Excitonic Collapse in Semiconducting Transition Metal Dichalcogenides

A. S. Rodin and A. H. Castro Neto

1Boston University, 590 Commonwealth Ave., Boston MA 02215
2Graphene Research Centre and Department of Physics, National University of Singapore, 117542, Singapore

(Dated: May 21, 2013)

PACS numbers: 73.20.Mf, 78.67.-n

Semiconducting transition metal dichalcogenides (STMDC) are two-dimensional (2D) crystals characterized by electron volt size band gaps, spin-orbit coupling (SOC), and d-orbital character of its valence and conduction bands. We show that these materials carry unique exciton quasiparticles (electron-hole bound states) with energy within the gap but which can “collapse” in the strong coupling regime by merging into the band structure continuum. The exciton collapse seems to be a generic effect in these 2D crystals.

Isolation of graphene in 2004[1] opened up a new field in condensed matter physics: 2D crystals[2]. Having found a stable 2D crystal, the scientific community began searching for other 2D systems with different properties[3]. A particularly interesting family of materials is the transition metal dichalcogenides (TMDC) with chemical formula $MC_2$, where $M$ is a transition metal and $C$ a chalcogen ($C = S, Se, Te$). At their thinnest, these systems are composed of three atomic layers: one layer of $M$ atoms is sandwiched by two $C$ atom layers. In each layer, atoms form a triangular lattice so that the material resembles graphene where instead of A and B sublattices one has different types of atoms found in different planes. These materials show a large class of ground states that vary from metallic or semiconducting to superconducting or charge density wave (CDW) depending on the same situation can occur in STMDC without the need of impurities but as a result of the internal electron-hole interaction in these materials.

The starting point of our discussion is the low-energy Hamiltonian used to model the band structure close to K and K’ points in the Brillouin zone[4].

$$H = at(r_k x \sigma_z + k_y \sigma_y) + \frac{\Delta}{2} \sigma_z - \lambda \frac{\sigma_z - 1}{2} s_z, \quad (1)$$

where $\tau$ is the valley index, $a$ is the lattice constant, $t$ is the hopping energy, $\Delta$ is the “mass” term, and $\lambda$ is the SOC parameter. In this model, the masses of electrons and holes are set to be equal and trigonal warping is neglected[11]. These effects can be added to (1) but they do not affect qualitatively the results here. The values for the constants in Eq. (1) are given in Ref.[8]. The Hamiltonian [1] is block-diagonal and the spins are uncoupled. This allows us to treat the spin sub-Hamiltonians separately, leading to a simplified $2 \times 2$ matrix:

$$H_{1/2} = \tau \begin{pmatrix} m_l & t \alpha e^{i \tau \theta} \\
-\alpha e^{-i \tau \theta} & -m_l \end{pmatrix} + \frac{\lambda}{2} - \mu = \tau H_{0,1} - \mu^* I,$$

where $\mu$ is the chemical potential (which can be tuned by an applied electric field), $m_l = (\Delta + \lambda \tau)/(2 \tau)$, and $\mu^* = \mu + \tau \lambda/2$. Equation (2) is the Hamiltonian of a “massive” Dirac fermion with a finite chemical potential where both $\mu^*$ and $m_l$ depend on valley and spin indices.

The mass term in (1) breaks the sublattice symmetry leading to a gap of size $2m_l$ between valence and conduction bands. The conduction, $s = +1$, and valence, $s = -1$, bands are given by $\varepsilon_{k,l,s} = \sqrt{m_l^2 + (ak)^2}$, revealing the emergent low energy Lorentz invariance of these materials close to the band edge. Furthermore, the
spin-up and spin-down Green’s functions are given by:
\[
\mathcal{G}_{\uparrow/\downarrow} = \left[ i \hbar \omega_n - H_{\uparrow/\downarrow} \right]^{-1} = \frac{1}{2} \sum_{s = \pm 1} \frac{\mathbb{1} + s \tau H_{0,1} / \varepsilon_{k,l}}{i \hbar \omega_n - \varepsilon_{k,l,s}}, \quad (3)
\]
\[
\varepsilon_{k,l,s} = s \varepsilon_{k,l} - \mu^*_s, \quad \varepsilon_{k,l} = \sqrt{m_l^2 + (t a k)^2} \quad (4)
\]

