Green's function theory for predicting device-to-device variability

Yu Zhu,[1,*] Lei Liu,[1] and Hong Guo[2,1]

[1]*NanoAcademic Technologies, Inc., 7005 Boulevard Taschereau, Brossard, Quebec J4Z 1A7, Canada*
[2]*Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada*
(Received 9 July 2013; revised manuscript received 1 August 2013; published 15 August 2013)

Due to random dopant fluctuations, the device-to-device variability is a serious challenge to emerging nanoelectronics. In this work we present theoretical formalisms and numerical simulations of quantum-transport variability, based on the Green's function technique and the multiple-scattering theory. We have developed a general formalism using the diagrammatic technique within the coherent-potential approximation that can be applied to a wide range of disorder concentrations. In addition, we have developed a method by using a perturbative expansion within the low-concentration approximation that is extremely useful for typical nanoelectronic devices having low dopant concentration. Applying both formalisms, transport fluctuations due to random impurities can be predicted without lengthy brute force computation of ensembles of device structures. Numerical implementations of the formalisms are demonstrated using both tight-binding models and first-principles models.

I. INTRODUCTION

A very important yet difficult issue of electronic device physics is to be able to predict fluctuations in quantum-transport properties due to atomic disorder.[1,2] In existing and emerging field-effect transistors with a channel length of ~10 nm or so, a serious source of property unpredictability is the random dopant fluctuation (RDF). RDF comes from the particular microscopic arrangement of the small number of dopant atoms inside the device channel. Experimentally it is extremely difficult, if not absolutely impossible, to control the precise location of each dopant atom; therefore transport properties vary from one device to another. It was even pointed out that nanowire transistors can suffer from RDF in the source/drain extension region even if the channel is dopant free.[3,4] The device-to-device variability is in fact a general phenomenon for device structures in the nanometer scale which compromises device performance and circuit functionality. From the theoretical point of view, incorporating disorder and randomness in nanoelectronics modeling is of great importance.[1,5] In particular, one is interested in predicting not only the average value of the transport property (e.g., conductance) but also the variance of it.

The device-to-device variability has so far been investigated by statistical analysis of a large number of simulations. For instance Reil *et al.* carried out classical drift-diffusion simulations for an ensemble of 10^5 dopant configurations under the combined influence of RDF and line edge roughness.[6] Martinez *et al.* did effective-mass nonequilibrium Green's function (NEGF) simulations of an ensemble of 30 dopant configurations to analyze statistical variability of quantum transport in gate-all-around silicon nanowires.[7] The contrast of the size of the statistical ensemble clearly shows the difficulty of quantum simulations. The difficulty in brute force computation becomes much more severe in full self-consistent atomistic modeling (as opposed to effective-mass modeling) such as the NEGF-based density functional theory (DFT).[8] There is an urgent need to develop viable theoretical methods that do not rely on brute force computation for predicting the device-to-device variability. It is the purpose of this work to present such a formalism.

We shall report a new theoretical approach to directly calculate statistical variations of quantum transport due to RDF without individually computing each and every impurity configuration by brute force. Our theory is composed of two formalisms: One is general but more complicated and the other is specialized but much simpler. The two formalisms are based on the Green's function technique and the multiple-scattering theory. The first formalism builds on the coherent-potential approximation (CPA) and can be applied to a wide range of impurity concentrations. The second formalism builds on the low-concentration approximation (LCA) and is extremely useful for situations involving low impurity concentration which is often the case for realistic semiconductor devices. Our theory and implementation have been checked with both analytical and numerical verification.

The basic physical model of a two-probe quantum coherent nanoelectronic device is schematically shown in any one of the subfigures of Fig. 1, which consists of a central channel region sandwiched by the left and right semi-infinite electrodes.[9] The electrodes extend to reservoirs at $z = \pm\infty$ where bias voltages are applied and electric current measured. We assume that the RDF occurs in the channel region of the system and each subfigure in Fig. 1 represents one dopant configuration. Clearly, due to different locations of the dopant atoms, every device exhibits slightly different transport behavior leading to the device-to-device variability. The transport current flowing through the device can be expressed[9] in terms of the transmission coefficient T (hereafter atomic units are assumed, $e = \hbar = 1$),

$$I = \int \frac{dE}{2\pi} T(E)[f_L(E) - f_R(E)], \tag{1}$$

where E is the electron energy, and $f_L(E)$ and $f_R(E)$ are the Fermi functions of the left/right electrodes. Without RDF, the electric current I is a definite number for a given bias voltage. In the presence of RDF, I depends on the particular impurity configuration and thus varies from one configuration to another. By calculating a large ensemble of configurations one can determine an average current and its associated variance δI. For our device model where RDF occurs inside
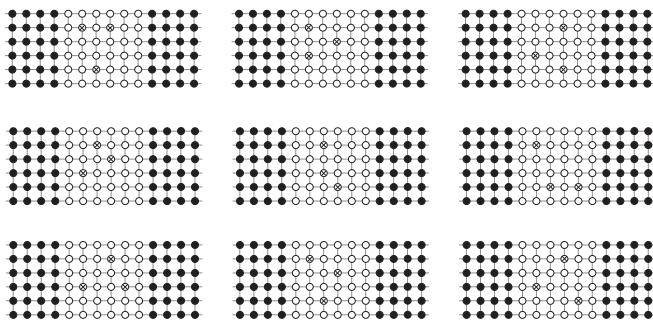
FIG. 1. Ensemble of two-probe devices with various disorder configurations. In each subfigure, the left and right electrodes extend to $z = \pm\infty$, respectively. The black dots are pure sites in the electrodes, the empty circles are pure sites in the central channel region, and the crossed empty circles are disorder sites (dopant or impurity) in the channel region.

the channel region, δI is obtained in terms of transmission fluctuation, δT, as follows:

$$\delta I \approx \int \frac{dE}{2\pi} \delta T(E) \left[f_L(E) - f_R(E) \right]. \tag{2}$$

By definition, the transmission fluctuation δT is obtained from the RDF ensemble average,

$$\delta T \equiv \sqrt{\overline{T^2} - \overline{T}^2}, \tag{3}$$

where $\overline{\cdots}$ refers to averaging over the disorder configurations. Notice that the transmission coefficient T can be expressed in terms of Green's functions. As a result the calculation of $\overline{T}$ involves evaluating a 2 Green's function correlator $\overline{G \cdot G}$. The calculation of transmission fluctuation $\delta T(E)$ which needs the quantity $\overline{T^2}$ involves a 4 Green's function correlator $\overline{G \cdot G \cdot G \cdot G}$.

In the literature, a well-known technique called the coherent-potential approximation[10,11] is available to evaluate the disorder average of a single Green's function $\overline{G}$. The CPA technique was generalized[12,13] to evaluate 2 Green's function correlators $\overline{G \cdot G}$ and 3 Green's function correlators $\overline{G \cdot G \cdot G}$ (albeit in other contexts). More recently, the generalized CPA technique for calculating the 2 Green's function correlator has been applied to study transmission[14] and nonequilibrium quantum transport[15] in disordered systems. This work will address how to evaluate 4 Green's function correlators and apply the technique to study device variability.

In a very recent paper,[16] Zhuang and Wang carried out an analysis of conductance fluctuation and shot noise in graphene by using a direct expansion approach. To some extent, their approach is complementary to the methods presented in this work with respect to accuracy and efficiency. Finally, there are large bodies of literature in mesoscopic physics to analyze such issues as the universal conductance fluctuation in bulk systems using the Kubo formula and δ-like short-range impurity potentials.[17] In contrast, the goal of this work is to formulate a theoretical approach for calculating the transmission fluctuation of two-probe systems where the disorder scattering is due to impurity atoms as opposed to δ-like models.

The rest of the paper is organized as follows. Section II reviews the multiple-scattering theory of the t-matrix formalism. Section III presents the first formalism, i.e., the CPA diagrammatic technique for calculating transmission fluctuation. Section IV presents the second formalism, i.e., the LCA perturbative expansion technique for calculating transmission fluctuation. Section V discusses a special but important situation where the device structure is periodic in transverse dimensions. Section VI presents some miscellaneous technical issues of the theory. Section VII presents three examples as applications of the CPA and LCA formalisms. Finally, the paper is concluded with a brief summary in Sec. VIII. Some technical details are enclosed in the two appendices.

II. THE t-MATRIX FORMALISM

To simulate disorder sites in the central region, the on-site energies are assumed to be discrete random variables. Namely, on a disorder site i the on-site energy ε_i can take the value ε_{iq} with the probability x_{iq} where $q = 1, 2, \ldots$ indicates the possible atomic species on that site and the normalization requires $\sum_q x_{iq} = 1$.

For a given disorder configuration $\{\varepsilon_i\}$, the transmission coefficient $T(E)$ can be derived in terms of the Green's functions of the central region[9,18]

$$T(E) = \mathrm{Tr}[G^r(E)\Gamma_L(E)G^a(E)\Gamma_R(E)], \tag{4}$$

where $G^{r,a}$ are the retarded and advanced Green's functions, and $\Gamma_{L,R}$ are the linewidth functions of the left and right electrodes. The retarded Green's function can be derived as[9,18]

$$G^r(E) = [E - H(\{\varepsilon_i\}) - \Sigma^r(E)]^{-1}, \tag{5}$$

where $H(\{\varepsilon_i\})$ is the Hamiltonian of the central region whose impurity configuration is $\{\varepsilon_i\}$; Σ^r is the retarded self-energy to take into account the influences of the semi-infinite electrodes on the central region. The linewidth function Γ in Eq. (4) is related to the self-energy

$$\Gamma_\beta(E) = i\left[\Sigma_\beta^r(E) - \Sigma_\beta^a(E)\right], \tag{6}$$

where $\beta = L, R$ labels the left or right electrode and $\Sigma^r(E) = \Sigma_L^r(E) + \Sigma_R^r(E)$ is the total retarded self-energy. The advanced Green's function and self-energy are Hermitian conjugates of their retarded counterparts,

$$G^a(E) = G^r(E)^\dagger, \quad \Sigma_\beta^a(E) = \Sigma_\beta^r(E)^\dagger.$$

To determine δT by Eq. (3), one needs to calculate the configuration averaged quantities $\overline{T^2}$ and $\overline{T}$. While vertex correction techniques have been applied successfully to calculate $\overline{T}$, the quantity $\overline{T^2}$ turns out to be extremely difficult to calculate and the goal of this work is to derive a necessary formulation for it. Since our approach is based on the t-matrix formalism, in the rest of this section, we briefly review the well-known t-matrix formalism following Ref. 12.

