Coulomb Impurity Problem in Graphene

Vitor M. Pereira, Johan Nilsson, and A. H. Castro Neto

Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA
(Received 27 June 2007; published 15 October 2007)

We address the problem of an unscreened Coulomb charge in graphene and calculate the local density of states and displaced charge as a function of energy and distance from the impurity. This is done nonperturbatively in two different ways: (1) solving the problem exactly by studying numerically the tight-binding model on the lattice and (2) using the continuum description in terms of the 2D Dirac equation. We show that the Dirac equation, when properly regularized, provides a qualitative and quantitative low energy description of the problem. The lattice solution shows extra features that cannot be described by the Dirac equation: namely, bound state formation and strong renormalization of the van Hove singularities.

Since the recent isolation of graphene [1], there has been an intense interest in understanding its electronic properties. The low energy electronic excitations in graphene are linearly dispersing, massless, chiral Dirac fermions, described by Dirac cones at the edges of the Brillouin zone (BZ) (at the K and K' points). Because of the vanishing density of states of these Dirac fermions in undoped graphene, the system is very sensitive to impurities and defects [2], which control its transport properties [3,4], and exhibits unconventional Friedel oscillations [5]. This also means that electrons in graphene screen poorly, and hence the issue of unscreened Coulomb interactions becomes paramount in the understanding of the experimental data [1].

In this Letter, we contrast the tight-binding approach (that we solve exactly with numerical techniques) with the continuum approach based on the Dirac equation. By obtaining the exact analytical solution, we show that the latter provides a good qualitative description of the problem at low energies, when properly regularized. We calculate the local density of states (LDOS) and induced charge around a Coulomb impurity as a function of energy and distance. These quantities are experimentally accessible through scanning tunneling spectroscopy (STS). We stress that calculations with long-range potentials are radically different from the short-range ones, which are exactly solvable within the T matrix [6].

Consider the problem of a single Coulomb impurity, with charge Ze, placed in the middle of a hexagon of the honeycomb lattice [7]. The tight-binding Hamiltonian for this problem, with nearest-neighbor hopping only, is given by (we use units such that $\hbar = 1$):

$$\mathcal{H} = t\sum_i (a_i^\dagger b_i + \text{H.c.}) + \frac{Ze^2}{\varepsilon_0}\sum_i \left(\frac{a_i^\dagger a_i}{r_i^A} + \frac{b_i^\dagger b_i}{r_i^B}\right), \quad (1)$$

where a_i (b_i) annihilates an electron at site $\mathbf{R}_i$ and sublattice A (B), $t \approx 2.7$ eV is the hopping energy, and $r_i^{A,B}$ is the distance between the carbon atoms and the impurity

(assumed to be at the origin of the coordinate system); ε_0 is the dielectric constant. We calculated numerically the spectrum of (1) using the methods of exact diagonalization and recursion used in Ref. [8] for the study of short-range unitary scatterers.

Close to the K point in the BZ, we can write an effective low energy Hamiltonian for (1) in terms of Dirac fermions with a spinor wave function $\Psi(\mathbf{r})$, whose components represent its weight on each sublattice. The wave function obeys the equation:

$$\nu_F(\boldsymbol{\sigma} \cdot \boldsymbol{p} - g/r)\Psi(\mathbf{r}) = E\Psi(\mathbf{r}), \quad (2)$$

with $\nu_F = 3at/2$ ($\approx 10^6$ m/s) the Fermi velocity, $\boldsymbol{p}$ the 2D momentum operator, σ_i the Pauli matrices, and $g \equiv Ze^2/(\nu_F\varepsilon_0)$ the dimensionless coupling constant. Henceforth, we shall take a (the C-C distance) and ν_F as distance and energy units. Notice that (2) does not involve intercone scattering because the unscreened Coulomb potential is dominated by small momentum transfers since, in Fourier space, it behaves like $1/q$ and is singular as $q \rightarrow 0$.

