Analytical study of edge states in a semi-infinite graphene nanoribbon

Liwei Jiang, Yisong Zheng,* Cuishan Yi, Haidong Li, and Tianquan Lü
Department of Physics, Jilin University, Changchun 130023, China
(Received 5 July 2009; revised manuscript received 7 August 2009; published 28 October 2009)

By means of the transfer-matrix technique, we present an analytical solution of the edge states localized at the lateral zigzag edge of a semi-infinite graphene nanoribbon. The electric field tuning on the energy level, the localized length, and the local electron probability distribution of an edge state is then studied in detail. The dependence of the edge state on the size of the ribbon, the presence of impurities, and the structural variation in the lateral edge is discussed. The physical natures of some previous numerical conclusions about the edge state are clarified. For example, it was previously expected that any edge state cannot survive while the width of the graphene nanoribbon becomes smaller than three times of the lattice constant and whenever such a width increases by triple lattice constants, one more edge state is added. The physical reasons of these issues can be intuitively seen in our analytical treatment on the edge states.

DOI: 10.1103/PhysRevB.80.155454 PACS number(s): 81.05.Uw, 71.55.−i, 73.23.−b, 73.25.+i

I. INTRODUCTION

Graphene, an atomically thin graphitic sheet,[1] is currently a research-intensive subject since it presents many unique electronic properties.[2–4] For example, the existence of localized states at zigzag edge is a typical electron characteristic of graphene structures.[5–16] This kind of edge state was theoretically anticipated[5,6] much earlier than the first acquirement of graphene sample.[1] And it is experimentally observed recently.[8,9] Numerous works demonstrate that the edge state appears commonly in graphene single[5–7] and multiple layers,[10,11] graphene nanostructures[12] or around a topological defect,[13] as long as an edge is terminated in a zigzag type. Many interesting phenomena in graphene prove to be intimately associated with the edge state, e.g., the valley-filtered transport[14] and magnetism.[12,15,16]

Although edge states in graphene have drawn much attention, some relevant issues deserve further investigation. For example, the size and edge-disorder dependence of the edge states in graphene nanostructures, the tunability of the energy level and localization length of an edge state by an exerted electric or magnetic field. Recently, the tight-binding and *ab initio* calculations agree on that an edge state appear only if the width of the zigzag edge exceeds three multiple of the lattice constant.[6,12,16] But a clear physical explanation about this numerical conclusion is yet lacking. In addition, to study the size dependence of edge states in graphene nanoribbon, the periodic boundary condition (PBC) is often employed to simplify the theoretical treatment.[6,7] However, such a boundary condition is only appropriate to a nanotube. Its applicability to a graphene nanoribbon needs a detailed inspection.

In this paper, we derive an analytical solution about the edge states localized at the lateral zigzag edge of a semi-infinite armchair-edged graphene nanoribbon (SIGNR), based on the transfer-matrix theory. Thus, the external field tuning, the size dependence, and edge-disorder effect on the edge state can be studied in a unified way. In addition, we will adopt two distinct kinds of boundary condition to describe the electron confinement in the transverse direction. The first, called the natural boundary condition (NBC), is to assume that just beyond the upper and bottom armchair edges there are the hard walls where the electron probability amplitudes vanish. The NBC has succeeded in describing the subband structures of infinite graphene nanoribbons.[17] The other one is the aforementioned PBC. It assumes that the uppermost and downmost atoms in one zigzag line in the vertical direction are exactly the same one. In fact, under the PBC, the graphene nanoribbon is equivalent to a carbon nanotube. Obviously, this is not a reasonable treatment, especially for the small size ribbon, though it is often employed to study the edge states.[6,7] By means of the transfer-matrix technique, we derive an analytical solution about the edge states. Thereby the quantitative difference of the edge state properties obtained by the two boundary conditions can be seen clearly. Meanwhile, the physical picture about the edge states can be seen intuitively.

II. TRANSFER MATRIX AND EDGE STATES IN A PERFECT SIGNR

The honeycomb lattice of the semi-infinite graphene nanoribbons considered by us is illustrated in Fig. 1(a). Its width is characterized by N, the number of the closed hexagons in a periodic unit. To explore the edge states localized in the vicinity of the lateral zigzag edge, we attempt to solve the Schrödinger equation $\mathcal{H}\boldsymbol{C}=E\boldsymbol{C}$ in the tight-binding representation, where E is the electron eigenenergy and the Hamiltonian in the nearest-neighbor approximation is given by

$$\mathcal{H}=\sum_{\mu}\varepsilon_{\mu}|\mu\rangle\langle\mu|+\sum_{\langle\mu,\mu'\rangle}(-t)|\mu\rangle\langle\mu'|, \tag{1}$$

where ε_{μ} denotes the on-site energy of a carbon atom labeled by the index μ, $|\mu\rangle$ stands for the Wannier state localized at the μth atom, the notation $\langle\mu,\mu'\rangle$ means the summation is restricted within the nearest-neighbor atomic pairs, and t denotes the corresponding hopping energy. The wave function ($\boldsymbol{C}$) is a column matrix consisting of all the electron probability amplitudes, denoted by c_{ln}^{L}, appearing at every atom. As labeled in Fig. 1(a), any atomic position is exclusively determined by the set of index (Lln), i.e., $\mu=(Lln)$. Thus the original Schrödinger equation can be separated into the following set of equations:

