Band-gap renormalization in carbon nanotubes: Origin of the ideal diode behavior in carbon nanotube *p-n* structures

Ji Ung Lee*
GE Global Research Center, Niskayuna, New York 12309, USA
(Received 15 August 2006; revised manuscript received 22 December 2006; published 9 February 2007)

We demonstrate dramatic modifications to the electronic structure in single-walled carbon nanotubes due to band gap renormalization, the many-body induced shrinkage of the fundamental band gap. This is examined within the framework of ideal *p-n* diodes formed along individual single-walled carbon nanotubes. A combination of photocurrent spectroscopy with detailed transport measurements provides a complete set of energy levels of the nanotube *p-n* structure. These energy levels confirm the large band gap shrinkage, consistent with enhanced many-body correction in one-dimensional confinement, and result in fundamental changes to the nanotubes diode transport properties as compared to their bulk counterparts. We show that the ideal diode behavior is a direct consequence of significant renormalization of the band gap at the doped *p* and *n* regions, resulting in formations of heterointerfaces, in stark contrast to a uniform band gap expected of a homogeneous material.

DOI: 10.1103/PhysRevB.75.075409 PACS number(s): 73.63.Fg, 78.67.Ch, 71.10.−w

Band gap renormalization (BGR), the shrinkage of the fundamental band gap due to many-body exchange-correlation properties of interacting electrons,[1] is of both fundamental and practical importance for understanding semiconductors at large carrier densities. BGR has been examined theoretically and experimentally on a wide range of different structures of successively lower dimensions.[2] With the continued emphasis on reduced features in modern electronics, quasi-one-dimensional (1D) semiconductors are expected to take on a greater technological importance with single-walled carbon nanotubes (SWNTs) representing the extreme limit of 1D confinement.[3] With greater confinement, however, a larger BGR is expected from enhanced many-body interaction.[4–6] Here, we provide an analysis of doping induced BGR in SWNTs. We show that BGR can be a large fraction of the intrinsic band gap, consistent with theoretical analysis of 1D semiconductors,[4,5] and much larger than BGR in higher dimensional structures.[2]

Similar to the analysis of BGR in bulk materials,[2] BGR is examined within the framework of ideal SWNT *p-n* diodes.[7,8] We examine the photocurrent (PC) spectroscopy with detailed transport measurements to provide a complete set of energy levels for determining the magnitude of BGR. As a bipolar device, the transport properties of a *p-n* diode depend on the minority carrier densities, which in turn depend sensitively on the band gap. In our diodes, we show that even though the diode current-voltage (*I-V*) characteristics follow the celebrated ideal diode equation derived for bulk materials, the underlying transport mechanism is fundamentally different than in bulk diodes. The ideal diode behavior in our nanotube diodes is a direct consequence of significant BGR in the doped *p* and *n* regions that enhances minority carrier densities, and results in the formations of heterointerfaces between the doped regions and the undoped region of a homogenous material.

The SWNT diode device is shown schematically in the inset of Fig. 1. Details of the fabrication process can be found elsewhere[7,8] with the following device dimensions: 1–1.5 μm for the split gate spacing *L* (between VG1 and VG2), 150 nm for the deposited SiO_2 on top of the split gate metal electrodes, and >1 μm for the lateral overlap between a gate and a SWNT to produce the doped regions. Unlike in bulk, our device is ideally suited for studying BGR since doping is not fixed but can be varied electrostatically. To form the *p-n* doping profile, opposite polarity bias is applied to the split gates to form *p*-type doping on one end of a SWNT and *n*-type doping on the other end. With the portion *L* of a SWNT suspended as shown, the band gap states induced from SiO_2 are removed, leading to the ideal behavior.[8] Previous results on *p-n* diodes based on chemical doping showed leaky diode characteristics with nonideal behavior due to tunneling and abrupt junction formation.[9]

