Supercritical Coulomb impurities in gapped graphene

Vitor M. Pereira, Valeri N. Kotov, and A. H. Castro Neto

Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA

(Received 16 June 2008; published 4 August 2008)

We study the problem of Coulomb field-induced charging of the ground state in a system of two-dimensional (2D) massive Dirac particles—gapped graphene. As in its 3D QED counterpart, the critical Coulomb coupling is renormalized to higher values compared to the massless case. We find that in gapped graphene, a different supercritical regime is possible, where the screening charge is comparable to the impurity charge; thus, leading to suppression of the Coulomb field at nanometer scales. We corroborate this with a numerical solution of the tight-binding problem on the honeycomb lattice.

I. INTRODUCTION

The successful experimental isolation of a single plane of sp^2 carbon atoms (graphene) has stirred our deepest assumptions about condensed-matter systems.[1–3] Not only is graphene unique in the realm of solid state, but it also has a vast potential as a test ground for many of the remarkable predictions of quantum electrodynamics (QED).[4,5] One remarkable characteristic of graphene is that its electronic properties can be tailored in many different ways, either by applying transverse electric and magnetic fields, by changing its geometry, or by modifying the substrate where graphene is deposited or grown.[3]

In fact, while graphene deposited in SiO$_2$ is well described by the two-dimensional (2D) massless Dirac equation,[1] graphene grown on SiC can be described in terms of massive 2D Dirac electrons.[6] Substrate induced potentials can break symmetries of the honeycomb lattice and generate gaps in the electronic spectrum. Hence, by suitable choice of substrates, one can tune the "rest mass" of the "relativistic" particles and explore phenomena beyond the ones characteristic of the massless case. Furthermore, the electronic properties of devices, such as carrier mobility, depend strongly on how the electronic degrees of freedom (massive or massless) interact with impurities. Of particular importance are charged impurities that naturally appear either on the substrate, on top of graphene, or between the substrate and graphene. Charged impurities play an important role in the transport characteristics of graphene deposited in SiO$_2$ (Refs. 7–9) and should play an important role in epitaxial graphene as well.[10]

In this paper we explore the nontrivial restructuring and charging of the vacuum that occur in the presence of a supercritical Coulomb center[11,12] if the system is described by a "massive" spectrum. The problem of an unscreened Coulomb impurity in undoped graphene has recently received considerable attention.[13–18] This is justified since, on the one hand, the vanishing density of states (DOS) at the Fermi energy of undoped graphene and the absence of backscattering suppress screening.[19] In addition, the Coulomb problem in graphene is the condensed-matter analog of supercritical nuclei ($Z\alpha > 1$) in QED whose rich and fundamental phenomena remain elusive to experimental testing.[12,13,15] We show here that this analogy achieves its fullest in gapped graphene, where one can resolve the incremental charging of

the vacuum with strong implications for the screening of the Coulomb center.

The rest of this paper is organized as follows. In Sec. II we introduce our model and in Sec. III its properties in the massless limit are reviewed. In Sec. IV the behavior of discrete energy levels in the massive case is discussed. We examine in detail the structure of the critical wave function and the corresponding renormalization of the critical Coulomb coupling in Sec. V. In Sec. VI we study the behavior of the vacuum charge across the critical point. Section VII contains the corresponding results for a finite-size system (where the energy gap is due to the finite size only). Section VIII contains a discussion of our results and conclusions.

II. MODEL

To address the Coulomb problem in graphene, we resort to the single-valley effective Dirac description of the electron dynamics. We start from the tight-binding Hamiltonian describing electrons in a honeycomb lattice and in the presence of a Coulomb center of strength Z. In addition, we assign a different local energy to each sublattice. The resulting Hamiltonian is

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

In the above equation, the sums run over unit cells and the operators $a_i(b_i)$ pertain to the A(B) sublattice. This system exhibits a band gap of Δ and, if undoped, has an insulating ground state with the Fermi level in the gap. Just as in its gapless counterpart, low-energy excitations can be addressed within an effective-mass approximation, consisting of a $\boldsymbol{k}\cdot\boldsymbol{p}$ expansion around the K and K' points in the Brillouin zone. The resulting effective Hamiltonian leads to the Dirac equation in 2D under a Coulomb field,

$$\left(-i\hbar v_F \boldsymbol{\sigma} \cdot \boldsymbol{\nabla} - \frac{Ze^2}{r} + m v_F^2 \sigma_z\right)\Psi(\boldsymbol{r}) = \epsilon\Psi(\boldsymbol{r}), \tag{2}$$

where the wave function $\Psi(\boldsymbol{r})$ has a spinor structure that carries the amplitude of the wave function on each sublattice,

$$\Psi(\mathbf{r}) = \begin{pmatrix} \psi^A(\mathbf{r}) \\ \psi^B(\mathbf{r}) \end{pmatrix}. \tag{3}$$

$v_F = 3ta/(2\hbar)$ is the Fermi velocity in graphene and a is the C-C distance. It is convenient to introduce g as the dimensionless coupling to the external Coulomb field $g \equiv Z\alpha$ and $\alpha \equiv e^2/(\hbar v_F)$ is the "fine-structure" constant of graphene. The gap in the spectrum is induced by the term proportional to σ_z, akin to a relativistic rest mass. The mass m is related to the local energy mismatch in the tight-binding formulation through $\Delta = 2mv_F^2$. To avoid a too cumbersome notation, throughout the paper we shall take a system of units in which $\hbar = v_F = 1$. Without loss of generality, the impurity will be considered attractive ($g > 0$) and the C-C distance ($a \simeq 1.42$ Å) will be used as the length unit.