Recall that H is the Hamiltonian of the central region which contains some disorder sites (see Fig. 1). Divide H into two parts, H_0 and V, where H_0 is the definite part of the

Hamiltonian and V is the random disorder potential,

$$H = H_0 + V, \qquad (7)$$

$$V = \sum_i \hat{V}_i, \qquad (8)$$

in which $\hat{V}_i$ is the random potential of disorder site i. $\hat{V}_i$ is a nearly all-zero matrix except for its ith diagonal element

$$\hat{V}_i = \text{diag}\,[0,\ldots,0,V_i,0,\ldots,0]\,,$$

where V_i is a discrete random variable which can take the value V_{iq} with the probability x_{iq}. V_{iq} is related to the on-site energy ε_{iq} by $V_{iq} = \varepsilon_{iq} - \varepsilon_i^0$, where ε_i^0 is a site-dependent arbitrary constant. Due to different choices of $\{\varepsilon_i^0\}$, the partition of H into H_0 and V is not unique. We shall exploit this freedom and adopt different partitions for CPA and LCA (see next two sections).

With the partition Eqs. (7) and (8), the retarded Green's function of Eq. (5) can be expressed in terms of the unperturbed Green's function G_0^r and the t matrix T^r,

$$G^r = G_0^r + G_0^r T^r G_0^r, \qquad (9)$$

where G_0^r and T^r are defined as

$$G_0^r \equiv [E - H_0 - \Sigma^r]^{-1}, \qquad (10)$$

$$T^r \equiv V\left(1 - G_0^r V\right)^{-1}. \qquad (11)$$

The t matrix T^r can be further expanded in terms of scattering amplitude $\hat{t}_i^r$,

$$T^r = \sum_i \hat{t}_i^r + \sum_i \sum_{j \neq i} \hat{t}_j^r G_0^r \hat{t}_i^r$$
$$+ \sum_i \sum_{j \neq i} \sum_{k \neq j} \hat{t}_k^r G_0^r \hat{t}_j^r G_0^r \hat{t}_i^r + \cdots, \qquad (12)$$

where $\hat{t}_i^r$ represents multiple disorder scattering on the site i:

$$\hat{t}_i^r \equiv \hat{V}_i + \hat{V}_i G_0^r \hat{V}_i + \hat{V}_i G_0^r \hat{V}_i G_0^r \hat{V}_i + \cdots$$
$$= \hat{V}_i\left(1 - G_0^r \hat{V}_i\right)^{-1}. \qquad (13)$$

Similar to $\hat{V}_i$, $\hat{t}_i^r$ is also a nearly all-zero matrix except for its ith diagonal element

$$\hat{t}_i^r = \text{diag}\big[0,\ldots,0,t_i^r,0,\ldots,0\big],$$

where t_i^r is a random variable which can take the value t_{iq}^r with the probability x_{iq}. t_{iq}^r is obtained as

$$t_{iq}^r \equiv V_{iq}\left(1 - G_{0,ii}^r V_{iq}\right)^{-1}, \qquad (14)$$

in which $G_{0,ii}^r$ means to take the ith diagonal element of G_0^r.

Inserting Eq. (12) into Eq. (9), G^r can be expanded in a series of scattering terms:

$$G^r = G_0^r + \sum_i G_0^r \hat{t}_i^r G_0^r + \sum_i \sum_{j \neq i} G_0^r \hat{t}_j^r G_0^r \hat{t}_i^r G_0^r$$
$$+ \sum_i \sum_{j \neq i} \sum_{k \neq j} G_0^r \hat{t}_k^r G_0^r \hat{t}_j^r G_0^r \hat{t}_i^r G_0^r + \cdots. \qquad (15)$$

In a diagrammatic language, Eq. (15) can be represented by Fig. 2 in which the thick line represents G^r, the thin line represents G_0^r, and the dotted line with a crossed dot represents $\hat{t}_i^r$ (random variable). It is required that adjacent $\hat{t}_i^r$ lines must
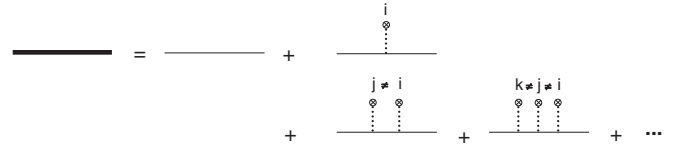

FIG. 2. Diagram representation of Eq. (15).

have different site indices. Clearly, a similar expansion can be carried out for the advanced Green's function G^a.

The t-matrix expansion in Eq. (15) is rigorous. By inserting Eq. (15) and its advanced counterpart into Eq. (4) and its square, after averaging over disorder configurations, $\overline{T(E)}$ and $\overline{T^2(E)}$ can be derived as a summation of products composed of x_{iq}, G_0^r and G_0^a, Γ_L and Γ_R, $\hat{t}_{iq}^r$ and $\hat{t}_{iq}^a$. In principle, one can calculate $\overline{T}$ and $\overline{T^2}$ by summing up these terms order by order, which is accurate but impractical for realistic device simulations. Alternatively, diagrammatic techniques will be developed in the following sections to evaluate the summation approximately.

III. THE COHERENT-POTENTIAL APPROXIMATION

In this section we present the formalism for calculating transmission fluctuation δT based on the CPA diagrammatic technique. The main idea is to expand $\overline{T}$ and $\overline{T^2}$ into a series of scattering terms each of which can be mapped into a diagram. At the CPA level, a subset of these diagrams (the noncrossing diagrams) can be collected and summed up. The diagrammatic technique was originally developed in Ref. 13 to calculate transport coefficients involving 3 Green's function correlators. Here, this technique is improved and generalized to calculate δT involving 4 Green's function correlators.

A. The Γ decomposition

As shown in Eq. (4), transmission coefficient T is a trace of matrix product, and hence T^2 is a product of two traces which is inconvenient to apply the diagrammatic technique. To proceed, we first rewrite T^2 into a proper matrix product form. Using the Γ-decomposition technique introduced in Ref. 19, the linewidth function of the right electrode, Γ_R, can be decomposed as $\Gamma_R = \sum_n |W_n\rangle\langle W_n|$, where $|W_n\rangle$ is the nth normalized eigenvector of the Γ_R matrix.[20] Consequently, using Eq. (4) T^2 can be rewritten in the following Γ-decomposition form:

$$T^2 = (\text{Tr}\,G^r \Gamma_L G^a \Gamma_R) \times (\text{Tr}\,G^r \Gamma_L G^a \Gamma_R)$$
$$= \sum_n \text{Tr}\,G^r \Gamma_L G^a |W_n\rangle\langle W_n| \sum_m \text{Tr}\,G^r \Gamma_L G^a |W_m\rangle\langle W_m|$$
$$= \sum_n \langle W_n|G^r \Gamma_L G^a |W_n\rangle \sum_m \langle W_m|G^r \Gamma_L G^a |W_m\rangle$$
$$= \sum_{nm} \text{Tr}\,G^r \Gamma_L G^a X_{nm} G^r \Gamma_L G^a X_{nm}^\dagger, \qquad (16)$$

where X_{nm} is defined as $X_{nm} \equiv |W_n\rangle\langle W_m|$.

So the calculations of $\overline{T}$ and $\overline{T^2}$ are reduced to the Green's function correlators $\overline{\text{Tr}\,G^r X_1 G^a X_2}$ and $\overline{\text{Tr}\,G^r X_1 G^a X_2 G^r X_3 G^a X_4}$ where X_k is a definite quantity which is referred to as the vertex of the correlator. Notice that

G^r and G^a always appear alternatively in $\overline{T}$ and $\overline{T^2}$; as such we shall omit the superscripts r,a in the CPA diagrammatic expansion without causing any ambiguity.

B. The CPA diagrams

Equation (16) indicates that we need to calculate various Green's function correlators such as

$$I_2 \equiv \mathrm{Tr}\overline{GX_1GX_2}, \tag{17}$$

$$I_3 \equiv \mathrm{Tr}\overline{GX_1GX_2GX_3}, \tag{18}$$

$$I_4 \equiv \mathrm{Tr}\overline{GX_1GX_2GX_3GX_4}. \tag{19}$$

To proceed we insert Eq. (15) into I_n ($n = 2,3,4$) to obtain a series expansion. In analogy to Eq. (15) and Fig. 2, each term in the I_n series expansion can be represented by a diagram: The thick line represents the full Green's function G; the thin line represents the unperturbed Green's function G_0; the blue dot represents the vertex X_n; the dotted line with a red dot represents the impurity scattering amplitude $\hat{t}_{iq}$. The trace operation is represented by a closed circle composed of G lines and X vertexes. If some impurity indices are identical in the disorder average, the corresponding impurity lines need to be contracted with each other. The major difference between the diagrams in this section and Fig. 2 is that the former diagrams represent terms after disorder average while the latter represents terms before disorder average.

Thus the lengthy series expansion of I_n is nicely organized into a diagrammatic fashion. One can sum up the diagrams in a perturbative manner up to some finite order as done in
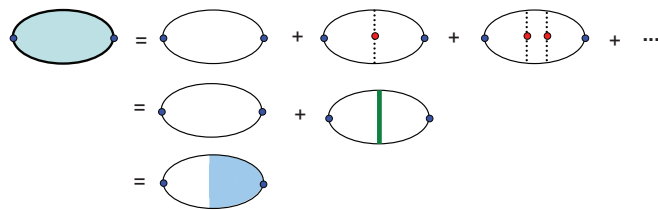

FIG. 3. (Color online) CPA diagrams of I_2.

