Analyzing Variability in Short-Channel Quantum Transport from Atomistic First Principles

Qing Shi* and Hong Guo
Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada

Yu Zhu and Lei Liu
Nanoacademic Technologies Inc., 7005 Boulevard Tachereau, Brossard, Quebec J4Z 1A7, Canada
(Received 23 March 2015; published 12 June 2015)

Effects of disorder scattering critically influence quantum-transport properties of nanostructures both fundamentally and practically. In this work, we report a theoretical analysis of the important issue of device-to-device quantum-transport variability (DDV) induced by random configurations of discrete dopants. Instead of calculating many impurity configurations by brute force, which is practically impossible to accomplish from first principles, here we use a state-of-the-art atomistic technique where the configurational average is carried out analytically, thereby, DDV can be predicted for any impurity concentration. The DDV we quantitatively analyze is the *off*-state tunnel conductance variability in Si nanosized field-effect transistor channels with channel lengths ranging from 6.5 to 15.2 nm doped with different concentrations of boron impurity atoms. The variability is predicted by varying the doping concentration, channel length, and the doping positions. We find that doping away from the source or drain contacts of the channel very significantly reduces variability, and doping close to the source or drain produces a nonintuitive outcome of increasing variability. The physics is understood by analyzing the microscopic details of the potential profile in the tunnel barrier. Finally, we organize the *ab initio* data by a Wentzel-Kramers-Brillouin model.

DOI: 10.1103/PhysRevApplied.3.064008

I. INTRODUCTION

An increasingly important piece of nanoelectronic device physics is the device-to-device variability (DDV) of transport properties due to atomistic disorder or the random discrete-dopant (RDD) effect [1]. When device size approaches tens of nanometers, DDV becomes significant because every individual configuration of dopant atoms makes the device property slightly different. DDV can cause serious issues in circuit design; namely, if every transistor behaves differently, it is hard to achieve a reliable circuit. For nanoscale transistors, DDV manifests in transport properties such as the conductance, the threshold voltage, subthreshold swing, and on:off ratio of the drain current [2–4]. RDD-induced DDV has already become a key problem in the integrated circuit of sub-50-nm technology [5,6] and is expected to be a critical issue in the sub-10-nm-scale nanoelectronics.

Theoretical investigation of DDV is very important for understanding its microscopic origin, thereby, rectifying it. To this end, the most widely applied approach is by solving classical, semiclassical, or effective mass quantum-transport equations [1,7–9]. These approaches rely on phenomenological material and device parameters, and for classical models, there are discrepancies between simulation results and experimental data [7] when the device size is approximately 10 nm. A more fundamental but less applied technique is based on first principles; namely, the device is analyzed atomistically, and the transport is calculated quantum mechanically [10–12]. However, a first-principles analysis of DDV often requires a prohibitively high computational cost. For instance, to atomistically model a dopant concentration of 10^{-4}, which is already extremely high for semiconductor devices, one will need to simulate 10 000 host atoms to just accommodate a single dopant atom, and there is a huge ensemble of dopant configurations to be calculated individually by brute force in order to determine DDV. Indeed, so far the first-principles method has been limited to simulate structures having a very limited number of atoms and unrealistically high dopant concentrations.

Clearly, to understand DDV, an approach that avoids brute-force computation of each and every dopant configuration and one that can deal with a wide range of impurity concentrations is required. To this end, we have recently reported a formalism of a nonequilibrium coherent potential approximation (NECPA) [13] based on which the RDD-induced DDV at the experimentally relevant impurity concentrations and device sizes can be calculated quantitatively from first principles parameter-free [14].

In this work, we apply this state-of-the-art first-principles formalism [14] to investigate the important

*qing.shi2@mail.mcgill.ca

device physics behind the DDV in a silicon nanosized field-effect transistor (nano-FET) induced by atomic impurities. In particular, we analyze the *off*-state tunnel conductance variability induced by RDD—a very significant device merit for Si nano-FET channels in the realistic ranges from 6.5 to 15.2 nm as a function of impurity concentration. In our calculations, both the host Si atoms and the impurity boron atoms are treated atomistically in equal footing, with the disorder average carried out by the NECPA formalism [13,14]. For these channels, the variability is found to be sensitive to the doping concentration C_d, channel length L, and the doping positions. Doping-position effects are investigated under the circumstance of localized doping [15], where dopant atoms are restricted within some spatial region. Variability is calculated by moving this narrow region from the source side to the drain side. While the dependence on C_d, L are as expected, the dependence on the doping position is quite nonintuitive. In particular, doping away from the source and drain contacts of the channel can very significantly reduce variability, but doping close to the source or drain has the nonintuitive outcome of increasing variability. The physics of this phenomenon can be understood by analyzing the microscopic details of the potential profile in the channel barrier. Finally, we organize the *ab initio* results by a Wentzel-Kramers-Brillouin (WKB) model.