In parallel with the exact analytical solution of Eq. (2), we solve the tight-binding problem of Eq. (1) exactly via direct numerical diagonalization on the lattice. This exact solution provides an important control of the validity of the Dirac approach and the results of the two approaches will be frequently compared.

III. COULOMB IMPURITIES IN MASSLESS GRAPHENE

Before delving into the details of the supercritical regime in the massive case, a brief overview of the main physics seen in the massless problem is pertinent.[13–15] When $m = 0$, Eq. (2) separates in cylindrical coordinates[20] and can be subsequently mapped into the radial Schrödinger equation of the usual Coulomb problem in three dimensions (3D).[15] One decisive peculiarity of this mapping is that, although the radial equation is formally the same as the radial equation in the Schrödinger Coulomb problem, the angular-momentum quantum number appears replaced by an "effective angular momentum" equal to $l = \sqrt{j^2 - g^2}$. Here, j is the quantum number associated with the total angular-momentum operator

$$J_z = L_z + \frac{1}{2}\sigma_z, \tag{4}$$

and takes the values $j = \pm 1/2, \pm 3/2, \dots$. The radial solutions are thus expressed in terms of the Coulomb functions[15,21] $F_l(-g \, \text{sign}(\epsilon), |\epsilon|r)$ and $G_l(-g \, \text{sign}(\epsilon), |\epsilon|r)$. Given that, as usual, l determines the asymptotic power-law behavior of the wave functions,

$$F_l(r \simeq 0) \sim r^{l+1}, \quad G_l(r \simeq 0) \sim r^{-l},$$

it is evident that $g = g_c = 1/2$ is a singular point for the lowest total angular-momentum channel ($j = \pm 1/2$). Below g_c the space of solutions is constrained by the requirement of regularity at the origin, as usual, and this selects $F_l(r)$ as the regular solution. But above g_c, l becomes imaginary and any linear combination of the solutions (F_l, G_l) becomes square integrable at the origin. At the same time, one sees from the above asymptotics that the solutions oscillate endlessly when $r \to 0$ as $\propto \exp[i \log(r)]$. This is a signature of the "fall to the center" characteristic of highly singular potentials.[22] In fact, the supercritical regime can be intuitively understood from a classical perspective; consideration of the classical equations

of motion for the Dirac Hamiltonian shows that, for a given angular momentum, there will be a critical coupling above which the classical orbits spiral and fall onto the potential source.[13] The potential has become too singular and there are no closed nor scattering orbits.

It is known that, in this case, the quantum-mechanical problem becomes uniquely defined only after an additional boundary condition is introduced, reflecting the physical cutoff of the potential at short distances.[11,23] For graphene the natural cutoff is the lattice spacing a. This regularization of the potential at short distances permits the exact solution to be extended to the supercritical regime ($g > g_c$), which (among other effects) is characterized by the presence of marked resonances in the spectral density of the hole channel.[15] Such spectral features are analogous to the positron resonances expected for a supercritical nucleus in QED.[12] But a fundamental difference exists between the two cases; whereas in QED, due to the finite mass, the emergence of positron resonances is an incremental process (as a function of g)—in massless graphene an infinite number of them instantly appears at $g = g_c$. Semiclassical considerations illustrate how these resonances emerge from an infinite number of quasi-bound-states embedded in the lower continuum.[13] Adding to these spectral peculiarities, the induced electronic charge behaves rather differently on the two sides of the critical point being localized close to the impurity for $g < g_c$ and otherwise decays algebraically[13,15] as $1/r^2$.

An important question is how will these nonperturbative effects contribute to screen the impurity potential. Adding to the ever present virtual vacuum polarization arising from virtual excitations,[18] in the supercritical regime one expects those quasilocalized states to partially screen the Coulomb center. It was argued, on the basis of a self-consistent treatment, that the impurity charge at large distances is screened down to its critical value $Z_c = 1/(2\alpha)$ for any $g > g_c$.[13] We will show that in the massive case of interest here, the situation is conceptually different and a different regime can be reached where the polarization charge is comparable to the bare charge Z. This leads to a strong tendency to neutralization of the impurity potential at a length scale set by the Compton wavelength in graphene. We examine in detail the conditions for this to occur.

IV. MASSIVE DIRAC FERMIONS IN GRAPHENE

The emergence of resonant solutions described above is best appreciated when the system acquires a mass and the electron dispersion becomes $\epsilon^2 = k^2 + m^2$. This mass can arise physically from sublattice symmetry breaking,[6] from spin-orbit coupling, or from reduced dimensionality, as in a finite-size mesoscopic sample. The one-particle solution for the massive case is straightforward and is discussed, for instance, in Refs. 14 and 24. The fundamental difference that a finite mass brings to the Coulomb problem is the immediate appearance of bound solutions corresponding to the hydrogenlike fine-structure spectrum in 2D,

$$\epsilon_{n,j} = m \, \text{sign}(g) \frac{n + \sqrt{j^2 - g^2}}{\sqrt{g^2 + (n + \sqrt{j^2 - g^2})^2}}. \tag{5}$$