Ref. 16. Alternatively, by selecting a subset of the diagrams, one can evaluate the diagrammatic summation to *infinite* order. In particular, the subset is called CPA diagrams selected by the following two rules. (i) An impurity line on one G line must contract with impurity line(s) of other G line(s), and no dangling impurity line is allowed. The reason is that in CPA[10,11] the partition of H_0 and V is chosen such that

$$\overline{t_i^r} = \overline{t_i^a} = 0, \tag{20}$$

and hence diagrams with dangling impurity lines vanish (see Appendix A for details). (ii) Contracted impurity lines do not cross each other. Namely, only the noncrossing diagrams are taken into account in the CPA diagrammatic summation. In the following subsections, CPA diagrams of I_2, I_3, and I_4 will be analyzed in detail.

C. I_2 diagrams

By inserting Eq. (15) into Eq. (17) and applying CPA diagram rules, the I_2 diagrams are obtained in Fig. 3. In the first row of Fig. 3, the diagram equation corresponds to the following algebraic equation:

$$\mathrm{Tr}\overline{GX_1GX_2} = \mathrm{Tr}G_0X_1G_0X_2 + \sum_{i_1q_1} x_{i_1q_1}\mathrm{Tr}G_0\hat{t}_{i_1q_1}G_0X_1G_0\hat{t}_{i_1q_1}G_0X_2$$

$$+ \sum_{i_1q_1}\sum_{\substack{i_2q_2 \\ i_2 \neq i_1}} x_{i_1q_1}x_{i_2q_2}\mathrm{Tr}G_0\hat{t}_{i_1q_1}G_0\hat{t}_{i_2q_2}G_0X_1G_0\hat{t}_{i_2q_2}G_0\hat{t}_{i_1q_1}G_0X_2 + \cdots. \tag{21}$$

The diagram representation in Fig. 3 significantly simplifies the algebraic expression of Eq. (21).

In the second and third rows of Fig. 3, the diagrams of I_2 are further simplified by using a bundled line (second row) and a dressed vertex (third row). The bundled line $\tilde{t}$ (green thick line) is a collection of ladder diagrams. The vertex correction Λ is the combination of a bundled line $\tilde{t}$ and a vertex X. The dressed vertex Π (cyan shadow) is a vertex X plus its vertex correction Λ. The meanings of the diagram elements $\tilde{t}$, Λ, and Π are explained in Fig. 4.

Given a vertex X, the corresponding vertex correction Λ is solved from the following equation:

$$\Lambda_i = \sum_q x_{iq}t_{iq}(G_0XG_0)_{ii}t_{iq}$$

$$+ \sum_{j \neq i}\sum_q x_{iq}t_{iq}(G_0)_{ij}\Lambda_j(G_0)_{ji}t_{iq}, \tag{22}$$

where $\Lambda = \mathrm{diag}([\Lambda_1, \Lambda_2, \ldots])$ is a diagonal matrix. Equation (22) is derived by the recursive relation illustrated in Fig. 5. Note that Eq. (22) is identical to Eqs. (49) and (50) in Ref. 12.
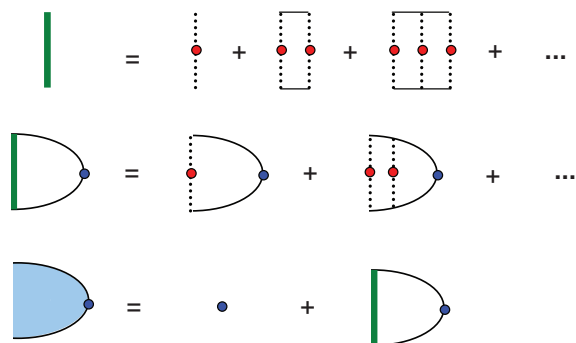

FIG. 4. (Color online) CPA diagram elements: the bundled line $\tilde{t}$, the vertex correction Λ, and the dressed vertex Π.
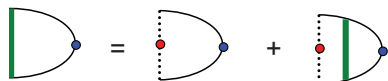
FIG. 5. (Color online) Diagram representation of Eq. (22).

D. I_3 diagrams

By inserting Eq. (15) into Eq. (18) and applying the CPA diagram rules, I_3 diagrams are obtained in Fig. 6. In the first three rows of Fig. 6, there are 16 diagrams constructed with bundled lines which are equivalent to Fig. 3 of Ref. 13. In the fourth row of Fig. 6, the diagram number is reduced to two by using the dressed vertex Π which has been defined in Fig. 4. In the fifth row of Fig. 6, the diagram number is reduced further to one by using the dressed vertex Π and the dressed double vertex Π_2 which is defined in Fig. 7.

E. I_4 diagrams

By inserting Eq. (15) into Eq. (19) and applying the CPA diagram rules, I_4 diagrams are obtained in Fig. 8. There are 256 diagrams if constructed only with the bundled lines (not shown). The diagram number is reduced to 16 if constructed with bundled lines and dressed vertexes, as shown in the first three rows of Fig. 8. The diagram number is reduced to 6 if constructed with bundled lines, dressed vertexes, and dressed double vertexes, as shown in the fourth row of Fig. 8. It is clear that the use of the dressed vertex and dressed double vertex greatly reduces the number of CPA diagrams.

F. The sum rules

How do we know that all the CPA diagrams have been included in the diagrammatic summation? There are some Ward's type identities in Green's functions which are helpful to verify the completeness of the CPA diagrams. The identities reduce a product of Green's functions to products of fewer Green's functions. By applying disorder average to both sides of the identity, the identity must remain valid if the average is done rigorously. This way the higher level correlators (e.g., 4 Green's function correlators) are related to lower level correlators (e.g., 2 Green's function correlators). The amazing feature of CPA is that the identity still holds even if approximations are made on both sides of the identity. In this sense, CPA is a consistent approximation for the Green's
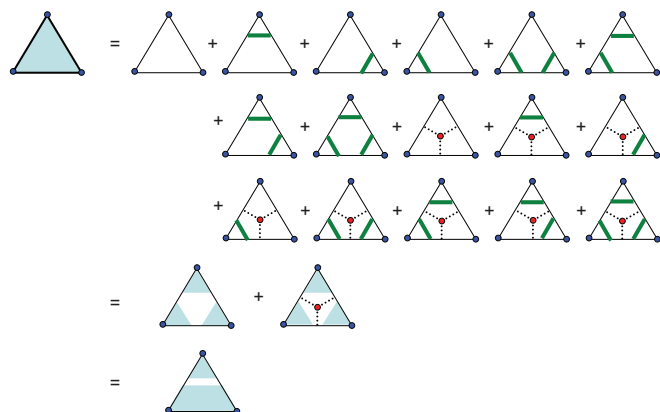

FIG. 7. (Color online) CPA diagram element: dressed double vertex Π_2.

function correlators. These identities can thus be used to verify theoretical derivations as well as numerical implementations. Missing a single diagram will make the identities unbalanced.

In particular, the identities for testing I_2, I_3, and I_4 are listed below:

$$\overline{G^r \Sigma^{ra} G^a} = \overline{G^r} - \overline{G^a}, \tag{23}$$

$$\overline{G^r \Sigma^{ra} G^a \Sigma^{ra} G^r} = \overline{G^r \Sigma^{ra} G^r} + \overline{G^a} - \overline{G^r}, \tag{24}$$

$$\overline{G^r \Sigma^{ra} G^a \Sigma^{ra} G^r \Sigma^{ra} G^a} = \overline{G^r \Sigma^{ra} G^r} + \overline{G^a \Sigma^{ra} G^a} \\ -2(\overline{G^r} - \overline{G^a}), \tag{25}$$

where $\Sigma^{ra} \equiv \Sigma^r - \Sigma^a$. Note that in these equalities, the left-hand side involves higher level correlators while the right-hand side involve lower level correlators. Our analytical formalism and numerical computation have been verified by confirming the equality to high precision. In Appendix B, we provide an analytical proof of the identity Eq. (23).

G. Summary of CPA diagram technique

In this section, CPA diagrams for evaluating Green's function correlators I_2, I_3, I_4 are presented. I_2 and I_3 have been investigated in Ref. 13 and are included here for completeness and improvement. As a main result of this work, we have derived the CPA diagrams for I_4 and reduced the diagram number from 256 to 6 by using the dressed vertex and dressed double vertex.

By using CPA diagrams of I_2 and I_4, transmission fluctuation $\delta T = \sqrt{\overline{T^2} - \overline{T}^2}$ can be calculated as follows: (i) Calculate G_0^r, t_{iq}^r, G_0^a, t_{iq}^a by solving CPA condition Eq. (20). The details are presented in Appendix A. (ii) Calculate $\overline{T^2}$ by using the disorder average of Eq. (16) and the CPA diagrams in Fig. 8. (iii) Calculate $\overline{T}$ by using the disorder average of Eq. (4) and the CPA diagrams in Fig. 3. The dressed vertexes in the CPA diagrams can be calculated by using Eq. (22). It is concluded that the RDF-induced transmission fluctuation can be calculated by the CPA diagrammatic technique presented in this section.
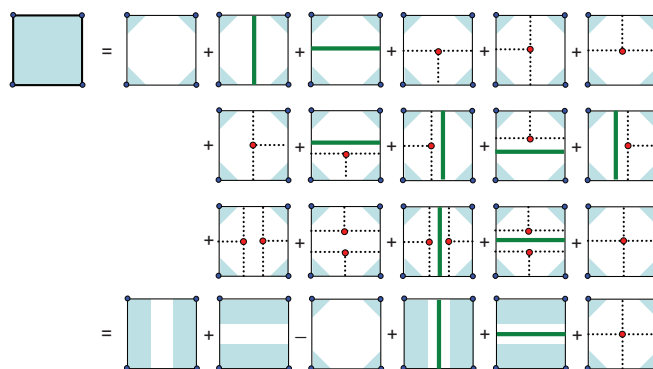

FIG. 6. (Color online) CPA diagrams of I_3.

