Intrinsic Current–Voltage Characteristics of Graphene Nanoribbon Transistors and Effect of Edge Doping

Qimin Yan,[†] Bing Huang,[†] Jie Yu,[†] Fawei Zheng,[†] Ji Zang,[‡] Jian Wu,[†] Bing-Lin Gu,[†] Feng Liu,*,[‡] and Wenhui Duan*,[†]

Department of Physics, Tsinghua University, Beijing 100084, China, and Department of Materials Science and Engineering, University of Utah, Salt Lake City, Utah 84112

ABSTRACT

We demonstrate that the electronic devices built on patterned graphene nanoribbons (GNRs) can be made with atomic-perfect-interface junctions and controlled doping via manipulation of edge terminations. Using first-principles transport calculations, we show that the GNR field effect transistors can achieve high performance levels similar to those made from single-walled carbon nanotubes, with ON/OFF ratios on the order of 10^3–10^4, subthreshold swing of 60 meV per decade, and transconductance of 9.5×10^3 Sm^{-1}.

The continued miniaturization of Si electronic device based on CMOS technology is approaching the physical and geometrical limits. Recent developments through innovations in strain engineering, structural design, and hybridization with new materials are expected to enable the current pace of downscaling for another decade. It is anticipated that future generations of nanoelectronics will likely go beyond CMOS technology to new device paradigms based on exotic nanoscale materials such as molecules and/or novel quantum transport mechanisms such as with spin current replacing charge current.

A proof-of-concept prototype molecular device, a field effect transistor (FET), has been demonstrated with single-walled carbon nanotubes (SWNTs).[1−5] The SWNT-FET has been shown to exhibit superior performance over the conventional Si-FET, characterized by large ON/OFF ratio, fast switching speed (high carrier mobility), small subthreshold swing, and high transconductance.[3−5] However, there remain some challenges before SWNT-based nanodevices can be realized in real applications. For example, common to all molecular devices, it is difficult to construct device junctions with atomically smooth interfaces and to dope the molecular device in a controllable and selective manner.

Graphene nanoribbons (GNRs) have recently attracted intensive interest because they are recognized as a new class of materials in the carbon family as promising building blocks for nanoelectronic and spintronic devices. Here, we illustrate architectural designs of molecular devices built on patterned graphene nanoribbons[6−9] and demonstrate the intrinsic current−voltage (I−V) characteristics of GNR-FETs from first-principles calculations.[10] We show that various device junctions can be constructed by connecting GNRs of different width and orientation with perfect atomic interface, and more importantly device units can be selectively doped by manipulating the edge terminations of GNRs. Furthermore, our calculations show that the GNR-FETs can exhibit high levels of performance similar to SWNT-FETs, with ON/OFF ratios on the order of 10^3-10^4, subthreshold swing of 60 meV per decade, and transconductance of 9.5×10^3 Sm^{-1}.

The conceptual basis of the GNR-based devices such as GNR-FETs originates from the recognition that the electronic properties of GNRs exhibit a dependence on the ribbon direction and width that is the same as those of SWNTs on the tube diameter and chirality.[11−16] A nanotube is often viewed as a rolled-up piece of graphene; conversely, a nanoribbon is nothing but an unrolled nanotube. The electron confinement, the physical origin that gives rise to the differentiation of semiconducting and metallic behavior, is equivalent in the tube and ribbon configurations, as illustrated in Figure 1a. In both cases, the electrons are confined to form standing waves along C_r, the rolling vector, albeit with slightly different boundary conditions. ($C_r = ma_1 + na_2$, where m and n are integers and a_1 and a_2 are the unit cell vectors of graphene lattice.) In a GNR, two nodal points of the standing wave must be at the ribbon edge (lower panel of Figure 1a), while in a SWNT, they can be anywhere, i.e., a circumferential periodic boundary condition (upper panel, Figure 1a).
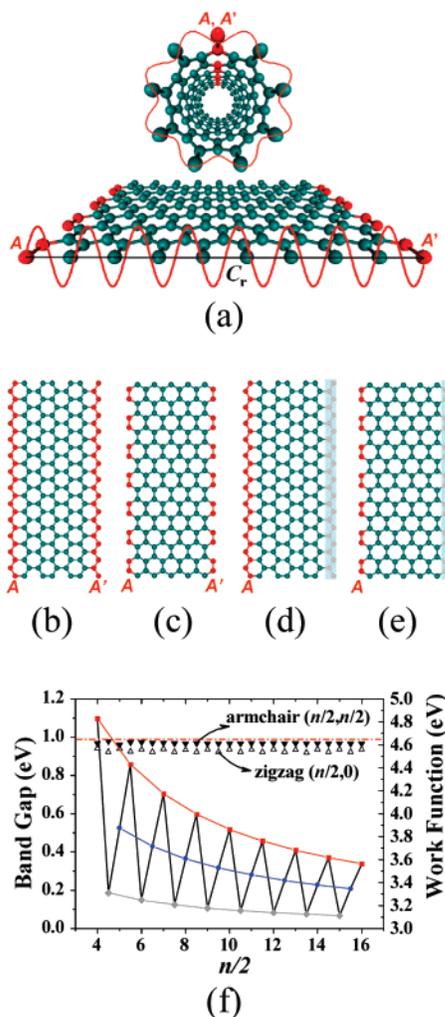
Figure 1. (a) Schematic illustration of a GNR as an unfolded SWNT. (b) A (3,3) asymmetric armchair GNR. (c) A (4.5,0) symmetric zigzag GNR. (d) A (2.5,2.5) symmetric armchair GNR. (e) A (4,0) asymmetric zigzag GNR. (f) Calculated band gaps of zigzag (left *y*-axis) and work functions (right *y*-axis) of zigzag (up triangles) and armchair (down triangles) GNRs as a function of the ribbon width. The horizontal red dashed line marks the work function of graphene at ~4.66 eV.