Equation (2) is separable in cylindrical coordinates. Resorting to eigenfunctions of the conserved angular momentum $J_z = L_z + \sigma_z/2$ [9],

$$\psi_j(r) = \frac{1}{\sqrt{r}}\begin{pmatrix} e^{i[j-(1/2)]\varphi}\varphi_j^A(r) \\ ie^{i[j+(1/2)]\varphi}\varphi_j^B(r) \end{pmatrix}, \quad (3)$$

and the radial equation for (2) reads ($j = \pm 1/2, \pm 3/2, \dots$)

$$\begin{bmatrix} \epsilon + g/r & -(\partial_r + j/r) \\ (\partial_r - j/r) & \epsilon + g/r \end{bmatrix}\begin{bmatrix} \varphi_j^A \\ \varphi_j^B \end{bmatrix} \equiv \mathcal{M}_j\varphi_j(r) = 0. \quad (4)$$

This equation can be solved by multiplication on the left by $\mathcal{M}_j' = \sigma_z\mathcal{M}_j\sigma_z$ and subsequent diagonalization. The eigenstates are then linear combinations of the type

$$\varphi_j(r) = \sum_{\lambda=\pm} C_\lambda u_\lambda f_\lambda(r), \qquad u_\pm = \sqrt{\frac{1}{2|j|}} \begin{pmatrix} \sqrt{|j \pm \alpha|} \\ s_{gj}\sqrt{|j \mp \alpha|} \end{pmatrix},$$

$$\tag{5}$$

where $s_x \equiv \text{sgn}(x)$, $\alpha = \sqrt{j^2 - g^2}$, and $f_\lambda(r)$ solves

$$\partial_r^2 f_\lambda(r) + [\epsilon^2 + 2g\epsilon/r - \alpha(\alpha - \lambda)/r^2]f_\lambda(r) = 0. \tag{6}$$

Introducing $\rho = |\epsilon|r$, the above becomes the familiar radial equation for the 3D Coulomb problem [10], and the presence of ϵ^2 (rather than ϵ) entails the absence of bound solutions in the Dirac problem. It is important to note that, when g is above $g_c = 1/2$, the parameter α in Eq. (6) becomes imaginary for some angular momentum channels. The nature of the solutions is then radically different, which, as will be seen, has dramatic consequences. We address the two regimes separately.

When $g < g_c$, Eq. (6) can be solved in terms of Coulomb wave functions [11,12]: $F_L(\eta, \rho), G_L(\eta, \rho)$. In fact, letting $\tilde{g} = s_\epsilon g$, it is straightforward to show that the appropriate linear combination in (5) that solves (4) is

$$\varphi_j(r)/\mathcal{N}_j = u_+ F_{\alpha-1}(-\tilde{g}, \rho) + s_{g\epsilon} u_- F_\alpha(-\tilde{g}, \rho), \tag{7}$$

where only the regular solution at the origin has been kept. Since $F_\alpha(-\tilde{g}, \rho)$ are the regular scattering solutions of the 3D Coulomb problem, they include the well-known logarithmic phase shift in the asymptotic expansion [10]:

$$F_\alpha(-\tilde{g}, \rho) \sim \sin[\rho + \tilde{g}\log(2\rho) + \vartheta_\alpha(\tilde{g})], \tag{8}$$

where $\vartheta_\alpha(\tilde{g}) = -\alpha\frac{\pi}{2} + \arg[\Gamma(1 + \alpha - i\tilde{g})]$. The logarithmic phase shift also carries to our case, for (7) can always be written asymptotically as

$$\sin[\rho + \tilde{g}\log(2\rho) + \arg(u_+ e^{i\vartheta_{\alpha-1}} + s_{g\epsilon} u_- e^{i\vartheta_\alpha})]. \tag{9}$$