The paper is organized as follows. In Sec. II, we briefly present the theoretical and computational technique for completeness and for ease of discussion. In Sec. III, we report the RDD-induced *off*-state tunnel conductance variability. The effects of various doping profiles on quantum transport are analyzed and microscopic physics revealed. Section IV presents a summary of the work.

II. METHOD

In this section, we summarize the theoretical and computational technique for predicting the RDD-induced variability. For device structures having no disorder, atomistic first-principles modeling of quantum transport is achieved by carrying out the density-functional theory (DFT) within the Keldysh nonequilibrium Green's function (NEGF) formalism, as first reported in Ref. [16] and now widely applied to analyze nonequilibrium quantum-transport problems in nanostructures. In the NEGF-DFT formalism, a DFT-like self-consistent field theory is used to calculate the Hamiltonian of the device model—an open structure consisting of a device scattering region sandwiched between electrodes extending to reservoirs at infinity where bias voltage is applied and electric current is collected [17]. With the Hamiltonian, the NEGF is calculated, which determines the nonequilibrium quantum-statistical information and the nonequilibrium density matrix of the device scattering region, which, in turn, provide inputs for calculating a new Hamiltonian. This process is repeated until self-consistency. Our work in this investigation builds on the NEGF-DFT formalism, and we refer interested readers to the original reference for further technical details [16].

When there are disorders such as impurity atoms in the device scattering region, one must perform a statistical average over the ensemble of the disorder configurations. For the NEGF-DFT formalism, the configurational average is achieved by the coherent potential approximation (CPA) at the single-particle Green's function and Hamiltonian level and by the nonequilibrium vertex correction (NVC) theory at the nonequilibrium density matrix level, as first reported in Ref. [18] and extended via the theory of NECPA in Ref. [13]. The essential physical ingredients of the CPA-NVC (and NECPA) approach are similar to the original equilibrium CPA [19,20]; namely, by solving a self-consistent coherent potential, one approximates the disordered system with an effective medium having translational invariance. The main difference is that CPA-NVC and NECPA work at nonequilibrium where the nonequilibrium statistical information plays an important role. We refer interested readers to the original papers [13,18] for further details of the CPA-NVC and NECPA theories.

Next, in order to predict the transport variability, one needs not only the average transmission coefficient but also its variance whose calculation involves the multiplication of four Green's functions in the NEGF-DFT formalism. The configuration average of the transmission variance gives rise to a more complicated vertex correction self-energy which must be evaluated. For atomistic modeling, we have recently reported [14] a theory based on NECPA to calculate the vertex correction self-energy involving four Green's functions so that multiple impurity scattering effects to the *variance* of the conductance can be predicted from atomic first principles parameter-free. In the following, we very briefly outline this method which is used in the rest of this work.

In the Landauer-Büttiker formulism of quantum transport [17], the energy (E)-resolved transmission coefficient is obtained from the retarded and advanced Green's functions $g^{r,a}$,

$$\overline{T(E)} = \mathrm{Tr}[\overline{g^r(E)\Gamma_L(E)g^a(E)\Gamma_R(E)}], \tag{1}$$

where the overbar indicates the random configuration average over the ensemble of impurity atoms. $\Gamma_{L(R)}$ are the linewidth functions of the left and right electrodes, respectively, and take the form $\Gamma_{L\,(R)} = i(\Sigma^r_{L\,(R)} - \Sigma^a_{L\,(R)})$. $\Sigma^{r,a}_{L\,(R)}$ are the retarded and advanced self-energies due to interactions of the left (L) and right (R) electrodes to the scattering region of the device, and i is the imaginary unit. In the subsequent numerical calculations of the Si nano-FET (see below), the impurity scattering occurs in the

device scattering region and not in the electrodes far from it. Therefore, self-energies $\Sigma_{L\,(R)}^{r,a}$ are nonrandom quantities and can be calculated by the standard transfer-matrix technique [16]. Therefore, the configuration average $\overline{T(E)}$ requires calculating only the average $\overline{g^r(E)\Gamma_L(E)g^a(E)}$ which involves the average over two Green's functions, and this average leads to the vertex correction self-energy which accounts for multiple impurity scattering, as discussed in Ref. [18]. The numerical computation for the configurationally averaged transmission function $\overline{T(E)}$ in Eq. (1) can be realized by atomistic first principles using the CPA-NVC theory [18] and/or the NECPA theory [13].

The DDV is quantified by the *variance* of the transmission function $T(E)$, namely,

$$dT \equiv \sqrt{\overline{T^2} - \overline{T}^2}, \tag{2}$$

whose evaluation requires the configurational average $\overline{T^2(E)}$. Clearly, since $\overline{T(E)}$ requires averaging over two Green's functions $\overline{g^r(E)\Gamma_L(E)g^a(E)}$, the calculation of $\overline{T^2(E)}$ involves averaging over four Green's functions of the type $\overline{T^2} \sim \overline{g^r\Gamma_L g^a\Gamma_R g^r\Gamma_L g^a\Gamma_R}$. This configurational average determines the physical effects of multiple impurity scattering to the transport variance through Eq. (2).

