Nonlinear Screening and Ballistic Transport in a Graphene *p-n* Junction

L. M. Zhang and M. M. Fogler

Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

(Received 11 August 2007; published 21 March 2008)

We study the charge density distribution, the electric field profile, and the resistance of an electrostatically created lateral *p-n* junction in graphene. We show that the electric field at the interface of the electron and hole regions is strongly enhanced due to limited screening capacity of Dirac quasiparticles. Accordingly, the junction resistance is lower than estimated in previous literature.

DOI: 10.1103/PhysRevLett.100.116804 PACS numbers: 73.63.−b, 73.40.Lq, 81.05.Uw

Unusual electron properties of graphene are an active topic of fundamental research and a promising source of new technology [1]. A monolayer graphene is a gapless two-dimensional (2D) semiconductor whose quasiparticles (electrons and holes) move with a constant speed of $v \approx 10^6$ m/s. The densities of these "Dirac" quasiparticles can be controlled by external electric fields. Recently, graphene *p-n* junctions (GPNJ) have been realized experimentally [2], in which electron density $\rho(x)$ changes *gradually* between two limiting values, $\rho_1 < 0$ and $\rho_2 > 0$, as a function of position x. This change occurs over a length scale D determined by the device geometry. For a junction created near an edge of a wide gate (Fig. 1), D is of the order of the distance to this gate.

In theory, GPNJ can exhibit a number of intriguing phenomena, including microwave-induced [3] and Andreev [4] reflection, lensing [5], and Klein tunneling [6,7]. Previously, Klein tunneling of massive Dirac quasiparticles was studied in tunnel diodes. In such semiconductor devices, the quasiparticle tunneling probability is given by [8] $T = \exp(-\pi\Delta^2/e\hbar v|F_{\rm pn}|)$, where $F_{\rm pn}$ is the electric field in the gapped region. The *single-particle* problem for a massless case of graphene is mathematically equivalent, except the role of the gap Δ is played by $\hbar v k_y$. Integrating $T(k_y)$ over the transverse momentum k_y to get conductance and then inverting it, one finds the resistance R per unit width of the GPNJ to be [7]

$$R = (\pi/2)(h/e^2)\sqrt{\hbar v/e|F_{\rm pn}|}. \tag{1}$$

Below, we show that a quantitative comparison of this formula with experiments [2] requires taking into account many-body effects, in particular, nonlinear screening. We analyze this problem in a controlled manner treating the strength of Coulomb interactions $\alpha = e^2/\kappa\hbar v$ as a small parameter. Here, κ is the effective dielectric constant. Small α can be realized using HfO_2 and similar large-κ substrates or in the presence of water (ice), $\kappa \sim 80$. For graphene on a SiO_2 substrate, $\alpha \approx 0.9$, in which case we expect corrections to our analytic theory to be ~25%.

Our main findings are as follows. First, the electric field at the *p-n* interface is given by

$$e|F_{\rm pn}| = 2.5\hbar v\alpha^{1/3}(\rho'_{\rm cl})^{2/3}, \tag{2}$$

where $\rho'_{\rm cl} > 0$ is the density gradient at the *p-n* interface computed according to classical electrostatics. Our result for $e|F_{\rm pn}|$ exceeds a naive estimate $e|F_{\rm pn}| = \hbar v k_F(\rho_1)/D$, where $k_F = \sqrt{\pi|\rho|}$ is the Fermi wave vector, by a parametrically large factor $(\alpha k_F D)^{1/3} \gg 1$ (which in practice may approach ~10). The enhancement is caused by the lack of screening at this interface where the quasiparticle density is very small. The calculation of $F_{\rm pn}$ requires going beyond the linear-response approach [9–11] to screening [12]. Our second main result is that Eq. (1) is rigorously valid if $\alpha \ll 1$, in which case we can substitute $F_{\rm pn}$ from Eq. (2) in Eq. (1) to obtain [13]

$$R = (1.0 \pm 0.1)(h/e^2)\alpha^{-1/6}(\rho'_{\rm cl})^{-1/3}. \tag{3}$$

This value of R is parametrically smaller than $(\pi/2) \times (h/e^2)\sqrt{k_F(\rho_1)/D}$ that one would get assuming the aforementioned naive estimate of $e|F_{\rm pn}|$ [14].