The normalization $\mathcal{N}_j$ is determined by imposing orthogonality on the energy scale $\int \psi_i(\epsilon, r)^\dagger \psi_j(\epsilon', r) dr = \delta_{ij}\delta(\epsilon - \epsilon')$, leading to $\mathcal{N}_j^{-2} = 2\pi^2\alpha^2/j^2$. With this choice, one conveniently recovers the free DOS per unit area and cone when $\tilde{g} = 0$. To see this, one notes that the LDOS $N(\epsilon, r) = \sum_E |\Psi_E(r)|^2 \delta(\epsilon - E)$ is given by $N(\epsilon, r) = \sum_{j=-\infty}^{\infty} n_j(\epsilon, r)$. Using $n_j(\epsilon, r) \equiv r^{-1}|\varphi_j^A(r)|^2 + r^{-1}|\varphi_j^B(r)|^2$, the contribution from each angular momentum channel is simply

$$n_j(\epsilon, r) = (\mathcal{N}_j^2/r)[F_{\alpha-1}^2 + F_\alpha^2 + 2\tilde{g}F_\alpha F_{\alpha-1}/|j|]. \tag{10}$$

In the limit $\tilde{g} \to 0$, the Coulomb wave functions reduce to Bessel functions [11], and one obtains $N(\epsilon, r) = |\epsilon|/2\pi$.

Of the several aspects encoded in (10), two are immediate: Particle-hole symmetry is lost, and the LDOS becomes singular as $E \to 0$. This last point follows from the fact that, in this limit, $N(\epsilon, r) \propto |\epsilon|^{2\alpha}$ and $\alpha < 1/2$ for $|j| = 1/2$; the asymmetry stems from the dependence of $\tilde{g}$ on the sign of the energy.

It is most instructive to compare the results derived within the Dirac approximation (2) with the results on

the lattice that one obtains using the full Hamiltonian in Eq. (1). In Fig. 1(a), one can observe that, at low energies ($E \lesssim 0.5t$), the result of (10) reproduces the LDOS on the lattice even at distances of the order of the lattice parameter (the two cases are barely distinguishable for most of the plotted range). Moreover, the attractive Coulomb potential brings locally a reduction of spectral weight in the lower band, the opposite happening to the upper band. The effect is strongest near the impurity and evolves towards the bulk behavior at larger distances.

This behavior of the spectrum near the Dirac point can be understood from an investigation of the quantized energies when the system is restricted to a region of finite radius R. A convenient way to confine 2D Dirac fermions is to introduce an infinite mass at the boundary [13], which translates into the boundary condition (BC) $\varphi_j^A(R) = \varphi_j^B(R)$. From (7) this expands into

$$Q_j \equiv F_{\alpha-1}(-\tilde{g}, R|\epsilon|) - s_j s_\epsilon F_\alpha(-\tilde{g}, R|\epsilon|) = 0. \tag{11}$$

This equation, whose roots determine the quantized energy levels, always has the trivial solution $\epsilon = 0$. As one can easily verify [see Fig. 1(c) and also the asymptotic phase shift in (9)], the other nodes are simply shifted to lower energies with increasing g — just as expected under an attractive potential — but they never cross $\epsilon = 0$. This means that no states are phase-shifted across the Dirac point, but they rather heap up close to it when $\epsilon > 0$ and conversely for $\epsilon < 0$. Hence, even though the gap in graphene is zero, no states will cross it while $g < g_c$, much like in a conventional semiconductor. In the spirit of Friedel's argument [14], this effect has profound consequences for the induced charge, which will be discussed later. Finally, we remark that, although the continuum
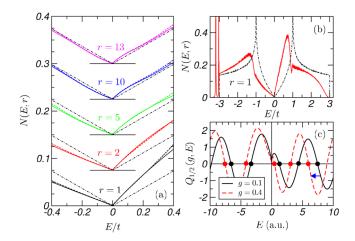

FIG. 1 (color online). (a) Comparison of the LDOS (solid line) with the numerical results in the lattice (dashed line), calculated at different distances from the impurity and $g = 1/6$. The DOS for $g = 0$ is also included for comparison (dotted-dashed line). For clarity, curves at different r have been vertically displaced. (b) LDOS at the site closest to the impurity in the lattice for $g = 1/3$. (c) Quantization condition (11) with $j = 1/2$, for $g = 0.1$ and 0.4.

approximation does not support bound solutions (unless a cutoff is introduced), they appear naturally in the lattice. This is seen in Fig. 1(b), where a bound state barely detached from the band is signaled by the sharp peak at the lower band edge. Furthermore, notice how the van Hove singularities are strongly renormalized by the presence of the impurity.