Universally, all diode *I-V* (current-voltage) characteristics are compared to the diode equation[10]

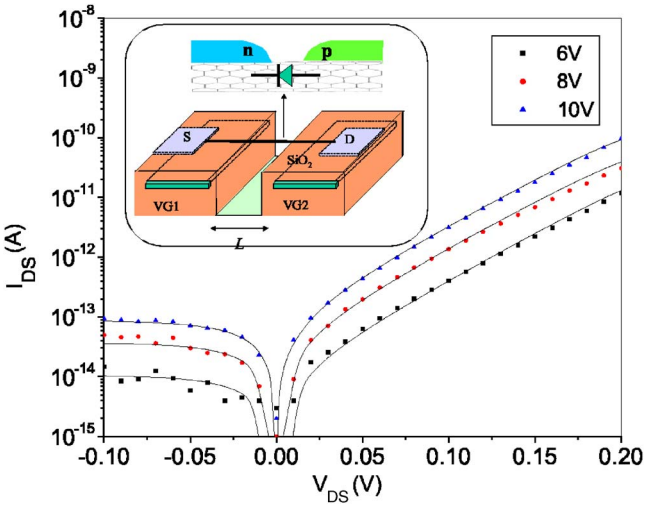

FIG. 1. (Color online) The inset shows a schematic diagram of the split gate diode device with a portion of SWNT suspended. Data are *I-V* curves of a SWNT diode at three different split gate bias (VG1,2) of ±6, 8, and 10 V, and measured at 300 K. The solid lines are fits to Eq. (1) with $n=1.04$ for all three *I-V* curves with I_0 of 10.5, 37.0, and 84.0 fA.

$$I_{DS} = I_0(e^{qV_{DS}/nK_BT} - 1), \quad (1)$$

where I_0 is the reverse-bias saturation (leakage) current, q the charge on an electron, V_{DS} the externally applied bias voltage, n the ideality factor, K_B Boltzman's constant, and T the absolute temperature. The parameters I_0 and n are determined by fundamental materials properties. Ideally, n is precisely equal to 1 but, depending on the nature of electron-hole generation and recombination (GR) mechanism, can take on a value >1 and also depend on voltage. For example, $n=2$ is observed when GR is completely mediated by midgap defect states.[10]

In elucidating the effects of BGR, we take advantage of the variability in doping that is established by different bias on the split gate electrodes VG1,2. A representative set of *I-V* curves from a single diode at different split gate bias voltages is shown in Fig. 1. Each curve follows Eq. (1) (solid curves) with n close to the theoretical value of 1. The increase in I_0 with gate bias is a direct result of BGR, and is one of the main subjects of our discussion.

The ideal diode behavior implies a very specific GR mechanism that must take into account effects of BGR. For a given linear doping density, a diameter dependent BGR (Ref. 4) is expected in SWNTs. In our diodes, BGR is restricted to the p and n doped regions where carrier densities exceed 5×10^6 cm^{-1} based on a calculated capacitance to the gate at bias voltages of ±10 V. This is a large doping density as compared to an equivalent bulk doping density. The resulting band structure is shown in Fig. 2(a) for a small reverse bias ($V_{DS}<0$) to illustrate better the origin of I_0. The suspended portion L is unperturbed by BGR since it is not doped and, as we will show, is also the region where the observed resonant optical absorption takes place. BGR primarily affects the band in which the doping occurs[11] and results in asymmetric heterointerfaces with the suspended region, and is in stark contrast to a uniform band gap expected of a homogenous material. Since a higher dielectric constant of SiO_2 than air confines the fields from the split gates to regions immediately above them, a sharper boundary forms between the regions modified by BGR and the unperturbed region L.