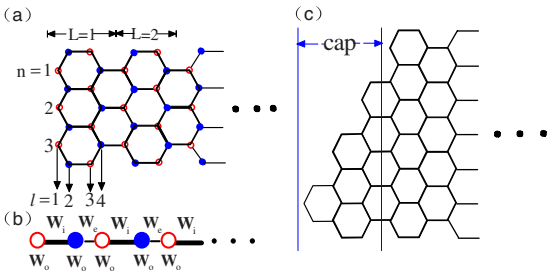

FIG. 1. (Color online) (a) Lattice structure of the semi-infinite armchair-edged graphene nanoribbon with a lateral zigzag edge. Each unit labeled by L has four layers denoted by $l=1,\ldots,4$. The index n labels the individual atom in a specific layer. The atoms belonging to two distinct sublattices (A and B) are indicated by ○ and ●, respectively. The lattice constant is denoted by a. (b) The reduced lattice where only the atoms in $l=1$ and 4 layers are left. W_α with $\alpha=i$, e, or o stands for the interaction matrix among these atoms. (c) An all-armchair-edged ribbon formed by prefixing a cap to the leftmost zigzag edge of the SIGNR as shown in (a).

$$[H_{11}^{11}](c_1^1)+[H_{12}^{11}](c_2^1)=0,$$

$$[H_{21}^{11}](c_1^1)+[H_{22}^{11}](c_2^1)+[H_{23}^{11}](c_3^1)=0,$$

$$[H_{32}^{11}](c_2^1)+[H_{33}^{11}](c_3^1)+[H_{34}^{11}](c_4^1)=0,$$

$$[H_{43}^{11}](c_3^1)+[H_{44}^{11}](c_4^1)+[H_{41}^{12}](c_1^2)=0,$$

$$[H_{14}^{21}](c_4^1)+[H_{11}^{22}](c_1^2)+[H_{12}^{22}](c_2^2)=0,$$

$$\ldots, \tag{2}$$

where the column matrix (c_l^L) consists of the electron probability amplitudes in the given layer (Ll) (below we can likewise use the notation (c^L) to denote the column matrix of the electron probability amplitudes in the Lth unit cell. As done above, we always use the notations $[\cdots]$ and $(\cdots)$ to denote, respectively, a matrix and a column matrix in expressions concerned for clarity). $[H_{ll'}^{LL'}]$ stands for a matrix describing the interaction between (Ll) and $(L'l')$ layers, whose matrix elements are given by $[H_{ll'}^{LL'}]_{nn'}=\langle Lln|(E-\mathcal{H})|L'l'n'\rangle$. Following such a definition, to take the $N=3$ SIGNR shown in Fig. 1(a), for example, under PBC we have

$$[H_{ll}^{LL}]=EI(3), \tag{3}$$

$$[H_{14}^{LL+1}]=[H_{23}^{LL}]=tI(3), \tag{4}$$

$$[H_{12}^{LL}]=[H_{43}^{LL}]=t\begin{bmatrix}1&1&0\\0&1&1\\1&0&1\end{bmatrix}, \tag{5}$$

On the other hand, these matrixes under NBC take forms as

$$[H_{11}^{LL}]=[H_{44}^{LL}]=EI(3), \tag{6}$$

$$[H_{22}^{LL}]=[H_{33}^{LL}]=EI(4), \tag{7}$$

$$[H_{23}^{LL}]=tI(4), \tag{8}$$

$$[H_{41}^{LL+1}]=tI(3), \tag{9}$$

$$[H_{12}^{LL}]=[H_{43}^{LL}]=t\begin{bmatrix}1&1&0&0\\0&1&1&0\\0&0&1&1\end{bmatrix}. \tag{10}$$

In above $I(N)$ denotes an $N\times N$ unity matrix. In addition, the Hermitian nature of the Hamiltonian guarantees the relationship $[H_{ll'}^{LL'}]^\dagger=[H_{l'l}^{L'L}]$.

According to the transfer-matrix theory,[18,19] we seek to establish a transfer matrix $[T]$ which connects the electron probability amplitudes between the adjacent unit cells. In doing so we start from the following relation:

$$\begin{pmatrix}c_3^1\\c_2^1\end{pmatrix}=\begin{bmatrix}-[H_{23}^{11}]^{-1}[H_{22}^{11}], & -[H_{23}^{11}]^{-1}[H_{21}^{11}]\\1, & 0\end{bmatrix}\begin{pmatrix}c_2^1\\c_1^1\end{pmatrix}. \tag{11}$$

By the same token we can then establish the connection of the wave functions between the layer of $(L=1,\ l=4)$ and its anterior layers, and finally we obtain the transfer matrix $[T]$ which connects the electron probability amplitudes in a way as $(c_2^2,c_1^2)^T=[T](c_2^1,c_1^1)^T$. However, such a plausible scheme fails as applied to the SIGNR shown in Fig. 1(a) since one cannot invert some Hamiltonian matrixes such as $[H_{12}^{22}]$ under NBC.

To circumvent such an embarrassment, we start our formulation from an equivalent lattice illustrated in Fig. 1(b), rather than the original one of the SIGNR. Such a reduced lattice is realized by working out the indirect interaction between (L,1) and (L,4) layers, thereby the layers (L,2) and (L,3) are dropped. Then the electron Schrödinger equation evolves into a set of recursive equation,

$$[W_{11}^{11}](c_1^1)+[W_{14}^{11}](c_4^1)=0,$$

$$[W_{41}^{LL}](c_1^L)+[W_{44}^{LL}](c_4^L)+[W_{41}^{LL+1}](c_1^{L+1})=0,$$

$$[W_{14}^{L+1L}](c_4^L)+[W_{11}^{L+1L+1}](c_1^{L+1})+[W_{14}^{L+1L+1}](c_4^{L+1})=0 \tag{12}$$

with $L=1,2,\ldots$ By a straightforward evaluation we obtain

$$[W_{11}^{11}]=[H_{11}^{11}]-[H_{12}^{11}][H_{22}^{11}]^{-1}[H_{21}^{11}]-[H_{12}^{11}][H_{22}^{11}]^{-1}[H_{23}^{11}]\times[F]^{-1}[H_{32}^{11}][H_{22}^{11}]^{-1}[H_{21}^{11}],$$

