Coulomb impurities in two-dimensional topological insulators

Jia-Lin Zhu,[1,*] Guo Li,[1] and Ning Yang[2]
[1]*Department of Physics, Tsinghua University, Beijing 100084, China, and*
State Key Laboratory of Low-Dimensional Quantum Physics, Tsinghua University, Beijing 100084, China
[2]*Institute of Applied Physics and Computational Mathematics, Beijing 100088, China*
(Received 2 December 2016; revised manuscript received 6 March 2017; published 22 March 2017; corrected 30 March 2017)

Introducing a powerful method, we obtain the exact solutions for a Coulomb impurity in two-dimensional infinite and finite topological insulators. The level order and zero-energy degeneracy of the spectra are found to be quite different between topological trivial and nontrivial phases. For quantum dots of topological insulator, the variation of the edge and Coulomb states with dot size, Coulomb potential, and magnetic field are clearly shown. It is found that for small dots the edge states can be strongly coupled with the Coulomb states and for large dots the edge states are insensitive to the Coulomb fields but sensitive to the magnetic fields.

DOI: 10.1103/PhysRevB.95.125431

I. INTRODUCTION

The topological insulators are narrow-gap semiconductors with topological protected edge (surface) states [1,2]. The peculiar properties of these states make topological insulators and their nanostructures useful for future applications ranging from spintronics to quantum computing [3–8].

As doping affects transport and optical properties of traditional semiconductors, the influence of impurities to topological insulators has also drawn extensive attention. Impurities in topological insulators can induce localized states [9–11], affect the transport properties [12,13], and give the quantized anomalous Hall phase [14,15]. So far, most theoretic works focus on the pointlike Gauss and Anderson impurities in topological insulators, while studies on the Coulomb impurities are rare although experimentally they have been confirmed to significantly change the surface band structures [16].

In topological insulators, the difference between the Coulomb impurity states under different topological phases is an interesting issue. Furthermore, in topological insulator nanostructures such as quantum dots, the edge states are much likely to interact with the impurities due to stronger spatial confinement. Knowing the interplay between the edge states and the Coulomb impurities as well as their magnetic field modulations are important for the designing of relevant nano devices.

The low-energy physics of topological insulators obey the modified Dirac equation [17,18], where a quadratic momentum mass term is added to the Dirac equation to distinguish different topological phases. Therefore, for topological insulators and their nano structures with impurities and external fields, one need to deal with the eigenvalue problems of the second-order differential equation array, while a method for exact solutions is still lacking so far.

In this work, we develop our series expansion method [19,20] to the case of the modified Dirac equation, to obtain the exact solutions of the single Coulomb impurity problem in infinite two-dimensional topological insulators (2DTIs), two-dimensional band insulators (2DBIs), and their quantum dots both with and without magnetic fields. The significant differences are shown between the spectra of infinite 2DTIs and 2DBIs. The coupling effects between the edge and Coulomb states are investigated in 2DTI quantum dots, and the variation of the coupling and spectrum characteristics with dot radius, Coulomb potential, and magnetic field are studied in detail.

II. METHOD

The modified Dirac Hamiltonian with an in-plane Coulomb potential is written as [18,21]

$$H_{\uparrow\downarrow} = \hbar v(\sigma_x k_x \pm \sigma_y k_y) + (M - Pk^2)\sigma_z - \frac{\hbar v\alpha}{r}, \qquad (1)$$

where v is the Fermi velocity, $\sigma_{x,y,z}$ are Pauli matrices describing the isospins, and α is a dimensionless value describing the Coulomb potential strength. Compared with the gapped graphene, the Pk^2 term is added to the mass M where $MP > 0$ and $MP < 0$ are for a 2DTI and a 2DBI, respectively. In the presence of a weak magnetic field, the model is still effective [22], and the substitution $\vec{k} = -i\vec{\nabla} + e\vec{A}/\hbar$ is used with the magnetic vector potential $\vec{A} = \vec{B} \times \vec{r}/2$.

Due to the rotation symmetry, the angular operator $\hat{j}_z = -i\hbar\partial_\theta + \frac{\hbar}{2}\sigma_z$ is conserved, so the eigenfunctions of $H_\uparrow$ can be expressed as $\psi(r,\theta) = (\varphi(r)e^{il_-\theta}, \chi(r)e^{il_+\theta})^T$ where $l_\pm = j \pm \frac{1}{2}$, and the half integer j is the eigenvalue of $\hat{j}_z$. Then we get the radial form of $H_\uparrow$

$$H_\uparrow(r) = \begin{bmatrix} M + Q_- - \frac{\hbar v\alpha}{r} & S_+ - \frac{i\hbar v r}{2l_b^2} \\ S_- + \frac{i\hbar v r}{2l_b^2} & -M - Q_+ - \frac{\hbar v\alpha}{r} \end{bmatrix}, \qquad (2)$$

where $Q_\pm = P(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} - \frac{{l_\pm}^2}{r^2} - \frac{r^2}{4l_b^4} - \frac{l_\pm}{l_b^2})$, $S_\pm = -i\hbar v(\frac{\partial}{\partial r} \pm \frac{l_\pm}{r})$, and $l_b = \sqrt{\frac{\hbar}{eB}}$ is the magnetic length. The radial form of $H_\downarrow$ satisfies the relation $H_\downarrow(r,j,B) = H_\uparrow(r,-j,-B)$. Therefore, each level is twofold degenerate for $B = 0$ while it splits for $B \neq 0$.