The strongest evidence for BGR comes from examination of the energy levels observed in the PC spectra along with the activation energy E_a of the diode dark *I-V* curves. Recently, we reported on the observation of a series of exciton and exciton-phonon peaks in the PC spectra,[12] including the lowest exciton peak E_{11}, along with the onset of continuum ($E_{\text{gap}}=E_{11}+E_B$ in region L), which measures the exciton binding energy E_B. To illustrate these features we show in Fig 2(b) three spectra from three different devices with similar first resonant energies. The PC is measured with zero bias across the diode ($V_{DS}=0$), i.e., the short circuit current I_{SC}. Each spectrum is normalized to the incident photon flux and to the first peak E_{11} and, for ease of comparison, is offset vertically by $\frac{1}{2}$ unit from each other. For each nanotube, the PC spectrum shows a series of narrow peaks consistent with optical absorption of a 1D semiconductor. The series of peaks are excitonic in nature and feature a Lorentzian resonance instead of the characteristic $1/\sqrt{E}$ tail predicted by one-particle van Hove singularity for joint density of states. A clear distinguishing feature is the wide dispersion for the third set of peaks compared to the first set, which is due to different (n,m) index nanotubes. This feature is similar to what has been observed in the photoluminescent data between excitation energy E_{22} and the emission energy E_{11} of SWNTs in aqueous solution.[13] Peaks 2 are sidebands of E_{11}, and are the exciton-phonon bound states related to the ~200 meV tangential vibrations.[12,14–16]

In addition to the peaks, an important feature in Fig. 2(b) is the observation of the onset of continuum (marked by arrows in the top two spectra). The onset of the continuum is evidenced by a weak increase in the PC with photon energy. Below this onset, the base line spectrum is flat, as indicated by the horizontal dotted lines. The weak onset is not always evident as is shown in the bottom spectrum, either due to a weak background signal or masked by the more dominant exciton resonance. Over half of the devices show a clear onset at different energies. The lack of any features and the weak oscillator strength at the onset of continuum are consistent with optical transitions in 1D semiconductors.[17–19] This is described through the Sommerfeld factor, the ratio of optical absorption intensity in the continuum with and without the Coulomb interaction, which is calculated to be less than 1 in 1D semiconductors.[20] This leads to higher oscillator strength being concentrated in the excitons rather than at the interband absorption edge, as demonstrated in our data. Our measured E_B is large (>0.25 eV) and consistent with general theories of excitons in SWNTs.[17–19]

In conjunction with the energy levels revealed in the PC spectrum, measurements of E_a from the *I-V* curves in the dark provide a complete set of energy levels for our nanotube diodes (E_a is measured by heating the diode in 10° increments from 300 to 340 K). Figure 2(a) illustrates these energy levels in more detail within our band model. The absolute magnitude of these levels is seen in Fig. 2(c), which shows that E_{11} and E_{gap}, measured over a wide range of leakage current diodes, are considerably above their E_a values. The solid lines are linear fits to their respective data and the dashed line is for $E=E_a$. The origin of E_a becomes clear when considering the possible mechanisms that give rise to an ideal diode behavior, for which there are only two. Again, to simplify our treatment, we examine the generation process shown in Fig. 2(a) (reverse bias case) rather than the full GR process.

An ideal diode behavior is either due to a direct band-to-band generation of electron-hole pairs in the field region L (Ref. 21) or due to diffusion of minority carriers (n_p and p_n) from the doped p and n regions[10] as shown in Fig. 2(a). In the latter process, thermally generated minority carriers, within a diffusion length of the edges of the field region L, diffuse and are swept by the built-in field; each p and n region contributes $I_0/2$ for a symmetrically doped diode. A direct band-to-band generation is expected in low defect materials where thermal generation is not mediated by midgap defect states. In this case, however, $E_a=E_{\text{gap}}=E_{11}+E_B$, which we do not observe. Therefore, we conclude that the origin of the ideal diode behavior is due to diffusion of minority carriers, which we describe in more detail below. Here, E_a is significantly modified by BGR, as illustrated in Fig 2(a), and is the difference between the Fermi energy E_F of a doped