In the above equation n is the principal quantum number and j is the total angular-momentum quantum number defined
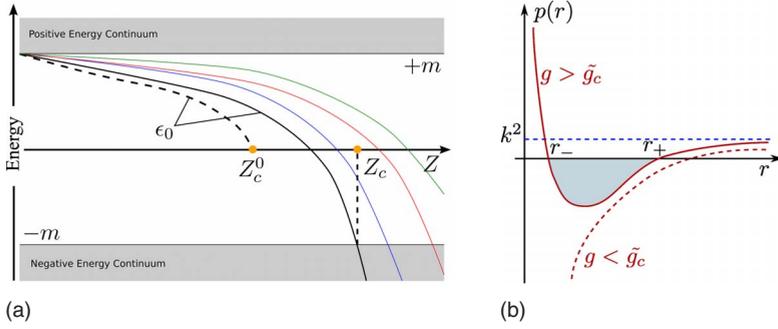
FIG. 1. (Color online) (a) Schematic depiction of the change of bound levels [Eq. (5)] with increasing coupling. The dashed line represents the evolution of ϵ_0 for a point nucleus and the solid one is for a regularized potential. (Adapted from Ref. 26.) (b) Effective semiclassical momentum of Eq. (8) when $\epsilon_0 \approx -m$ and $g > \tilde{g}_c$ (solid) or $g < \tilde{g}_c$ (dashed).

above. These states reside in the gap $-m < \epsilon < m$ and are described by wave functions that decay exponentially for distances larger than the "Bohr" radius $a_0 = \lambda_C/(Z\alpha)$, where $\lambda_C = \hbar/(m v_F)$ is the "Compton" wavelength in graphene. For an attractive potential the lowest bound state $(n=0, j=1/2)$ has an energy given by

$$\epsilon_0 = m\sqrt{1 - (2g)^2}, \qquad (6)$$

and hence reaches zero singularly at $g = g_c$—becoming imaginary beyond this coupling strength. Similarly to what we have discussed regarding the massless case, this imaginary energy reflects the fact that a description of a point nucleus is impossible for $g > g_c$. A physical regularization procedure, where one cuts the potential off at small r, relaxes this constraint and the bound levels are then allowed to sink further[12,25] through negative energies until the point $\epsilon = -m$ is reached. The diving of the bound solutions with increasing coupling is schematically depicted in Fig. 1(a). The vicinity of this point where the lowest bound state is about to merge (or has just merged) with the continuum is the focus of all our subsequent discussion.

V. CRITICAL WAVE FUNCTION AND CRITICAL COUPLING RENORMALIZATION

Given that in a regularized Coulomb potential the bound levels are allowed to dive into negative energies, the relevance is transferred from the particular coupling $g = g_c = 1/2$ to the value $\tilde{g}_c$ at which $\epsilon_0 = -m$. Necessarily $\tilde{g}_c > g_c$ and the critical coupling is thus renormalized to higher values. This is one of the immediate consequences of the presence of a gap. The merging of the bound solution with the continuum is nonetheless peculiar and, hence, it is worth careful consideration. We begin with a semiclassical argument. The first important thing to notice is that the merging state does not become completely delocalized, as one would expect in typical Schrödinger problems, but remains localized as it dives into the lower continuum. This is somewhat counter intuitive and is best appreciated by considering the problem semiclassically within WKB. An alternative to the full WKB approximation is to consider the *classical relativistic* momentum

$$p(r)^2 = [\epsilon - U(r)]^2 - m^2, \qquad (7)$$

where $U(r)$ is the Coulomb potential. With this definition, the radial momentum p_r is written as

$$p_r(r)^2 = k^2 + 2g\epsilon/r + (g^2 - \ell^2)/r^2, \qquad (8)$$

with ℓ as the classical effective angular momentum and $k^2 = \epsilon^2 - m^2$. The solid line in Fig. 1(b) shows the profile of $p_r(r)^2$ slightly after the merging has taken place. There is a classically forbidden region $[p(r)^2 < 0]$ between r_- and r_+ and the long-range Coulomb tail implies that $r_+ \gg r_-$. In fact, precisely at $\epsilon_0 = -m$ we have $r_+ = \infty$ and the barrier is infinite. Close to the critical point we can write $\epsilon_0 \simeq -m - \beta(g - \tilde{g}_c)$ and examine the penetrability of this barrier within WKB. It reads

$$w = e^{-2S_{cl}} = \exp\left(-\frac{2\pi m \tilde{g}_c}{\sqrt{2\beta m}\sqrt{g - \tilde{g}_c}}\right), \qquad (9)$$

and becomes exponentially small as $g \to \tilde{g}_c^+$. Such wide barrier (absent when $\epsilon_0 \simeq +m$) ensures that the state remains appreciably localized even after merging into the lower continuum.

We now turn to the exact quantum-mechanical solution of Eq. (2) for the particular regularization of the Coulomb potential described by[25]

$$V(r) = \begin{cases} -g/r, & r > a \\ -g/a, & r \leq a \end{cases}. \qquad (10)$$

This regularization represents the physical situation in which there is a Coulomb impurity at the center of a hexagon in the honeycomb lattice. In this case the lattice parameter a is the closest distance that an electron hopping among carbon sites can be from the potential source. It is also a good approximation for an impurity placed slightly above the graphene plane. Cylindrical symmetry is preserved by this regularization and Eq. (2) naturally separates in cylindrical coordinates. We define the radial spinor components as $(j = \pm 1/2, \pm 3/2, \ldots)$

$$\Psi_j(r, \varphi) = \frac{1}{\sqrt{r}} \begin{pmatrix} e^{-i(j-1/2)\varphi} A(r) \\ i e^{-i(j+1/2)\varphi} B(r) \end{pmatrix}, \qquad (11)$$

after which the radial equations for Eq. (2) become $(r > a)$

$$\begin{bmatrix} \left(\epsilon + \dfrac{g}{r} - m\right) & -\left(\partial_r + \dfrac{j}{r}\right) \\ \left(\partial_r - \dfrac{j}{r}\right) & \left(\epsilon + \dfrac{g}{r} + m\right) \end{bmatrix} \begin{bmatrix} A(r) \\ B(r) \end{bmatrix} = 0. \quad (12)$$