As an application of Eq. (3), let us consider a prototypical geometry depicted in Fig. 1. The voltage difference $-V_g$ between graphene and the semi-infinite gate with the edge at $x = x_g$ determines the total density drop $\rho_2 - \rho_1 = -\kappa V_g/4\pi eD$. The density ρ_2 is assumed to be fixed by other means, e.g., a global "backgate" on the opposite

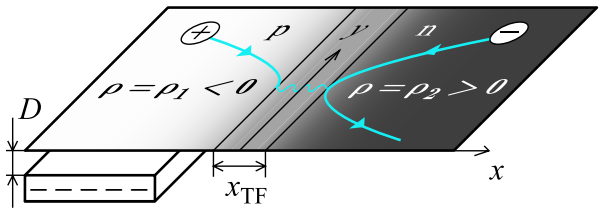

FIG. 1 (color online). Device geometry. The semi-infinite gate on the left side (beneath the graphene sheet) controls the density drop $\rho_2 - \rho_1$ across the junction, while another infinite backgate above the sheet (not shown) fixes the density ρ_2 at far right. The smooth curves with the arrows depict typical ballistic trajectories of an electron (−) and a hole (+). The wavy curve symbolizes their recombination via quantum tunneling.

side of the graphene sheet (not shown in Fig. 1). This model is a reasonable approximation to the available experimental setups [2]. An analytical expression for $\rho_{\rm cl}(x)$ follows from the solution of a textbook electrostatics problem, Eq. (10.2.51) of Ref. [15]. It predicts that function $\rho_{\rm cl}(x)$ crosses zero at the point $x_{\rm pn} = x_g + (D/\pi) \times [1 + (|\rho_1|/\rho_2) + \ln(|\rho_1|/\rho_2)]$. Thus, for obvious physical reasons, the position $x_{\rm pn}$ of the p-n interface is gate-voltage dependent [16]. Taking the derivative of $\rho_{\rm cl}$ at $x = x_{\rm pn}$ and substituting the result into Eq. (3), we obtain

$$R = \frac{0.7}{\alpha^{1/6}} \frac{h}{e^2} \left(1 - \frac{\rho_1}{\rho_2}\right)^{2/3} \left|\frac{D}{\rho_1}\right|^{1/3}, \qquad |\rho_1| \gg \frac{1}{D^2}. \quad (4)$$

At fixed ρ_2, $R(\rho_1)$ has an asymmetric minimum at $\rho_1 = -\rho_2$. Away from this minimum, the more dramatic $R(\rho_1)$ dependence (of potential use in device applications) occurs at the $\rho_1 \to 0$ side where R diverges. The reason for this behavior of R is vanishing of the density gradient $\rho'_{\rm cl}(x)$ at far left (above the gate). Equation (4) becomes invalid at $|\rho_1| \lesssim 1/D^2$ where the gradual junction approximation breaks down. At this point, $R \sim (h/e^2)D$.

Let us now turn to the derivation of the general formula (3). Our starting point is the basic principle of electrostatics, according to which we can replace the potential due to the external gates with that created by the fictitious in-plane charge density $\rho_{\rm cl}(x)$. Shifting the origin to $x = x_{\rm pn}$, we have the expansion $\rho_{\rm cl}(x) \simeq \rho'_{\rm cl} x$ for $|x| \ll D$. The induced charge density $\rho(x)$ attempts to screen the external one to preserve charge neutrality; thus, a p-n interface forms at $x = 0$. We wish to compute the deviation from the perfect screening $\sigma(x) \equiv \rho_{\rm cl}(x) - \rho(x)$ caused by the quantum motion of the Dirac quasiparticles.