In order to obtain the exact solutions of the eigenequations $H_\uparrow\psi = E\psi$, we develop our method introduced earlier [19,20] with use of the theory of first-order differential

*zjl-dmp@tsinghua.edu.cn

equation array [23,24]. To obtain the first-order equation array, a four-component spinor $w = (f_1, f_2, f_3, f_4)^{\mathrm{T}}$ is defined as $f_1 = \frac{\varphi}{r}, f_2 = \frac{d\varphi}{dr}$, $f_3 = i\frac{\chi}{r}$, and $f_4 = i\frac{d\chi}{dr}$. Using $f_{1(3)} + r\frac{df_{1(3)}}{dr} = f_{2(4)}$, and $H_\uparrow\psi = E\psi$, the equation array is

$$\frac{dw}{dr} = \sum_{k=-1}^{3} A_k r^k w, \tag{3}$$

where A_k are 4×4 real coefficient matrices. Then the series form of the exact solutions in the regular $(0,r_0)$, Taylor (r_i,r_{i+1}) with $i = 0,1,\ldots,I-1$ and irregular (r_I,∞) regions for infinite 2DTIs (2DBIs) are found as follows:

$$w = \begin{cases} r^\rho \sum_{k=0}^{\infty} a_k r^k & r \in (0,r_0) \\ \sum_{k=0}^{\infty} b_k (r-x_i)^k & r \in (r_i,r_{i+1}). \\ e^{q(r)} r^\eta \sum_{k=0}^{\infty} c_k r^{-k} & r \in (r_I,\infty) \end{cases} \tag{4}$$

In the regular region, $\rho = \rho_{1,2}$ with $\rho_1 = \rho_2 + 1 = l_-$ for $j > 0$ and $\rho_1 = \rho_2 + 1 = -l_+$ for $j < 0$, and in the irregular region,

$$q(r) = \begin{cases} -\lambda_\pm r & \text{for} \quad B = 0 \\ -\frac{r^2}{4l_b^2} & \text{for} \quad B \neq 0' \end{cases} \tag{5}$$

where $\lambda_\pm = [(\hbar v)^2 - 2MP \pm \sqrt{\Delta}]^{1/2}/\sqrt{2}|P|$ with $\Delta = (\hbar v)^4 - 4MP(\hbar v)^2 + 4P^2E^2$ and the corresponding $\eta = \eta_\pm$ can be directly determined. For $B \neq 0$, two $\eta = \eta_{1,2}$ are determined by a quadratic equation. In each Taylor regions, the expansion are made about the center $x_i = (r_i + r_{i+1})/2$. The expansion coefficients a_k, b_k, and c_k have been obtained by deducing the recurrence relations. In the Appendix, we give the details of the series expansion process.

Using the continuity of the wave functions and its derivative at $r = r_i$, the exact energy levels and then wave functions for infinite 2DTIs and 2DBIs can be fixed by exact numerical calculations. In the calculations, we need properly choose $r_0, r_1, \ldots, r_{I-1}$ and r_I. For 2DTI and 2DBI quantum dots with the radius R, the vacuum boundary conditions at $r = r_I = R$, i.e., $\varphi(R) = \chi(R) = 0$, are used and here only solutions in the regular and Taylor regions are needed. The method is suitable for the eigenvalue problem of the high-order ordinary differential equation arrays when the external fields allow series solutions in the subregions. Besides topological insulators, multiband systems such as bilayer graphene and semiconductor light-heavy hole coupling are also examples where the method can be used to get exact solutions.

The Coulomb states for $H_\uparrow$ and $H_\downarrow$ are respectively labeled by $(n,j)_\uparrow$ and $(n,j)_\downarrow$, with the radial quantum number $n = 0,1,2,\ldots$. Different from the general bound states in 2DTI quantum dots, the helical edge states are only dependent on the spins and the angular numbers, and can be labeled by $(\uparrow,j)$ and $(\downarrow,j)$. For convenience, we use the lowercase (uppercase) letters to denote the states with up (down) spins and the superscript $\pm$ to distinguish the states $(n,\pm|j|)$. Then the denotation of states used in this paper are $a^\pm(0,\pm 1/2)_\uparrow$, $A^\pm(0,\pm 1/2)_\downarrow$, $b^\pm(1,\pm 1/2)_\uparrow$, $c^\pm(0,\pm 3/2)_\uparrow$, and $d^\pm(2,\pm 1/2)_\uparrow$. Additionally, $e^\pm(\uparrow,\pm 1/2)$ and $E^\pm(\downarrow,\pm 1/2)$ are used for the edge states in 2DTI quantum dots.

In this paper, we choose $\hbar v = 3.65$ eV Å, $M = 30$ meV, while $P = -100$ eV Å^2 and $P = 100$ eV Å^2 are used for 2DBIs and 2DTIs, respectively. The chosen parameters satisfy $\hbar^2 v^2 > 2MP$, which are in agreement with those of the HgTe/CdTe quantum wells and in this condition the energy gap is $(-|M|,|M|)$.

III. COULOMB STATES IN INFINITE CASE

We first investigate the ideal case that the 2DTIs are infinite large, where the edge states are absent as there is not an edge. We will focus on the states bounded near the Coulomb center and study their differences between 2DTI and 2DBI.