We will be interested in the states at the threshold with energy $\epsilon = -m$. In that case, the equation for the $A(r)$ component reads ($r > a$)

$$r^2(\sqrt{r}A(r))'' + \left[g^2 - j^2 + \dfrac{1}{4} - 2gmr\right]\sqrt{r}A(r) = 0, \quad (13)$$

whose solution is given in terms of the modified Bessel functions $I_\nu(z)$ and $K_\nu(z)$. By imposing a vanishing boundary condition at infinity, one obtains ($\mathcal{N}$ is a normalization constant),

$$A(r) = \mathcal{N}K_{i\nu}(\sqrt{8gmr}), \quad (14a)$$

where $\nu = 2\sqrt{g^2 - j^2}$. We note that this solution has the same form as the corresponding spinor component in the 3D QED problem.[11,25] The B component follows directly from Eq. (12),

$$B(r) = \dfrac{\mathcal{N}}{g}\left[\left(j + \dfrac{i\nu}{2}\right)K_{i\nu}(\sqrt{8gmr}) + \sqrt{2gmr}K_{i\nu-1}(\sqrt{8gmr})\right]. \quad (14b)$$

The solutions (14a) and (14b) pertain to the region $r > a$, where the potential (10) has the Coulomb form. For distances smaller than the regularization distance, the solution for the radial spinor components of Eq. (11) is given in terms of cylindrical waves,

$$\begin{bmatrix} A(r) \\ B(r) \end{bmatrix} = \mathcal{N}'\sqrt{r}\begin{bmatrix} gJ_{j-1/2}(\tilde{k}r) \\ \tilde{k}aJ_{j+1/2}(\tilde{k}r) \end{bmatrix}, \quad (15)$$

where we have introduced $\tilde{k}a = \sqrt{g^2 - 2mga}$. The solutions (14a), (14b), and (15) are at energy $\epsilon = -m$ and need to be matched at $r = a$,

$$\left(\dfrac{A^<(g)}{B^<(g)}\right)\bigg|_{r=a} = \left(\dfrac{A^>(g)}{B^>(g)}\right)\bigg|_{r=a}. \quad (16)$$

This completes the determination of the *critical spinor* wave function $\Psi_c(\mathbf{r})$. Since the Dirac (2) is of the first order in the gradient operator, one is required to match only the spinor components (and not, additionally, their derivatives as would be the case for the Schrodinger equation) and this is enough to guarantee the continuity of the probability current density across the region $r = a$. The matching procedure generates a transcendental equation for g, with an infinite number of solutions, on account of the oscillating character of the functions $K_{i\nu}(z)$. For each angular momentum j, the multiple solutions correspond to the values of g for which the higher bound states (i.e., those levels having the same j symmetry) reach the continuum. To determine $\tilde{g}_c$ we concentrate on the s level ($j = 1/2$) and solve Eq. (16). Its smallest solution for g yields $\tilde{g}_c$, the renormalized critical coupling. A plot of $\tilde{g}_c$ as a function of ma is shown in Fig. 2. It can be seen that the departure of $\tilde{g}_c$ from the value $g_c = 1/2$ is singular at the
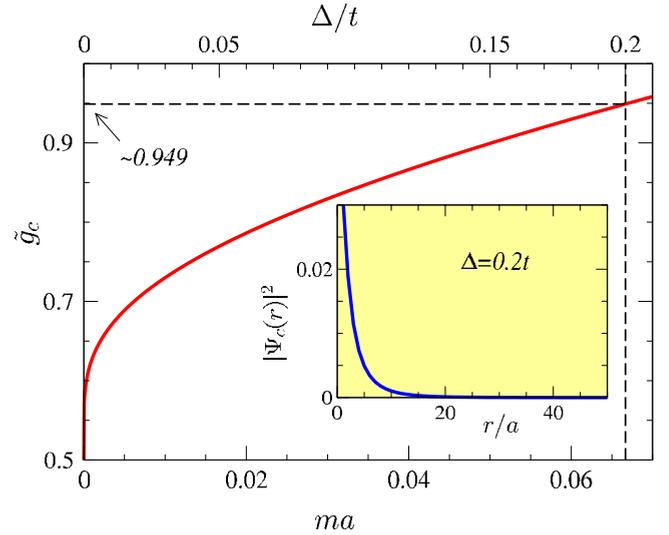
FIG. 2. (Color online) Critical coupling $\tilde{g}_c$ as a function of the mass/gap obtained from the solution of Eq. (16). The top horizontal axis shows ma in the units natural for the tight-binding calculation $ma \to a/\lambda_C = \Delta/(3t)$. The inset shows the probability density associated with the critical wave function at $\epsilon = -m$ [cf. Eq. (17)].

origin, which means that even an arbitrarily small mass causes a significant change in the critical coupling. In the figure we also point out (dashed lines) the value of $\tilde{g}_c \simeq 0.949$ that corresponds to a gapped graphene spectrum with $\Delta = 2mv_F^2 = 0.2t$ (in units of the tight-binding hopping parameter). We have chosen this value for illustration purposes (to be discussed later). The critical probability amplitude, defined in terms of the critical wave function as $\rho(r) = \Psi_c^\dagger(r)\Psi_c(r)$, is shown in the inset. The crux of our argument resides in the fact that Ψ_c is clearly localized, as can be inspected from the asymptotic behavior of $\rho(r)$ for $r \gg 1/m$,

$$\rho(r) = \Psi_c^\dagger(r)\Psi_c(r) \sim r^{-1/2}\exp(-2\sqrt{8m\tilde{g}_c r}). \quad (17)$$

Restoring the units, the Compton wavelength emerges as the characteristic localization length $mr \to mv_F r/\hbar = r/\lambda_C$.