For $g > g_c$, α can become purely imaginary, and we introduce $\beta = -i\alpha = \sqrt{g^2 - j^2}$ for those j's such that $|g| > |j|$. In general, linearly independent solutions of (6) are [15]

$$\lambda = +1: \quad F_{\alpha-1}(-\tilde{g}, \rho) \quad \text{and} \quad F_{-\alpha}(-\tilde{g}, \rho), \quad (12)$$

$$\lambda = -1: \quad F_{\alpha}(-\tilde{g}, \rho) \quad \text{and} \quad F_{-\alpha-1}(-\tilde{g}, \rho). \quad (13)$$

When $\alpha \in \mathbb{R}$, the ones with negative index are divergent at the origin, and thus only the first were kept in (7). But when $\alpha \in i\mathbb{R}$, the solutions are well-behaved at the origin (albeit oscillatory), and two linearly independent solutions emerge. One is analogous to (7):

$$\bar{\varphi}_{i\beta}(r) = \bar{u}_+ F_{i\beta-1}(-\tilde{g}, \rho) + s_{jg\epsilon}\bar{u}_- F_{i\beta}(-\tilde{g}, \rho), \quad (14)$$

apart from a normalization factor, and where now

$$\bar{u}_\pm = \sqrt{\frac{1}{2|g|}} \begin{pmatrix} \sqrt{j \pm i\beta} \\ s_g\sqrt{j \mp i\beta} \end{pmatrix}. \quad (15)$$

The other solution is simply $\bar{\varphi}_{-i\beta}(r)$. The general solution is therefore of the type $\bar{\varphi}_j(r) = C_1\bar{\varphi}_{i\beta}(r) + C_2\bar{\varphi}_{-i\beta}(r)$, where $C_{1,2}$ are to be set by the BC at short distances. Since we seek the effective low energy description of a problem defined in a lattice, a natural BC is to have an infinite mass at some short cutoff distance $a_0 \simeq a$. This has the effect of forbidding the penetration of electrons to distances shorter than a_0 [13], thus reflecting the physical situation, while, at the same time, naturally curing the divergence in the potential at the origin. This translates again into a BC $\varphi_j^A(a_0) = \varphi_j^B(a_0)$, and, given that $C_{1,2}$ can always be chosen so that $C_1/C_2 = \exp[2i\delta_j(\epsilon)]$, one then obtains the phase $\delta_j(\epsilon)$:

$$e^{i2\delta_j(\epsilon)} = s_g \frac{F_{-i\beta-1} - s_{\epsilon j}F_{-i\beta}}{F_{i\beta-1} - s_{\epsilon j}F_{i\beta}}\bigg|_{\rho=\epsilon a_0}. \quad (16)$$

We can follow the same procedure as before to normalize the states in the energy scale and then extract the contribution of the overcritical j's to the LDOS:

$$\bar{n}_j(\epsilon, r) = \frac{1}{2\pi^2 r} \frac{\varrho_j^I(\rho) + s_{\epsilon j}\text{Re}[e^{i2\delta_j}\varrho_j^{II}(\rho)]}{\langle \varrho_j^I(\infty) + s_{\epsilon j}\text{Re}[e^{i2\delta_j}\varrho_j^{II}(\infty)]\rangle_r}, \quad (17)$$

where, for readability, we defined

$$\varrho_j^I \equiv |F_{i\beta}|^2 + |F_{i\beta-1}|^2 + \frac{2|j|}{\tilde{g}}[F_{i\beta}F_{-i\beta-1}],$$

$$\varrho_j^{II} \equiv 2F_{i\beta}F_{i\beta-1} + \frac{|j|}{\tilde{g}}(F_{i\beta}^2 + F_{i\beta-1}^2),$$

and $\langle\cdots\rangle_r$ stands for the constant term as $r \to \infty$. Equations (10) and (17) determine the LDOS for any coupling strength g, which can be summarized as

$$N(\epsilon, r) = \sum_{|j|<|g|} \bar{n}_j(\epsilon, r) + \sum_{|j|>|g|} n_j(\epsilon, r). \quad (18)$$