The energy spectra as a function of α are shown as the solid lines both for an infinite 2DBI ($PM < 0$) and an infinite 2DTI ($PM > 0$) in Fig. 1. One can see the Coulomb states can be present in the range $E^2 < M^2$. In fact, for $E^2 > M^2$ the index

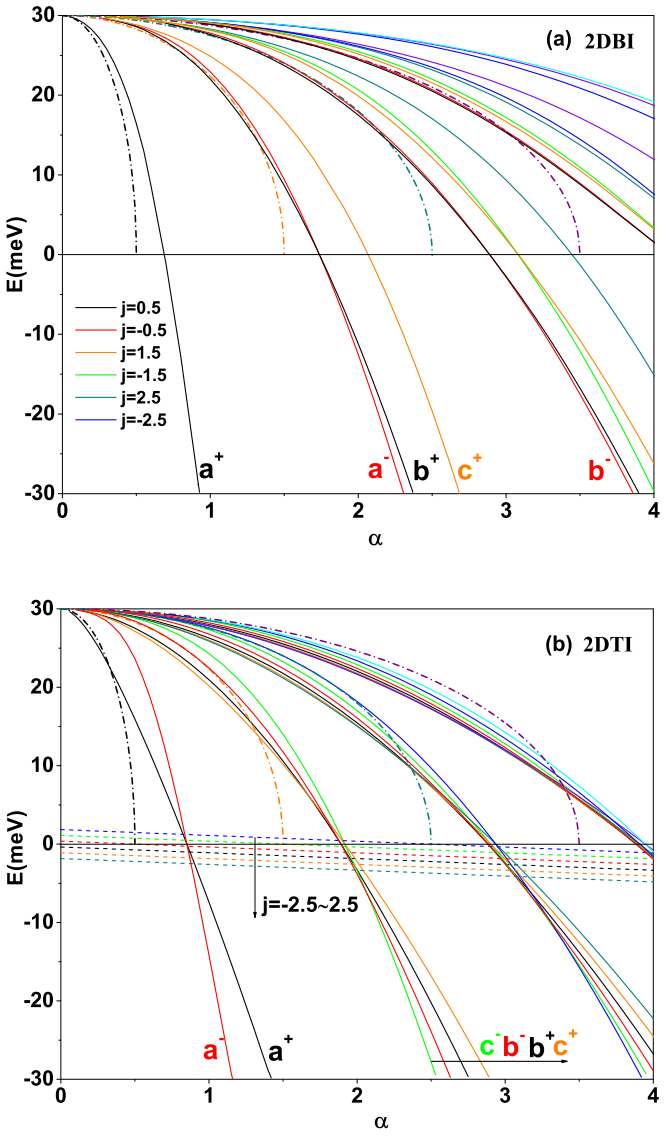

FIG. 1. The Coulomb states of $H_\uparrow$ as a function of α for an infinite 2DBI (a) and 2DTI (b). The signals $a^\pm$, $b^\pm$, and $c^\pm$ stand for the Coulomb states $(0,\pm 1/2)_\uparrow$, $(1,\pm 1/2)_\uparrow$, and $(0,\pm 3/2)_\uparrow$, respectively. The dash-dotted lines stand for the states $(0,j)$ of gapped graphene. The dashed lines in panel (b) show the edge states $(\uparrow,j)$ for a 2DTI quantum dot with $R = 500$ nm.

λ_- in the irregular solutions are purely imaginary, which allow the electrons to tunnel to infinity. It is quite different from the gapped graphene ($P = 0$), where the Coulomb collapse occurs [25] when $\alpha^2 = j^2$. Due to the Pk^2 term, the index ρ in the regular solutions is independent of α rather than $\rho = \sqrt{j^2 - \alpha^2}$ in gapped graphene, then α^2 of the $(0, j)_\uparrow$ states at $E = 0$ in an infinite 2DTI (2DBI) can be much larger than j^2, as shown in Fig. 1.

Let us note the spectra in Fig. 1 in detail. For $P = 0$, the states $(n + 1, |j|)_\uparrow$ and $(n, -|j|)_\uparrow$ are degenerate [26]. For $PM < 0$, such degeneration is only conserved at $E = 0$. The energy order is quite same as the case of $P = 0$, which can be understand since both of them describe topological trivial systems.

However, for $PM > 0$, the spectra are quite different. The energy levels can generally be grouped into different bunches, with each bunch described by the quantum number $N_q = n + |j| + \frac{1}{2}$. The number of states in the N_q bunch is determined by the all possible combinations of n and j. The energy levels do not cross between different bunches, but cross near $E = 0$ within the same bunches. By a closer look of Fig. 1(b) one can find the degeneracy between $\pm j$ states at $E = 0$ are still strict, but they are now between $(n, |j|)_\uparrow$ and $(n, -|j|)_\uparrow$ states, which is in contrast with the case of $PM < 0$.

As analyzed above, the level order and zero-energy degeneracy of the Coulomb states in 2DTIs and 2DBIs are quite different. In fact, our calculations show that the spectra with $\alpha \neq 0$ are discontinuous when P changes across zero, which supports that the above differences are the manifestation of the topological phase in the Coulomb spectra.

IV. EDGE AND COULOMB STATES IN FINITE CASE

In this section, we consider the Coulomb impurities in 2DTI and 2DBI quantum dots. By introducing a finite edge, the edge states will be present in 2DTI, and then an interesting problem is how the edge states are affected by the Coulomb impurity. We will show that the edge states are insensitive to the Coulomb impurity, unless the dot size is small and the Coulomb potential strength is strong with which the edge states can be strongly coupled to the Coulomb states.

The levels of the edge states in a large quantum dot with $R = 500$ nm are shown by the dashed lines in Fig. 1(b). A remarkable characteristic of the edge states is that their energies decrease linearly with α. The slope is $\frac{\partial E}{\partial \alpha} = \langle \frac{\hbar v}{r} \rangle$, which indicates that the wave functions of edge states are almost unchanged for different α. In other words, the edge states are insensitive to the attraction of the Coulomb potential for a large dot size, which should be attributed to the topological nature of these states.