Thomas-Fermi domain.—Consider the region $|x| \gg x_s$,

$$x_s \equiv (1/\pi)(\alpha^2 \rho'_{\rm cl})^{-1/3}. \quad (5)$$

At such x, the screening is still very effective, $|\sigma(x)| \ll |\rho_{\rm cl}(x)|$ because the local screening length $r_s(x)$ is smaller than the characteristic scale over which the background charge density $\rho_{\rm cl}(x)$ varies, in this case $\min\{|x|, D\}$. Indeed, the Thomas-Fermi (TF) screening length for graphene is [12] $r_s = (\kappa/2\pi e^2)(d\mu/d\rho) \sim 1/\alpha\sqrt{|\rho|}$, where μ is the chemical potential

$$\mu(\rho) = \mathrm{sgn}(\rho)\sqrt{\pi}\hbar v |\rho|^{1/2} \quad (6)$$

appropriate for the 2D Dirac spectrum. Substituting $\rho_{\rm cl}(x)$ for ρ, we obtain $r_s \sim |\alpha^2 \rho'_{\rm cl} x|^{-1/2}$ at $|x| \ll D$. Therefore, at $|x| \gg x_s$, the condition $r_s \ll |x|$ that ensures the nearly perfect screening is satisfied.

The behavior of the screened potential $V(x)$ and the electric field $F(x) = -dV/dx$ at $|x| \gg x_s$ can now be easily calculated within the TF approximation,

$$\mu[\rho(x)] - eV(x) = 0. \quad (7)$$

It leads to the relation

$$eF(x) \simeq -\hbar v \sqrt{\pi/4} (\rho'_{\rm cl}/|x|)^{1/2}, \qquad x_s \ll |x| \ll D, \quad (8)$$

which explicitly demonstrates the aforementioned enhancement of $|F(x)|$ near the junction. The TF approximation is valid if $k_F^{-1}(x) \ll \min\{|x|, D\}$. For $\alpha \sim 1$, this criterion is met if $|x| \gg x_s$. For $\alpha \ll 1$, the TF domain extends down to $|x| = x_{\rm TF} \sim \sqrt{\alpha} x_s$, see below.

A more formal derivation of the above results is as follows. From 2D electrostatics [15], we know that

$$\sigma(x) \equiv \rho_{\rm cl}(x) - \rho(x) = \frac{\kappa}{2\pi^2 e} \mathcal{P} \int \frac{dz}{z - x} F(z). \quad (9)$$

Combined with Eqs. (6) and (7), this yields

$$\rho(x) - \rho'_{\rm cl} x = \sqrt{\rho'_{\rm cl} x_s^3}\, \mathcal{P} \int_0^\infty \frac{x dz}{z^2 - x^2} \frac{d}{dz} \sqrt{|\rho(z)|}. \quad (10)$$

Here, the upper integration limit was extended to infinity, which is legitimate if $D \gg x_s$. The solution for $\rho(x)$ can now be developed as a series expansion in $1/x$. The leading correction to the perfect screening is obtained by substituting $\rho(x) = \rho'_{\rm cl} x$ into the integral, yielding $\sigma(x)/\rho(x) \simeq (\pi/4)|x_s/x|^{3/2}$. In accord with the arguments above, this correction is small at $|x| \gg x_s$. Furthermore, it falls off sufficiently fast with x to ensure that to the order $\mathcal{O}(x_s/D)$, the results for $V(x)$ and $\rho(x)$ at the origin are insensitive to the large-x behavior. In the opposite limit, $|x| \ll x_s$, one can show that

$$\rho_{\rm TF}(x) \simeq c^2 \rho'_{\rm cl} \frac{x^2}{x_s}, \qquad e|F_{\rm TF}| \simeq c\pi\hbar v \alpha^{1/3} (\rho'_{\rm cl})^{2/3}, \quad (11)$$

where $c \sim 1$ is a numerical coefficient. (The subscripts serve as a reminder that these results are obtained within the TF approximation.)