FIG. 8. (Color online) CPA diagrams of I_4.

IV. THE LOW-CONCENTRATION APPROXIMATION

In the last section we have presented a general formalism based on the CPA diagrammatic technique to evaluate the transmission fluctuation δT. It is general in the sense that RDF is calculated for arbitrary impurity concentration x. Nevertheless, in semiconductor devices (e.g., transistors) the doping concentration is always very low. Even for heavily doped Si at a doping level 10^{20} cm^{-3}, the impurity concentration amounts to $x \sim 2 \times 10^{-3}$ which is a small parameter. Therefore one can carry out a perturbative expansion to the lowest order of the small parameter x to evaluate δT, which is referred to as the low-concentration approximation. This is especially useful for analyzing RDF-induced device-to-device variability in semiconductor nanoelectronics. This section is devoted to presenting the LCA formalism.

Let $q = 0$ represent the host material atom species and $q > 0$ impurity atom species. Low concentration means that the concentration of host material atoms is much larger than that of impurity atoms; i.e., $x_{i,q=0} \gg x_{i,q>0}$. The main idea of LCA is to collect the lowest order terms in δT^2 which are proportional to $x_{i,q>0}$. Because the impurity concentration is small, in the partition of the total Hamiltonian Eqs. (7) and (8), we naturally choose H_0 to be the Hamiltonian of the host material and V to be the difference between impurity atoms and host atoms. Consequently the disorder scattering potential V_{iq} is

$$V_{iq} = \varepsilon_{iq} - \varepsilon_{i0}, \tag{26}$$

where ε_{iq} is the on-site energy of the impurity atom and ε_{i0} is the on-site energy of the host atom. This is in contract to the CPA diagrammatic formalism of the last section in which H_0 and V have been chosen such that the CPA condition $\overline{t_i^r} = \overline{t_i^a} = 0$ is satisfied.

The simplicity of LCA is that it does not need Γ decomposition as in CPA. One can directly substitute Eq. (15) and its advanced counterpart into Eq. (4) and its square to obtain a series expansion for T and T^2. Averaging over disorder configurations and collecting the terms up to the first order of $x_{i,q>0}$, $\overline{T}$ and $\overline{T^2}$ can be obtained and represented by the LCA diagrams in Fig. 9 and Fig. 10, respectively.

The meaning of LCA diagrams is similar to that of CPA diagrams: The thin line represents unperturbed Green's function G_0^r or G_0^a; the blue dot represents vertex Γ_L or Γ_R; the dotted line with a red dot represents impurity scattering amplitude $\hat{t}_i^r$ or $\hat{t}_i^a$. The closed Green's function circle means to carry out the trace operation. The contraction of impurity lines means that the disorder site indices are the same. To be specific, the LCA diagrams (1), (2), (3), (4) in Fig. 9 correspond to the following algebraic expressions in order:

$$\mathrm{Tr}\left(G_0^r \Gamma_L G_0^a \Gamma_R\right),$$

$$\sum_{iq} x_{iq} \mathrm{Tr}\left(G_0^r \Gamma_L G_0^a \hat{t}_{iq}^a G_0^a \Gamma_R\right),$$

$$\sum_{iq} x_{iq} \mathrm{Tr}\left(G_0^r \hat{t}_{iq}^r G_0^r \Gamma_L G_0^a \Gamma_R\right),$$

$$\sum_{iq} x_{iq} \mathrm{Tr}\left(G_0^r \hat{t}_{iq}^r G_0^r \Gamma_L G_0^a \hat{t}_{iq}^a G_0^a \Gamma_R\right).$$
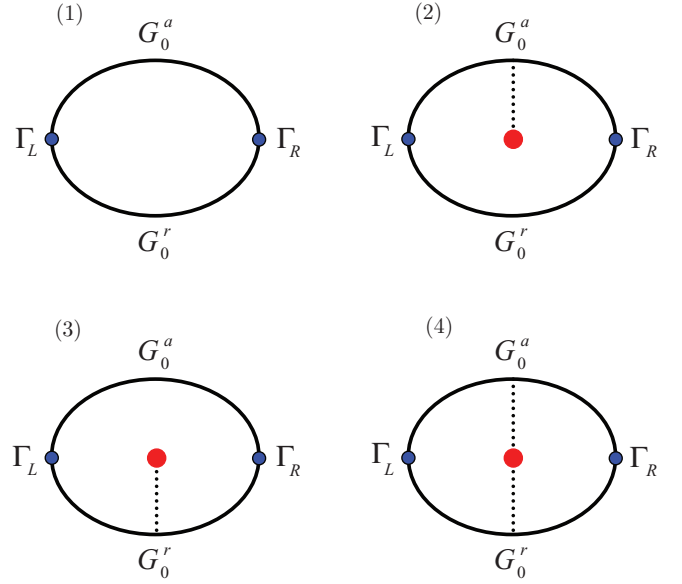

FIG. 9. (Color online) LCA diagrams of $\overline{T}$.

The LCA diagrams (1), (6), (11), (16) in Fig. 10 correspond to the following algebraic expressions in order:

$$\mathrm{Tr}\left(G_0^r \Gamma_L G_0^a \Gamma_R\right) \mathrm{Tr}\left(G_0^r \Gamma_L G_0^a \Gamma_R\right),$$

$$\sum_{iq} x_{iq} \mathrm{Tr}\left(G_0^r \hat{t}_{iq}^r G_0^r \Gamma_L G_0^a \hat{t}_{iq}^a G_0^a \Gamma_R\right) \mathrm{Tr}\left(G_0^r \Gamma_L G_0^a \Gamma_R\right),$$

$$\sum_{iq} x_{iq} \mathrm{Tr}\left(G_0^r \hat{t}_{iq}^r G_0^r \Gamma_L G_0^a \Gamma_R\right) \mathrm{Tr}\left(G_0^r \hat{t}_{iq}^r G_0^r \Gamma_L G_0^a \Gamma_R\right),$$

$$\sum_{iq} x_{iq} \mathrm{Tr}\left(G_0^r \hat{t}_{iq}^r G_0^r \Gamma_L G_0^a \hat{t}_{iq}^a G_0^a \Gamma_R\right)$$

$$\times \mathrm{Tr}\left(G_0^r \hat{t}_{iq}^r G_0^r \Gamma_L G_0^a \hat{t}_{iq}^a G_0^a \Gamma_R\right).$$

Of the 16 LCA diagrams for $\overline{T^2}$, 7 diagrams [from (1) to (7) in Fig. 10] are unconnected and will cancel with the 7 LCA diagrams from $\overline{T}^2$ in calculating δT^2. The summation of the remaining 9 diagrams [from (8) to (16) in Fig. 10] can be further simplified as

$$\delta T^2 = \sum_{i,q>0} x_{iq}\left(Y_{iq}^\alpha + Y_{iq}^\beta + Y_{iq}^\gamma\right)^2, \tag{27}$$

where

$$Y_{iq}^\alpha = \mathrm{Tr}\left\{t_{iq}^a\left[G_0^a \Gamma_R G_0^r \Gamma_L G_0^a\right]_{ii}\right\}, \tag{28}$$

$$Y_{iq}^\beta = \mathrm{Tr}\left\{t_{iq}^r\left[G_0^r \Gamma_R G_0^a \Gamma_L G_0^r\right]_{ii}\right\}, \tag{29}$$

$$Y_{iq}^\gamma = \mathrm{Tr}\left\{t_{iq}^r\left[G_0^r \Gamma_L G_0^a\right]_{ii} t_{iq}^a\left[G_0^a \Gamma_R G_0^r\right]_{ii}\right\}, \tag{30}$$

in which $(Y_{iq}^\alpha)^* = Y_{iq}^\beta$ and $(Y_{iq}^\gamma)^* = Y_{iq}^\gamma$. It follows that $\delta T^2 > 0$ which is consistent with the physical meaning of this quantity. Note that the summation over i and q in Eq. (27) clearly identifies the contribution of each impurity species and disorder site to the total transmission fluctuation. Equation (27) and the definitions of G_0^r in Eq. (10), t_{iq}^r in Eq. (14), and V_{iq} in Eq. (26) are the central results of this section.
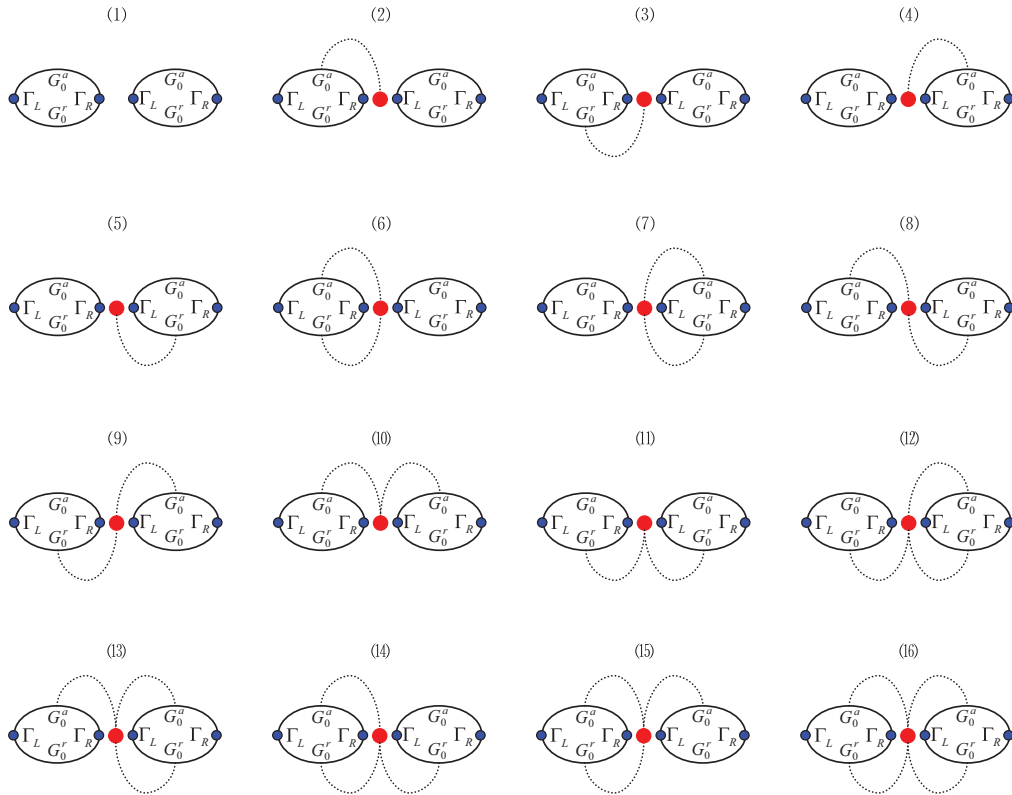
FIG. 10. (Color online) LCA diagrams of $\overline{T^2}$.