Then we seek for the coupling effects between the impurities and the edge states under different dot size and potential strength. Since there is no coupling between states with different j, we will focus on states with a fixed j below.

The levels of $a^+(0, \frac{1}{2})_\uparrow$ and e^+ $(\uparrow, \frac{1}{2})$ as a function of R are shown in Fig. 2. Both in 2DBI and 2DTI quantum dots, the a^+ states are always present, and their levels converge quickly to the infinite case with increasing R. The choice of vacuum boundary condition allows the bound states to exist in the region $E^2 > M^2$. Due to the bulk-boundary correspondence,

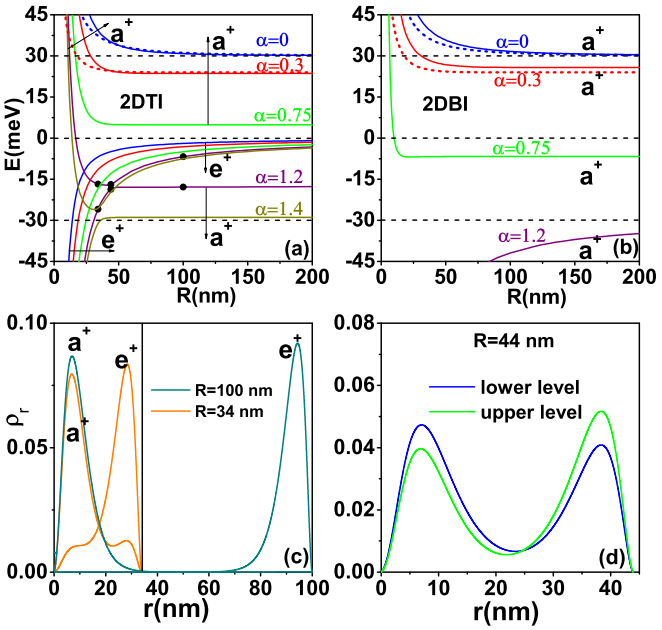

FIG. 2. The levels of Coulomb states a^+ and edge states e^+ of $H_\uparrow$ under different α as a function of R for 2DTI (a) and 2DBI (b). The red and blue dot lines are the levels of a^+ for gapped graphene. The radial density distribution ρ_r of a^+ and e^+ states for $\alpha = 1.2$ at $R = 34$ nm and 100 nm (c) and at $R = 44$ nm (d) in 2DTI quantum dots.

the e^+ states can only exist in 2DTI quantum dots. When $R \to \infty$, the levels of e^+ for different α always trend to zero. In Figs. 2(a) and 2(b), the corresponding a^+ levels of the gapped graphene dots are also shown by the dot lines, where the infinite mass boundary condition [27] is used in the calculations. It is seen that the a^+ levels for $P = 0$, both with and without Coulomb potentials, are lower than those for $P \neq 0$ with the increasing of size confinement. Besides, for $P = 0$ the edge states are absent, and when $\alpha^2 > j^2$ the Coulomb states also disappear due to the Coulomb collapse.

Note that for 2DTI dots the levels of states a^+ at large R fall below zero for strong Coulomb potential of $\alpha > 0.85$; see Fig. 1(b). With decreasing R, the levels of states a^+ and e^+ increase and decrease, respectively. The levels of a^+ and e^+ have to meet at some R, but the crossings of levels with the same j are generally avoided, and, an anticrossing occurs as shown by the case of $\alpha = 1.2$ in Fig. 2(a). Through the anticrossing, the energy order of the a^+ and e^+ states is changed. We notice that for $R = 34$ nm, an additional peak near the edge (center) on ρ_r for a^+ (e^+) arises in Fig. 2(c). It is an effect of coupling between a^+ and e^+. When R is more closed to the coupling region center $R_c = 44$ nm, where the two levels are nearest, a^+ and e^+ will be strongly coupled. It is hard to distinguish the edge and Coulomb states according to their wave functions, as shown in Fig. 2(d). Besides, as shown in Fig. 2(a), with larger α, R is smaller where the coupling occurs.

To further study the coupling effects, the energy levels of states $(n, \frac{1}{2})_\uparrow$ and e^+ as a function of α with fixed $R = 50$ nm are shown in Fig. 3(a). The levels of a^+ overlap with the corresponding infinite levels for $\alpha > 0.4$, and they decrease

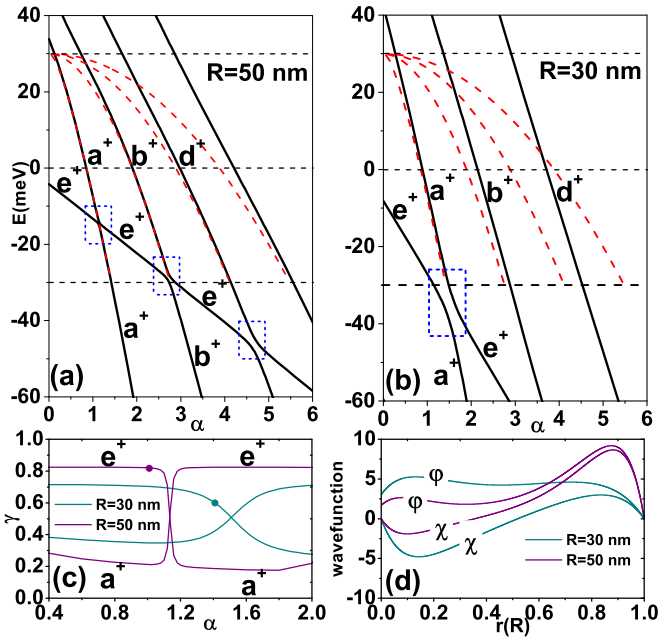

FIG. 3. The energy levels of Coulomb states $(n,\frac{1}{2})_\uparrow$ and edge states e^+ in a 2DTI dot as a function of α with $R = 50$ nm (a) and $R = 30$ nm (b). The red dashed lines are the corresponding Coulomb levels in an infinite 2DTI. (c) The values of $\gamma = \langle\hat{r}\rangle/R$ for a^+ and e^+ as a function of α. (d) The wave functions of the states e^+ respectively for $R = 30$ nm and $R = 50$ nm when $\alpha = \alpha_c - 0.1$, as labeled by the dots in panel (c).