Unsuccessful in finding c analytically, we turned to numerical simulations. To this end, we reformulated the problem as the minimization of the TF energy functional

$$E[V(x)] = E_0 + \int eV(x) \left[\frac{1}{2}\sigma(x) - \rho_{\rm cl}(x)\right] dx, \quad (12)$$

$$E_0 = \frac{e^3}{3\pi\hbar^2 v^2} \int |V(x)|^3 dx, \quad (13)$$

where $\sigma(x)$ is defined by Eq. (9). The convolution integral in that equation was implemented by means of a discrete Fourier transform (FT) over a finite interval $-D \le x < D$. Similarly, the integral in Eq. (12) was implemented as a discrete sum. Since the FT effectively imposes the periodic boundary conditions on the system, we chose the background charge density in the form

$$\rho_{\rm cl}(x) = \rho_0 \sin(\pi x/D), \quad (14)$$

so that the p-n interfaces occur at all $x = nD$, where n is an integer. Starting from the initial guess $\sigma \equiv 0$, the solution for $\rho(x)$ and $V(x)$ within a unit cell $-D \le x < D$ was found by a standard iterative algorithm [17]. As shown in

Fig. 2(a), at large x, the TF density profile is close to Eq. (14). At small x, it is consistent with Eq. (11) using $c = 0.8 \pm 0.05$, cf. Fig. 2(c).

Dirac domain.—Let us now discuss the immediate vicinity of the p-n interface, $|x| < x_{\mathrm{TF}} \sim \sqrt{\alpha} x_s$ (the precise definition of x_{TF} is given below). At such x, the TF approximation is invalid and instead we have to use the true quasiparticle wave functions to compute ρ and V. For a gradual junction, the two inequivalent Dirac points ("valleys") of graphene [1] are decoupled, and the wave functions can be chosen to be two-component spinors $\exp(ik_y y)[\psi_1(x)\psi_2(x)]^T$ (their two elements represent the amplitudes of the wave function on the two sublattices of graphene). Here we already took advantage of the translational invariance in the y-direction and introduced the conserved momentum k_y. The effective Hamiltonian we need to diagonalize has the Dirac form

$$H = \hbar v(-i\tau_1 \partial_x + \tau_2 k_y) - eV(x), \tag{15}$$

where τ_1 and τ_2 are the Pauli matrices. At the end of the calculation, we will need to multiply the results for $\rho(x)$ by the total spin-valley degeneracy factor $g = 4$.

The solution of this problem can be obtained analytically under the condition $\alpha \ll 1$. In this case, it is legitimate to directly substitute the TF result for the electrostatic potential $V(x)$ into Eq. (15). In particular, we can assume that the electric field $F(x)$ is uniform at $|x| \ll x_s$.

In order to understand this statement, one needs to distinguish between the electron density $\rho(x)$ and the total charge density $\sigma(x) = \rho_{\mathrm{cl}}(x) - \rho(x)$. As mentioned above, $\rho(x)$ is correctly described by the TF approximation only at $|x| \gg \sqrt{\alpha} x_s$. Nevertheless, the leading-order behavior of $\sigma(x)$ is given by this approximation at *all* x. The deviation of the true $\rho(x)$ from $\rho_{\mathrm{TF}}(x)$ appears only when both of them are so small that the total charge density is dominated by the external one: $\sigma(x) \simeq \rho'_{\mathrm{cl}} x$. Since the electric field is determined by the total charge, at any x, it can be safely taken from the TF solution.