V. FORMULATION IN FOURIER SPACE

Having presented two theoretical methods for computing δT, i.e., the CPA diagrammatic formalism and the LCA diagrammatic formalism, we now consider an important special situation where two-probe systems are "periodic" in the transverse dimensions. When there is no disorder, periodicity is well defined, and one can identify a unit cell in the transverse dimensions and apply the Bloch theorem by Fourier transform. In disordered two-probe systems, one can also identify a unit cell but the situation is more complicated. On the one hand, the Hamiltonian does not have translational symmetry in the presence of random disorder and thus the Bloch theorem breaks down. On the other hand, the disorder-averaged physical quantities are still periodic and can be Fourier transformed. The formalisms developed in the previous two sections need to be modified slightly to adapt to such disordered "periodic" two-probe systems.

Assume that a disordered two-probe system has periodicity in one transverse dimension. Define the dimensionless crystal momentum k as $k = \mathbf{k} \cdot \mathbf{a}$ where $\mathbf{k}$ is the wave vector and $\mathbf{a}$ is the unit cell vector of the periodic dimension.[21] A periodic physical quantity Y as a function of unit cell indices I_1 and I_2 should be only dependent on the index difference $I_1 - I_2$. Therefore $Y_I \equiv Y_{I_1 - I_2}$ can be transformed into the Fourier space

$$Y(k) = \sum_I e^{-ikI} Y_I.$$

For example, H_0 and Σ_β^r do not contain randomness and can be Fourier transformed into $H_0(k)$ and $\Sigma_\beta^r(k)$. Consequently

$$G_0^r(k) = [E - H_0(k) - \Sigma^r(k)]^{-1},$$
$$\Gamma_\beta(k) = i[\Sigma_\beta^r(k) - \Sigma_\beta^a(k)].$$

To obtain the on-site quantity Y_{ii}, one needs to integrate over k (inverse Fourier transform)

$$Y_{ii} = \int_{-\pi}^{+\pi} \frac{dk}{2\pi} Y_{ii}(k).$$

To carry out Γ decomposition, the summation over the electrode conducting channel n should be replaced by an integral over k in addition to the summation over n; i.e., $\sum_n \longrightarrow \int_{-\pi}^{+\pi} \frac{dk}{2\pi} \sum_n$. The necessary modifications of CPA formalism and LCA formalism are presented explicitly as follows.

For the CPA diagrammatic formalism presented in Sec. III, the Γ decomposition Eq. (16) should be modified as

$$T^2 = \int_{-\pi}^{+\pi} \frac{dk}{2\pi} \int_{-\pi}^{+\pi} \frac{dk'}{2\pi} \sum_{nn'} \text{Tr} G^r(k) \Gamma_L(k) G^a(k) X_{nk,n'k'} G^r(k') \Gamma_L(k') G^a(k') X_{nk,n'k'}^\dagger, \tag{31}$$

where $X_{nk,n'k'}$ is defined as

$$X_{nk,n'k'} \equiv |W_n(k)\rangle \langle W_{n'}(k')|,$$

in which the eigenvector $|W_n(k)\rangle$ comes from the k-dependent Γ decomposition of $\Gamma_R(k)$

$$\Gamma_R(k) = \sum_n |W_n(k)\rangle\langle W_n(k)|.$$

Moreover, the vertex correction Eq. (22) needs to be modified as

$$\Lambda_i = \sum_q x_{iq} t_{iq} \left\{ \int_{-\pi}^{+\pi} \frac{dk}{2\pi} [G_0(k)X(k)G_0(k)]_{ii} \right\} t_{iq} + \sum_q x_{iq} t_{iq} \left\{ \int_{-\pi}^{+\pi} \frac{dk}{2\pi} [G_0(k)\Lambda G_0(k)]_{ii} \right\} t_{iq}$$

$$- \sum_q x_{iq} t_{iq} \left[\int_{-\pi}^{+\pi} \frac{dk}{2\pi} [G_0(k)]_{ii} \right] \Lambda_i \left[\int_{-\pi}^{+\pi} \frac{dk}{2\pi} [G_0(k)]_{ii} \right] t_{iq}, \tag{32}$$

in which $X(k)$ is the Fourier transform of X.

For the LCA diagrammatic formalism presented in Sec. IV, Y_{iq}^α, Y_{iq}^β, Y_{iq}^γ, and t_{iq}^r in Eq. ((27)) should be modified as

$$Y_{iq}^\alpha = \text{Tr} \left\{ t_{iq}^a \left[\int_{-\pi}^{+\pi} \frac{dk}{2\pi} G_0^a(k)\Gamma_R(k)G_0^r(k)\Gamma_L(k)G_0^a(k) \right]_{ii} \right\}, \tag{33}$$

$$Y_{iq}^\beta = \text{Tr} \left\{ t_{iq}^r \left[\int_{-\pi}^{+\pi} \frac{dk}{2\pi} G_0^r(k)\Gamma_R(k)G_0^a(k)\Gamma_L(k)G_0^r(k) \right]_{ii} \right\}, \tag{34}$$

$$Y_{iq}^\gamma = \text{Tr} \left\{ t_{iq}^r \left[\int_{-\pi}^{+\pi} \frac{dk}{2\pi} G_0^r(k)\Gamma_L(k)G_0^a(k) \right]_{ii} t_{iq}^a \left[\int_{-\pi}^{+\pi} \frac{dk'}{2\pi} G_0^a(k')\Gamma_R(k')G_0^r(k') \right]_{ii} \right\}, \tag{35}$$

and

$$t_{iq}^r = \left[(\varepsilon_{iq} - \varepsilon_{i0})^{-1} - G_{0,ii}^r \right]^{-1} \quad (q > 0), \tag{36}$$

where $G_{0,ii}^r$ is obtained as

$$G_{0,ii}^r = \int_{-\pi}^{+\pi} \frac{dk}{2\pi} [G_0^r(k)]_{ii}.$$

VI. FURTHER DISCUSSIONS

Several important issues are worth further discussions including the scaling behavior of the transmission fluctuation δT, the comparison of CPA and LCA diagrammatic formalisms, the generalization of CPA and LCA to atomic models of nanoelectronics, the application of CPA and LCA to compute other physical quantities, and the procedure to determine the variation of threshold voltage for field effect transistors.

A. Scaling

In two-probe systems with transverse periodicity, transmission coefficient T and transmission fluctuation δT are calculated for a single unit cell in the transverse dimensions, as discussed in Sec. V. It should be emphasized that T and δT have very different scaling behaviors with respect to the cross section area. Suppose a cross section contains $\mathcal{N}$ unit cells in the transversion dimensions, and transmission is proportional to $\mathcal{N}$ but transmission fluctuation is proportional to $\sqrt{\mathcal{N}}$. In the limit of infinitely large transverse cross section, the ratio of δT over T goes to zero which is the thermodynamic limit. It is therefore clear that the device variability due to RDF is most significant in nanoscale systems whose cross section area is not sufficiently large to exhibit self-averaging of the disorder configurations.

B. CPA vs LCA

We have so far presented two diagrammatic formalisms, CPA and LCA, for calculating δT. A comparison of CPA and LCA is as follows. (i) In principle CPA is more accurate than LCA, because from the diagram point of view LCA only considers the lowest order diagrams while CPA considers all noncrossing diagrams to infinite order. As a result LCA is applicable to the low-concentration limit while CPA is applicable to a wider concentration range. Numerically we shall compare the two methods in Sec. VII. (ii) To apply CPA formalism to calculate δT, one has to carry out Γ decomposition to rewrite T^2 into a proper matrix product form [see Eq. (16)]. In contrast, the LCA formalism does not require Γ decomposition and can be applied directly to calculate δT. The Γ decomposition leads to double summation and double k integral [see Eq. (31)] over conducting channels of the electrode and significantly increase the computational cost. (iii) CPA is far more complicated to implement than LCA, because the former needs to solve the CPA equations as well as several vertex correction equations iteratively as discussed at the end of Sec. III G. In contrast, LCA provides an explicit formula, Eq. (27), to calculate the transmission fluctuation directly. (iv) To reduce the computational cost in modeling nanoelectronic devices, it is often desirable to partition a two-probe system into many slices along the transport direction and apply a numerical trick—the principal layer algorithm—in the Green's

function calculation.[22] This very useful algorithm can be easily integrated into the LCA formalism but it is incompatible with the CPA diagrams. In short, the CPA diagrammatic formalism is much more complicated and costly than LCA to calculate δT due to the reasons listed in (ii) to (iv), although CPA is more accurate and applicable to a wider concentration range.