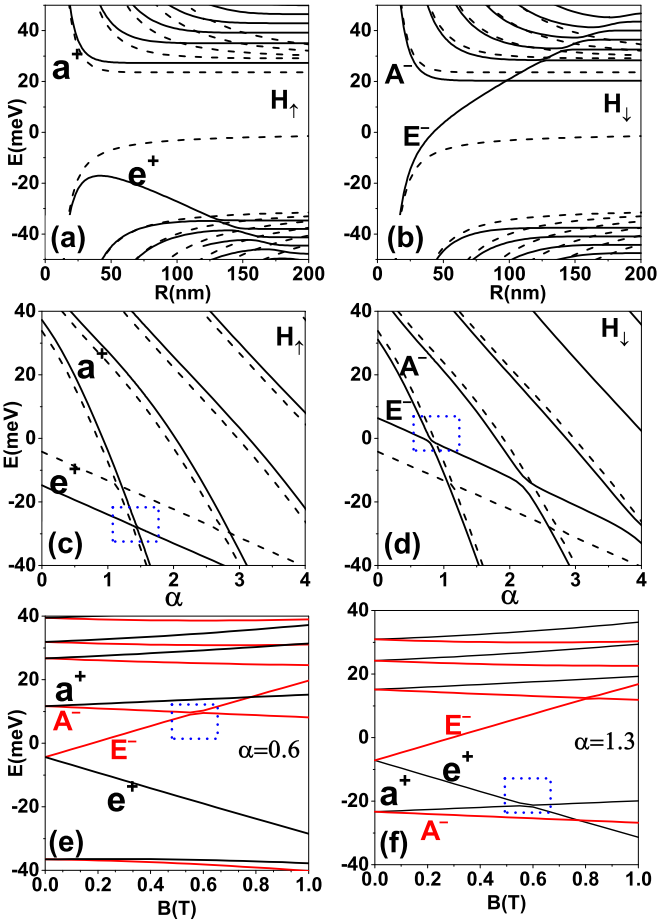

FIG. 4. The energy levels of Coulomb and edge states in a 2DTI quantum dot with $B = 1$ T as a function of R at fixed $\alpha = 0.3$ for $j = \frac{1}{2}$, $H_\uparrow$ (a) and $j = -\frac{1}{2}$, $H_\downarrow$ (b). The levels as a function of α at fixed $R = 50$ nm for $j = \frac{1}{2}$, $H_\uparrow$ (c) and $j = -\frac{1}{2}$, $H_\downarrow$ (d). Both levels of $H_\uparrow$ (black lines) and $H_\downarrow$ (red lines) as a function of B at fixed $R = 150$ nm for $\alpha = 0.6$ (e) and $\alpha = 1.3$ (f). The dashed lines are levels for $B = 0$ in panels (a)–(d).

more quickly than levels of e^+ with increasing α. Therefore, the coupling between a^+ and e^+ is inevitable at some α, and it further occurs between e^+ and b^+, $d^+(2,\frac{1}{2})_\uparrow$,..., with increasing α. It is found that for a smaller dot size the level space at the coupling region center is larger, as shown in Fig. 3(b). To understand this, we define a dimensionless value $\gamma = \langle\hat{r}\rangle/R \in (0,1)$ to describe the distribution of the states and show its plot as a function of α in Fig. 3(c). The γ values of the upper and lower levels cross near the coupling region center, i.e., $\alpha_c = 1.14$ and $\alpha_c = 1.51$ for $R = 50$ nm and $R = 30$ nm, respectively. Compared with the case of $R = 50$ nm, γ values for $R = 30$ nm change much gently in a larger region of α. Meanwhile, it is shown in Fig. 3(d) that when α is away from α_c with the same amplitude, the changes of the edge state wave functions for $R = 30$ nm are obviously larger. It means that the coupling is stronger for a smaller dot because the overlapping of edge and Coulomb states is larger.

From the above discussion, the coupling of the Coulomb and edge states manifest via the anticrossing of the levels. Strong Coulomb potentials can make the levels closer, and smaller dot sizes induce large wave function overlaps. With the changes of levels and transition probabilities, the coupling effects in 2DTI quantum dots are expected to be observed experimentally.

V. MAGNETIC FIELD EFFECTS

Now we consider the magnetic field modulations of the edge states, Coulomb states, and their coupling in 2DTI quantum dots. The distributions of the edge states have much larger magnetic moments, especially for the case of a large dot, and therefore edge states will be quite sensitive to the magnetic fields. Since $H_\uparrow(r,j,B) \neq H_\downarrow(r,-j,B)$ for $B \neq 0$, we calculate both levels of $H_\uparrow$ for $j = \frac{1}{2}$ and levels of $H_\downarrow$ for $j = -\frac{1}{2}$.