In our earlier work, Ref. [14], an *analytical* expression was deduced for evaluating $\overline{T^2}$ in terms of the configurationally average Green's functions $\overline{g^{r,a}}$ and various impurity scattering self-energies. While the derivation is complicated and tedious, the spirit is as follows. One starts from the scattering form of the total Green's function

$$g^r = \bar{g}^r + \sum_{\mathbf{R}} \bar{g}^r t^r_{\mathbf{R}} \bar{g}^r + \sum_{\mathbf{R}} \sum_{\mathbf{R}' \neq \mathbf{R}} \bar{g}^r t^r_{\mathbf{R}'} \bar{g}^r t^r_{\mathbf{R}} \bar{g}^r + \sum_{\mathbf{R}} \sum_{\mathbf{R}' \neq \mathbf{R}} \sum_{\mathbf{R}'' \neq \mathbf{R}'} \bar{g}^r t^r_{\mathbf{R}''} \bar{g}^r t^r_{\mathbf{R}'} \bar{g}^r t^r_{\mathbf{R}} \bar{g}^r + \cdots, \tag{3}$$

where $\mathbf{R}$ labels the atomic site, and $t^r_{\mathbf{R}}$ the retarded scattering matrix represents multiple impurity scattering at site $\mathbf{R}$. Within CPA (and NECPA), the $t^r_{\mathbf{R}}$'s are random variables which can take the value $t^{r,Q}_{\mathbf{R}}$ with probability of $c^Q_{\mathbf{R}}$ for the impurity species labeled by Q; $c^Q_{\mathbf{R}}$ is the probability of site $\mathbf{R}$ occupied by atom species Q and is generally related to the concentration of species Q. After inserting Eq. (3) and its advanced counterpart into the expression of $\overline{T^2}$ and noting that typically the concentration of defects and/or dopants is rather small, a compact form for dT of Eq. (2) can be obtained [14],

$$\begin{aligned}
(dT)^2 &= \sum_{\mathbf{R},Q>0} c^Q_{\mathbf{R}} (Y^{\alpha,Q}_{\mathbf{R}} + Y^{\beta,Q}_{\mathbf{R}} + Y^{\gamma,Q}_{\mathbf{R}})^2, \\
Y^{\alpha,Q}_{\mathbf{R}} &= \mathrm{Tr}\{t^{a,Q}_{\mathbf{R}} [\bar{g}^a \Gamma_R \bar{g}^r \Gamma_L \bar{g}^a]_{\mathbf{R},\mathbf{R}}\}, \\
Y^{\beta,Q}_{\mathbf{R}} &= \mathrm{Tr}\{t^{r,Q}_{\mathbf{R}} [\bar{g}^r \Gamma_R \bar{g}^a \Gamma_L \bar{g}^r]_{\mathbf{R},\mathbf{R}}\}, \\
Y^{\gamma,Q}_{\mathbf{R}} &= \mathrm{Tr}\{t^{r,Q}_{\mathbf{R}} [\bar{g}^r \Gamma_L \bar{g}^a]_{\mathbf{R},\mathbf{R}} t^{a,Q}_{\mathbf{R}} [\bar{g}^a \Gamma_R \bar{g}^r]_{\mathbf{R},\mathbf{R}}\},
\end{aligned} \tag{4}$$

where $Q = 0$ refers to the host atom species whose concentration is large compared to dopants or defects. This expression can be explicitly calculated by the CPA-NVC theory [18] and/or NECPA theory [13] since it involves only configurationally averaged single-particle Green's functions. We refer interested readers to the original Ref. [14] for details of the derivation.

In the rest of the paper, we apply Eq. (4) within the NEGF-DFT first-principles formalism to analyze RDD-induced DDV for Si nano-FET channels. It is worth mentioning that Eq. (2) or, equivalently, Eq. (4), enables us to determine DDV by a single self-consistent NEGF-DFT-NECPA calculation.