C. Generalization to atomic model

It is straightforward to generalize both CPA and LCA formalisms to the atomic model of nanoelectronic devices. Assume that each atom is represented by M atomic orbitals; the on-site energy ε_{iq} should be replaced by an $M \times M$ matrix block. Correspondingly, the variables V_{iq}, t_{iq}^r, $G_{0,ii}^r$, $\tilde{\varepsilon}_{iq}^r$, Λ_i also become $M \times M$ matrix blocks. Meanwhile the formulation should be adapted according to the definition of the Green's functions in the specific method.

For example, in the first-principles model implementing the linear muffin-tin orbital (LMTO) method,[23–25] the on-site energy ε_{iq} should be replaced by the potential function $-P_{iq}(E)$ which is a $(L_{\max} + 1)^2 \times (L_{\max} + 1)^2$ diagonal matrix block where $L_{\max}$ is the maximum angular momentum quantum number. The coherent potential $\tilde{\varepsilon}_i^r$ should be replaced by the LMTO coherent potential $-\tilde{P}_i^r(E)$ which is a $(L_{\max} + 1)^2 \times (L_{\max} + 1)^2$ full matrix block. Moreover, the definition of the auxiliary Green's function in the LMTO method is very different from that of standard Green's function presented in Secs. III and IV, and hence the formulation need to be modified accordingly. In the CPA formalism, $E - T - \tilde{\varepsilon}^r$ in the fifth row of Eq. (A1) needs to be replaced by $\tilde{P}^r(E) - S(k)$ where $\tilde{P}^r(E)$ is the LMTO coherent potential and $S(k)$ is the Fourier-transformed structure constant. In the LCA formalism, $E - H_0$ in Eq. (10) should be replaced by $P_0(E) - S(k)$, where $P_0(E)$ is the potential function of the host material. For technical details of the LMTO method, we refer interested readers to the monographs of Refs. 23–25.

In this way, we have implemented a transmission fluctuation analyzer based on the LCA diagrammatic formalism and the LMTO method in the first-principles nanoscale device simulation package NANODSIM,[26] which will be applied in an example in Sec. VII.

D. Other physical quantities

In the NEGF approach, to calculate a physical quantity, the general idea is to first express the quantity in terms of Green's functions and then evaluate these Green's functions. In Ref. 27, disorder-averaged Green's functions $\overline{G^r}$ and $\overline{G^<}$ have been solved from the equations of the nonequilibrium coherent potential approximation (NECPA). Therefore if a quantity can be expressed as a linear combination of G^r and $G^<$, the disorder average of this quantity can be readily calculated with NECPA. It has been shown in Ref. 27 that electric current and occupation number belong to this category.

Some physical quantities, however, involve Green's function correlators which are beyond the scope of NECPA. The CPA and LCA formalisms presented in this work can systematically calculate disorder-averaged Green's function correlators and related physical quantities. In addition to transmission fluctuations studied here, CPA and LCA techniques can also

be applied to investigate other quantities. For example, the shot noise can be expressed as[16]

$$S = \text{Tr}[G^r \Gamma_L G^a \Gamma_R - (G^r \Gamma_L G^a \Gamma_R)^2],$$

and the disorder-averaged shot noise $\overline{S}$ can be readily evaluated with CPA or LCA formalism.

E. Variation of the threshold voltage

For field effect transistors it is relevant to predict the variation of threshold voltage in addition to the variations of on-state and off-state current due to RDF. This can be done with the following procedure. (i) Calculate the disorder-averaged current as a function of gate voltage $\overline{I} = F(V_g)$. (ii) Determine the averaged threshold voltage $\overline{V_T}$ from $F(V_g)$. (iii) Calculate the current fluctuation δI at $\overline{V_T}$ by using the CPA or LCA formalism of this work. (iv) Estimate the variation of the threshold voltage by the slope of $F(V_g)$ at $\overline{V_T}$:

$$\delta V_T \approx \frac{\delta I}{|F'(\overline{V_T})|}.$$

VII. NUMERICAL EXAMPLES

In this section, CPA and LCA formalisms are applied to tight-binding (TB) models and an atomic model to investigate transmission fluctuation induced by RDF. Three examples are provided: a TB model with finite cross section, a TB model with periodic transverse cross section, and an atomic model with periodic transverse cross section.

A. Tight-binding model: Finite cross section

This example investigates transmission fluctuation in a one-dimensional (1D) tight-binding nanoribbon. The system is shown in the inset of Fig. 11(a) where the yellow sites represent host sites whose on-site energies are set to zero. The red sites represent impurity sites whose on-site energies are either zero with the probability $1 - x$ or 0.5 with the probability x. Only the nearest neighbors have interactions with a coupling strength set to unity. Figure 11(a) also shows transmission coefficient $T(E)$ in the clean limit ($x = 0$). As expected, $T(E)$ is an integer steplike curve which coincides with the number of the conducting channels at the energy E.

For this simple example the exact solution is available by brute force enumeration. Namely, $T(E)$ can be calculated for all disorder configurations and δT can be evaluated exactly. This example sets a benchmark to check the validity and accuracy of CPA and LCA. In Fig. 11(b) to Fig. 11(h), δT is calculated by using three different methods: exact, LCA, and CPA. The disorder concentration is increased systematically from $x = 0.001$ to $x = 0.5$.

A few observations are in order. (1) For $x \leqslant 0.01$, both LCA and CPA give very satisfactory results in comparison to the exact solution. For $x \geqslant 0.2$, both LCA and CPA solutions become less accurate. The reason is that LCA neglects higher order terms of concentration x while CPA neglects crossing diagrams. (2) The LCA solution is always physical in the sense of $\delta T^2 > 0$ which is actually expected from Eq. (27). The
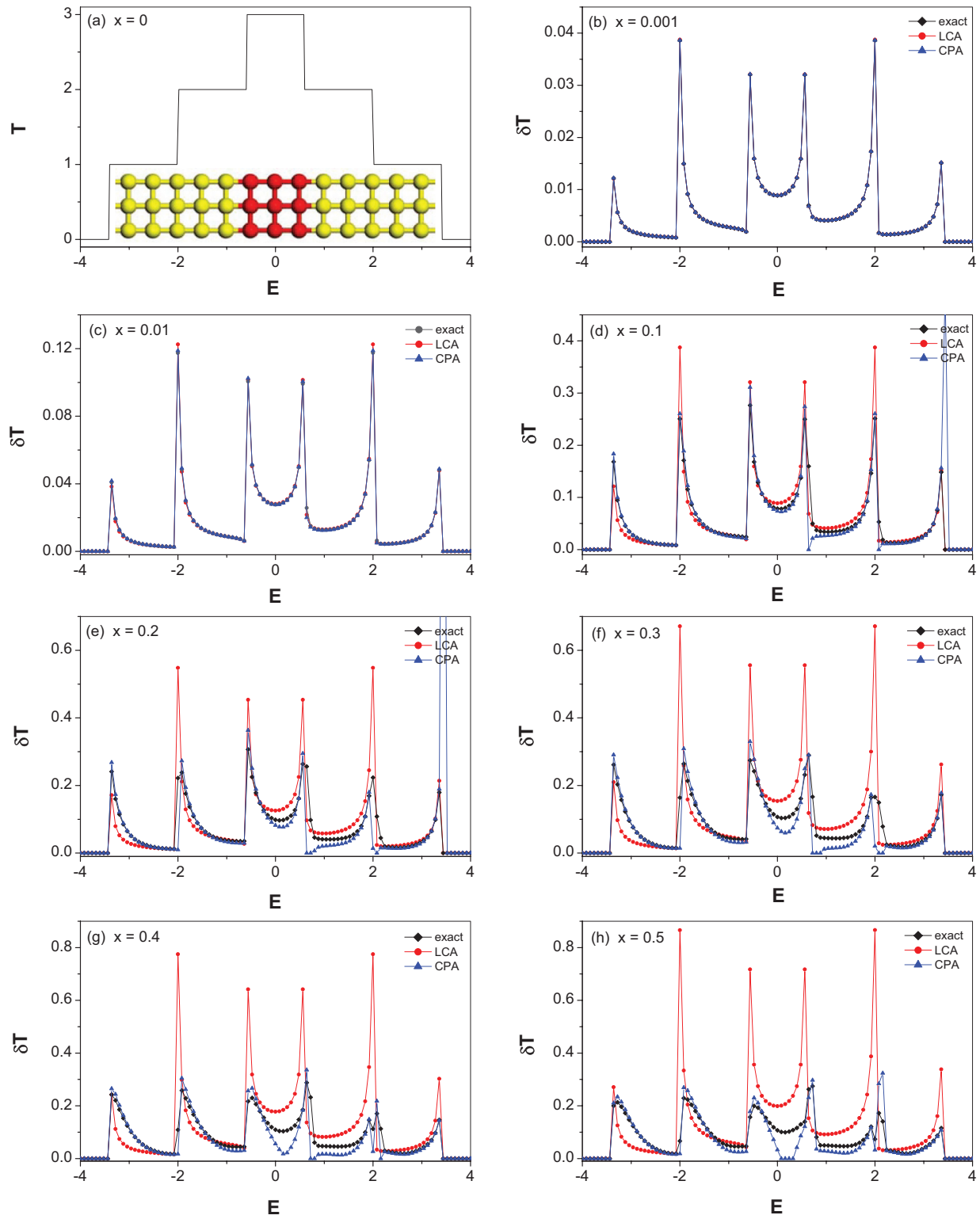
FIG. 11. (Color online) Transmission fluctuation in the tight-binding model with finite cross section shown in the inset of (a). (a) Transmission $T(E)$ in the clean limit. (b) to (h) Transmission fluctuation $\delta T(E)$ at different doping concentrations x. For comparison, δT is calculated with three methods: exact, LCA, and CPA.

CPA solution, however, may give nonphysical results in some energies where $\delta T^2 < 0$. In Fig. 11, the nonphysical points have been reset to $\delta T = 0$. (3) Large transmission fluctuation occurs at energies where the transmission channel number

changes drastically. It implies that current fluctuation can be suppressed if the bias voltage window is tuned to locate in an energy plateau with slowly varying conducting channel number.