There can be, in principle, two ways to unfold SWNTs that result in two different classes of GNRs by cutting open the SWNT along the axial direction through a row of atoms and then splitting the atom row onto both edges of the resulting GNR, marked as A and A′ in parts b and c of Figure 1, respectively, or by cutting open the SWNT through a row of C−C bonds, and then a row of atoms will be missing on one edge of the resulting GNR as the missing A′ in Figure 1d and e. So, any given (*m,m*) armchair SWNT can unfold into either a (*m,m*) or (*m*-1/2,*m*-1/2) armchair GNR, and any given (*m*,0) zigzag SWNT can unfold into either a (*m*-1,0) or (*m*-1/2,0) zigzag GNR. There can be twice as many GNRs as SWNTs. Also, a GNR may have either a left−right symmetric (Figure 1c and d) or asymmetric (Figure 1b and e) structure. Therefore, in analogy to naming SWNTs, we can label all the GNRs in a unified sequence of (*m*/2,*n*/2), where *m* and *n* are integers, with uniquely defined ribbon orientation and width [(*m*/2)a_1 + (*n*/2)a_2]. We note that a

different notation of GNRs has been used in the literature[12−16] based on the ribbon edge structures, which do not use the direct correspondence to the SWNTs.[17]

Our first-principles band calculations[10] of GNRs agree well with previous studies.[12−16] All of the (*n*/2,0) zigzag GNRs are semiconducting, and their band gaps are shown in Figure 1f as a function of ribbon width, $W = [(n + 1)/2] \times (0.246)$ nm. Similar to zigzag SWNTs, the band gaps of zigzag GNRs are divided into three groups, with the $n = 3i$ group having the smallest gap,[18] the $3i+1$ group having a medium gap, and the $3i+2$ group having the largest gap, where i is an integer and each group of gaps decrease with increasing width as $1/W$. We found that even a zigzag GNR of 15 nm wide still has a gap of ~0.1 eV, which can be a useful feature for employing the zigzag GNRs as semiconducting channels of FETs.

We also found the zero-temperature ground state of armchair GNRs to be spin-polarized, as predicted by previous calculations.[14,15] However, the energy of the spin-polarized state is only ~20 meV per edge atom lower than the nonspin-polarized state.[15] Furthermore, magnetization is strictly forbidden in 1D and 2D systems at finite temperatures.[19] Therefore, for our investigation of GNR-FETs, only the nonspin-polarized metallic state of armchair GNRs will be considered. We also note that the band gaps of zigzag GNR channels in the FETs are underestimated by the local density approximation.[10] This will affect some of the device characteristics, such as the threshold voltage, but it will not alter any of the main conclusions.

Also shown in Figure 1f are the calculated work functions (WFs). All the zigzag and armchair GNRs (semiconducting or metallic) are found to have almost the same WF of ~4.58 eV, slightly lower than the WF of the graphene (4.66 eV). This is different from the case of SWNTs, where the curvature effect induces surface dipole that makes the WF of SWNTs dependent on tube size.[20] The constancy of the GNR WF turns out to be a very useful property for the design of GNR-based devices. It indicates that when GNRs of different type and of different size are brought into contact, their energy levels will be easily aligned.