$$[W_{44}^{11}]=[H_{44}^{11}]-[H_{43}^{11}][F]^{-1}[H_{34}^{11}],$$

$$[W_{41}^{12}]=[H_{41}^{12}],$$

$$[W_{14}^{11}]=[H_{12}^{11}][H_{22}^{11}]^{-1}[H_{23}^{11}][F]^{-1}[H_{34}^{11}],$$

$$\ldots, \tag{13}$$

where $[F]=[H_{33}^{11}]-[H_{32}^{11}][H_{22}^{11}]^{-1}[H_{23}^{11}]$. For the perfect SIGNR, we find there are only three different matrixes which appear in the Schrödinger equation of the reduced lattice

illustrated by Fig. 1(b). Thus Eq. (13) reduces to

$$[W_o](c_1^1) + [W_i](c_4^1) = 0,$$

$$[W_i](c_1^1) + [W_o](c_4^1) + [W_e](c_1^2) = 0,$$

$$[W_e](c_4^1) + [W_o](c_1^2) + [W_i](c_4^2) = 0,$$

$$\cdots \tag{14}$$

In the case of NBC the interaction matrixes in the above equations are defined as $[W_e]=tI(N)$, $[W_i]=t^3(E^2-t^2)^{-1}[M]$, and $W_o=EI(N)-t^2E(E^2-t^2)^{-1}[M]$, where the matrix $[M]$ is given by

$$[M] = \begin{bmatrix} 2 & 1 & 0 & \cdots & \delta \\ 1 & 2 & 1 & \cdots & 0 \\ 0 & 1 & 2 & & 0 \\ & & & \ddots & \\ \delta & 0 & 0 & \cdots & 2 \end{bmatrix}_{N\times N} \tag{15}$$

with $\delta=0$ and 1 for the NBC and PBC, respectively.

One can find out that under the NBC and PBC the matrix $[M]$ is equivalent to the tight-binding Hamiltonian of an N-atomic chain and ring, respectively. By virtue of such a mapping, we can solve analytically the eigenproblems of the matrix $[M]$. Under NBC we have

$$[M](\boldsymbol{\xi}_j) = \lambda_j(\boldsymbol{\xi}_j) \tag{16}$$

with the component of the eigenvector defined as

$$(\boldsymbol{\xi}_j)_n = \sqrt{\frac{2}{N+1}}\sin\left(\frac{j\pi n}{N+1}\right) \quad \text{for} \quad n=1,\ldots,N. \tag{17}$$

And corresponding to each eigenvector $(\boldsymbol{\xi}_j)$, the eigenvalue is

$$\lambda_j = 4\cos^2\left(\frac{j\pi}{2N+2}\right). \tag{18}$$

The above eigenequation can be rewritten as $[\boldsymbol{\xi}]^\dagger[M][\boldsymbol{\xi}]=[\boldsymbol{\lambda}]$, where $[\boldsymbol{\xi}]$, being an $N\times N$ matrix, is defined as $[\boldsymbol{\xi}]=[(\boldsymbol{\xi}_1),\ldots,(\boldsymbol{\xi}_N)]$. And $[\boldsymbol{\lambda}]=\mathrm{diag}[\lambda_1,\ldots,\lambda_N]$ is a diagonal matrix, consisting of all the eigenvalues of M. In contrast, under PBC the eigensolution of $[M]$ is given by

$$(\boldsymbol{\xi}_j)_n = \sqrt{\frac{1}{N}}\exp\left(i\frac{2j\pi n}{N}\right) \tag{19}$$

and

$$\lambda_j = 4\cos^2\left(\frac{j\pi}{N}\right). \tag{20}$$

Noting that all the interaction matrixes can be diagonalized in a way as $[\boldsymbol{\xi}]^\dagger[W_\alpha][\boldsymbol{\xi}]=[\tilde{W}_\alpha]$ with $\alpha=i,e$, or o, we can diagonalize the recursive electron equations of the reduced lattice given above. It yields

$$\tilde{w}_o(j)\tilde{c}_1^1(j) + \tilde{w}_i(j)\tilde{c}_4^1(j) = 0,$$

$$\tilde{w}_i(j)\tilde{c}_1^1(j) + \tilde{w}_o(j)\tilde{c}_4^1(j) + \tilde{w}_e(j)\tilde{c}_1^2(j) = 0,$$

$$\tilde{w}_e(j)\tilde{c}_4^1(j) + \tilde{w}_o(j)\tilde{c}_1^2(j) + \tilde{w}_i(j)\tilde{c}_4^2(j) = 0,$$

$$\cdots \tag{21}$$

for $j=1,\ldots,N$. The coupling coefficients are given by $\tilde{w}_e(j)=t$, $\tilde{w}_i(j)=t^3(E^2-t^2)^{-1}\lambda_j$, and $\tilde{w}_o(j)=E-t^2E(E^2-t^2)^{-1}\lambda_j$. Notice that $\tilde{c}_l^L(j)=(\boldsymbol{\xi}_j)^T(c_l^L)$ is a number, rather than a column vector. Evidently, corresponding to a given eigenmode of the matrix $[M]$, Eq. (21) describes a one-dimensional tight-binding diatomic chain.