III. TRANSPORT VARIABILITY IN SILICON NANO-FET CHANNELS

The theory outlined in Sec. II makes a first-principles analysis of DDV possible, and in this section we investigate the conductance variability due to RDD effects in the technologically important Si nano-FET channels schematically shown in Fig. 1. Our goal is to predict the source-to-drain conductance variability in the *off*-state for channels ranging from 6.5 to 15.2 nm, doped with different concentrations of boron impurity atoms. Note that these length scales are technologically relevant for current and near-future FETs. *Off*-state conductance is an extremely important short-channel property which measures the source-to-drain tunnel leakage that plays an increasingly

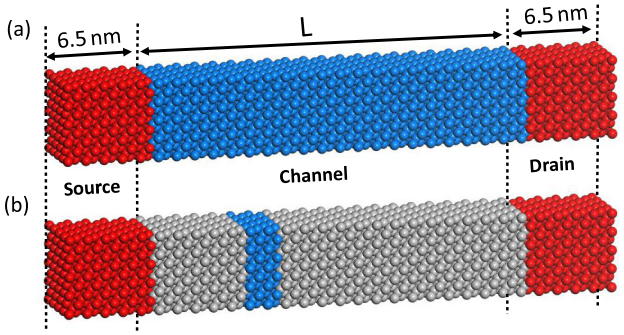

FIG. 1. Schematic of *n*-*p*-*n* Si nano-FET channel with length L and p doped by boron atoms. The left and right regions in red are the heavily doped source and drain contacts, respectively. (a) Uniform doping where dopants are randomly distributed in the whole channel indicated by blue. (b) Localized doping where dopants are randomly distributed in a narrow region indicated by blue, and the region in gray indicates intrinsic Si regions.

dominating role for power consumption of the device operation.

A. Device model

Figure 1 schematically shows the two-probe structure of the nano-FET where the heavily doped Si source and drain electrodes sandwich a Si channel of length L. In the NEGF-DFT-NECPA calculations, the cross section of the supercell is set to one unit cell of Si crystal, 5.43×5.43 Å^2, and is periodically extended in the transverse dimensions. We attach 6.5-nm buffer layers to both sides of the device to ensure a smooth matching of the potential at the simulation box edges. For L ranges from 6.5 to 15.2 nm, the number of atoms per supercell within the scattering region of the device varies from 288 to 416. In the NEGF-DFT framework, the device Hamiltonian $H_{\rm dev}$ takes the following form:

$$H_{\rm dev} = T + V_H + V_{\rm ext} + V_{\rm XC}, \tag{5}$$

where $T\,(=-\nabla^2/2)$ is the kinetic energy operator of the electron, V_H is the Hartree potential, $V_{\rm ext}$ the external potential, and $V_{\rm XC}$ the exchange-correlation potential. In this work, we employ the tight-binding linear muffin-tin orbitals (TB LMTO) as the basis [21–23]. The main advantage of using TB LMTO comes from two aspects: (i) the device Hamiltonian matrix becomes nearly orthogonal under such basis thereby greatly speeding up the computation [24]; (ii) it is compatible with the CPA formulism [24]. To describe the electronic properties of the valence electrons (four for silicon and three for boron), nine atomic orbitals (s, p, d) are used for each atom, localized within each muffin-tin sphere.

We consider two doping profiles: uniform doping where dopants are randomly distributed in the whole channel as depicted in Fig. 1(a) and localized doping where dopants are randomly distributed in a restricted spatial region as shown by the blue atoms in Fig. 1(b). Boron atoms are the dopants to the channel whose concentration C_d is related to the doping probability c^d of each atomic site as

$$N_c c^d = C_d V_c, \tag{6}$$

where N_c denotes the total number of atoms in the device channel and V_c the channel volume. c^d is identical to $c^Q_{\mathbf{R}}$ mentioned in the previous section, and index $\mathbf{R}$ is omitted for the sake of simplicity. The source and drain are uniformly and degenerately doped to a concentration of 5×10^{19} cm^{-3} using the technique of virtual crystal approximation [25].

For the semiconductor devices, it is crucial to correctly determine the band gap. To this end, we employ the modified Becke-Johnson semilocal exchange [26] in our NEGF-DFT calculation; for bulk Si, we obtain an indirect band gap of 1.11 eV in excellent agreement with the experimental value. We carry out the first-principles device modeling using the NANODSIM quantum-transport package [27] which implements the theory of DDV [13,14,18] described in Sec. II.

B. Variability of uniformly doped channels

The uniform doping concentration of boron atoms is fixed to 5×10^{18} cm^{-3}, and channel length L varies from 6.5 to 15.2 nm. For each L, after the self-consistent NEGF-DFT-NECPA calculation of the two-probe structure is converged, we calculate the average transmission coefficient $\bar{T}$ as well as its variability dT as discussed in Sec. II. The configuration-averaged conductance G and its variability dG are then obtained, $G = G_0\bar{T}$ and $dG = G_0 dT$, where $G_0 \equiv e^2/\pi\hbar$ is the conductance quanta with e the carrier charge and $\hbar$ the reduced Planck constant. Other numerical details can be found in Ref. [28].

Figure 2(a) plots G (black down triangles) and dG (blue up triangles) versus L. As expected, G reduces in some exponential fashion due to the tunneling physics consistent with the behavior of the potential along the channel shown in Fig. 2(b); namely, devices with longer L have a higher tunneling potential $V(z)$. The variability dG also has an exponential behavior with L, which exemplifies the significance of DDV when L is small.