It is shown in Figs. 4(a) and 4(b) that the differences between the levels of $B = 0$ and $B = 1$ T are quite small for $R < l_b = 25.66$ nm. With increasing R, the levels of Coulomb states turn flat for both $B = 0$ and $B = 1$ T. However, the levels of the edge states are quite different between $B = 0$ and $B = 1$ T for $R > l_b$. For $B = 0$ the levels of $e^+(\uparrow,\frac{1}{2})$ and $E^-(\downarrow,-\frac{1}{2})$ states trend to zero with increasing R, while for $B = 1$ T they respectively decrease and increase quickly to out of gap. The levels of e^+ and E^- states meet and couple with the positive-energy and negative-energy states for $R \geqslant 131$ nm and $R \geqslant 97$ nm, respectively. The coupling effects become strong for large R, leading to the missing of the edge state characteristics. It is because the magnetic moments of the edge states increase with R and have opposite signs for e^+ and E^-.

The sensitivity of the edge states to magnetic fields for large dot size stands in contrast with their insensitivity to Coulomb fields.

The coupling region center α_c of the Coulomb and edge states is tunable with magnetic fields which make it convenient to observe the coupling effect experimentally. The levels of a 2DTI quantum dot with fixed $R = 50\,\text{nm}$ and $B = 1$ T are shown in Figs. 4(c) and 4(d). Compared with the case of $B = 0$, the levels of e^+ and E^- are respectively lowered and lifted while the levels $a^+(0,\frac{1}{2})_\uparrow$ and $A^-(0,-\frac{1}{2})_\downarrow$ only changes slightly. The coupling region centers are moved right and left for $H_\uparrow$ and $H_\downarrow$, respectively.

The level order of the Coulomb and edge states can also be modulated by magnetic fields. In Figs. 4(e) and 4(f), we show the splitting of the Kramers' pairs with B for a large dot size. Clearly, the levels of the edge states changes more sharply than the Coulomb states. For $\alpha = 0.6$, the level order of A^- and E^- states changes at $B = 0.58\,\text{T}$. When α increases to 1.3, the decreases of the edge states (compared with $\alpha = 0.6$) are much smaller than the Coulomb states, which leads to change of the energy order of a^+ and e^+ at $B = 0.59\,\text{T}$.

Under weak magnetic fields, the main difference between gapped graphene and 2DTI is the absence of edge states such as e^+ and E^- in Figs. 4(a) and 4(b) in graphene dots. With increasing of magnetic fields, the levels of 2DTI and graphene dots will transit from the quantum dot levels to the corresponding Landau levels. When $\alpha = 0$, the zero Landau levels [28] depend linearly on B for $P \neq 0$, i.e., $L_0^\pm = \mp(M - ePB/\hbar)$, while they are independent of B for $P = 0$, i.e., $L_0^\pm = \mp M$. When $\alpha \neq 0$ the degenerate Landau levels split whereas the above dependence with B can still hold approximately. This may be the most notable differences between gapped graphene and 2DTI dots under stronger magnetic fields.

VI. SUMMARY

Developing our series expansion method, the exact spectra of Coulomb impurities in infinite 2DTIs and 2DBIs are obtained and compared. For 2DBIs, the level order of Coulomb states are similar to those of gapped graphene, while for 2DTIs the spectra show bunchy structures. The degeneracies at $E = 0$ are quite different between 2DBIs and 2DTIs.

In 2DTI quantum dots, the edge states can be coupled to the Coulomb states in strong Coulomb potentials. Aside the coupling region center the level order of the two kind states change while near the coupling region center the states turn into mixtures. The coupling are stronger for smaller dots. For larger dots, the edge states are insensitive to Coulomb potentials while they are sensitive to magnetic fields.

Finally, the mathematical method is powerful for ordinary differential equation arrays, which can describe more complicated Dirac-like systems. The results deepen the understanding of the topological phases and the properties of the edge states in the present of Coulomb impurities. Based on the exact energy levels and wave functions, the related properties such as optical spectra of 2DTI quantum dots can be further explored.

ACKNOWLEDGMENTS

Financial supports from NSF China (Grants No. 10974108 and No. 11174170) and Open Research Fund Program of the State Key Laboratory of Low-Dimensional Quantum Physics (Grant No. KF201507) are gratefully acknowledged.

APPENDIX: SERIES EXPANSIONS IN DIFFERENT REGIONS

We will take the $H_\uparrow$ block as an example to illustrate the series expansion methods for the modified Dirac equation. The coefficient matrix A_{-1} in Eq. (3) is

$$A_{-1} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ l_-^2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & l_+^2 & -1 \end{pmatrix}. \tag{A1}$$

In the regular region $(0,r_0)$ the special solutions can be expressed as

$$w = r^\rho \sum_{n=0}^{\infty} a_n r^n. \tag{A2}$$

Substitute w into Eq. (3) and compare the coefficients, we get

$$(\rho + n - A_{-1})a_n = \sum_{k=0}^{3} A_k a_{n-k-1} \tag{A3}$$

with $n = 0,1,2,\ldots$, and we have assumed $a_n = 0$ for $n < 0$ here and after. By setting $n = 0$ we find that ρ have to be the eigenvalues of A_{-1} and a_0 are the corresponding eigenvectors. We need the convergence solutions near $r = 0$, so only two of the four eigenvalues is preserved, which are $\rho = \rho_{1,2}$ with $\rho_1 = \rho_2 + 1 = l_-$ for $j > 0$ and $\rho_1 = \rho_2 + 1 = -l_+$ for $j < 0$. When $\rho = \rho_1$, a_n for $n > 0$ can be directly determined via left multiplying Eq. (A3) by $(\rho_1 + n - A_{-1})^{-1}$.

