13}$$

where

$$\Xi_{\alpha,\beta}^\lambda(E) = \frac{1}{2\pi} \sum_{\alpha,\beta} \frac{\mathrm{Tr}\left[\left[|\beta\rangle\langle\alpha| \right]_C \boldsymbol{B}_C \mathbf{\Gamma}_C \boldsymbol{B}_C^{-1\dagger} \right]}{(E^+ - E_\alpha)(E^- - E_\beta)} \tag{14}$$

and

$$\boldsymbol{B}_C = \mathbf{1}_C - \mathbf{\Gamma}_C \boldsymbol{G}_C^0. \tag{15}$$

Importantly, all matrices involved in Eqs. (10)-(14) are now of size ($N_C \times N_C$), where $N_{D_i} >> N_C$, considerably reducing the computational cost of NEGF operations.

Furthermore, it has been shown in [10] that by imposing generalized Neumann boundary conditions to the decoupled device, an incomplete set of eigenstates is sufficient to represent the true open-system solution. The idea is to rewrite the total self-energy matrix as

$$\boldsymbol{\Sigma}(E) = \sum_{\lambda} \boldsymbol{\Sigma}_{\lambda}(E) = -\boldsymbol{K} + \boldsymbol{\Sigma}^{N}(E), \tag{16}$$

with

$$\boldsymbol{K} = \begin{cases} W_{\lambda}\delta_{ij}, & i,j \in C_{\lambda}, \lambda = \lambda_1, ..., \lambda_L \\ 0, & i,j \notin C_{\lambda}, \lambda = \lambda_1, ..., \lambda_L, \end{cases} \tag{17}$$

where W_{λ} is the coupling between i and j grid-points and C_{λ} represents the boundary grid space of lead λ. Then, the retarded Green's function matrix can be rewritten as

$$\begin{aligned} \boldsymbol{G}_D(E) =& (\boldsymbol{E}^{+} - \boldsymbol{H}_D^0 - \boldsymbol{\Sigma}(E))^{-1} \\ =& (\boldsymbol{E}^{+} - \boldsymbol{H}_D^0 - \boldsymbol{K} - \boldsymbol{\Sigma}^{N}(E))^{-1} \\ =& (\boldsymbol{E}^{+} - \boldsymbol{H}_D^N - \boldsymbol{\Sigma}^{N}(E))^{-1}, \end{aligned} \tag{18}$$

where $\boldsymbol{H}_D^N$ is independent of the energy and $\boldsymbol{\Sigma}^{N}(E)$ tends toward zero for values of E that lies not far from the band edge. This enables us to solve the Dyson equation with $\boldsymbol{H}_D^N$ instead of $\boldsymbol{H}_D^0$ and to use an incomplete set of eigenstates.

A summary of the algorithm implemented in our quantum transport simulator is shown in Fig. 1. The Schrödinger equation is solved for the closed systems taking into account the Hartree potential $\psi^{H}(\boldsymbol{r}_i)$ and the exchange and correlation potential $\psi^{XC}(\boldsymbol{r}_i)$. The electron charge density $n(\boldsymbol{r}_i)$ is computed by Eq. (13). The potential and the carrier density are then used to calculate the residuum F of the Poisson equation

$$\left|\left|\boldsymbol{F}[\psi^{H}(\boldsymbol{r}_i)]\right|\right| = \left|\left|\boldsymbol{A}\psi^{H}(\boldsymbol{r}_i) - \boldsymbol{n}(\boldsymbol{r}_i) + \boldsymbol{N}_D(\boldsymbol{r}_i) - \boldsymbol{N}_A(\boldsymbol{r}_i)\right|\right|, \tag{19}$$

where $\boldsymbol{A}$ is the matrix derived from the discretization of the Poisson equation and $\boldsymbol{N}_D$ and $\boldsymbol{N}_A$ 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 and the Anderson mixing scheme [9]. Using the updated Hartree potential and the corresponding carrier density, which is a function of the change in the Hartree potential, the exchange-correlation is computed again for the next step. Finally, we emphasize that the convergence of highly conductive systems is very challenging. Therefore, the implementation of the predictor-corrector method, together with the Anderson mixing scheme is crucial for these systems to improve the convergence.