B. Tight-binding model: Periodic cross section

This example investigates transmission fluctuation in a two-dimensional (2D) tight-binding lattice. The system is shown in the inset of Fig. 12(a) where the yellow sites represent host atoms whose on-site energies are set to zero; the red sites represent impurities whose on-site energies are either zero with the probability $1 - x$ or 0.5 with the probability x. Only the nearest neighbors have interactions with a coupling strength set to unity. Figure 12(a) also shows the transmission coefficient $T(E)$ in the clean limit ($x = 0$). $T(E)$ has a sharp peak at $E = 0$ which can be well understood by the corresponding band structure of this lattice.

For this example an exact solution is unavailable due to the infinite degrees of freedom. The CPA solution is very expensive due to double summation and double k integral in the Γ decomposition Eq. (31). Since the LCA solution of finite cross section has been checked in the previous subsection, the LCA solution of periodic cross section will be checked against it by using a large finite cross section containing 1000 rows.

For this 2D model, the LCA solution of the periodic cross section agrees very well with that of large finite cross section, as expected. Note that the solution for the finite cross section model must be rescaled with a proper scaling factor $\sqrt{1000}$ as discussed in Sec. VI A. The transmission fluctuation shows a sharp peak around $E = 0$ where the transmission also has

a spike. An impression is that the transmission fluctuation is more pronounced in the energy regime where the transmission coefficient changes rapidly.

C. Atomic model: Periodic cross section

This example investigates the transmission fluctuation in a three-dimensional (3D) Cu lattice having 1% random atomic vacancies by using an atomic implementation of the LCA formalism. The system has a periodic cross section and the transport is perpendicular to the Cu (111) direction. In the atomic model, the left and right semi-infinite Cu electrodes are connected to a central region which consists of 5 perfect Cu layers (buffer layer), 15 disordered Cu layers in the alloy model of $Cu_{0.99}Vac_{0.01}$ ("Vac" indicates vacancy), and another 5 perfect Cu layers (buffer layer). Namely, the central region can be represented by the formula $[Cu]_5$-$[Cu_{0.99}Vac_{0.01}]_{15}$-$[Cu]_5$.

The calculation proceeds in two steps. First, we self-consistently solve the device Hamiltonian of the open two-probe system using the NECPA-LMTO method as implemented in the NANODSIM package.[26] Second, we calculate the transmission fluctuation using the LCA formalism combined with the LMTO method which has been implemented into the NANODSIM package as a post-analysis tool. The result for $[Cu]_5$-$[Cu_{0.99}Vac_{0.01}]_{15}$-$[Cu]_5$ is presented in Fig. 13.
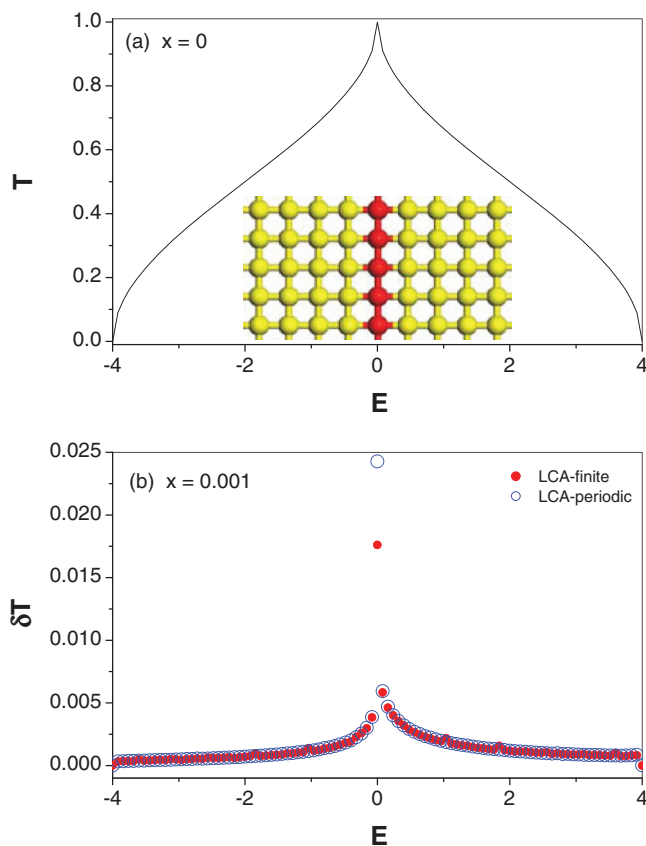

FIG. 12. (Color online) Transmission fluctuation in the tight-binding model with periodic cross section shown in the inset of (a). (a) Transmission $T(E)$ in the clean limit. (b) Transmission fluctuation $\delta T(E)$ for the doping concentration $x = 0.001$. For comparison, δT is calculated with two methods: LCA-finite and LCA-periodic.
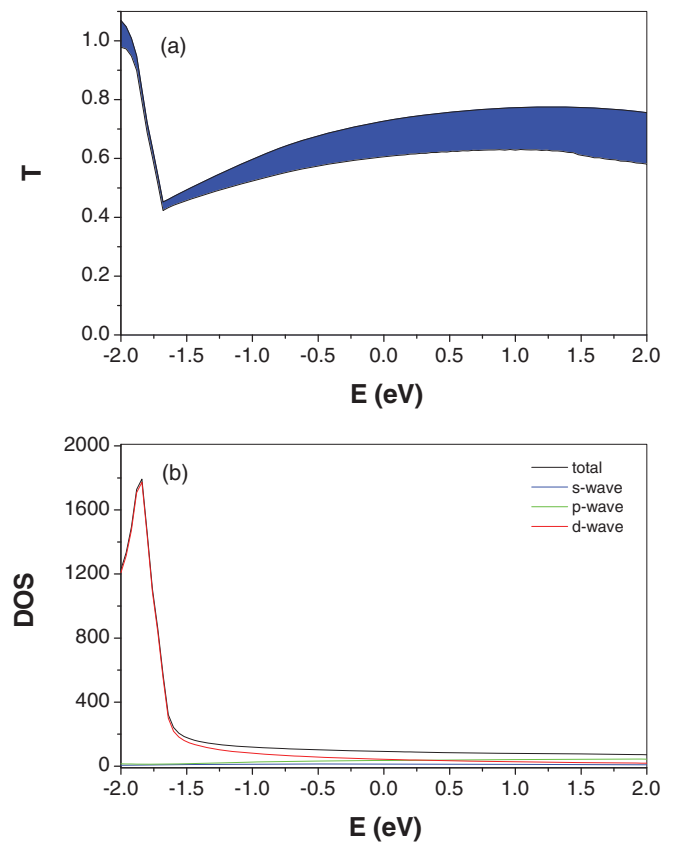
FIG. 13. (Color online) Transmission fluctuation of the 3D Cu two-probe lattice with 1% random vacancy defects. (a) Transmission fluctuation δT on top of transmission T versus energy E. The area of the unit cell cross section is 5.64 Å^2. (b) Total and angular momentum resolved density of states versus E.

The transmission fluctuation exhibits a strong energy dependence. This is quite interesting since it means the conductance fluctuation can be effectively suppressed by shifting the Fermi energy. In the vicinity of $E = -1.64$ eV, δT is rather small although T changes rapidly, which seems to be different than the observations in the tight-binding examples. Further analysis shows that the density of states is dominated by the d wave in the energy regime around $E = -1.64$ eV and is a mixture of s wave, p wave, and d wave well above this energy. Transmission fluctuation is enhanced due to disorder scattering among different angular momentum states. This indicates that the transmission fluctuation is not only affected by the number of the conducting channels (as in the TB models) but also by the angular momentum states of the channels when realistic atomic models are considered.

VIII. CONCLUSION

In this work, we have developed two theoretical formalisms based on CPA and LCA to predict device-to-device variability induced by random dopant fluctuation. The advantage of our theory is that statistical averaging due to RDF is carried out analytically to avoid a large number of dopant configuration samplings in device simulations.

The numerical accuracy of CPA and LCA formalisms depends on the doping concentration x. For $x \leqslant 0.01$, both CPA and LCA solutions are satisfactory, as shown in the comparison to the exact solution of the 1D TB model. For $x \geqslant 0.2$, both CPA and LCA become numerically less accurate even though they still capture a rough trend of the transmission fluctuation as demonstrated by the 1D TB model. In LCA we have neglected high order terms in the x expansion, while in CPA we have neglected the crossing diagrams. These approximations limit the accuracy of the theory to the relatively low impurity concentrations. We note that for essentially all the practical semiconductor devices, the dopant concentration is well within the applicability range of our formalisms. In numerical modeling, the LCA is easier and perhaps more practical for realistic nanoelectronic devices because an explicit formula Eq. (27) is available and the computational cost is much cheaper than that of CPA. We have also implemented the LCA theory into the first-principles device modeling package NANODSIM so that first-principles analysis of device-to-device variability can now be carried out without any phenomenological parameters.

Preliminary studies indicate that transmission fluctuation is most pronounced in the energy regime where the number of the conducting channels varies rapidly. In addition, angular momentum states of the conducting channels also play an essential role. Since the fluctuation strongly depends on the electron energy, our numerical simulation suggests that the RDF-induced transmission fluctuation could be suppressed by engineering the bias voltage window to a proper energy regime. Finally, we have so far focused on investigating the RDF-induced transmission fluctuation in nanostructures; our theory and numerical implementation can be applied to study many other physical quantities such as the shot noise, the fluctuation of threshold voltage, and the device variability in spintronics. We hope to report these and other investigations in future publications.

APPENDIX A: THE CPA CONDITION

In this Appendix, we present how to calculate quantities G_0^r, t_{iq}^r, G_0^a, t_{iq}^a by using the CPA condition Eq. (20). As mentioned in Sec. II, there is some freedom to partition H into H_0 and V, i.e., Eqs. (7) and (8). CPA takes the advantage of this freedom and chooses a special partition such that the disorder-averaged scattering vanishes, i.e., Eq. (20).