In order to ascertain the validity of these results in the context of the original tight-binding problem, we explicitly compare the above with the exact numerical solution in the honeycomb lattice. Of particular interest are the renormalization of the critical coupling and the character of the diving states. In Fig. 3 we present the exact tight-binding spectrum for a Coulomb center in the honeycomb lattice with a sublattice energy mismatch (energy gap) of $\Delta = 0.2t$. Inspection of the main panel reveals several features, among which, (i) the lowest positive level (which corresponds to ϵ_0) immediately detaches from the continuum and sinks rapidly, followed by other states at higher g; (ii) the level ϵ_0 reaches $-m$ at $g = 0.83$ and touches the lower continuum at $g = 0.87$, which is in very good agreement with the result $\tilde{g}_c \simeq 0.9$ predicted in Fig. 2 from the solution of the Dirac equation;[27] (iii) as g increases past $\tilde{g}_c$ this level sinks further, as is evident from the level avoidances highlighted by the dashed/shaded region; and (iv) the critical wave function (shown in the inset) is highly concentrated around the origin, as ex-
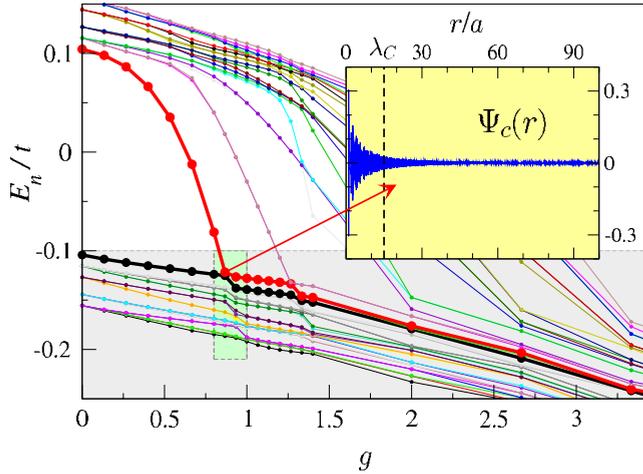
FIG. 3. (Color online) Low-energy close-up of the exact energy spectrum (E_n) as a function of the coupling g for the tight-binding problem (1). The inset contains the exact numerical wave function $\Psi_c(r)$, with the dashed line marking λ_C for this gap. The sublattice gap is $\Delta = 0.2t$ and a lattice of 124^2 sites with a central impurity has been used.

pected from the foregoing—namely Eq. (17). The lowest bound level in Fig. 3 is doubly degenerate. This is expected since although the lowest bound solution of the Dirac Eq. (6) is nondegenerate, there is a valley degeneracy to account for in the Dirac description.

While the continuum solutions are sensitive to the finite size of the numerical system,[27] one does not expect the lowest bound solution to be markedly sensitive for the sizes used in this calculation. Hence, we can extrapolate the trajectory of the lowest bound state in Fig. 3 to the thermodynamic limit. In that case, the critical coupling in the lattice would be at $\simeq 0.83$, which is smaller than the value $\widetilde{g}_c = 0.949$, and indicates that the effective cutoff distance for the regularization (10) is slightly smaller than a.

VI. VACUUM POLARIZATION AND VACUUM CHARGING

It is clear, either from Eq. (17) or from the exact results in the lattice [Fig. 3(inset)], that the merging state has a localized character. The characteristic length scale λ_C is related to the gap through $\lambda_C/a = 3t/\Delta$. For the gap $\Delta = 0.26$ eV (reported in Ref. 6) this corresponds to $\lambda_C \simeq 30a$. However, the size of the merging bound state, measured as the average radius and calculated by using the critical wave functions, is smaller $\langle r \rangle = 13.9a$. Beyond this scale, the impurity potential is screened by essentially one charge unit times the product of the spin and valley degeneracies ($N=4$). Of course this screening is meaningful only in a very diluted impurity configuration ($n_i \ll 10^{12}$ cm^{-2} for the quoted experimental gap). We can appreciate this more clearly by studying the induced charge, defined as

$$\delta\rho(r) = \sum_{E \leq E_F} \chi_E^\dagger(r)\chi_E(r) - \sum_{E \leq -m} \chi_E^{0\dagger}(r)\chi_E^0(r), \quad (18)$$

where $\chi_E(r)$ are the continuum wave functions in the presence of the potential and $\chi_E^0(r)$ stand for the continuum wave

functions in the absence of potential. We work with a *constant number of electrons* as the potential is turned on and, hence, $E_F \neq -m$ in general.[28] We also need to consider the intermediate situation in which there is a bound state just about to merge with the lower band. The continuum wave functions in this case are labeled as $\chi_E^c(r)$. For the sake of the argument, assume that there is only one bound state that we denote by $\Psi_c(r)$ and, furthermore, let us accept that we can project everything onto the states $E < m$. In other words, we accept that the states $E < m$ approximately form a complete set in each circumstance,

$$\sum_{E \leq -m} \chi_E^{0\dagger}(r)\chi_E^0(r') \simeq \delta(r - r'), \quad (19a)$$

$$\sum_{E \leq -m} \chi_E^\dagger(r)\chi_E(r') \simeq \delta(r - r'), \quad (19b)$$

$$\sum_{E \leq -m} \chi_E^{c\dagger}(r)\chi_E^c(r') + \Psi_c^\dagger(r)\Psi_c(r') \simeq \delta(r - r'). \quad (19c)$$

From these relations it follows that above $\widetilde{g}_c$, when one state has dived onto the continuum, the induced charge at constant density defined in Eq. (18) can be approximated as

$$\delta\rho(r) \simeq |\Psi_c(r)|^2 + \delta\rho_{\text{pol}}(r) - |\chi_{-m}^c(r)|^2, \quad (20)$$

where $\delta\rho_{\text{pol}}(r)$ represents the polarization charge caused by the deformation of the continuum wave functions only,

$$\delta\rho_{\text{pol}}(r) = \sum_{E \leq -m} |\chi_E^c(r)|^2 - |\chi_E^0(r)|^2, \quad (21)$$

with

$$\int \delta\rho_{\text{pol}}(r)dr = 0,$$

and $\chi_{-m}^c(r)$ represents the topmost plane wave that appears because we keep the number of electrons constant as g is varied. Expression (20) is valid for one bound level merging into the continuum but can be generalized for the additional subsequent divings.