Similar to the synthesis of SWNTs that requires better control over their diameter and chirality, fabrication of GNRs awaits for experimental progress,[6−9] in particular the nano-patterning techniques, to achieve control over their width (equivalent to diameter of SWNTs) and directionality (equivalent to chirality of SWNTs). In principle, GNRs can be fabricated directly into device structures and even integrated circuits by a single process of patterning a graphene sheet, as has been demonstrated by recent experiments.[21,22] Parts a−c of Figure 2 illustrate three basic device building blocks: (i) a metal−semiconductor junction, (ii) a $p-n$ junction, and (iii) a heterojunction, which can be, respectively, made by patterned GNRs (i) along different direction, (ii) with different edge doping, and (iii) with different width. We envision a variety of devices can be constructed from these building blocks. For example, a FET can be made simply by two metal−semiconductor junctions, as shown in Figure 2d.
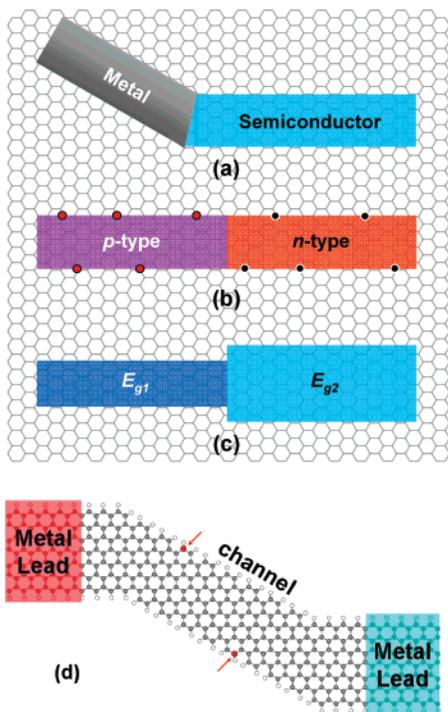
Figure 2. Schematics of three device building blocks: (a) a metal−semiconductor junction between an armchair and a zigzag GNR, (b) a p−n junction between two zigzag GNRs with different edge doping, and (c) a heterojunction between two zigzag GNRs of different width (band gap). (d) Schematics of a GNR-FET, made from one (4,0) zigzag semiconductor GNR channel and two (3,3) armchair metallic GNR leads connected to two external metal electrodes.

There are some key potential advantages in designing and constructing device architectures based on GNRs. First, all the junctions between GNRs of different width and directionality (Figure 2a and c) have perfect atomic interface, a feat that is difficult to achieve for interfacing nanotubes of different diameter and chirality. Second, it is generally difficult to find a robust method to make contact with the molecular device unit because there exists usually a large contact resistance between the metal electrodes and molecules (e.g., SWNTs) due to a very small contact area. This difficulty may be circumvented by using GNRs, because the GNR-based devices can be connected to the outside circuits exclusively via metallic GNRs, as illustrated in Figure 2d, which serve as extensions of metal electrodes to make contact with the semiconducting GNRs so that an atomically smooth metal−semiconductor interface is maintained with minimum contact resistance. Last, doping a molecular device has been extremely difficult. The GNRs have two free edges, which may serve as effective means for doping by manipulating their edge terminations. Simply by replacing the C atoms at the ribbon edge with different types of dopants at different sections, one can selectively dope one single GNR to form a p−n junction, as shown in Figure 2b. The dopant concentration should also be tunable via control of GNR size and amount of dopant atoms (molecules) introduced.