Apparently, G and dG in Fig. 2(a) do not quite follow a simple e^{-L} behavior, which is true if the tunnel barrier $V(z)$ has no z dependence. For our Si channel, the calculated $V(z)$ is shown in Fig. 2(b) which has a strong z dependence. To abstract a general behavior from the *ab initio* data, we consider a WKB model of the transmission coefficient,

$$T \sim e^{-2\gamma}, \qquad \gamma = \frac{1}{\hbar}\int \sqrt{2m^*|V(z) - E_F|}\,dz, \tag{7}$$

where z is along the tunneling direction, $V(z)$ is the potential, m^* the effective mass, and the integration is limited to the region of $V(z) - E_F > 0$. We approximate $V(z)$ of Fig. 2(b) by a triangle whose height is the peak value of $V(z)$ and whose base is an *equivalent channel*

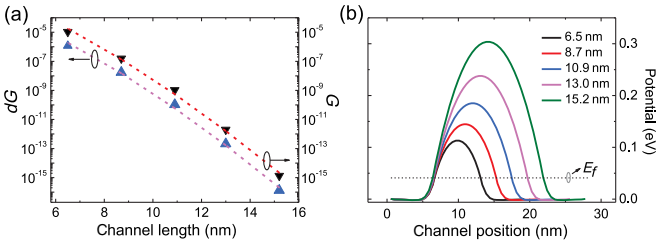

FIG. 2. (a) The calculated average conductance G and variability dG in units of G_0 per unit cell versus $L = 6.5$, 8.7, 10.9, 13.0, and 15.2 nm at doping concentration $C_d = 5 \times 10^{18}$ cm^{-3}. Symbols are *ab initio* data points and dashed lines are the WKB fitting. (b) The calculated potential along the channel (averaged over the cross section of the supercell). E_f indicates the Fermi level.

length equal to the width of $V(z)$ above the Fermi energy. Because the source and drain doping is much larger than that in the channel (by a factor of 10), the *equivalent channel width* is almost the same as L. In addition, as shown in Fig. 2(b), the height of $V(z)$ is also proportional to L. With these rough approximations, it is straightforward to derive the conductance from Eq. (7):

$$G = C_1 e^{-A_1 L^{\frac{3}{2}}}, \tag{8}$$

where C_1, A_1 are constants. As shown in Fig. 2(a), this WKB formula fits the *ab initio* data very well [29]. This exercise suggests that the G versus L curve in Fig. 2(a) is controlled by the potential shape such that $G \sim \exp(-L^{3/2})$.

Next, to organize the *ab initio* data for the variability dG by a WKB formula, we make the following ansatz:

$$dG \propto \sqrt{\left|\frac{\partial G}{\partial l}\right|_{\bar{l}}^2 + \left|\frac{\partial G}{\partial h}\right|_{\bar{h}}^2}, \tag{9}$$

where l and h are parameters characterizing $V(z)$ such as its width and height [30]. The letter with the overbar means evaluating at the average of l or h. The spirit of this ansatz is that due to RDD, the potential profiles of individual dopant configurations deviate from some "average potential," thereby, inducing a conductance variability dG. For uniform doping, as shown in Fig. 2(b), the potential height is proportional to L, i.e., $h \sim L$; hence, Eq. (9) reduces to only one variable l. Using Eq. (8), we arrive at

$$dG = B_1 L^{1/2} e^{-A_1 L^{3/2}} \tag{10}$$

for uniform doping, where B_1 is another constant. Figure 2(a) shows a very good agreement between Eq. (10) and the *ab initio* data [29], suggesting a scaling of $dG \sim \exp(-L^{3/2})$. Equation (10) also indicates a strong short-channel non-self-averaging effect; namely, the normalized quantity dG/G scales with $L^{1/2}$ for short channels.

It is interesting to investigate dG by varying the doping concentration C_d, as shown in Fig. 3(a). L is fixed at 10.9 nm, and the channel doping varies from 0.5 to 5×10^{19} cm^{-3}. Figure 3(b) plots the calculated potential $V(z)$, which gets higher and wider as the doping concentration increases, and, accordingly, dG drastically reduces [see Fig. 3(a)]. Again, we can organize the *ab initio* data by the WKB formula Eqs. (7) and (9). Approximating $V(z)$ by a triangle, we obtain

$$G = C_2 e^{-A_2 L H^{1/2}},$$
$$\ln(dG) = \frac{1}{2}\ln\left(H + \frac{L^2}{4H}\right) - A_2 L H^{\frac{1}{2}} + B_2, \tag{11}$$

where H is the height of $V(z)$, and A_2, B_2, C_2 are constants. These expressions fit the *ab initio* data very well as shown by the dotted lines in Fig. 3(a) [29]. Finally, the inset of Fig. 3(a) plots the normalized variability dG/G versus doping concentration C_d, showing the expected behavior that a smaller C_d generates weaker DDV.