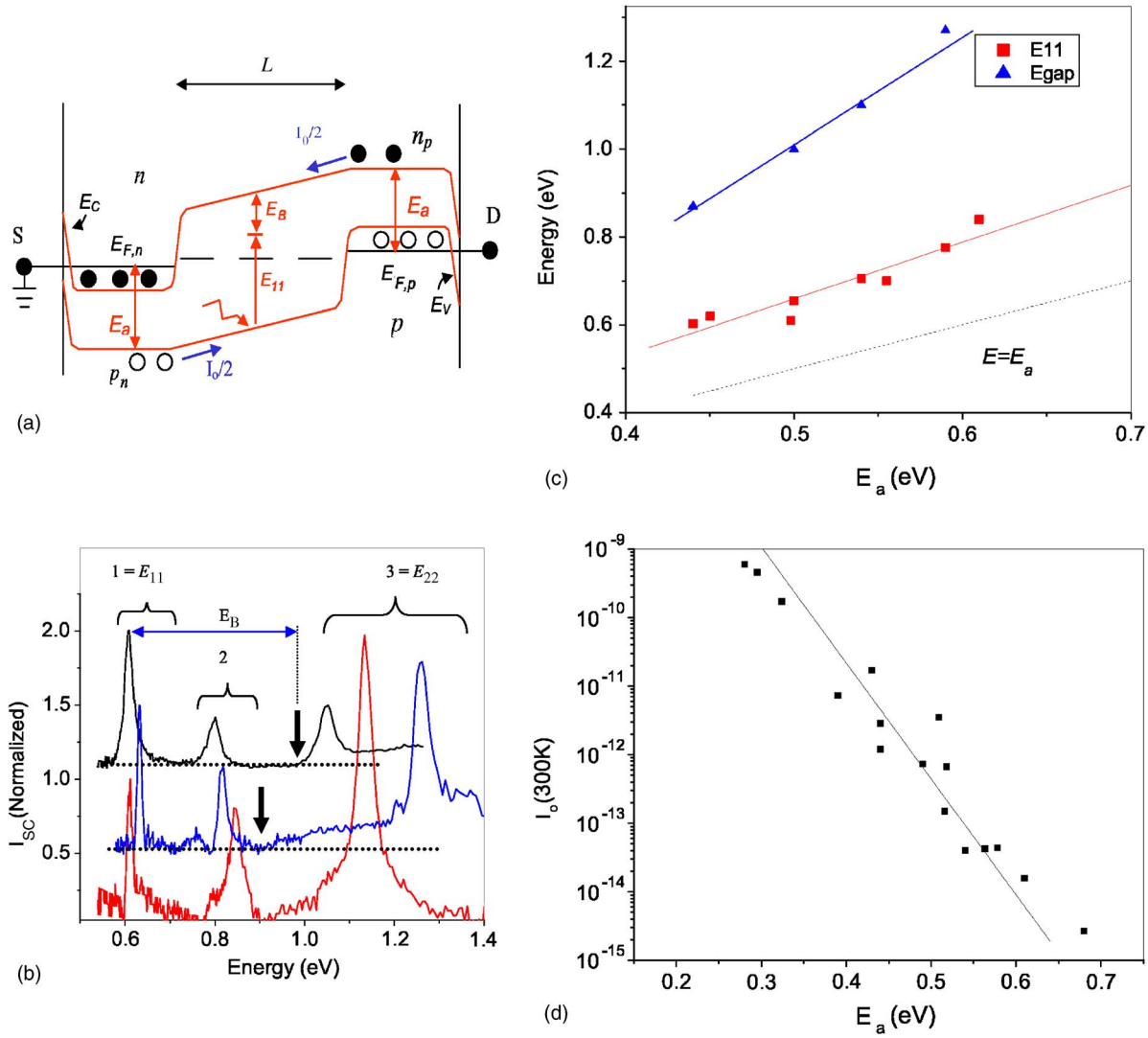

FIG. 2. (Color online) (a) The band diagram of our diode showing the doped regions (●-electron, ○-hole) from the split gates, and the formation of heterointerfaces between the doped and the suspended portion (undoped) L of a SWNT. Each doped region contributes $I_0/2$ in leakage current due to diffusion of minority carriers n_p and p_n. In region L, the optical excitation at E_{11} creates the lowest excitons with a binding energy E_B. Band bending at S/D interface is due to Schottky contact formation with the metal electrodes. (b) Normalized photocurrent spectra measured at 300C showing the first three peaks to emphasize the onset of continuum (mark by arrows). Each spectra is offset by $\frac{1}{2}$ unit in intensity. The horizontal dotted lines help to highlight the onset of the continuum as shown by the arrows. (c) Comparison between E_{11} and E_{gap} vs E_a where E_a is the activation energy of the doped regions. (d) Measured I_0 vs E_a. The solid curve is from model for minority carrier diffusion [Eq. (6)].

region ($E_{F,n}$ or $E_{F,p}$) and its minority carrier band edge.

Between different materials in bulk, diffusion current dominates over direct band-to-band generation for a smaller band gap semiconductor, where a larger minority carrier density forms for a given majority carrier doping, than those with a larger band gap. In our diodes, however, the diffusion current dominates because of a larger than expected minority carrier densities in the doped regions, arising from a significant reduction in the band gap, and is largely band gap (diameter) independent. On the other hand, the magnitude of BGR is diameter dependent since this changes the carrier confinement and hence the many-body interaction (the capacitive coupling is weakly dependent on diameter). The larger slope of the fitted lines in Fig. 2(c) are consistent with a larger BGR expected of smaller diameter (larger E_{11}) SWNTs.

Referring back to Fig. 1, since E_F moves deeper into the majority carrier subband with increased doping, in the absence of BGR, minority carriers n_p and p_n should decrease and result in a corresponding decrease in I_0. Precisely the opposite behavior is observed because of the shrinkage of the band gap. Here, BGR brings $E_{F,n}$ and $E_{F,p}$ closer to their respective minority carrier band edges, resulting in increases of n_p and p_n. We examine the doping dependence of I_0 in more detail in Figs. 3(a) and 3(b) and quantify the magnitude of BGR based on the analysis of minority carrier generation rate as derived below.