Assume that the Hamiltonian matrix is written as $H = T + \varepsilon$ where T is the off-diagonal part of the Hamiltonian and ε the diagonal part. T is a definite matrix and does not have any randomness. In contrast, the diagonal matrix ε contains discrete random variables; the ith diagonal element ε_i can take the value ε_{iq} with the probability x_{iq} and $\sum_q x_{iq} = 1$. One can introduce a diagonal quantity called coherent potential $\tilde{\varepsilon}^r \equiv \mathrm{diag}([\tilde{\varepsilon}_1^r, \tilde{\varepsilon}_2^r, \ldots])$ and define H_0 and V as

$$H_0 = T + \tilde{\varepsilon}^r, \quad V = \varepsilon - \tilde{\varepsilon}^r.$$

By imposing CPA condition Eq. (20) to the above partition of H_0 and V, $\tilde{\varepsilon}^r$ can be solved from the following CPA equations:

$$\overline{t_i^r} = \sum_q x_{iq} t_{iq}^r = 0, \quad t_{iq}^r = V_{iq}\big[1 - \overline{G_i^r} V_{iq}\big]^{-1},$$

$$V_{iq} = \varepsilon_{iq} - \tilde{\varepsilon}_i^r, \quad \overline{G_i^r} = [\overline{G^r}]_{ii}, \quad \text{(A1)}$$

$$\overline{G^r} = (E - T - \tilde{\varepsilon}^r - \Sigma^r)^{-1}.$$

Once $\tilde{\varepsilon}^r$ is solved, $G_0^r = \overline{G^r}$ and t_{iq}^r are also known from Eq. (A1). Finally, G_0^a and t_{iq}^a are simply Hermitian conjugates of G_0^r and t_{iq}^r, respectively.

APPENDIX B: PROOF OF EQUATION (23)

In this Appendix, we provide an analytical proof of Eq. (23). By using the vertex correction, the left-hand side of Eq. (23) can be obtained as

$$\overline{G^r \Sigma^{ra} G^a} = \overline{G^r}(\Sigma^{ra} + \Lambda)\overline{G^a}, \quad \text{(B1)}$$

where Λ is the vertex correction determined by Eq. (22) with $X = \Sigma^{ra}$. By using the expressions of $\overline{G^r}$ and $\overline{G^a}$ in CPA, the right-hand side of Eq. (23) can be transformed into a form similar to that of the left-hand side:

$$\overline{G^r} - \overline{G^a} = \overline{G^r}(\Sigma^{ra} + \tilde{\Lambda})\overline{G^a}, \quad \text{(B2)}$$

where $\tilde{\Lambda} \equiv \tilde{\varepsilon}^r - \tilde{\varepsilon}^a$. $\overline{G^r}$ and $\overline{G^a}$ are determined by the coherent potential $\tilde{\varepsilon}^r$ and $\tilde{\varepsilon}^a$ [see Eq. (A1)],

$$\overline{G^r} = (E - H_0 - \tilde{\varepsilon}^r - \Sigma^r)^{-1},$$

$$\overline{G^a} = (E - H_0 - \tilde{\varepsilon}^a - \Sigma^a)^{-1}.$$

Comparing Eq. (B1) and Eq. (B2), it is inferred that $\tilde{\Lambda}$ and Λ must be identical. Also note that the vertex correction Eq. (22) for Λ is an inhomogeneous linear equation and thus has a unique solution. Hence the identity is proved if and only if $\tilde{\Lambda}$ satisfies Eq. (22).

By using the CPA condition Eq. (A1) and its Hermitian conjugate

$$t_{iq}^r = \left[(V_{iq} - \tilde{\varepsilon}^r)^{-1} - \overline{G_i^r} \right]^{-1},$$
$$t_{iq}^a = \left[(V_{iq} - \tilde{\varepsilon}^a)^{-1} - \overline{G_i^a} \right]^{-1},$$

one can derive the equation for $\tilde{\Lambda}$ by eliminating V_{iq}

$$\left(t_{iq}^{a-1} + \overline{G_i^a} \right)^{-1} - \left(t_{iq}^{r-1} + \overline{G_i^r} \right)^{-1} = \tilde{\Lambda}. \qquad (B3)$$

After some algebra, the equation of $\tilde{\Lambda}$ can be simplified as

$$t_{iq}^a - t_{iq}^r + t_{iq}^r \left(\overline{G_i^r} - \overline{G_i^a} \right) t_{iq}^a$$
$$= \left(1 + t_{iq}^r \overline{G_i^r} \right) \tilde{\Lambda}_i \left(1 + \overline{G_i^a} t_{iq}^a \right). \qquad (B4)$$

By using Eq. (B2), one obtains

$$t_{iq}^a - t_{iq}^r + t_{iq}^r [\overline{G^r}(\Sigma^{ra} + \tilde{\Lambda})\overline{G^a}]_{ii} t_{iq}^a$$
$$= \left(1 + t_{iq}^r \overline{G_i^r} \right) \tilde{\Lambda}_i \left(1 + \overline{G_i^a} t_{iq}^a \right). \qquad (B5)$$

Notice that $\sum_q x_{iq} = 1$ due to normalization, and $\sum_q x_{iq} t_{iq}^r = 0$ and $\sum_q x_{iq} t_{iq}^a = 0$ due to the CPA condition. Applying the weighed summation $\sum_q x_{iq}$ on both sides of Eq. (B5), one derives

$$\sum_q x_{iq} t_{iq}^r [\overline{G^r}(\Sigma^{ra} + \tilde{\Lambda})\overline{G^a}]_{ii} t_{iq}^a$$
$$= \tilde{\Lambda}_i + \sum_q x_{iq} t_{iq}^r \overline{G_i^r} \tilde{\Lambda}_i \overline{G_i^a} t_{iq}^a, \qquad (B6)$$

which is equivalent to Eq. (22) and thus proves Eq. (23).

---

*eric@nanoacademic.com

[1] International Technology Roadmap for Semiconductors, http://public.itrs.net/.

[2] A. Asenov, IEEE Trans. Electron Devices **45**, 2505 (1998).

[3] A. Brown, A. Asenov, and J. Watling, IEEE Trans. Nanotechnol. **1**, 195 (2002).

[4] R. Wang *et al.*, IEEE Trans. Electron Devices **58**, 2317 (2011).

[5] See, for example, the *Special Issue on Characterization of Nano CMOS Variability by Simulation and Measurements*, IEEE Trans. Electron Devices **58**, 2190 (2011).

[6] D. Reid *et al.*, IEEE Trans. Electron Devices **58**, 2257 (2011).

[7] A. Martinez *et al.*, IEEE Trans. Electron Devices **58**, 2209 (2011).

[8] J. Taylor, H. Guo, and J. Wang, Phys. Rev. B **63**, 121104(R) (2001).

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

[10] P. Soven, Phys. Rev. **156**, 809 (1967).

[11] D. W. Taylor, Phys. Rev. **156**, 1017 (1967).

[12] B. Velický, Phys. Rev. **184**, 614 (1969).

[13] K. Levin, B. Velický, and H. Ehrenreich, Phys. Rev. B **2**, 1771 (1970).

[14] K. Carva, I. Turek, J. Kudrnovsky, and O. Bengone, Phys. Rev. B **73**, 144421 (2006).

[15] Y. Ke, K. Xia, and H. Guo, Phys. Rev. Lett. **100**, 166805 (2008).

[16] J. N. Zhuang and J. Wang, J. Appl. Phys. **114**, 063708 (2013).

[17] S. Washburn and R. A. Webb, Adv. Phys. **35**, 375 (1986); B. L. Altshuler, Pis'ma Zh. Eksp. Teor. Fiz. **41**, 530 (1985) [JETP Lett. **41**, 648 (1985)]; P. A. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985).

[18] H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductor* (Springer-Verlag, Berlin, 1996).

[19] J. Wang and H. Guo, Phys. Rev. B **79**, 045119 (2009).

[20] The linewidth function Γ is a positive-definite Hermitian matrix due to its physical meaning. By eigenvalue decomposition, Γ can be factorized as $\Gamma = \Phi \Lambda \Phi^\dagger$ with $\Phi \Phi^\dagger = 1$, where $\Phi = (|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_N\rangle)$ is the eigenvector matrix and $\Lambda = \text{diag}([\lambda_1, \lambda_2, \ldots, \lambda_N])$ is the eigenvalue matrix. Define $|W_n\rangle \equiv \sqrt{\lambda_n}|\phi_n\rangle$; it follows that $\Gamma = \sum_n |W_n\rangle\langle W_n|$.

[21] For systems with periodicity in two transverse dimensions, k is defined as $(k_1, k_2) = (\mathbf{k} \cdot \mathbf{a}_1, \mathbf{k} \cdot \mathbf{a}_2)$ where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the two unit cell vectors of the periodic transverse dimensions. Correspondingly, $\int_{-\pi}^{+\pi} \frac{dk}{2\pi}$ should be understood as $\int_{-\pi}^{+\pi} \frac{dk_1}{2\pi} \int_{-\pi}^{+\pi} \frac{dk_2}{2\pi}$.

[22] D. Waldron, L. Liu, and H. Guo, Nanotechnology **18**, 424026 (2007).

[23] O. K. Anderson, O. Jepsen, and D. Glötzel, in *Highlights of Condensed-Matter Theory*, edited by F. Bassani, F. Fumi, and M. P. Tosi (North-Holland, New York, 1985).

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

[25] I. Turek *et al.*, *Electronic Structure of Disordered Alloys, Surfaces, and Interfaces* (Kluwer Academic Publishers, 1997).

[26] For software information of NANODSIM, please visit http://www.nanoacademic.com. For technical details of NANODSIM, please see Ref. 15 and Ref. 27.

[27] Y. Zhu, L. Liu, and H. Guo, arXiv:1305.2515 [cond-mat.mes-hall].