To explore the viability and potential of GNR-based devices, we have carried out extensive first-principles transport calculations[10] to characterize the performance
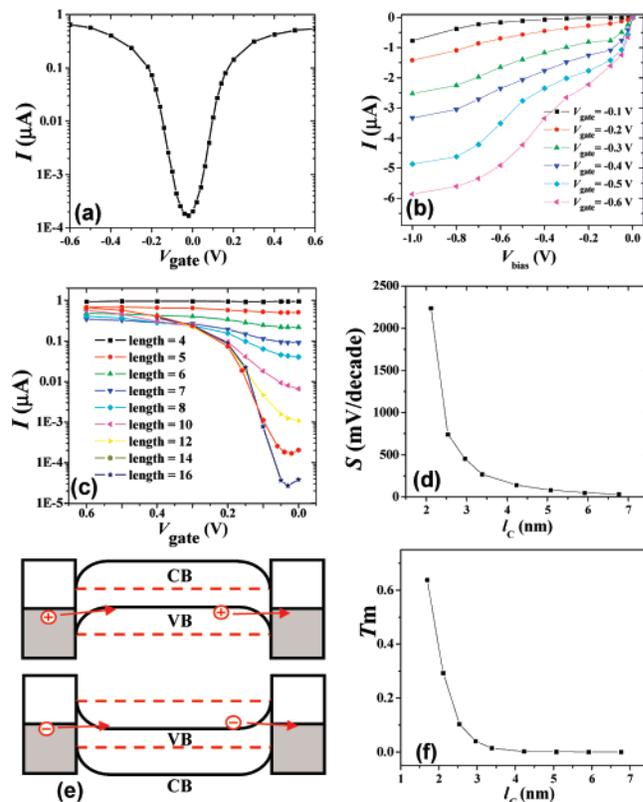

Figure 3. (a−d) Calculated performance characteristics of the intrinsic GNR-FET as shown in Figure 2f. (a) $I-V_{gate}$ curve for a 5.91 nm long channel (V_{bias} = 20 mV). (b) $I-V_{bias}$ curves under different gate voltage (V_{gate}). (c) $I-V_{gate}$ curves as a function of channel length (V_{bias} = 20 mV). (d) The subthreshold swing (S) as a function of channel length. (e) Schematic illustration of band bending under positive (upper panel) and negative (lower panel) gate voltage, illustrating the electron and hole tunneling mechanisms. (f) Transmission coefficients as a function of channel length.

parameters of a series of GNR-FETs as a function of channel length and width for both intrinsic and doped channels. Figure 3a shows the typical $I-V_{gate}$ curve of a GNR-FET made with a 5.91 nm long intrinsic (4,0) zigzag semiconducting channel connected to two (3,3) armchair metallic leads (source and drain). In the voltage window from −0.6 to 0.6 V, the GNR-FET exhibits a near-symmetric $I-V_{gate}$ curve, characteristic of an excellent ambipolar transistor. The on-current I_{on} is about 0.7 μA. At the off-state of zero gate voltage, the minimum leakage current is $\sim 10^{-4}$ μA. This translates to a very large ON/OFF ratio of $I_{on}/I_{off} \sim 2000$. The large ON/OFF ratio manifests again the "perfect" atomic interface between the metal and semiconductor GNRs with minimum contact resistance. Using the $I-V_{gate}$ data and the effective channel width of $W \sim 1.10$ nm, we have calculated the transconductance of the GNR-FET to be $G = (dI/dV_{gate}|_{V_{bias}=1.0eV})/W \sim 9500$ S/m, which can be compared with the best value of 5000−7000 S/m achieved experimentally with SWNT-FETs.[3−5]

Figure 3b shows the calculated the $I-V_{bias}$ curves under different gate voltages. At the zero gate voltage (V_{gate} = 0), the current is suppressed and a pronounced gaplike nonlinearity develops around V_{bias} = 0. The $I-V_{bias}$ curves at low gate voltages exhibit a typical power-law behavior of

semiconductor transport and gradually change into an Ohmic behavior when a large negative V_{gate} is applied. The changing $I-V_{bias}$ characteristics with changing gate voltage demonstrates that the transport properties of the GNR-FETs can be effectively modulated by the gate voltage.

We have also examined the switching characteristics of the GNR-FETs by calculating the subthreshold swing, $S = dV_{gate}/d \log I \sim 60$ mV/decade. This value, which is close to the theoretical limit of conventional Si-based FETs,[23] was also seen in some SWNT-FETs.[4,5] The subthreshold swing is a key parameter for device miniaturization, and generally a small value of S is desired for low threshold voltage and low-power operation when the FETs are made small. Thus, we have further examined S values against the channel length (l_c) of GNR-FETs. Figure 3c shows the $I-V_{gate}$ curves of the GNR-FETs made from the same (4,0) channel with its length varying from 1.69 to 6.76 nm, from which the values of S are derived as a function of l_c, as shown in Figure 3d. Clearly, S decreases with increasing l_c and gradually approaches to ~ 60 mV/decade when l_c becomes longer than 6 nm. We found that the ON-current stays the same, independent of l_c, but the OFF-state leakage current increases rapidly with decreasing l_c, which gives rise to a large S.