To facilitate a more quantitative analysis of the diffusion

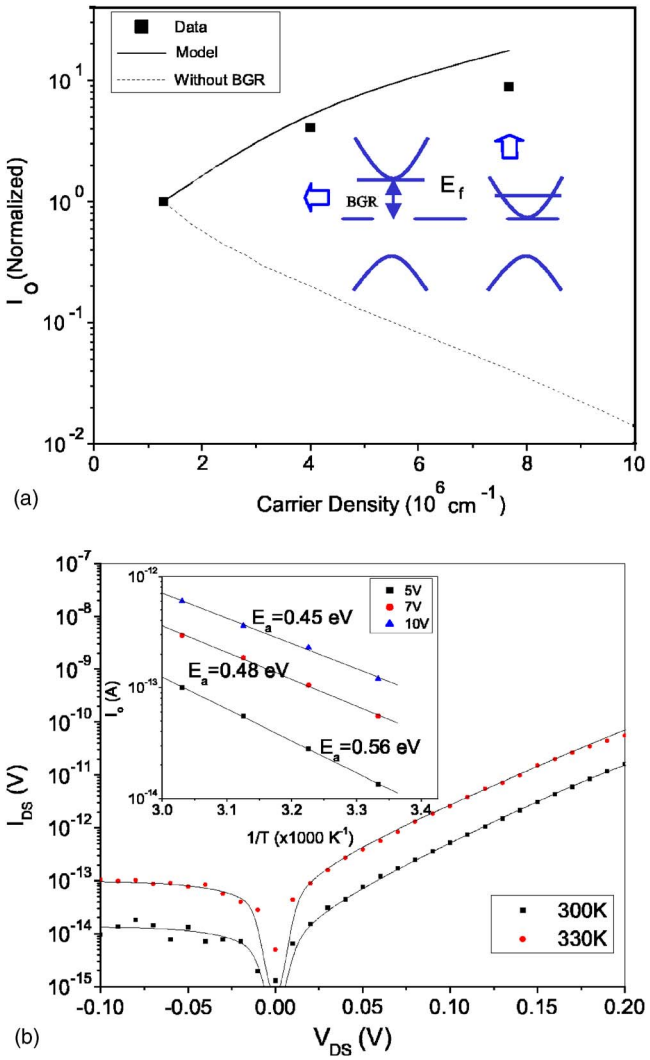

FIG. 3. (Color online) (a) Effects of BGR on I_0 with increased doping from the split gates. The observed increase in I_0 with doping is consistent with large BGR as shown in the inset for the n doped subband and predicted by the solid curve, and is contrary to a large decrease expected in the absence of BGR (dotted curve). (b) The diode I-V curves measured at T=300 and 330 K for split gate bias of ±5 V along with their fits (solid curves) to Eq. (1) using n =1.05. The inset shows the measured E_a vs the split gate bias voltage. The decrease in E_a with doping is also consistent with a large BGR with increased doping. The carrier density in (a) corresponds to the gate doping density in (b) for a measured threshold voltage of 4 V. The calculated band gaps are 0.59, 0.47, and 0.39 eV for the three doping densities.

current, we plot I_0 from different devices as a function of E_a in Fig. 2(d) for fixed doping of ±10 V. I_0 is expected to scale as e^{-E_a/K_BT}, which we derive below. The diffusion current is expressed as[10]

$$I_0 = q(n_p + p_n)L_D/\tau = qD_{\text{iff}}(n_p + p_n)/\sqrt{D_{\text{iff}}\tau}, \tag{2}$$

where n_p and p_n are the minority carrier densities in p and n doped regions, respectively, D_{iff} is the diffusion constant, and τ is the minority carrier lifetime. $L_D=\sqrt{D_{\text{iff}}\tau}$ is the diffusion length, and represents the portion of a doped region where minority carriers, which are generated at the rate $1/\tau$, are able to diffuse to the edge of the field region L. n_p and p_n depend strongly on the position of Fermi level E_F, especially with large doping which places E_F within the majority carrier subbands as shown in Fig. 2(a). We calculate the minority carrier density assuming, uncritically, the density of states (DOS) is unperturbed by doping. For the p-doped region, DOS for the first conduction subband is given as[22]

$$D(E,E_c) = \begin{cases} D_0 \dfrac{E}{\sqrt{E^2 - E_c^2}}, & E > E_c, \\ 0, & 0 < E < E_c, \end{cases} \tag{3}$$

where D_0 depends on the chirality of the nanotube. An analogous argument holds for the n-doped region based on symmetry of DOS. $D(E,E_c)$ exhibits a divergent van Hove singularity (vHs) at the conduction band edge E_c and $2E_c$ is the electronic band gap of the lowest vHs. Here, it is sufficient to consider only the first vHs since $E_{F,p}$ is far from the minority carrier band edge. The minority carrier density is

$$n_p = 2\int_0^\infty D(E,E_c)f(E,E_{F,p})dE, \tag{4}$$

where $f(E,E_{F,p})=(1+e^{(E-E_{F,p})/K_BT})^{-1}$ is the Fermi-Dirac distribution function. The factor 2 is for spin degeneracy. Within our notation $E_{F,p}$ is negative, which is set by the majority carrier density. Because $f(E,E_{F,p})$ is a slowly varying function compared to $D(E,E_c)$ near E_c, using the Boltzman approximation for $f(E,E_F)$ at E_c, Eq. (4) can be approximated as

$$n_p \approx \frac{D_0}{4} e^{-(E_c - E_{F,p})/K_BT} = N_C e^{-E_a/K_BT}. \tag{5}$$