III. Results and Discussion

We apply our quantum transport framework to investigate 2D Si:P δ-layer TJ systems. The 2D TJ model, shown in Fig. 2, consists of two semi-infinite contacts, source and drain, in close contact with the channel. The channel is composed of a very lightly doped Si body and Si cap with acceptors, which embed a P δ-layer with an intrinsic gap.

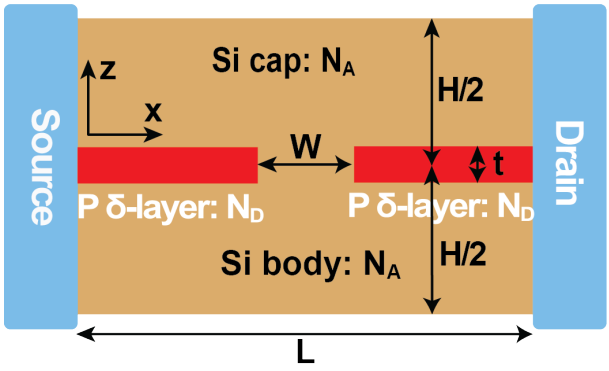


Fig. 2. The Si:P δ-layer TJ system is composed of a Si body, a P-δ layer with an tunnel gap, and a Si cap

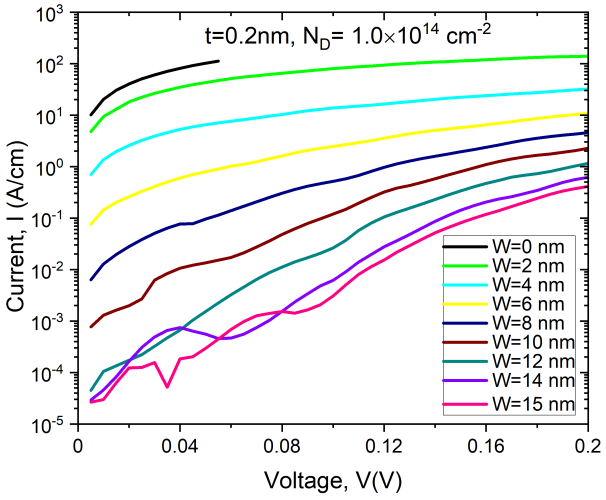
We first start analyzing the conductive properties for ideal TJs. Figure 3(a) shows the I-V curves for different tunnel gap widths, and Figure 3(b) shows the dependence of current on the tunnel gap width for different applied voltages. We observe that the slope in Fig. 3(b) depends not only on the gap width, but also on the applied voltage. This result indicates that the potential barrier height is not constant and varies with the tunnel gap width and the applied bias.

In our previous work for Si:P δ-layer wires [5,6], we showed that free electrons were distributed in energy in different sub-bands (e.g. 1Γ and 2Γ). For very confined δ-layers, the contribution of these sub-bands on the current was equal. However, our simulations reveal a different behavior for TJs: higher-energy electrons (i.e. electrons allocated in 2Γ sub-band) contribute more to the current as shown in Fig. 4. Additionally, the contribution of the lower-energy electrons decreases rapidly with the increase of the tunnel gap width. Indeed, this contribution is minimal for tunnel gap widths $W > 5$ nm, as seen in Fig. 4(b).

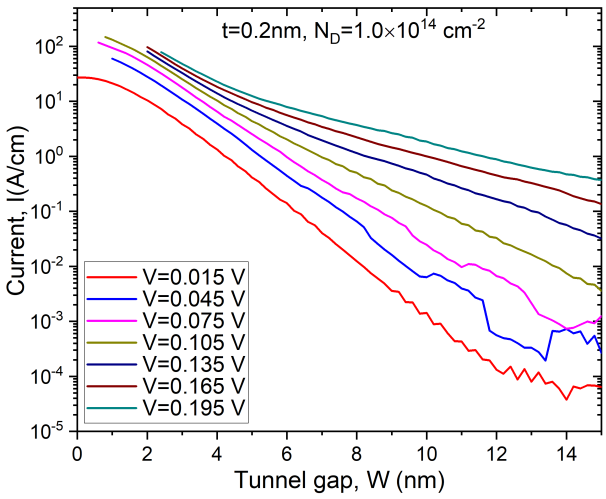
Finally, we investigate the effect of unintentional charged impurities in the tunnel gap (Fig. 5). Interestingly, our simulations predict a strong asymmetry with the impurity electrical sign. Positively charge impurities (e.g. P atoms) in the intrinsic semiconductor gap dramatically affect the current value, increasing the current by an order of magnitude for all considered values of W. At the same time, negatively charged impurities (B or Al atoms) reduce the current in a significantly smaller manner. However, for both cases we can conclude that impurities in the tunnel junction can dramatically alter the value of the tunneling current and hence cannot be ignored in realistic quantum and classical computing applications.