The result (20) shows that, although after the merging there is (strictly) no localized state, the total induced charge (20) remains essentially determined by the profile of the critical state $\Psi_c(r)$ plus the background polarization charge. The last term in Eq. (20) is a phase-shifted wave and, thus, it is the smallest and negligible contribution to $\delta\rho(r)$. The constant number of electrons and the fact that all states are normalized evidently implies that the integral of $\delta\rho(r)$ over the entire volume vanishes. This allows for the definition of $Q(R)$ (the total charge inside a radius $r=R$),

$$Q(R) = N \int_{|r| < R} \delta\rho(r)dr. \quad (22)$$

Here N represents the degeneracies of the problem; $N = 4$(spin $\times$ valley) for the solution of the single-valley Dirac equation and $N=2$ (spin) for the numerical solution in the honeycomb lattice. This function $Q(R)$ will rise quickly at small R, rapidly attaining a maximum on account of the lo-
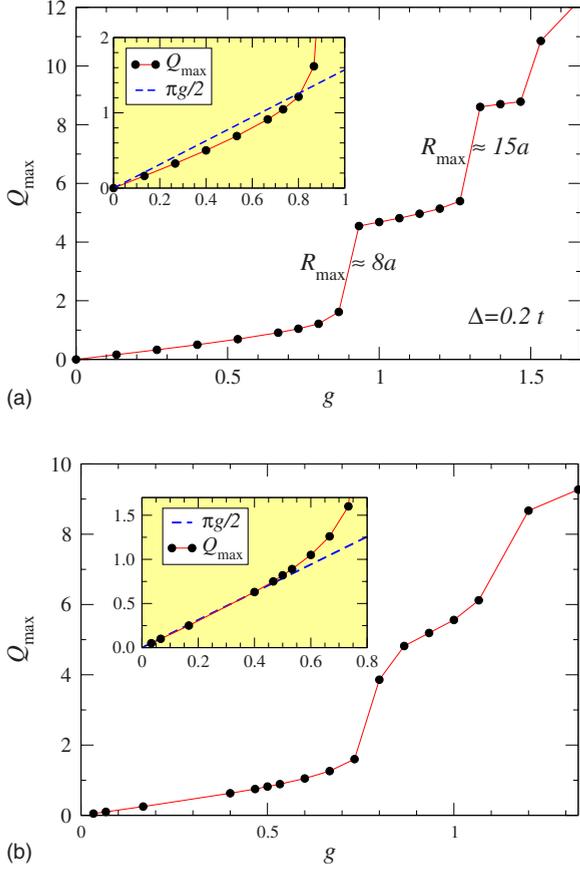
FIG. 4. (Color online) (a) Plot of Q_{max} (the amount of charge pulled to the vicinity of the impurity) (see text) versus g from the exact numerical diagonalization of the tight-binding problem (Fig. 3). The steps signal the incremental diving of bound levels into the lower continuum. The distances at which $Q(R)=Q_{max}$ are indicated for the first two steps. (b) The same finite system without an explicit gap ($\Delta=0$). In both cases the system has 124×124 carbon sites and linear dimensions of $107a\times186a$. The insets amplify the low-g region, which is compared with the RPA result for $m=0$: $Q(R)=\frac{\pi}{2}g\int_{|r|<R}\delta(r)dr=\frac{\pi}{2}g$ (dashed line).

calized profile of $\Psi_c(\mathbf{r})$.[29] $Q(R)$ then returns slowly to zero as R is increased further because of the normalization of the states and the fact that we work at a constant number of electrons. The maximum in $Q(R)$ (which we designate as Q_{max}) is thus a measure of the amount of charge pulled to the vicinity of the impurity, which is our quantity of interest. For $g<\tilde{g}_c$, the only contribution to $Q(R)$ comes from the polarization charge $\delta\rho_{pol}(\mathbf{r})$; but whenever a new level dives, a contribution of the type $|\Psi_c(\mathbf{r})|^2$ adds to the total induced charge and one should observe discontinuous jumps in Q_{max} with increasing g.

Once again this expectation is confronted with exact results on the lattice. We compute Eq. (18) numerically using the exact wave functions of the tight-binding model. In Fig. 4(a) we plot the resulting Q_{max}. Up to $g\simeq0.9$ the only contribution comes from $\delta\rho_{pol}$ and the variation of Q_{max} is smooth. At the point where in Fig. 3 the first level merges with the continuum, the first discontinuity occurs in Q_{max}, which jumps by roughly four units ($N\times1=4$), just as expected. Also as expected, this charge is concentrated within a

region of radius $\sim\lambda_C$. Therefore, the approximations used to obtain expression (20) are legitimate and the induced charge beyond $\tilde{g}_c$ is mostly determined by the profile of the critical states.