It is noticeable that $\rho_2 < \rho_1$ differs by an integer with ρ_1, which means that when $\rho = \rho_2$ one should add the $r^{\rho_2}\ln r \sum_{n=0}^{\infty} a_n' r^n$ term to Eq. (A2) to get the other independent solution. Fortunately, we have proved that this term is vanishing. Then Eq. (A3) is also suitable for $\rho = \rho_2$. For $n > 1$ one can similarly left multiply Eq. (A3) by $(\rho_2 + n - A_{-1})^{-1}$ to get a_n. For $n = 1$, however, the matrix $(\rho_2 + 1 - A_{-1})$ is singular and we obtain

$$a_1 = \left[\frac{\hbar v\alpha}{(2\rho_2+3)P}, \frac{(\rho_2+2)\hbar v\alpha}{(2\rho_2+3)P}, \delta, \delta(\rho_2+2)\right]^{\mathrm{T}} \tag{A4}$$

for $j > 0$ and

$$a_1 = \left[\delta, \delta(\rho_2+2), \frac{\hbar v\alpha}{(2\rho_2+3)P}, \frac{(\rho_2+2)\hbar v\alpha}{(2\rho_2+3)P}\right]^{\mathrm{T}} \tag{A5}$$

for $j < 0$ where δ is arbitrary. Once δ is chosen, two linearly independent solutions corresponding to ρ_1 and ρ_2 are obtained, and any different choices of δ will give solutions linearly dependent with the obtained two.

Away from $r = 0$, we divide the region (r_0,r_I) into I small regions (r_i,r_{i+1}) with $i = 0,1,\ldots,I-1$. In region (r_i,r_{i+1}), the special solutions of Eq. (3) are expanded at $x_i = \frac{r_i+r_{i+1}}{2}$ as

the Taylor series

$$w = \sum_{n=0}^{\infty} b_n (r - x_i)^n. \tag{A6}$$

By substituting this expression into Eq. (3) we get the recurrence relations for $n \geqslant 1$:

$$b_n = \frac{1}{n x_i} \left[\sum_{m=-1}^{3} \sum_{k=0}^{m+1} A_m C_{m+1}^k x_i^{m-k+1} b_{n-k-1} - (n-1) b_{n-1} \right]. \tag{A7}$$

For $n = 0$, b_0 are completely underdetermined and can be chosen as the column vectors of the 4×4 unit matrix, which correspond to four independent solutions.

In the irregular region (r_I, ∞), we need treat respectively the case of $B = 0$ and $B \neq 0$ as the magnetic field changes the properties of Eq. (3) at $r = \infty$. We will illustrate the techniques of the asymptotic expansion in case of $B = 0$, while the case of $B \neq 0$ is similar.

When $B = 0$, we use the transformation

$$\xi = U w = \begin{pmatrix} r & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} w, \tag{A8}$$

and then ξ is proved to satisfy the differential equation array

$$\frac{d\xi}{dr} = \sum_{k=-2}^{0} B_k r^k \xi. \tag{A9}$$

Compared with Eq. (3), the highest power of r decreases to the lowest in Eq. (A9). The coefficient matrix B_0 is

$$B_0 = \frac{1}{P} \begin{pmatrix} 0 & 0 & 0 & 0 \\ E - M & 0 & 0 & \hbar v \\ 0 & 0 & 0 & 0 \\ 0 & \hbar v & -(E + M) & 0 \end{pmatrix}. \tag{A10}$$

The formal solutions of Eq. (A9) are written as

$$\xi = e^{\sum_{i=1}^{s} q_i r^i} r^{\widetilde{\eta}} \sum_{n=0}^{\infty} \widetilde{c}_n r^{-n}. \tag{A11}$$

By substituting Eq. (A11) into Eq. (A9) and comparing the coefficients we find $s = 1$ and

$$(B_0 - q_1)\widetilde{c}_n = (\widetilde{\eta} - n - B_{-1})\widetilde{c}_{n-1} - B_{-2}\widetilde{c}_{n-2}. \tag{A12}$$

By setting $n = 0$ in Eq. (A12) it is seen that

$$q_1 = \pm \sqrt{\frac{(\hbar v)^2 - 2MP \pm \sqrt{(\hbar v)^4 - 4MP(\hbar v)^2 + 4P^2E^2}}{2P^2}} \tag{A13}$$

are the four eigenvalues of B_0, and $\widetilde{c}_0$ are the corresponding eigenvectors. When $q_1 > 0$, Eq. (A11) is divergent at $r = \infty$; therefore we only chose the two solutions corresponding to $q_1 = -|q_1|$. For determination of $\widetilde{\eta}$, we need notice that the rank of $(B_0 - q_1)$ is three. It allows us to find a matrix T so that $T(B_0 - q_1)$ is in the form of row echelon. Left multiplying Eq. (A12) by T, we get

$$C\widetilde{c}_n = D\widetilde{c}_{n-1} - F\widetilde{c}_{n-2}, \tag{A14}$$

where $C = T(B_0 - q_1)$, $D = T(\widetilde{\eta} - 1 - B_{-1})$, and $F = TB_{-2}$. Remember that the fourth row of the matrix C is zero, and by setting $n = 1$ in Eq. (A14) we have

$$\mathrm{row}\{D,4\}\widetilde{c}_0 = 0, \tag{A15}$$

where $\mathrm{row}\{D,i\}$ denotes the row vector formed by the i's row of the matrix D. Equation (A15) fixes $\widetilde{\eta}$ for each chosen q_1.