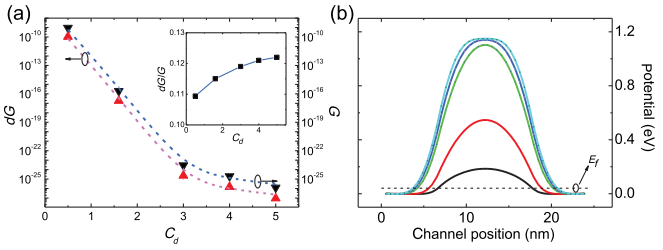

FIG. 3. Average conductance G and its variability dG versus channel-doping concentration C_d, with $L = 10.9$ nm. (a) Black and red triangles are *ab initio* data of G and dG (units G_0 per unit cell); red and purple dashed curves are the WKB fitting by Eq. (11). Unit of C_d is 1×10^{19} cm^{-3}. The inset plots dG/G versus C_d. (b) Calculated $V(z)$ along the channel with different C_d: from low to high $C_d = 0.5, 1, 3, 4$, and 5×10^{19} cm^{-3}. Black dashed line indicates the Fermi level.

C. Variability of locally doped channels

Having presented a DDV for uniformly doped channels, we now investigate dG for localized doping. Localized doping is interesting since certain doping profiles significantly reduce the source-to-drain *off*-state tunnel leakage [15]. Here, we expect dG to also reduce, since by definition, localized doping has less randomness in dopant positions. Similar to Ref. [15], we put dopant atoms (boron) randomly in a narrow region of 1.1 nm inside a channel of $L = 10.9$ nm, and dG is calculated by moving this narrow region from the source side to the drain side. To compare to uniform doping, the localized doping concentration is fixed by Eq. (6) to 5×10^{19} cm^{-3}.

Figure 4(a) plots G (blue down triangles) and dG (red up triangles) versus localized doping position. Figure 4(b) gives $V(z)$ along the channel including that of uniform doping for a comparison. In agreement with Ref. [15], localized doping generates higher potential barriers $V(z)$ than uniform doping unless the localized doping position is close to the source or drain. As an outcome of $V(z)$, dG is drastically reduced when the doping position is away from the source or drain as shown in Fig. 4(a). Figure 4(c) shows the normalized DDV dG/G: the variability has a relatively large value of approximately 12.5% of the tunnel conductance when the doping position nears the source or drain, and it is reduced to approximately 6% when the doping position is in the middle of the channel.

Again, we may organize the *ab initio* data by our WKB approximation. The curves in Fig. 4(b) indicate that $V(z)$ above the Fermi level can be approximated by triangles with a nearly identical base but different heights. For each triangle [to fit each $V(z)$], its vertex is at the midpoint of the

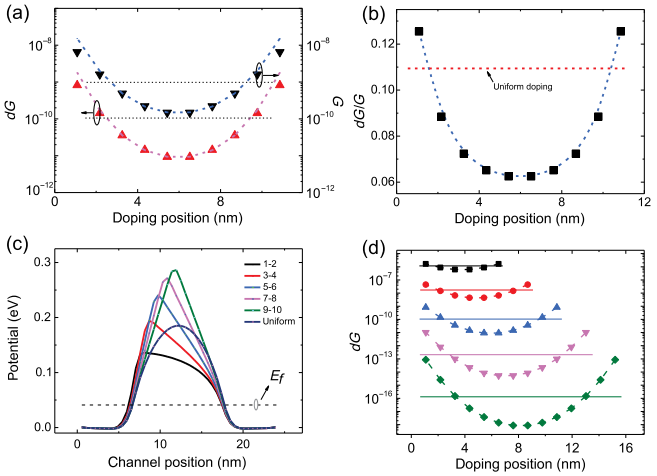

FIG. 4. (a) G and dG versus doping position for $L = 10.9$ nm. The black (red) triangles are the *ab initio* data, and the blue (purple) dashed line is the WKB fitting. Horizontal black dotted lines are for uniform doping. (b) Potential $V(z)$ along the channel for various localized doping positions. (c) dG/G versus doping positions. (d) dG versus doping positions for channel lengths, from low to high $L = 6.5, 8.7, 10.9, 13.0$ and 15.2 nm. Horizontal straight lines refer to the corresponding uniform doping case.