This is the precise analog of the quantum electrodynamics (QED) prediction for the charging of the vacuum. In a super-heavy nucleus with high enough Z, the hydrogenlike bound levels can merge with the positron continuum. When that happens, the Schwinger mechanism of electron-positron creation becomes spontaneous since the bound level has become degenerate with the positron states and consequently there is no energy cost involved in creating an electron-positron pair. The electron occupies the bound level close to the nucleus while the positron escapes to infinity.[12] Since the positron continuum constitutes the vacuum of QED, the spontaneous creation of an electron in a level below $-m$ leads to a restructuring of the vacuum that acquires a charge $Q=-2e$.[30] The curve in Fig. 4(a) mimics entirely the QED prediction found in Fig. 1.5 of Ref. 12. The striking difference rests in the magnitude of the effect. In QED, due to the smallness of the real fine-structure constant ($\alpha_{QED}\simeq1/137$), the nuclear charge required to reach the critical regime is very large ($Z_c\simeq170$) and the jumps in the polarization charge lead to an effective atomic number $Z^{eff}=Z_c-Q$, with $Q\ll Z_c$ (a very small correction).

In graphene, for estimate purposes, we assume a SiC substrate (with dielectric constant $\varepsilon\simeq10$) as used in Ref. 6 leading to $\alpha\simeq0.4$. For the gap $\Delta=0.26$ eV seen in this experiment, we obtain $g_c=(Z\alpha)_c=0.84$ and, consequently, the critical valence is $Z_c\simeq2.1$. This means that, according to our discussion, impurities with valence $Z\gtrsim2$ are expected to be completely screened beyond the scale $\langle r\rangle=13.9a\approx2$ nm, since at this distance the amount of accumulated charge in the vacuum is $Q=4$ and, thus, $Z^{eff}(r>\langle r\rangle)=Z-Q\approx0$. In fact, for $Z<4$ overscreening takes place and we expect many-body interactions to be important, effectively removing the overscreening tendency; however, such calculations are very complex and beyond the scope of this work. In addition, we note that substrates with smaller dielectric constants lead to higher values of $Z\alpha$ and are, in principle, capable of reducing the valences Z needed to observe the effect to $Z=1,2$.

Electron-electron interactions can also lead to a change of the critical coupling. Estimated perturbatively ($\alpha\ll1$) this change is $\delta g_c\sim\alpha g_c$, which would not affect significantly the physics discussed here (but would clearly affect the precise value of Z_c for a given α). We recall also that the above theory assumes the chemical potential to be at $\mu=-m$, while the experiments of Ref. 6 in fact have $\mu>+m$. Tuning the Fermi level to the gap via chemical doping or gating and varying α by changing the substrate's dielectric constant would provide a possible way of exploring our predictions experimentally.

VII. VACUUM CHARGING IN A FINITE SYSTEM

The discrete eigenstates of a finite graphene sheet and the fact that the density of states (DOS) vanishes at the Dirac point lead to another realization of a gapped spectrum and the question of whether the effects discussed above carry to

this case arises naturally. We have diagonalized the same lattices as above, this time without any sublattice mismatch (i.e., $\Delta = 0$ or $m = 0$), corresponding to the massless Dirac problem on a finite system. In this case we found that the gap induced by the finite size of the system is $\Delta \simeq 0.06t$, $\lambda_C \approx 50a$, and $\langle r \rangle \approx 21a$. This leads to a renormalized $\tilde{g}_c \simeq 0.76$, representing a 50% increase with respect to g_c just from the fact that, although massless, the system is constrained to a finite size. For brevity we show only the integrated polarization charge Q_{max} in Fig. 4(b). The curve of Q_{max} displays the characteristic step discontinuity at $0.7 < g < 0.8$, which is in agreement with the estimate for $\tilde{g}_c$. Of course, in this case the situation is peculiar since the effective λ_C is tied to the linear size of the system L and the gap is now $\Delta \propto L^{-1}$. Consequently λ_C—although still a fraction of L—grows with the system size. Nevertheless, most importantly, the picture of vacuum charging discussed above remains valid, with the vacuum charge residing at a distance $\langle r \rangle \approx 21a$ significantly smaller than the linear size $L \approx 107a$. Thus, the vacuum charging can be potentially observed in mesoscopic samples as well.

VIII. DISCUSSION AND CONCLUSIONS

The fact that gapped graphene is described in terms of massive Dirac quasiparticles makes it a rather unique solid-state system. We have seen, in particular, that effects of supercritical vacuum charging are expected when undoped graphene is exposed to low-valence charged impurities.

One of the key steps in our calculation is the explicit regularization of the potential (10) that allows the diving of the bound levels into the lower continuum. Since an impurity located slightly out of the plane would still lead to a regularized Coulomb potential, our results are not altered qualitatively because details of the regularized potential at short distances are not important.[11] At the quantitative level, the regularization distance determines the renormalization of $\tilde{g}_c$. Since the typical distances to the plane expected for chemisorbed impurities in graphene are still in the subnanometer range,[9] we do not expect a prohibitive increase in $\tilde{g}_c$ nor, consequently, in Z_c. Therefore, the charged impurity does not need to be embedded in the graphene plane and can be simply adsorbed to its surface.

The theory predicts a strong tendency toward screening of the external Coulomb potential on nanometer scales, with the actual values depending on the details of the particular situation (notably the magnitude of the gap). We mention also

that, despite our initial assumption of attractive impurities ($g > 0$), this is only for convenience. The particle-hole symmetry of the problem ensures that the results remain valid for repulsive charges (negative ions), in which case, the (positron) bound levels merge with the upper continuum of states at precisely the same critical couplings.