IV. Conclusion

We presented an efficient self-consistent implementation of the NEGF formalism, based on the CBR method, the predictor-corrector approach and the Anderson mixing scheme. We applied this framework to investigate horizontal Si:P δ-layer TJs. First, we characterized ideal TJs and revealed that the



(a)



(b)

Fig. 3. Characteristic curves for an ideal Si:P δ-layer TJ systems: (a) Current versus applied voltage for different tunnel gap widths (W); (b) Current versus tunnel gap widths for different applied voltages. $N_D = 10^{14}\text{cm}^{-2}$, $N_A = 10^{17}\text{cm}^{-3}$, $t = 0.2$ nm, $L = 50$ nm and $H = 40$ nm.

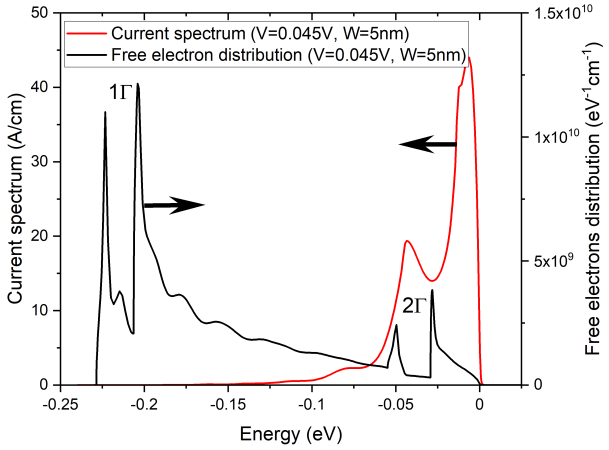
potential barrier height is not constant and varies with the tunnel gap width and the applied bias. Second, we studied the influence of unintentional charged impurities in the tunnel gap. Our simulations revealed that the sign of these impurities plays an important role in the electrical current: B-type impurities reduce slightly the current; P-type impurities increase considerably the current.

REFERENCES

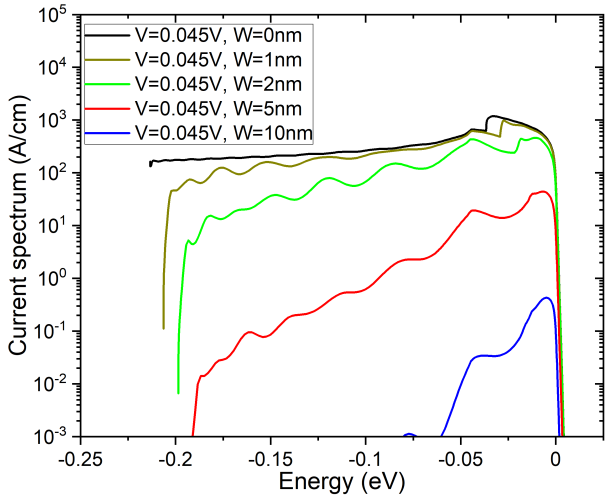
[1] J. G. Keizer, S. Koelling, P. M. Koenraad, and M. Y. Simmons, "Suppressing segregation in highly phosphorus doped silicon monolayers," *ACS Nano*, vol. 9, no. 12, pp. 12 537–12 541, 2015, pMID: 26568129. [Online]. Available: https://doi.org/10.1021/acsnano.5b06299

[2] F. Mazzola, C.-Y. Chen, R. Rahman, X.-G. Zhu, C. M. Polley, T. Balasubramanian, P. D. King, P. Hofmann, J. A. Miwa, and J. W. Wells, "The sub-band structure of atomically sharp dopant profiles in silicon," *npj Quantum Mater.*, vol. 5, no. 34, 2020.

[3] A. J. Holt, S. K. Mahatha, R.-M. Stan, F. S. Strand, T. Nyborg, D. Curcio, A. K. Schenk, S. P. Cooil, M. Bianchi, J. W. Wells,



(a)



(b)

Fig. 4. (a) Contribution of the sub-bands on the current for TJ systems; (b) Current spectrum for different tunnel gap widths (W). $N_D = 10^{14}\text{cm}^{-2}$, $N_A = 10^{17}\text{cm}^{-3}$, $t = 0.2$ nm, $L = 50$ nm and $H = 40$ nm.