localized doping region. Therefore, the quantity of γ in Eq. (7) can be expressed as a function of the potential height H, and Eqs. (7) and (9) give

$$G = C_3 e^{-A_3 H^{1/2}}, \qquad dG = B_3 H^{-1/2} e^{-A_3 H^{1/2}}, \tag{12}$$

where A_3, B_3, C_3 are fitting parameters. As indicated by the dashed lines in Fig. 4(a), these WKB expressions fit the *ab initio* data very well [29]. Several observations are in order. (i) Localized doping away from the source or drain suppresses not only the tunnel conductance G as found before [15] but also suppresses its variability dG, in comparison to that of uniform doping. Suppression of dG agrees with the intuitive expectation since dopant positions are less random in localized doping. (ii) On the other hand, localized doping near the source or drain can, in fact, enhance dG over the uniform doping; this completely nonintuitive outcome is due to the microscopic details of potential change. Indeed, from Eq. (12) we conclude that dG increases if the potential height H decreases, which is the case for doping positions close to the source or drain. Clearly, a reduction of H near the source or drain is due to the screening effects since the source and drain are heavily doped [15].

Figure 4(d) presents dG versus the localized doping position for different channel lengths from 6.5 to 15.2 nm, where the horizontal lines are dG of uniform doping with concentration 5×10^{18} cm^{-3}. The corresponding localized doping concentrations are obtained from Eq. (6) to be 3, 4, 5, 6 and 7×10^{19} cm^{-3} for $L = 6.5, 8.7, 10.9$ (plotted again for comparison purposes), 13.0, and 15.2 nm channels. For all L, the general trend is the same as that of Fig. 4(a); namely, doping away from the source or drain reduces dG. Again, we observe a clear short-channel behavior; namely, the value of dG increases with decreasing L.

Finally, we mention in passing the computational cost of the above analysis. All the calculations were done by a computer cluster with seven Intel X5650 processors (each processor has six cores, 2.66 GHz, 12-MB cache). For an $L = 10.9$ nm device, the self-consistent NEGF-DFT-NECPA calculation takes approximately 7.6 h. In addition, the transmission and variability calculation takes 12 and 5 m per energy point, respectively.

IV. SUMMARY

Using an atomistic first-principles technique, we analyze the variability of tunnel conductance in Si nano-FET channels of various lengths L and impurity doping concentration, for both uniform doping and localized doping profiles. The variability comes about due to the well-known random discrete-dopant effect, and in this work, it is due to the randomness in positions of the boron dopants. When the channel is doped uniformly, the variability dG is found to scale exponentially as $\exp(-L^{3/2})$ showing a significant short-channel effect. dG also increases significantly with respect to the decrease of the channel doping concentration. We find that by localized doping, the tunnel conductance G can be not only drastically reduced as found before [15], but dG can also be significantly reduced. While naively this is expected since localized doping is less random than uniform doping, it is surprising that if the doping region is close to the source or drain, dG actually increases over that of the uniform doping. The reason lies in the microscopic details of the tunnel potential profile $V(z)$ which is reduced for doping near the heavily doped source and drain, causing both G and dG to increase. Nevertheless, if one can control the doping region to be somewhat near the middle of the channel away from the source or drain, dG is predicted to be controllable to a reasonable fraction of G (i.e., approximately 6%). Finally, our *ab initio* data can be organized by a WKB formulation of the tunneling physics. In particular, after fitting a few constants in the WKB expressions, very good agreement to the first-principles results is obtained by the WKB formula, which reveals interesting trends of dG against device parameters L and the potential height H. We conclude that for Si nano-FET channels around 10 nm lengths, localized doping should provide an interesting approach to rectify the RDD-induced variability in the *off*-state tunnel conductance.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support of the NSERC (H. G.) and IRAP (Y. Z. and L. L.) of Canada. We thank CalcuQuebec and Compute-Canada for providing computation facilities.

[1] A. Asenov, in *Proceedings of the 2007 Symposium on VLSI Technology, Kyoto, 2007* (IEEE, New York, 2007), p. 86.

[2] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs, IEEE Trans. Electron Devices **50**, 1837 (2003).

[3] H.-W. Su, Y. Li, Y.-Y. Chen, C.-Y. Chen, and H.-T. Chang, in *Proceedings of the Device Research Conference (DRC)*, 2012, p. 131.

[4] Y. Li, K.-F. Lee, C.-Y. Yiu, Y.-Y. Chiu, and R.-W. Chang, Dual-material gate approach to suppression of random-dopant-induced characteristic fluctuation in 16 nm metal-oxide-semiconductor field-effect-transistor devices, Jpn. J. Appl. Phys. **50**, 04DC07 (2011).

[5] K. Takeuchi, T. Fukai, T. Tsunomura, A. T. Putra, A. Nishida, S. Kamohara, and T. Hiramoto, in *Proceedings of the International Electron Devices Meeting*, 2007, p. 467.

[6] F.-L. Yang, J.-R. Hwang, H.-M. Chen, J.-J. Shen, S.-M. Yu, Y. Li, and D. D. Tang, in *Proceedings of the 2007 Symposium on VLSI Technology, Kyoto, 2007* (IEEE, New York, 2007), p. 208.