Whereas in gapless graphene, the supercritical regime is characterized by an infinite number of resonances in the hole (positron) channel,[13,15] the phenomenon of vacuum polarization in gapped graphene is an incremental process. Each level dives at successively higher values of the coupling $g = Z\alpha$, leading to the step-wise shape of the curves of Q_{max} shown in Fig. 4. This translates to a quite dissimilar screening of the supercritical impurity in the massless and massive cases; for the latter we can have complete screening at very short distances in a system with no carriers.

Finally, the estimates of the critical valence $Z_c \simeq 2$ made in this paper are based on the specific parameters of the epitaxial samples used in Ref. 6 (to wit the gap Δ and the dielectric constant of the substrate). The supercritical regime is determined by $Z\alpha \gtrsim \tilde{g}_c$ to which several players contribute: Z itself, the dielectric medium (via α), the mass/gap, and the regularization distance (via $\tilde{g}_c$). Some of these are more manageable to experimental control than others. Encouraging experiments are emerging that seek control over some of them. In Ref. 9 controlled doping with monovalent ions was achieved and presumably the same could be done with divalent alkaline ions. In Ref. 31 the fine-structure constant α in exfoliated graphene could be controlled through changes in the dielectric environment. And in Ref. 32 a gap was found for graphene flakes on Ni surfaces.

The charging of the vacuum in the presence of a strong Coulomb center is a long-standing prediction of QED in strong fields that remains unconfirmed through a direct experiment. This is an obvious consequence of the difficulties imposed by the nuclear charge of $Z \simeq 170$ required to reach the supercritical regime in QED. Our calculations suggest that, under the conditions discussed, the analogous effect might be observable in a solid-state context with low-valence ions.

[1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, and A. A. Firsov, Nature (London) **438**, 197 (2005).

[2] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, arXiv:0709.1163, Rev. Mod. Phys. (to be published).

[3] A. K. Geim and K. S. Novoselov, Nat. Mater. **6**, 183 (2007).

[4] A. H. Castro Neto, F. Guinea, and N. M. R. Peres, Phys. World **19**, 33 (2006).

[5] M. I. Katsnelson and K. S. Novoselov, Solid State Commun. **143**, 3 (2007).

[6] S. Y. Zhou, G.-H. Gweon, A. V. Fedorov, P. N. First, W. A. de Heer, D.-H. Lee, F. Guinea, A. H. Castro Neto, and A. Lanzara, Nat. Mater. **6**, 770 (2007).

[7] K. Nomura and A. H. MacDonald, Phys. Rev. Lett. **98**, 076602 (2007).

[8] S. Adam, E. H. Hwang, V. M. Galitski, and S. D. Sarma, Proc. Natl. Acad. Sci. U.S.A. **104**, 18392 (2007).

[9] J. H. Chen, C. Jang, M. S. Fuhrer, E. D. Williams, and M. Ishigami, Nat. Phys. **4**, 377 (2008).

[10] W. A. de Heer *et al.*, Solid State Commun. **143**, 92 (2007).

[11] Y. B. Zeldovich and V. S. Popov, Sov. Phys. Usp. **14**, 673 (1972).

[12] W. Greiner, B. Muller, and J. Rafelski, *Quantum Electrodynamics of Strong Fields*, 2nd ed. (Springer, New York, 1985).

[13] A. V. Shytov, M. I. Katsnelson, and L. S. Levitov, Phys. Rev. Lett. **99**, 246802 (2007).

[14] D. S. Novikov, Phys. Rev. B **76**, 245435 (2007).

[15] V. M. Pereira, J. Nilsson, and A. H. Castro Neto, Phys. Rev. Lett. **99**, 166802 (2007).

[16] R. R. Biswas, S. Sachdev, and D. T. Son, Phys. Rev. B **76**, 205122 (2007).

[17] M. M. Fogler, D. S. Novikov, and B. I. Shklovskii, Phys. Rev. B **76**, 233402 (2007).

[18] I. S. Terekhov, A. I. Milstein, V. N. Kotov, and O. P. Sushkov, Phys. Rev. Lett. **100**, 076803 (2008).

[19] T. Ando, J. Phys. Soc. Jpn. **75**, 074716 (2006).

[20] D. P. DiVincenzo and E. J. Mele, Phys. Rev. B **29**, 1685 (1984).

[21] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).

[22] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Pergamon, New York, 1981).

[23] K. M. Case, Phys. Rev. **80**, 797 (1950).

[24] V. R. Khalilov and C.-L. Ho, Mod. Phys. Lett. A **13**, 615 (1998).

[25] V. S. Popov, Sov. J. Nucl. Phys. **12**, 235 (1971).

[26] B. Müller and J. Rafelski, Phys. Rev. Lett. **34**, 349 (1975).

[27] In Fig. 3 the boundaries of the continuum also change with g. This is due to the fact that, in such a finite system as the one analyzed there, all the energy levels are phase-shifted by the potential. This remains true for the levels close to the gap and therefore the merging of the bound solution with the continuum in a finite-sized lattice appears at a larger coupling. This phase shift effect vanishes when the size of the system becomes infinite.

[28] The chemical potential (the last occupied level) follows the bold black line in Fig. 3.

[29] There is an additional localized contribution that comes from $\delta\rho_{\text{pol}}$ which is reminiscent of the localized screening charge $\propto \delta(\mathbf{r})$ in the massless case (Ref. 18). This term yields the smoothly varying part in Fig. 4, but is not our focus here.

[30] The factor of 2 comes from the spin degeneracy.

[31] C. Jang, S. Adam, J.-H. Chen, E. D. Williams, S. D. Sarma, and M. S. Fuhrer, arXiv:0805.3780 (unpublished).

[32] A. Gruneis and D. V. Vyalikh, Phys. Rev. B **77**, 193401 (2008).