Let us elaborate. Since the potential $V(x)$ is small near the interface and the spectrum is gapless, $\rho(x)$ must be smooth and have a regular Taylor expansion at $x \to 0$,

$$\rho(x) = a_1 x + a_3 x^3 + \ldots \tag{16}$$

Requiring the leading term to match with the TF Eq. (11) at the common boundary $x = x_{\mathrm{TF}} \sim \sqrt{\alpha} x_s$ of their validity, we get $a_1 \sim \sqrt{\alpha} \rho'_{\mathrm{cl}}$. This means that the net charge per unit length of the interface on the n-side of the junction is somewhat smaller than the TF approximation predicts, by the amount of $\Delta Q = e \int_0^\infty [\rho(x) - \rho_{\mathrm{TF}}(x)] dx \sim \sqrt{\alpha} \rho'_{\mathrm{cl}} x_{\mathrm{TF}}^2$. In turn, the true $|F_{\mathrm{pn}}|$ is lower than $|F_{\mathrm{TF}}|$ by $\sim \Delta Q / \kappa x_{\mathrm{TF}}$. However, for $\alpha \ll 1$, this is only a small, $\mathcal{O}(\alpha)$ relative correction.

As soon the legitimacy of the linearization $V(x) \simeq -F_{\mathrm{pn}} x$ is established, wave functions ψ_1 and ψ_2 for arbitrary energy ϵ are readily found. Since ϵ enters the Dirac equation only in the combination $-eV(x) - \epsilon = eF_{\mathrm{pn}}(x - x_\epsilon)$, the energy-$\epsilon$ eigenfunctions are the $\epsilon = 0$ eigenfunctions shifted by $x_\epsilon \equiv \epsilon/(eF_{\mathrm{pn}})$ in x. In turn, these are known from the literature [7,8]. They are expressed in terms of confluent hypergeometric functions $\Phi(a; b; z)$ [18]. The sought electron density $\rho(x)$ can now be obtained by a straightforward summation over the occupied states ($\epsilon \le 0$), which leads us to [19]

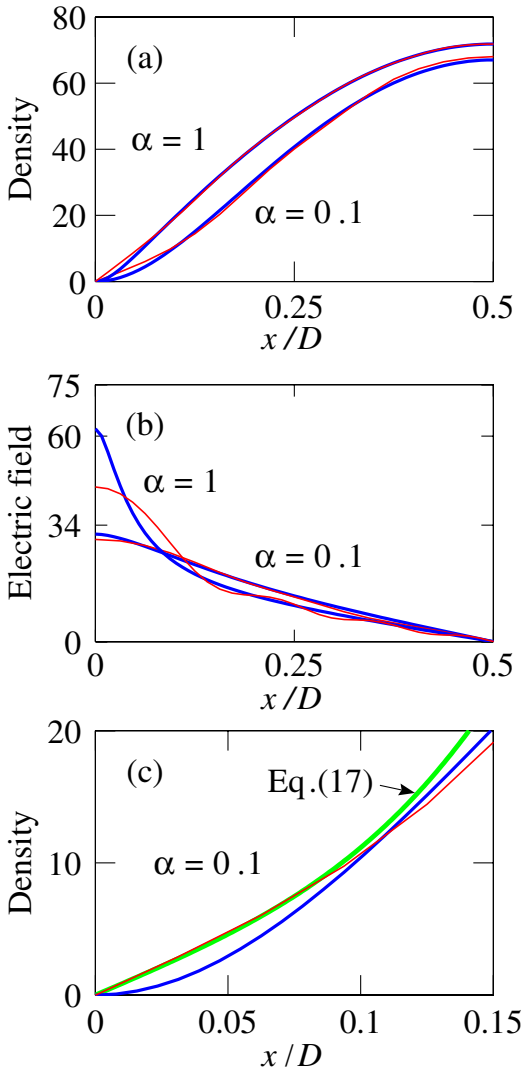

FIG. 2 (color online). (a) Electron density in units of $4/D^2$ for $\alpha = 1$, $\rho_0 = 75$ and $\alpha = 0.1$, $\rho_0 = 100$. Thicker curves are from minimizing the TF functional, Eqs. (12)–(14); thinner lines are from replacing E_0 in this functional by the ground-state energy of Hamiltonian (15). The p-n interface is at $x = 0$. (b) Magnitude of the electric field in units of $4\hbar v/eD^2$ for the same parameters. Numerical values "34" and "60" are the predictions of Eq. (2) for the nearby TF (thick) curves [20]. (c) Enlarged view of the $\alpha = 0.1$ data from the panel (a) and the numerically evaluated Eq. (17).