The performance of the ambipolar GNR-FETs made of intrinsic semiconductor channels can be understood in terms of metal−semiconductor tunneling junctions within the semiclassical band-bending model. At the OFF state of zero gate voltage, the Fermi level is located at the midgap of the semiconducting channel because the WFs of both the metal leads and the semiconducting channel are the same as shown in Figure 1f. As such, the carriers cannot transport through the channel because both electrons and holes experience an energy barrier at the metal−semiconductor junction, which equals to one-half of the band gap of the channel. For the (4,0) channel, this barrier is about 0.56 eV. When a gate voltage is applied, the electrostatic potential in the channel is raised or lowered. This will result in a thinning of barrier at the metal−semiconductor interface due to band bending, as illustrated in Figure 3e. For the case of electrostatic potential being raised in the channel (Figure 3e, upper panel), holes (electrons) will tunnel from lead (channel) into channel (lead). Conversely, for the case of electrostatic potential being lowered in the channel (Figure 3e, lower panel), electrons will tunnel from metal lead into the semiconductor channel. In this way, the FET stays in the OFF state without gate voltage and will be switched on when a large enough gate voltage is applied, leading to an ambipolar FET with large ON/OFF ratio.

The increase of leakage current at the OFF state (also the S value) with decreasing channel length can be understood within the same tunneling picture. From the leakage current, we have calculated the transmission probability (T_m) of carriers through the channel as a function of l_c, as shown in Figure 3f. On the other hand, we may estimate T_m using the semiclassical theory of tunneling through a finite square potential barrier as $T_m = 1/[1 + C \sinh^2(\kappa x)]$, where C is a parameter related to the carrier energy, and $\kappa = \sqrt{2m^*\Delta}/\hbar$, m^* is the effective mass of carrier, and Δ is the barrier height.
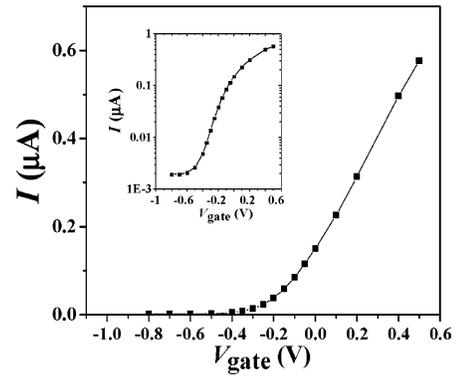

Figure 4. $I-V_{gate}$ curve calculated for a GNR-FET made of a 5.91 nm long N-doped n-type channel, using a bias voltage of 20 mV.

By a nonlinear fitting to the data in Figure 3f, we obtained $m^* \sim 0.08\ m_e$, where m_e is the mass of free electron, which is comparable to the effective mass in the ultrathin graphite films ($0.06\ m_e$).[6]

In addition to the undoped channels, we have characterized an n-type GNR-FET by edge doping with N atoms to demonstrate the doping effect on GNRs. We again used the (3,3)/(4,0)/(3,3) structure, as shown in Figure 2d, but substituted two edge C atoms with N (indicated by arrows in Figure 2d) in the 5.91 nm long (4,0) channel. It corresponds to a dopant concentration of $\sim 2.4 \times 10^{13}/cm^2$. Figure 4 shows the calculated $I-V_{gate}$ curves under $V_{bias} = 20$ mV, exhibiting the typical behavior of an n-type FET. In comparison with the intrinsic undoped FET made from the same GNR structures, we found that the doping has had a small effect on the ON-state current but increased the minimum leakage current by about 1 order of magnitude from $\sim 10^{-4}$ to $10^{-3}\ \mu A$. Consequently, the ON/OFF ratio is decreased by about six times from ~ 2000 to ~ 300, as shown in the inset of Figure 4. For the same reason, the switching characteristic of the n-type GNR-FET is modified, with an increase of subthreshold swing up to ~ 200 mV/decade. In general, using a single type of p- or n-doped GNR channel will decrease the overall performance of GNR-FETs because doping will increase the leakage current.

In conclusion, we have performed extensive first-principles transport calculations to demonstrate the intrinsic current−voltage characteristics of GNR-FETs. The FETs made from intrinsic semiconductor zigzag ribbons can exhibit very high levels of performance, with ON/OFF ratio up to 10^4, subthreshold swing as low as 60 meV per decade, and transconductance of 9.5×10^3 Sm^{-1}. The performance of GNR-FETs can be generally improved by increasing channel length and/or decreasing channel width. Our calculations have laid the basis for establishing the theoretical limits of an important class of GNR-based nanoelectronic devices, providing useful guidelines for future experimental exploration. We illustrate that the GNR-based devices can be made with the atomic-perfect-interface junctions and with controlled doping through edge termination. One may envision a variety of device architectures as well as complete integrated circuits to be fabricated by nanopatterning of a single graphene sheet into the networks of GNRs, opening up a new direction of nanoelectronics.