The presence of the first term in Eq. (18) brings a profound rearrangement of the spectrum close to the impurity, with much more striking consequences than in the weak coupling regime. In Figs. 2(a)–2(c), we plot the LDOS obtained numerically in the lattice with $g = 4/3$, together with the first contribution in (18). For such g, it comes only from $\bar{n}_{\pm 1/2}(\epsilon, r)$, and we used $a_0 = 0.55a$ to impose the BC. It is clear that the analytical result captures quite accurately the behavior of the LDOS in the lattice. Most importantly, both results exhibit 3 *marked resonances* in the negative (holelike) energy region, which decay away from the impurity. Indeed, their amplitude is such that they dominate the profile of the LDOS at low energies. Increasing g will cause the resonances to migrate downwards in energy and their number to increase. This is rather peculiar and has to do with the fact that, in reality, the Dirac point is an accumulation point of infinitely many resonances [inset in Fig. 2(e)]. One can appreciate the origin of this from the fact that $F_{L-1}(\eta, \rho \simeq 0) \sim \rho^L$. Since $L \in \mathbb{C}$ in Eqs. (14) and (16), it implies that the wave functions oscillate with logarithmically diverging frequency as $\epsilon \to 0$. This situation is akin to the *fall of a particle to the center* [10], and the effect carries to the LDOS with the consequences shown in Figs. 2(a)–2(c). In Fig. 2(d), we present the remainder contribution (second term) to the total LDOS in Eq. (18). It is evident that in the region $\epsilon \lesssim 0$, dominated by the resonances, this contribution is highly suppressed, whereas, for positive energies, the LDOS exhibits an oscillating behavior around the bulk limit.
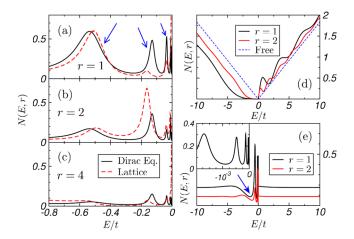

FIG. 2 (color online). (a)–(c) The LDOS in the lattice (dashed line, recursion method) is compared with the first contribution in (18) (solid line) for different distances from the impurity. (d) The second contribution in (18) (solid line) and the free, linear, DOS for reference. (e) First contribution in (18); the inset is a magnification for $\epsilon \simeq 0$. In all panels, $g = 4/3$.

As is customary, it is also of interest here to understand how the electronic density readjusts itself in the presence of this charged impurity. Even though interactions are not included (and thus there is no real screening), one can obtain important insights from the noninteracting problem in the spirit of Friedel [14]. The charge density $n(r)$ is straightforwardly obtained from integration of (18) in energy from an energy cutoff $-\Lambda$ up to $E_F = 0$. Since it involves integrals of $F_\alpha(-\tilde{g}, \rho)$, this can be done exactly. For example, when $g < g_c$, one has (for each j channel)

$$n_j(r) = [\rho n_j(\epsilon, r)/r - |j|F_\alpha F_{\alpha-1}/(\pi^2 r^2)]|_{\epsilon=-\Lambda}. \quad (19)$$

Although the above needs only to be evaluated at $\epsilon = -\Lambda$, it is not free from difficulties yet, for there is an infinite sum over j to be performed. Expanding (19) asymptotically, the induced charge per channel reads

$$\delta n_j(r) = n_j(r) - n_j^0(r) \sim (1/r)[\Lambda - g/r - \Lambda^0 + \mathcal{O}(r^{-2})], \quad (20)$$