$E_a=E_c-E_{F,p}$ is the activation energy, and is larger than the band gap by $E_V-E_{F,p}$ as show in Fig. 2(a). From symmetry, $p_n=N_Ve^{-E_a/K_BT}$, where $N_V=N_C$. The difference between Eqs. (4) and (5) is at most 30% for E_a ranging from 0.2–0.7 eV, and n_p ranging over nine decades. This error is less than the scatter in the data, therefore, a good approximation. We note the similarity in Eq. (5) to the effective density of states notation (N_C) commonly used to describe carrier densities in bulk semiconductors.[10] To simplify our analysis, we consider the DOS for arm-chair SWNTs $(n,0)$ (Ref. 22) for which we calculate $D_0=16/3\pi dV_{pp\pi}$, where d=1.44 Å is the carbon-carbon bond distance and $V_{pp\pi}$=2.5 eV is the nearest neighbor interaction energy. I_0 now can be written as

$$I_0 = qD_{\text{iff}}(D_0 e^{-E_a/K_BT})/2\sqrt{D_{\text{iff}}\tau} \tag{6}$$

We use the Einstein relation $D_{\text{iff}}/\mu=K_BT/q$, where μ is the minority carrier mobility. Using $\mu=2\times10^4$ cm^2/V s,[23] the best fit is obtained for $\tau=0.51\times10^{-12}$ s [see Fig. 2(c)]. These values correspond to L_D=0.16 μm which is much smaller than the length of the doped regions (>1 μm), validating the use of Eq. (2) with L_D as the characteristic length. This τ is at least 2× smaller than the lifetime of an equivalently doped bulk direct band gap GaAs,[24,25] and suggests a possible decrease due to carrier confinement. In our simple treatment, we have assumed DOS, τ and μ to be constants although it is

anticipated that they are doping density and (n,m) chiral index dependent, which may explain some of the scatter in Fig. 2(d) data.

We now examine BGR in more detail by using Eq. (6) to analyze the rate of BGR with doping. Fig. 3(a) shows the I_0 dependence on gate bias, converted to linear doping density, and Fig. 3(b) shows the corresponding E_a at these gate biases (see inset). Note that E_a decreases with increased doping as expected from a significant BGR. In converting the gate bias to linear doping density, we used an electrostatic gate capacitance of 205 fF/cm obtained from a computer simulation, with a measured threshold voltage of 4 V. The dotted curve is the expected I_0 without BGR and the solid curve is derived from Eq. (6). Deviations from the dotted curve have also been observed in bulk diodes[26] but to a lesser extent because of lesser significance of BGR in bulk. Here, we observe a significant deviation with increased doping because of larger BGR than the opposing change in E_F, as illustrated in the inset of Fig. 3(a) for the *n*-doped subband. In our analysis, we use the measured E_a to calculate the band gap and the position of E_F. This is then used to calculate n_p and p_n at each doping density. The mobility dependence on doping is based on Ref. 23. The solid curve represents BGR of 200 meV as the doping is increased from 1.3×10^6 to 7.7×10^6 cm^{-1}. Clearly, the rate of BGR with doping is large.

The magnitude of BGR can be calculated from Fig. 2(c) data and the position of E_F. BGR with respect to the undoped region is a sum of three terms: E_B, $E_{11}-E_a$, and $E_{F,n}-E_C$. For example, a nanotube with $E_{11}=620$ meV, $E_B=290$ meV, and $E_{F,n}-E_C=130$ meV,[12] this sum is 540 meV at a doping density of 7×10^6 cm^{-1}, or more half the intrinsic band-gap, and confirms the formations of the heterointerfaces shown in Fig. 2(a). A similarly large BGR is also predicted in nanowires.[4,5] By contrast, BGR in bulk materials is about an order of magnitude less at an equivalent doping density.[2]

This work was supported by M. L. Blohm and the Nanotechnology Program at GE Global Research Center. J.U.L. would like to thank A.K. Swan, G.E. Possin, and J. Bray for useful discussions.