Supporting Information Available: Methods. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tans, S. J.; Verschueren, A. R. M.; Dekker, C. *Nature* **1998**, *393*, 49−52.
(2) Bachtold, A.; Hadley, P.; Nakanishi, T.; Dekker, C. *Science* **2001**, *294*, 1317−1320.
(3) Rosenblatt, S.; Yaish, Y.; Park, J.; Gore, J.; Sazonova, V.; McEuen, P. L. *Nano Lett.* **2002**, *2*, 869−872.
(4) Javey, A.; Kim, H.; Brink, M.; Wang, Q.; Ural, A.; Guo, J.; Mcintyre, P.; Mceuen, P.; Lundstrom, M.; Dai, H. *Nat. Mater.* **2002**, *1*, 241−246.
(5) Javey, A.; Guo, J.; Farmer, D. B.; Wang, Q.; Wang, D.; Gordon, R. G.; Lundstrom, M.; Dai, H. *Nano Lett.* **2004**, *4*, 447−450.
(6) Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V.; Firsov, A. A. *Science* **2004**, *306*, 666−669.
(7) Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Grigorieva, I. V.; Dubonos, S. V.; Firsov, A. A. *Nature* **2005**, *438*, 197−200.
(8) Zhang, Y.; Tan, Y.-W.; Stormier, H. L.; Kim, P. *Nature* **2005**, *438*, 201−204.
(9) Berger, C.; Song, Z.; Li, X.; Wu, X.; Brown, N.; Nafud, C.; Mayo, D.; Li, T.; Hass, J.; Marchenkov, A. N.; Conrad, E. H.; First, P. N.; de Heer, W. A. *Science* **2006**, *312*, 1191−1196.
(10) Supporting Information on methods is available online at http://pubs.acs.org.
(11) Saito, R.; Dresselhaus, G.; Dresselhaus, M. S. *Physical Properties of Carbon Nanotubes*; Imperial College Press: London, 1998.
(12) Miyamoto, Y.; Nakada, K.; Fujita, M. *Phys. Rev. B* **1999**, *59*, 9858−9861.
(13) Nakada, K.; Fujita, M.; Dresselhaus, G.; Dresselhaus, M. S. *Phys. Rev. B* **1996**, *54*, 17954−17961.
(14) Son, Y.; Cohen, M. L.; Louie, S. G. *Nature* **2006**, *444*, 347−349.
(15) Son, Y.; Cohen, M. L.; Louie, S. G. *Phys. Rev. Lett.* **2006**, *97*, 216803.
(16) Barone, V.; Hod, O.; Scuseria, G. *Nano Lett.* **2006**, *6*, 2748−2754.
(17) The convention used so far such as in refs 12−16 names GNRs by edge structure, so an armchair (zigzag) tube would unfold into a zigzag (armchair) ribbon. It lacks a unifying nomenclature for all the ribbons. The size of an armchair ribbon is counted by number of dimer rows, while the size of a zigzag ribbon by number of zigzag rows and the size of a "chiral" ribbon would be hard to count.
(18) Ideally, the $n = 3i$ group would exhibit a zero gap to be metallic like their counterpart SWNTs. However, because the "effective ribbon width", i.e., the distance between the two nodal points at the edge, is not exactly the integer times of $(3/2)|a_1|$, the wavevector of the confined electrons will always miss the Fermi point of the corresponding graphene to open a small gap. In a tight-binding model, this can be partly understood in terms of different hopping integrals at the edge.[15]
(19) Mermin, N. D.; Wagner, H. *Phys. Rev. Lett.* **1966**, *17*, 1133−1136.
(20) Shan, B.; Cho, K. *Phys. Rev. Lett.* **2005**, *94*, 236602.
(21) Chen, Z.; Lin, Y.-M.; Rooks, M. J.; Avouris, P. **2007**, cond-mat/0701599.
(22) Han, M. Y.; Oezyilmaz, B.; Zhang, Y.; Kim, P. **2007**, cond-mat/0702511.
(23) Sze, S. M. *Physics of Semiconductor Devices*; Wiley: New York, 1981.

NL070133J