From Eq. (21) we can readily establish a 2×2 transfer matrix $[T_j]$ which connects the wave-function components of the adjacent units in a way as $(\tilde{c}^{L+1})=[T_j](\tilde{c}^L)$ with the shorthand notation $(\tilde{c}^L)=(\tilde{c}_4^L,\tilde{c}_1^L)^T$. It is given by

$$[T_j] = \begin{bmatrix} \tilde{w}_i^{-1}\tilde{w}_o^2\tilde{w}_e^{-1} - \tilde{w}_i^{-1}\tilde{w}_e, & \tilde{w}_o\tilde{w}_e^{-1} \\ -\tilde{w}_o\tilde{w}_e^{-1}, & -\tilde{w}_i\tilde{w}_e^{-1} \end{bmatrix}. \tag{22}$$

Noting that we will often drop the j index in the relevant quantities if it does not cause any confusion. To solve the eigenproblem of the transfer matrix, we obtain $[T_j][V]=[V][\boldsymbol{\nu}]$ with $[V]=[\boldsymbol{v}_1,\boldsymbol{v}_2]$ and $[\boldsymbol{\nu}]=\mathrm{diag}[\nu_1,\nu_2]$, where $\boldsymbol{v}_i$ and ν_i with $i=1,2$ are the eigenvectors and eigenvalues of the transfer matrix. It should be emphasized that both $\boldsymbol{v}_i$ and ν_i depend on the index j and the electron energy E. From the eigenequation of $[T_j]$ and the relation $(\tilde{c}^{L+1})=[T_j]^L(\tilde{c}^1)$, we obtain $(\phi^{L+1})=(\boldsymbol{\nu})^L(\phi^1)$ with $(\phi^L)=[V]^{-1}(\tilde{c}^L)=(\phi_1^L,\phi_2^L)^T$. Alternatively, we can write it as $\phi_i^{L+1}=\nu_i^L\phi_i^1$. Up to now we can see that the essential condition for the existence of a localized edge state is that at least one of the two eigenvalues, say ν_1, obeys $|\nu_1|<1$. Moreover, from the definition of T_j we can readily prove that the condition $|\nu_1|<1$ leads consequentially to $|\nu_2|>1$. Thus, the existence of an edge state implies a constraint $\phi_2^1=0$, owing to the requirement of the wave-function convergence, which can be clearly seen from $\phi_2^{L+1}=\nu_2^L\phi_2^1$ when $L\to\infty$. The constraint is equivalent to $[V]_{21}^{-1}\tilde{c}_4^1+[V]_{22}^{-1}\tilde{c}_1^1=0$, where $[V]_{nn}^{-1}$, denotes a specific element of matrix $[V]^{-1}$. To combine this constraint with the first relation in Eq. (21) we have

$$\begin{bmatrix} [V]_{21}^{-1}, & [V]_{22}^{-1} \\ \tilde{w}_i, & \tilde{w}_o \end{bmatrix}\begin{pmatrix} \tilde{c}_4^1 \\ \tilde{c}_1^1 \end{pmatrix} = [D]\begin{pmatrix} \tilde{c}_4^1 \\ \tilde{c}_1^1 \end{pmatrix} = 0. \tag{23}$$

Then we find that $\det[D]=0$ is the sufficient condition for the existence of an edge state. Our calculation indicates that such a condition can be satisfied only at the Dirac point. In other words, all possible edge states only appear at $E=0$, regardless of the width of the ribbon. Such a conclusion is consistent with the prediction by the Dirac equation,[20] from which a transcendental equation, $q=p\tanh(qL_y)$, is obtained for the possible edge state in a finite graphene segment. When $L_y\to\infty$ it changes into our semi-infinite structure. As a result, we can infer that $E=\gamma\sqrt{p^2-q^2}=0$.

To solve the edge state at the Dirac point, the derivation becomes very simple since the transfer matrix is now diagonal $[T_j]=\mathrm{diag}[\lambda_j^{-1},\lambda_j]$. By a straightforward derivation, we

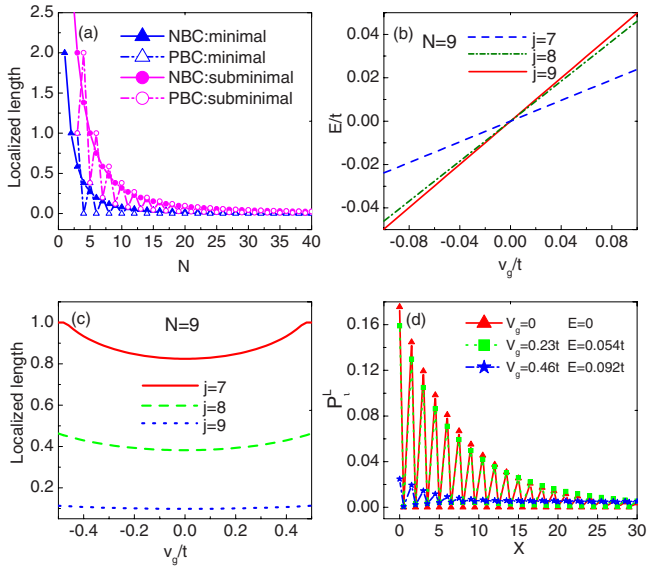

FIG. 2. (Color online) (a) Comparison of the minimal and subminimal localized lengths corresponding to two edge states calculated by using the NBC and PBC, respectively. (b) Edge state levels of the $N=9$ ribbon tuned by a gate voltage. (c) Localized lengths of the edge states as a function of the gate voltage for the $N=9$ ribbon. (d) The electron probability distribution $P_l^L=\Sigma_n|c_l^L|^2$ in some layers near the zigzag edge as a function of the distance of the layer away from the zigzag edge (denoted by x in units of C-C bond length).

find that corresponding to $|\lambda_j|<1$ there is an edge state at $E=0$, whose analytical wave function is given by

$$(\boldsymbol{C}(j))=\begin{cases}(c_1^1)=\sqrt{1-\lambda_j}(\boldsymbol{\xi}_j)\\(c_1^L)=\lambda_j(c_1^{L-1})\\(c_3^L)=-(H_{32}^{LL})^{-1}(H_{21}^{LL})(c_1^L)\\(c_2^L)=(c_4^L)=0.\end{cases}. \tag{24}$$

It can be seen that the localization length and the electron probability amplitude of an edge state are characterized, respectively, by the eigenvalue λ_j and eigenvector $(\boldsymbol{\xi}_j)$ of the matrix $[M]$ or $[M']$. Whenever more than one eigenvalue satisfy $|\lambda_j|<1$, the edge states at $E=0$ become degenerate. But with the relation $[\boldsymbol{\xi}]^\dagger[\boldsymbol{\xi}]=1$ one can readily prove the orthogonality of these degenerate edge states, namely, $(\boldsymbol{C}(j))^\dagger(\boldsymbol{C}(j'))=\delta_{jj'}$.