$$\rho = \frac{g}{x_{\mathrm{TF}}^2}\int\frac{dk_y}{2\pi}\int_0^x \frac{dz}{\pi e^{2\pi\nu}}\left[\left|\Phi\left(i\nu;\frac{1}{2};\frac{z^2}{ix_{\mathrm{TF}}^2}\right)\right|^2-\frac{1}{2}\right] \quad (17)$$

where $\nu = k_y^2 x_{\mathrm{TF}}^2/4$ and $x_{\mathrm{TF}} \equiv \sqrt{\hbar v/|F_{\mathrm{pn}}|} \sim \sqrt{\alpha} x_s$. This formula is fully consistent with Eq. (14): the Taylor expansion of the integrand yields, after a simple algebra, $a_1 = g/(\sqrt{2}\pi^2 x_{\mathrm{TF}}^3)$, $a_3 = g\sqrt{2}/(3\pi^3 x_{\mathrm{TF}}^5)$, etc. Using the known integral representations of the function Φ [18], one can also deduce $\rho(x)$ at $x \gg x_{\mathrm{TF}}$. The leading term is precisely the TF result $\rho_{\mathrm{TF}}(x) = gx^2/4\pi x_{\mathrm{TF}}^4$. Therefore, Eq. (17) seamlessly connects to Eq. (11) at $x \sim x_s$. [At such x, corrections to $\rho_{\mathrm{TF}}(x)$, including Friedel-type oscillations, are suppressed by extra powers of parameter α.] We conclude that for $\alpha \ll 1$, we have obtained the complete and rigorous solution for $\rho(x)$, $V(x)$, and F_{pn} [Eq. (2)], in particular. As discussed in the beginning, it immediately justifies the validity of Eq. (1) and leads to our result for the ballistic resistance, Eq. (3). However, in current experiments $\alpha \sim 1$ and in the remainder of this Letter, we offer a preliminary discussion of what can be expected there.

Since it is the strip $|x| < x_{\mathrm{TF}}$ that controls the ballistic transport across the GPNJ [7], the constancy of the electric field in this strip is crucial for the accuracy of Eq. (1). This is assured if $\alpha \ll 1$, but at $\alpha \sim 1$, the buffer zone between x_{TF} and x_s vanishes, and so we expect $F(x_{\mathrm{TF}})$ and $F(0) = F_{\mathrm{pn}}$ to differ by some numerical factor.

To investigate this question, we again turned to numerical simulations. We implemented a lattice version of the Dirac Hamiltonian by replacing $-i\partial_x$ in Eq. (13) with a finite difference on a uniform grid. We also replaced E_0 in Eq. (15) by the ground-state energy of H, taken with the negative sign: $E_0 = -L_y^{-1}\sum_j \epsilon_j/[1 + \exp(\beta\epsilon_j)]$. Here, ϵ_j are the eigenvalues of H (computed numerically) and the β is a computational parameter (typically, 4 orders of magnitude larger than $1/\max|V|$). We have minimized thus modified functional E by the same algorithm [17], which produced the results shown in Fig. 2. As one can see, for $\alpha = 0.1$, the agreement between analytical theory and simulations is very good. However, for $\alpha = 1$, we find that $|F_{\mathrm{pn}}|$ is approximately 25% smaller than given by Eq. (2). Note also that for $\alpha = 1$, the electric field is noticeably nonuniform near the junction, in agreement with the above discussion [20]. Therefore, Eq. (1) should also acquire some corrections. In principle, we could compute numerically the transmission coefficients $T(k_y)$ for this more complicated profile of $F(x)$. However, this would not be the ultimate answer to this problem. Indeed, at $\alpha \sim 1$ electron interactions are not weak, and so exchange and correlation effects are likely to produce further corrections to the self-consistent single-particle scheme we employed thus far, which may be quite nontrivial inside the Dirac strip $|x| < x_{\mathrm{TF}}$. We leave this issue for future investigation.