where the remainder is oscillating with frequency Λ and convergent with j. Clearly, if $\Lambda = \Lambda^0$, the sum over j diverges. We regularize this by locally changing the cutoff: $\Lambda = \Lambda^0 + g/r$, whereupon the leading contribution is $\sim r^{-3}$ and oscillating with frequency Λ [Fig. 3(d)]. Nonetheless, despite accidentally reproducing the lattice behavior, the oscillation itself is tied to the cutoff procedure. We point out that any charge oscillation decaying faster than $1/r^2$ on the lattice appears, in the continuum theory, as a Dirac delta function at the origin, in agreement with perturbative studies of this problem [16], but differs from the self-consistent calculation in Ref. [9]. Interestingly, the behavior of the induced charge in the lattice is indeed $\sim r^{-3}$ and oscillating, as seen in Fig. 3(a), wherein exact numerical results in the lattice are plotted. An analogous analytical procedure can be undertaken for $g > g_c$, leading to an induced charge decaying as $\sim r^{-2}$. Figure 3(b) shows that this agrees with the numerical data in the lattice, where $\delta n(r) \sim r^{-2}$, and nonoscillating. One thus concludes that the induced charge behaves *quite differently* below and above g_c, as had been hinted following Eq. (11), on account of the peculiar behavior of the phase shifts below g_c. This last point can be confirmed by inspecting the behavior of the numerical energy levels (E_n) as a function of g shown in Fig. 3(c), being evident the difference between the two regimes: Below g_c, the E_n never cross $E_F = 0$.

We studied the problem of a Coulomb charge in graphene via exact numerical methods on the lattice and the Dirac Hamiltonian. Calculating the LDOS and local charge as a function of energy and distance from the impurity, we found that the Dirac equation provides a qualitative description of the problem at low energies. We found new features in the lattice description that are naturally beyond the Dirac equation: bound states and strong renormalization of the van Hove singularities. We have also shown the
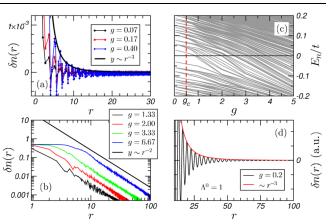

FIG. 3 (color online).   (a) Induced charge numerically obtained from exact diagonalization on a lattice with 124^2 sites ($g < g_c$). (b) *Idem* ($g > g_c$). (c) Evolution of the ordered numerical spectrum with g, E_n being the nth energy level. The symmetries of the lattice cause some E_n to merge as $g \to 0$. (d) Analytical $\delta n(r)$ obtained using (19) and regularization discussed after (20).

existence of a critical g_c separating the weak and strong coupling regimes, with radical differences in the features of the LDOS and induced charge. These results can be tested experimentally through STS.

Note added.—While preparing the manuscript, we became aware of two preprints [17,18] with a similar approach to this problem.

[1] A. K. Geim and K. S. Novoselov, Nat. Mater. **6**, 183 (2007).
[2] N. M. R. Peres *et al.*, Phys. Rev. B **73**, 125411 (2006).
[3] K. Nomura *et al.*, Phys. Rev. Lett. **98**, 076602 (2007).
[4] S. Adam *et al.*, arXiv:0705.1540.
[5] V. V. Cheianov *et al.*, Phys. Rev. Lett. **97**, 226801 (2006).
[6] N. M. R. Peres *et al.*, arXiv:0705.3040.
[7] This geometry is chosen to preserve sublattice symmetry.
[8] Vitor M. Pereira *et al.*, Phys. Rev. Lett. **96**, 036801 (2006).
[9] D. P. DiVincenzo *et al.*, Phys. Rev. B **29**, 1685 (1984).
[10] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Pergamon, New York, 1981).
[11] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[12] F. L. Yost *et al.*, Phys. Rev. **49**, 174 (1936).
[13] M. V. Berry *et al.*, Proc. R. Soc. A **412**, 53 (1987).
[14] J. Friedel, Philos. Mag. **43**, 153 (1952).
[15] If $2L \notin \mathbb{Z}$, one can express $G_L(\eta, \rho)$ in terms of $F_L(\eta, \rho)$ and $F_{-L-1}(\eta, \rho)$, much like as in the Bessel functions.
[16] A. Kolezhuk *et al.*, Phys. Rev. B **74**, 165114 (2006).
[17] A. V. Shytov *et al.*, arXiv:0705.4663.
[18] D. S. Novikov, arXiv:0706.1391.