[7] A. Asenov, Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFETs: A 3-D "atomistic" simulation study, IEEE Trans. Electron Devices **45**, 2505 (1998).

[8] C. Shin, X. Sun, and T.-J. K. Liu, Study of random-dopant-fluctuation (RDF) effects for the trigate bulk MOSFET, IEEE Trans. Electron Devices **56**, 1538 (2009).

[9] R. Dave, M. Campbell, R. Gareth, R. Scott, and A. Asen, Analysis of threshold voltage distribution due to random dopants: A 100 000-sample 3-D simulation study, IEEE Trans. Electron Devices **56**, 2255 (2009).

[10] T. Markussen, R. Rurali, A.-P. Jauho, and M. Brandbyge, Scaling Theory Put into Practice: First-Principles Modeling of Transport in Doped Silicon Nanowires, Phys. Rev. Lett. **99**, 076803 (2007).

[11] E. Cruz-Silva, Z. M. Barnett, B. G. Sumpter, and V. Meunier, Structural, magnetic, and transport properties of substitutionally doped graphene nanoribbons from first principles, Phys. Rev. B **83**, 155445 (2011).

[12] S. Mandal and R. Pati, Codoping in a single molecular junction from first principles, Phys. Rev. B **83**, 195420 (2011).

[13] Y. Zhu, L. Liu, and H. Guo, Quantum transport theory with nonequilibrium coherent potentials, Phys. Rev. B **88**, 205415 (2013).

[14] Y. Zhu, L. Liu, and H. Guo, Green's function theory for predicting device-to-device variability, Phys. Rev. B **88**, 085420 (2013).

[15] J. Maassen and H. Guo, Suppressing Leakage by Localized Doping in Si Nanotransistor Channels, Phys. Rev. Lett. **109**, 266803 (2012).

[16] J. Taylor, H. Guo, and J. Wang, *Ab initio* modeling of quantum transport properties of molecular electronic devices, Phys. Rev. B **63**, 245407 (2001).

[17] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, New York, 1995).

[18] Y. Ke, K. Xia, and H. Guo, Disorder Scattering in Magnetic Tunnel Junctions: Theory of Nonequilibrium Vertex Correction, Phys. Rev. Lett. **100**, 166805 (2008).

[19] B. Velický, Theory of electronic transport in disordered binary alloys: Coherent-potential approximation, Phys. Rev. **184**, 614 (1969).

[20] P. Soven, Coherent-potential model of substitutional disordered alloys, Phys. Rev. **156**, 809 (1967).

[21] H. L. Skriver, *The LMTO Method* (Springer, Berlin, 1984).

[22] O. K. Anderson and O. Jepsen, Explicit, First-Principles Tight-Binding Theory, Phys. Rev. Lett. **53**, 2571 (1984).

[23] O. K. Anderson, Z. Pawlowska, and O. Jepsen, Illustration of the linear-muffin-tin-orbital tight-binding representation: Compact orbitals and charge density in Si, Phys. Rev. B **34**, 5253 (1986).

[24] I. Turek, V. Drchal, J. Kudrnovský, M. Sob, and P. Weinberger, *Electronic Structure of Disorder Alloys Surfaces and Interfaces* (Kluwer Academic Publishers, Boston, 1997).

[25] L. Nordheim, Zur Elektronentheorie der Metalle. I, Ann. Phys. (Berlin) **401**, 607 (1931).

[26] F. Tran and P. Blaha, Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential, Phys. Rev. Lett. **102**, 226401 (2009).

[27] See http://www.nanoacademic.ca/index.jsp.

[28] For the dT calculation, the number of $k_\perp$ points is 64×64. $\bar{T}$ and dT are calculated at the Fermi energy of the device $E = E_f$ at 0 K temperature. Other parameters for controlling the numerical calculation are essentially the same as those in Ref. [15].

[29] In our simulations, a 6.5-nm-long buffer layer is added to both sides of the devices' scattering region to ensure a smooth matching of the potential at the boundaries of the simulation domain. Devices in the equilibrium state mean zero bias voltage between the source and drain. The WKB fitting parameters in this paper are as follows: $A_1 = 0.5353\ \mathrm{nm}^{-1.5}$, $B_1 = 6.361 \times 10^{-3}\ \mathrm{nm}^{-0.5}$, $C_1 = 0.1282$, $A_2 = 2.729\ \mathrm{nm}^{-1}\,\mathrm{eV}^{-0.5}$, $B_2 = -15.548$, $C_2 = 4.498 \times 10^{-5}$, $A_3 = 24.761\,\mathrm{eV}^{-0.5}$, $B_3 = 1.548 \times 10^{-6}\,\mathrm{eV}^{0.5}$, and $C_3 = 3.173 \times 10^{-5}$.

[30] Q. Shi, Master's thesis, McGill University, 2013.