With the well-developed formula given above, we can clarify some issues about the edge states. First of all, we can assert that the PBC is inappropriate to describe the edge states in a narrow ribbon. For example, for $N=5$ ribbon the NBC gives only one edge state but the PBC says there are two edge states. A quantitative comparison is made in Fig. 2(a) where the minimal and the subminimal localized lengths corresponding to two specific edge states are plotted for the NBC as well as PBC. When $N<10$ the difference between the two cases is very notable. Besides, for any even N the PBC always gives an edge state with $\lambda_j=0$. It is evidently an unreasonable result. Second, we find that only when $N\geq 3$ an eigenvalue of $[M]$ satisfying $|\lambda_j|<1$ begins to appear. This indicates that it is impossible for an edge state to exist in a narrower ribbon, which accounts for unambiguously the relevant results from the first-principle calculations.[12,16] In addition, we find that the edge states begin to be degenerate when $N\geq 6$. And a new edge state is added whenever N increases by 3. A peculiar case is that when $N+1$ being a multiple of 3, the eigenvalue $\lambda_j=1$ appears. This implies that an extended state exists at the Dirac point. Consequently, such a ribbon is gapless, hence it presents metallic behavior. This result agrees completely with the previous theoretical prediction.[6,17] However, density-functional calculation performed recently claimed that the armchair graphene nanoribbons always possess finite band gap,[21,22] regardless of the ribbon width. This is due to the C-C bond distortion at the armchair edges. It has been demonstrated that such an effect can be modeled by the tight-binding approach by simply adjusting the hopping energy between the adjacent carbon atoms at the armchair edges, slightly different from the interior ones. As a result, the nonzero band gap is produced by the tight-binding approach, which is quantitatively consistent with the density-functional result. In Sec. IV, we will discuss the effect of the C-C bond modification at the armchair edges on the edge states.

III. EDGE GREEN'S FUNCTIONS OF THE SIGNR

Following the method developed above, we can work out the analytical expressions about the Green's functions located at the lateral zigzag edge of the SIGNR, referred to as the edge Green's functions hereafter. The edge Green's functions are very useful for the study of electronic transport through some graphene nanodevices when the present SIGNR acts as the leads. These Green's functions are associated with the self-energy of the device Green's function produced by the semi-infinite graphene leads,[23,24] which is responsible to the finite linewidth of the resonance peaks in the conductance spectrum. Usually, they are evaluated by means of the recursive method which is much time consuming for the relatively large size structures. Now we try to derive the analytical results about the edge Green's functions. The Green's function of the SIGNR, defined formally as $\hat{G}(E)=[E-\mathcal{H}]^{-1}$, takes a matrix form in the atomic representation. The matrix element is denoted as $[G(E)_{ll'}^{LL'}]_{nn'}=\langle Lln|G(E)|L'l'n'\rangle$. We can thus define the Green's function submatrix, $[G(E)_{ll'}^{LL'}]$, between two layers (Ll) and $(L'l')$. And the edge Green's functions refer to the elements of the submatrix concerning the zigzag edge, i.e., $[G(E)_{11}^{11}]_{nn'}$. By a straightforward derivation we find the Green's function submatrixes satisfy the following equations, similar to Eq. (14)

$$[W_o][G_1^1]+[W_i][G_4^1]=1,$$

$$[W_i][G_1^1]+[W_o][G_4^1]+[W_e][G_1^2]=0,$$

$$[W_e][G_4^1]+[W_o][G_1^2]+[W_i][G_4^2]=0,$$

$$\ldots, \tag{25}$$

where we have used the abbreviation $[G(E)_{l1}^{L1}]=[G_l^L]$. When we diagonalize matrix $[W_\alpha]$ into $[\boldsymbol{\xi}]^\dagger[W_\alpha][\boldsymbol{\xi}]=[\tilde{W}_\alpha]$ as before, we can obtain

$$\begin{bmatrix} \tilde{w}_o(j) & \tilde{w}_i(j) & & & & \\ \tilde{w}_i(j) & \tilde{w}_o(j) & \tilde{w}_e(j) & & & \\ & \tilde{w}_e(j) & \tilde{w}_o(j) & \tilde{w}_i(j) & & \\ & & \tilde{w}_i(j) & \tilde{w}_o(j) & \tilde{w}_e(j) & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \begin{pmatrix} \tilde{G}_1^1(j) \\ \tilde{G}_4^1(j) \\ \tilde{G}_1^2(j) \\ \tilde{G}_4^2(j) \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix} \tag{26}$$

in the above equation, $\tilde{G}_l^L(j)=[\boldsymbol{\xi}^\dagger G_{ll'}^{LL'}\boldsymbol{\xi}]_{jj}$. The coefficient matrix is tridiagonal. For short we omit the index j, then the Green's functions are expressed as

$$\begin{pmatrix} \tilde{G}_1^1 \\ \tilde{G}_4^1 \\ \tilde{G}_1^2 \\ \tilde{G}_4^2 \\ \vdots \end{pmatrix} = \begin{bmatrix} \tilde{w}_o & \tilde{w}_i & & & & \\ \tilde{w}_i & \tilde{w}_o & \tilde{w}_e & & & \\ & \tilde{w}_e & \tilde{w}_o & \tilde{w}_i & & \\ & & \tilde{w}_i & \tilde{w}_o & \tilde{w}_e & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}. \tag{27}$$