*Present address: College of NanoScale Science and Engineering, SUNY-Albany, Albany, NY 12203.

[1] G. D. Mahan, *Many-Particle Physics* (Plenum, New York, 1990).
[2] R. A. Abram, in *Semiconductors and Semimetals*, edited by R. Ahrenkiel and M. S. Lundstrom (Academic Press, New York, 1993), Vol. 39, Chap. 5.
[3] *Carbon Nanotubes: Synthesis, Structure Properties and Applications*, edited by M. Dresselhaus, G. Dresselhaus, and Ph. Avouris (Springer, Berlin, 2001).
[4] Ben Yu-Kuang Hu and S. Das Sarma, Phys. Rev. Lett. **68**, 1750 (1992).
[5] E. H. Hwang and S. Das Sarma, Phys. Rev. B **58**, R1738 (1998).
[6] S. Das Sarma and D. W. Wang, Phys. Rev. Lett. **84**, 2010 (2000).
[7] J. U. Lee, P. P. Gipp, and C. M. Heller, Appl. Phys. Lett. **85**, 145 (2004).
[8] J. U. Lee, Appl. Phys. Lett. **87**, 073101 (2005).
[9] C. Zhou, J. Kong, E. Yenilmez, and H. Dai, Science **290**, 1552 (2000).
[10] S. M. Sze, *Physics of Semiconductor Devices* (Wiley-Interscience, New York, 1981).
[11] B. E. Sernelius, Phys. Rev. B **33**, 8582 (1986).
[12] J. U. Lee, P. J. Codella, and M. Pietrzykowski (unpublished).
[13] S. M Bachilo, M. S. Strano, C. Kittrell, R. H. Hauge, R. E. Smalley, and R. B. Weisman, Science **298**, 2361 (2002).
[14] V. Perebeinos, J. Tersoff, and Ph. Avouris, Phys. Rev. Lett. **94**, 027402 (2005).
[15] F. Plentz, H. B. Ribeiro, A. Jorio, M. S. Strano, and M. A. Pimenta, Phys. Rev. Lett. **95**, 247401 (2005).
[16] X. Qui, M. Freitag, V. Perebeinos, and Ph. Avouris, Nano Lett. **5**, 749 (2005).
[17] T. Ando, J. Phys. Soc. Jpn. **66**, 1066 (1997).
[18] V. Perebeinos, J. Tersoff, and Ph. Avouris, Phys. Rev. Lett. **92**, 257402 (2004).
[19] H. Zhao and S. Mazumdar, Phys. Rev. Lett. **93**, 157402 (2004).
[20] T. Ogawa and T. Takagahara, Phys. Rev. B **43**, 14325 (1991).
[21] W. Shockley and H. J. Queisser, J. Appl. Phys. **32**, 510 (1961).
[22] J. W. Mintmire and C. T. White, Phys. Rev. Lett. **81**, 2506 (1998).
[23] T. Durkp, S. A. Getty, E. Cobas, and M. S. Fuhrer, Nano Lett. **4**, 35 (2004).
[24] S. Tiwari and S. L. Wright, Appl. Phys. Lett. **56**, 563 (1990).
[25] R. K. Ahrenkiel, in *Semiconductors and Semimetals*, edited by R. Ahrenkiel and M. S. Lundstrom (Academic Press, New York, 1993), Vol. 39, Chap. 2.
[26] M. E. Klausmeier-Brown, M. S. Lundstrom, M. R. Melloch, and S. P. Tobin, Appl. Phys. Lett. **52**, 2255 (1988).