Having known $\widetilde{c}_{n-2}$ and $\widetilde{c}_{n-1}$, one cannot get $\widetilde{c}_n$ directly from Eq. (A12) since $(B_0 - q_1)$ is singular, which is different from the regular and Taylor series expansions. We again use Eq. (A14) and take the first three rows as three independent equations to fix $\widetilde{c}_n$. Then we set $n \to n + 1$ and take the fourth row of Eq. (A14) as the other independent equation. The above argument leads to the following recurrence relation

$$\widetilde{c}_n = \begin{bmatrix} \mathrm{row}\{C,1\} \\ \mathrm{row}\{C,2\} \\ \mathrm{row}\{C,3\} \\ \mathrm{row}\{D,4\} \end{bmatrix}^{-1} \times \left\{ \begin{bmatrix} \mathrm{row}\{D,1\} \\ \mathrm{row}\{D,2\} \\ \mathrm{row}\{D,3\} \\ \mathrm{row}\{F,4\} \end{bmatrix} \widetilde{c}_{n-1} - \begin{bmatrix} \mathrm{row}\{F,1\} \\ \mathrm{row}\{F,2\} \\ \mathrm{row}\{F,3\} \\ 0 \end{bmatrix} \widetilde{c}_{n-2} \right\}. \tag{A16}$$

For the asymptotic solutions, one can establish the relation of Eq. (4) and Eq. (A11) through Eq. (A8). For example we have $\lambda_{\pm} = -|q_1|$ and $\eta = \widetilde{\eta} - 1$. Instead of deducing c_k and then calculating $w(r = r_I)$ to connect the asymptotic solutions and the Taylor solutions, it is more convenient to first calculate $\xi(r = r_I)$ and then use Eq. (A8) to obtain $w(r = r_I)$ in the numerical calculations.

[1] X. L. Qi and S. C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[2] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
[3] A. Roth, C. Brüne, H. Buhmann, L W. Molenkamp, J. Maciejko, X.-L. Qi, and S.-C. Zhang, Science **325**, 294 (2009).
[4] M. Korkusinski and P. Hawrylak, Sci. Rep. **4**, 4903 (2014).
[5] L. Fu and C. L. Kane, Phys. Rev. Lett. **102**, 216403 (2009).
[6] G. J. Ferreira and D. Loss, Phys. Rev. Lett. **111**, 106802 (2013).
[7] K. Chang and W. K. Lou, Phys. Rev. Lett. **106**, 206802 (2011).
[8] P. Michetti and P. Recher, Phys. Rev. B **83**, 125420 (2011).
[9] J. Lu, W.-Y. Shan, H.-Z. Lu, and S.-Q. Shen, New. J. Phys. **13**, 103016 (2010).
[10] V. A. Sablikov and A. A. Sukhanov, Phys. Rev. B **91**, 075412 (2015).
[11] Q. Liu and T. Ma, Phys. Rev. B **80**, 115216 (2009).

[12] T. L. Schmidt, S. Rachel, F. von Oppen, and L. I. Glazman, Phys. Rev. Lett. **108**, 156402 (2012).
[13] G. Tkachov and E. M. Hankiewicz, Physica E (Amsterdam, Neth.) **44**, 900 (2012).
[14] R. Yu, W. Zhang, H.-J. Zhang, S.-C. Zhang, X. Dai, and Z. Fang, Science **329**, 61 (2010).
[15] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang, Z.-Q. Ji, Y. Feng, S. Ji, X. Chen, J. Jia, X. Dai, Z. Fang, S.-C. Zhang, K. He, Y. Wang, L. Lu, X.-C. Ma, and Q.-K. Xue, Science **340**, 167 (2013).
[16] L. A. Wray, S.-Y. Xu, Y. Xia, D. Hsieh, A. V. Fedorov, Y. S. Hor, R. J. Cava, A. Bansil, H. Lin, and M. Z. Hasan, Nat. Phys. **7**, 32 (2011).
[17] B. A. Bernevig, T. L. Hughes, and S. C. Zhang, Science **314**, 1757 (2006).
[18] W.-Y. Shan, J. Lu, H.-Z. Lu, and S.-Q. Shen, Phys. Rev. B **84**, 035307 (2011).
[19] J.-L. Zhu, Phys. Rev. B **39**, 8780 (1989).
[20] J.-L. Zhu, S. Y. Sun, and N. Yang, Phys. Rev. B **85**, 035429 (2012).
[21] D. G. Rothe, R. W. Reinthaler, C.-X. Liu, L. W. Molenkamp, S.-C. Zhang, and E. M. Hankiewicz, New J. Phys. **12**, 065012 (2010).
[22] M. J. Schmidt, E. G. Novik, M. Kindermann, and B. Trauzettel, Phys. Rev. B **79**, 241306 (2009).
[23] E. A. Coddington and N. Levionson, *Theory of Ordinary Differential Equations* (TaTa McGraw-Hill, New Delhi, 1955).
[24] L. Sirovich, *Techniques of Asymptotic Analysis* (Springer, New York, 1971).
[25] A. V. Shytov, M. I. Katsnelson, and L. S. Levitov, Phys. Rev. Lett. **99**, 236801 (2007).
[26] D. S. Novikov, Phys. Rev. B **76**, 245435 (2007).
[27] S. Schnez, K. Ensslin, M. Sigrist, and T. Ihn, Phys. Rev. B **78**, 195427 (2008).
[28] M. König, H. Buhmann, L. W. Molenkamp, T. L. Hughes, C. X. Liu, X. L. Qi, and S. C. Zhang, J. Phys. Soc. Jpn. **77**, 031007 (2008).