To define D_i the determinant of the tridiagonal submatrix except the former i columns and rows of the coefficient matrix, we obtain the following recursive equations:

$$D_0/D_1 = \tilde{w}_o - \tilde{w}_i^2 D_2/D_1,$$

$$D_1/D_2 = \tilde{w}_o - \tilde{w}_e^2 D_3/D_2,$$

$$\cdots \tag{28}$$

Considering that the coefficient matrix is infinite, the relations $D_0/D_1=D_2/D_3$ and $D_1/D_2=D_3/D_4$ hold true. Then from Eqs. (27) and (28), we can readily get the retarded Green's function,

$$\tilde{G}_1^1(E+i0^+) = D_1/D_0 = \frac{(\tilde{w}_o^2-\tilde{w}_i^2+\tilde{w}_e^2)+\sqrt{(\tilde{w}_o^2-\tilde{w}_i^2+\tilde{w}_e^2)^2-4\tilde{w}_o^2\tilde{w}_e^2}}{2\tilde{w}_o}. \tag{29}$$

Here $\tilde{G}_1^1$ is the jth diagonal element of the diagonal matrix $[\tilde{G}_1^1]$. Finally we obtain the analytical expressions about the retarded edge Green's functions in matrix form. It is given by

$$[G_1^1(E+i0^+)] = [\boldsymbol{\xi}][\tilde{G}_1^1(E+i0^+)][\boldsymbol{\xi}^\dagger]. \tag{30}$$

IV. TUNING OF GATE VOLTAGE AND INFLUENCE OF IMPURITIES ON EDGE STATES

The edge states can be tuned by applying a local electric field around the lateral zigzag edge. Such an effect can be studied by our theory developed above. To do so, we mimic the effect of a gate voltage by changing the on-site energies of the atoms in some initial units. To be concrete, we set the on-site energy variation as $\epsilon_{ln}^L=2^{-L}V_g$ for $L=1$ to k, with V_g being the strength of a gate voltage. As a result, the coupling coefficients $\tilde{w}_\alpha^L$ in these units are different from those of the sequent units in that E should be replaced by $E-2^{-L}V_g$. Consequently, the transfer matrix should be established from the $(k+1)$th unit. Thus the $[D]$ matrix in Eq. (21), with $\det[D]=0$ as the sufficient condition to justify whether the edge state is allowed, needs to be redefined as the coefficient matrix of the following set of equations:

$$\tilde{w}_o^1(j)\tilde{c}_1^1(j)+\tilde{w}_i^1(j)\tilde{c}_4^1(j)=0,$$

$$\tilde{w}_i^L(j)\tilde{c}_1^L(j)+\tilde{w}_o^L(j)\tilde{c}_4^L(j)+\tilde{w}_e^{L+1}(j)\tilde{c}_1^{L+1}(j)=0,$$

$$\tilde{w}_e^L(j)\tilde{c}_4^L(j)+\tilde{w}_o^{L+1}(j)\tilde{c}_1^{L+1}(j)+\tilde{w}_i^{L+1}(j)\tilde{c}_4^{L+1}(j)=0,$$

$$[V^{-1}]_{21}\tilde{c}_4^{k+1}+[V^{-1}]_{22}\tilde{c}_1^{k+1}=0 \tag{31}$$

for $L=1,\ldots,k$.

Then, from the condition $\det[D]=0$ we can analyze the energy-level variation in the edge states driven by the gate voltage V_g. Such a result is shown in Fig. 2(b), from which we can see that a notable effect of the gate voltage is to relax the degeneracy of these edge states. And the shift of these edge state levels depends linearly on the gate voltage. The shorter the localization length is, the farther the energy level deviates from the Dirac point. This is a reasonable result since for an edge state with a shorter localization length than the others, it is mainly influenced by the gate voltage very near the zigzag edge where the gate voltage is relatively stronger. In Fig. 2(c) the localization lengths of the degenerate edge states as functions of the gate voltage are shown. In general, the effect of the gate voltage is to make the edge states delocalized. In particular, when V_g is large enough, the edge state with the largest localized length will disappear (for the case of $V_g>0.46t$ in the figure). This indicates that a sufficiently large gate voltage can eliminate the localized edge state. After determining the energy level of an edge state in the presence of a gate voltage, we can derive the corresponding wave function in the following way. At first, the electron probability amplitudes in the first layer are assumed to take a form as $(c_1^1)=c_0(\boldsymbol{\xi}_j)$, then, we can obtain $(c_4^1)=-(c_1^1)\tilde{w}_o^1/\tilde{w}_i^1$ from the recursive electron eigenequation in the presence of a gate voltage. Just repeating this procedure, we can obtain the electron probability amplitudes in the sequent layers. Finally, by summing up all the local electron probabilities, the normalization constant c_0 can be determined. With the obtained wave function, we can calculate the electron probability distributions in any layer. In Fig. 2(d) we compare the electron probability distributions of a specific edge state at the initial layers, which vary as the tuning of a gate voltage. We can see that the electron is allowed to

occupy the B sublattice and the localization is weakened as the increase in the gate voltage.