We are grateful to V. I. Falko, D. S. Novikov, and B. I. Shklovskii for valuable comments and discussions, to NSF DMR and UCSD ACS for support, and to the Aspen Center for Physics and W. I. Fine TPI for hospitality (M. F.).

[1] For a review, see A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, arXiv:0709.1163.
[2] B. Huard, J. A. Sulpizio, N. Stander, K. Todd, B. Yang, and D. Goldhaber-Gordon, Phys. Rev. Lett. **98**, 236803 (2007); B. Özyilmaz, P. Jarillo-Herrero, D. Efetov, D. A. Abanin, L. S. Levitov, and P. Kim, Phys. Rev. Lett. **99**, 166804 (2007); J. R. Williams, L. DiCarlo, and C. M. Marcus, Science **317**, 638 (2007).
[3] M. V. Fistul and K. B. Efetov, Phys. Rev. Lett. **98**, 256803 (2007).
[4] A. Ossipov, M. Titov, and C. W. J. Beenakker, Phys. Rev. B **75**, 241401(R) (2007).
[5] V. V. Cheianov, V. I. Falko, and B. L. Altshuler, Science **315**, 1252 (2007).
[6] M. I. Katsnelson, K. S. Novoselov, and A. K. Geim, Nature Phys. **2**, 620 (2006).
[7] V. V. Cheianov and V. I. Fal'ko, Phys. Rev. B **74**, 041403(R) (2006).
[8] E. O. Kane and E. I. Blount, in *Tunneling Phenomena in Solids*, edited by E. Burstein and S. Lundqvist (Plenum, New York, 1969), pp. 79–91.
[9] T. Ando, J. Phys. Soc. Jpn. **75**, 074716 (2006).
[10] B. Wunsch, T. Stauber, F. Sols, and F. Guinea, New J. Phys. **8**, 318 (2006).
[11] E. H. Hwang and S. Das Sarma, Phys. Rev. B **75**, 205418 (2007).
[12] See also M. M. Fogler, D. S. Novikov, and B. I. Shklovskii, Phys. Rev. B **76**, 233402 (2007), and references therein.
[13] Disorder, unavoidable in real systems, may produce correction to this formula, see M. M. Fogler, L. I. Glazman, D. S. Novikov, and B. I. Shklovskii, Phys. Rev. B **77**, 075420 (2008).
[14] These incorrect values of F and R could be inferred from Ref. [7] if D is incautiously identified with parameter d therein, as Fig. 1 of that paper prompts one to do.
[15] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[16] One can manipulate x_{pn} by the backgate voltage, which shifts $\rho_{\mathrm{cl}}(x)$ by a constant affecting neither $\rho'_{\mathrm{cl}}(x)$ nor R.
[17] Function fminunc of MATLAB, ©MathWorks, Inc.
[18] I. S. Gradshteyn and I. M. Ryzhik, in *Table of Integrals, Series, and Products*, edited by A. Jeffrey and D. Zwillinger (Academic, San Diego, 2000), 6th ed..
[19] A similar expression was derived in A. V. Andreev, Phys. Rev. Lett. **99**, 247204 (2007) in the context of carbon nanotube *p-n* junctions. The only difference is that no integration over k_y is present there.
[20] The undulations of $F(x)$ seen on the $\alpha = 1$ curves in Fig. 2(b) may be the aforementioned Friedel oscillations, but we cannot exclude numerical artifacts either.