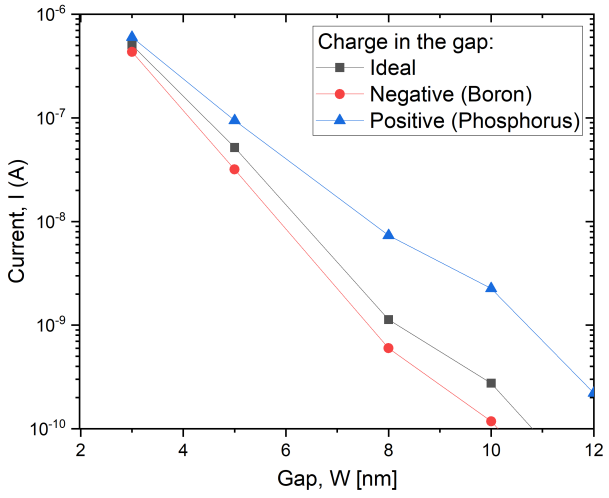


Fig. 5. Effect of charged impurities in the tunnel gap on the current for an applied bias of 1 mV. $N_D = 10^{14}$ cm^{-2}, $N_A = 10^{17}$ cm^{-3}, $t = 0.2$ nm, $L = 50$ nm and $H = 20$ nm.

P. Hofmann, and J. A. Miwa, "Observation and origin of the Δ manifold in Si:P δ layers," *Phys. Rev. B*, vol. 101, p. 121402, 2020.

[4] S. Lee, H. Ryu, H. Campbell, L. C. L. Hollenberg, M. Y. Simmons, and G. Klimeck, "Electronic structure of realistically extended atomistically resolved disordered Si:P δ-doped layers," *Phys. Rev. B*, vol. 84, p. 205309, 2011.

[5] D. J. Carter, O. Warschkow, N. A. Marks, and D. R. McKenzie, "Electronic structure models of phosphorus δ-doped silicon," *Phys. Rev. B*, vol. 79, p. 033204, 2009.

[6] M. G. House, E. Peretz, J. G. Keizer, S. J. Hile, and M. Y. Simmons, "Single-charge detection by an atomic precision tunnel junction," *Applied Physics Letters*, vol. 104, no. 11, p. 113111, 2014. [Online]. Available: https://doi.org/10.1063/1.4869032

[7] J. P. Mendez, D. Mamaluy, X. Gao, E. M. Anderson, D. M. Campbell, J. A. Ivie, T.-M. Lu, S. W. Schmucker, and S. Misra, "Quantum transport in si:p δ-layer wires," in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 181–184.

[8] D. Mamaluy, J. P. Mendez, X. Gao, and S. Misra, "Revealing quantum effects in highly conductive δ-layer systems," *Communications Physics (accepted for publication)*, 2021.

[9] X. Gao, D. Mamaluy, E. Nielsen, R. W. Young, A. Shirkhorshidian, M. P. Lilly, N. C. Bishop, M. S. Carroll, and R. P. Muller, "Efficient self-consistent quantum transport simulator for quantum devices," *J. Appl. Phys.*, vol. 115, no. 13, p. 133707, 2014.

[10] D. Mamaluy, M. Sabathil, and P. Vogl, "Efficient method for the calculation of ballistic quantum transport," *J. Appl. Phys.*, vol. 93, no. 8, pp. 4628–4633, 2003.

[11] D. Mamaluy, D. Vasileska, M. Sabathil, T. Zibold, and P. Vogl, "Contact block reduction method for ballistic transport and carrier densities of open nanostructures," *Phys. Rev. B*, vol. 71, p. 245321, 2005.

[12] H. R. Khan, D. Mamaluy, and D. Vasileska, "Quantum transport simulation of experimentally fabricated nano-finfet," *IEEE T. Electron Dev.*, vol. 54, no. 4, pp. 784–796, 2007.

[13] D. Mamaluy and X. Gao, "The fundamental downscaling limit of field effect transistors," *Applied Physics Letters*, vol. 106, no. 19, p. 193503, 2015. [Online]. Available: https://doi.org/10.1063/1.4919871