Now we turn to discuss the influence on the edge states of the impurities appearing in the initial units and the structural variation in the lateral edge. The impurities are theoretically mimicked by introducing random on-site energy of the lattice points in a few units near the zigzag edge. The structural variation in the lateral edge is realized by prefixing a specific cap to the zigzag terminal. For example, by such a connection, we can obtain a lateral edge of armchair type, as shown in Fig. 1(c). The presence of the impurities or the structural variation at the lateral edge lowers the symmetry of the ribbon. As a result, we cannot map the ribbon into a single mode chain as done above. Instead, we can only establish a transfer matrix based on the reduced lattice, as shown in Fig. 1(b). It connects the wave function in a way as $(c^{L+1})=[T](c^L)$ with the notation $(c^L)=[(c_4^L),(c_1^L)]^T$ and is defined as

$$[T]=\begin{bmatrix}[W_i]^{-1}[W_o][W_e]^{-1}[W_o]-[W_i]^{-1}[W_e]^{\dagger}, & [W_i]^{-1}[W_o][W_e]^{-1}[W_i]^{\dagger}\\ -[W_e]^{-1}[W_o], & -[W_e]^{-1}[W_i]^{\dagger}\end{bmatrix}. \tag{32}$$

Noting that we can only use $[T]$ to link the adjacent sublayers beyond the surface region where the impurities or the edge structural variation occurs. In the surface region we have to use the original electron eigenequation in recursive form. For example, when impurities occur only in the first unit, the relevant equations are given by

$$\begin{aligned}
&[H_{11}^{11}](c_1^1)+[H_{12}^{11}](c_2^1)=0,\\
&[H_{21}^{11}](c_1^1)+[H_{22}^{11}](c_2^1)+[H_{23}^{11}](c_3^1)=0,\\
&[H_{32}^{11}](c_2^1)+[H_{33}^{11}](c_3^1)+[H_{34}^{11}](c_4^1)=0,\\
&[H_{43}^{11}](c_3^1)+[H_{44}^{11}](c_4^1)+[W_e](c_1^2)=0,\\
&[W_e](c_4^1)+[W_o](c_1^2)+[W_i](c_4^2)=0,\\
&\ldots,
\end{aligned} \tag{33}$$

where the tight-binding Hamiltonian matrixes $[H_{ll'}^{LL'}]$ can be readily obtained following the lattice structure. When we obtain the eigensolution of the transfer matrix, denoted as $[U]^{-1}[T][U]=[R]$, and inserting it to the relation $(c^{L+2})=[T]^L(c^2)$, we can obtain the constraints $[U]_j^{-1}c^2=0$ corresponding to the eigenvalues of $[T]$ which satisfy $|r_j|\geq 1$. In fact, these constraints provide us with some new equations involving (c^2). To combine them with the above equations in the surface region, we obtain the sufficient condition for any edge state is allowed, that is, $\det[D]=0$ with $[D]$ being the coefficient matrix of these combined equations.

In Fig. 3(a), the determinant $\det[D]$ is numerically calculated as a function of energy E when impurities are present in the first unit. The zero points of $\det[D]$ imply the existence of an edge state at such an energy position. We can see that the presence of impurities causes the deviation of the edge state levels from the Dirac point. And the degeneracy between the edge states is relaxed if there are many edge states in a wide ribbon. The energy-level splitting of the edge states is about 10 percent of the average strength of the impurity potential $\Delta\varepsilon$. When the lateral edge is also of armchair type, as shown in Fig. 3(b), $\det[D]$ is always nonzero. This implies that no edge state is allowed in such an all-armchair-edged ribbon, which is consistent with the assertion that the edge state is intimately associated with the zigzag edge.[17]

Finally, as another application of our theory, we discuss the effect of the C-C bond distortion at the armchair edges on the edge states localized at the lateral zigzag edge. From the

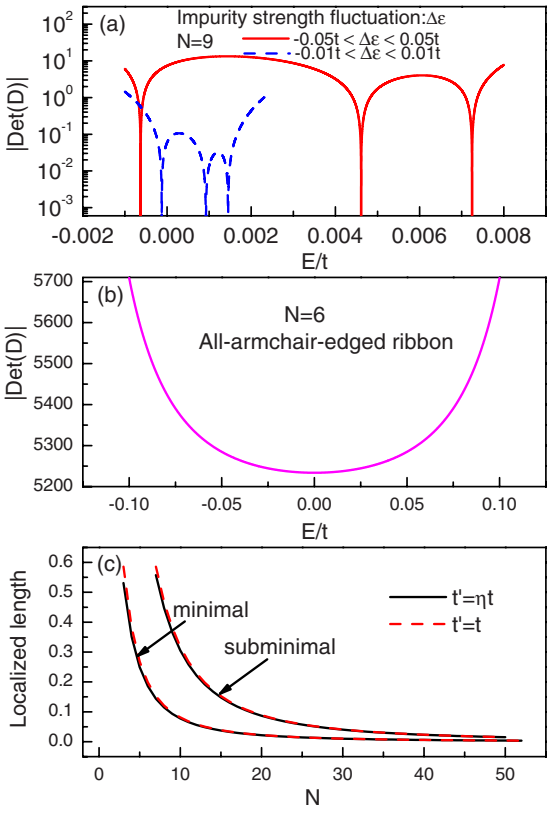

FIG. 3. (Color online) The absolute value of the determinant $\det(D)$ as a function of energy E whose zero points give the edge state levels. (a) The case in the presence of impurities in the first unit. (b) The case for an all-armchair-edged ribbon. (c) The minimal and subminimal localized lengths of the edge states with and without the edge bond distortion.

previous calculations, we know that an appropriate C-C hopping energy at the armchair edges should be modified as $t' = \eta t$ with $\eta = 1.12$.[21,22] And such a modification does not alter the symmetry between the electron-hole subbands of the armchair nanoribbon, from which we infer that the edge bond distortion cannot cause the energy-level deviation of the edge states from the Dirac point. Then, we find that a couple of Hamiltonian submatrix, $[H_{23}^{LL}]$ and $[H_{32}^{LL}]$, are affected by the edge bond distortion. For the case of $N=3$, we have

$$[H_{23}^{LL}] = \begin{bmatrix} t' & 0 & 0 & 0 \\ 0 & t & 0 & 0 \\ 0 & 0 & t & 0 \\ 0 & 0 & 0 & t' \end{bmatrix}. \tag{34}$$

By a simple analysis, we find that our theory developed in Sec. II still hold true for the present case. The edge bond distortion only alters the M matrix in a way as

$$[M] = \begin{bmatrix} \frac{1+\eta}{\eta} & 1 & 0 & \cdots & 0 \\ 1 & 2 & 1 & \cdots & 0 \\ 0 & 1 & 2 & & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \cdots & \frac{1+\eta}{\eta} \end{bmatrix}_{N\times N}. \tag{35}$$

Noting here we need only to consider the case of NBC. As a result, the localization length of the edge states is expected to be influenced by the edge bond distortion to some extent. In Fig. 3(c) we compare the calculated localization length of two typical edge states for the cases with and without the edge bond distortion. We can see that the change in the localization length caused by the edge bond distortion is trivial and can be ignored safely.

V. SUMMARY

We have investigated the edge state properties located at the lateral zigzag edge of a semi-infinite graphene nanoribbon. By mapping the SIGNR onto a reduced lattice, we are allowed to establish the transfer matrix which connects the electronic probability amplitudes of the adjacent unit cells. Then based on the transfer-matrix theory, we can obtain the analytical solution about the wave functions of the edge states under the natural boundary condition. In contrast, we find that the periodic boundary condition employed in some previous work is inappropriate to describe the edge states of the graphene nanoribbons of small size. From our theory the physical natures of some previous numerical conclusions about the edge state are clarified, for example, an edge state is allowed only when the width of the nanoribbon $N \geq 3$ and whenever N increases by three lattice constants a new edge state comes into being. With the same theoretical framework, we derive the analytical expression about the edge Green's functions which are desirable for the study on the electronic transport through some graphene nanostructures. In addition, we also find that a localized edge states can be eliminated by exerting a gate voltage at the lateral zigzag edge. The impurities in the region near the zigzag edge have the effect to relax the energy degeneracy of the edge states. Noting that the localized edge states are responsible to the spontaneous magnetic ordering of many graphene nanostructures, our findings concerning the gate tuning and impurity effect on the edge states will influence the magnetic properties of graphene nanostructures to some extent. Finally, we would like to point out that our theory can be used to analyze other effects on the edge state, such as the magnetic field tuning and the edge roughness. Some interesting results will be reported in future.

ACKNOWLEDGMENTS

This work was financially supported by the National Nature Science Foundation of China under Grant No. NNSFC10774055, the Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. SRFDP20070183130), and National Found for Fostering Talents of Basic Science (Grant No. J0730311).

*Author to whom correspondence should be addressed; zys@mail.jlu.edu.cn

[1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).
[2] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, and A. A. Firsov, Nature (London) **438**, 197 (2005).
[3] Y. B. Zhang, Y.-W. Tan, H. L. Stormer, and Ph. Kim, Nature (London) **438**, 201 (2005).
[4] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).
[5] M. Fujita, K. Wakabayashi, K. Nakada, and K. Kusakabe, J. Phys. Soc. Jpn. **65**, 1920 (1996).
[6] K. Nakada, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, Phys. Rev. B **54**, 17954 (1996).
[7] K. Sasaki, S. Murakami, and R. Satio, Appl. Phys. Lett. **88**, 113110 (2006).
[8] Y. Niimi, T. Matsui, H. Kambara, K. Tagami, M. Tsukada, and H. Fukuyama, Appl. Surf. Sci. **241**, 43 (2005).
[9] Y. Kobayashi, F. Ken-ichi, T. Enoki, K. Kusakabe, and Y. Kaburagi, Phys. Rev. B **71**, 193406 (2005).
[10] Eduardo V. Castro, N. M. R. Peres, J. M. B. Lopes dos Santos, A. H. Castro Neto, and F. Guinea, Phys. Rev. Lett. **100**, 026802 (2008).
[11] Eduardo V. Castro, N. M. R. Peres, and J. M. B. Lopes dos

Santos, EPL **84**, 17001 (2008).
[12] J. Fernández-Rossier and J. J. Palacios, Phys. Rev. Lett. **99**, 177204 (2007).
[13] Vitor. M. Pereira, F. Guinea, J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Phys. Rev. Lett. **96**, 036801 (2006).
[14] A. Rycerz, J. H. Bardarson, A. Rycerz, and C. W. J. Beenakker, Nat. Phys. **3**, 172 (2007).
[15] L. Pisani, J. A. Chan, B. Montanari, and N. M. Harrison, Phys. Rev. B **75**, 064418 (2007).
[16] O. Hod, V. Barone, and G. E. Scuseria, Phys. Rev. B **77**, 035411 (2008).
[17] L. Brey and H. A. Fertig, Phys. Rev. B **73**, 235411 (2006).
[18] D. H. Lee and J. D. Joannopoulos, Phys. Rev. B **23**, 4988 (1981).
[19] D. H. Lee and J. D. Joannopoulos, Phys. Rev. B **23**, 4997 (1981).
[20] C. L. Tang, W. Yan, Y. Zheng, G. Li, and L. Li, Nanotechnology **19**, 435401 (2008).
[21] Y.-W. Son, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **97**, 216803 (2006).
[22] Huaixiu Zheng, Z. F. Wang, Tao Luo, Q. W. Shi, and Jie Chen, Phys. Rev. B **75**, 165414 (2007).
[23] F. Munoz-Rojas, D. Jacob, J. Fernandez-Rossier, and J. J. Palacios, Phys. Rev. B **74**, 195417 (2006).
[24] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995).