In order to understand the electron-electron interactions in these materials it is important to study the behavior of the dielectric function. In the random phase approximation (RPA), the polarization function is given by the density-density correlation:
\[
P_l(q, i \omega_n) = \text{Tr} \left[ \sum_{k,i k_n} \mathcal{G}(k,i k_n) \mathcal{G}(k+q,i k_n+i \omega_n) \right]. \quad (5)
\]

In the low temperature regime, \(T \to 0\), and long wavelength limit, \(q \to 0\), we find:
\[
P_l(q \to 0, \omega) = \frac{q^2}{8 \pi |m_l|} \left[ \frac{E_l}{2 \omega^2} - \frac{1 + \tilde{\omega}_l^2}{4 \omega^2} \ln \left( \frac{E_l + \tilde{\omega}_l}{E_l - \tilde{\omega}_l} \right) \right], \quad (6)
\]
where \(E_l = \max[\mu^*_l/m_l, 1]\), \(\tilde{\omega}_l = \hbar \omega/(2|m_l|)\), and \(\mu^*_l > 0\).

For a vanishing gap, \(2|m_l| \to 0\), the result approaches the graphene polarization function\[{12}\]. One important difference is the fact that the imaginary part of the product \(\omega_l P(q, \tilde{\omega}_l)\), proportional to the low-\(q\) conductivity, is not constant for \(\tilde{\omega}_l > E_l\). Instead, it has a maximum at \(E_l = \tilde{\omega}_l\) which is the consequence of a diverging density of states at the extremal points of the bands. For undoped or ungated crystals, the chemical potential is located in the gap region, the conduction band is empty, and the conductivity becomes finite only for \(\hbar \omega = 2|m_l|\). As the system becomes \(n\)-doped, \(\tilde{\omega}_l P(q, \tilde{\omega}_l)\) is nonzero at \(\hbar \omega > 2|\mu^*_l|\) since interband transitions with smaller \(\tilde{\omega}_l\) are forbidden. We plot the polarization function for several values of \(E_l\) in Fig. 1. If the Fermi level is in the gap, the real part of polarization is always negative. This means that the dielectric function \(\epsilon_{\text{RPA}}(q, \omega) = \epsilon_0 - 2\pi e^2/q \times P(q, \omega)\), where \(\epsilon_0\) is the external dielectric constant, never changes sign and there are no collective charge excitations such as plasmons. Conversely, allowing electrons to populate the conduction band allows \(\epsilon_{\text{RPA}}\) to pass through zero, giving rise to plasmonic behavior.

The screened Coulomb interaction between charged particles is given by the static polarization function. Since we are interested in excitons, we treat the case where the Fermi level is in the middle of the gap. Straightforward manipulations of Eq. (5) at \(T = 0\) yield:
\[
\begin{align*}
P(q, 0) &= -\frac{m_l}{4 \pi^2 (t a)^2} F \left( \frac{\tau a}{m_l} \right), \\
F(Q) &= \int_0^\infty d\theta \int dx \left( 1 - \frac{1 + x(x + Q \cos \theta)}{\sqrt{1 + x^2} \sqrt{1 + x^2 + Q^2 + 2xQ \cos \theta}} \right) \left( \frac{1}{\sqrt{1 + x^2} \sqrt{1 + x^2 + Q^2 + 2xQ \cos \theta}} \right). \quad (7)
\end{align*}
\]

The asymptotic behavior of \(F(Q)\) is given by:
\[
F(Q) = \begin{cases} 
\frac{\pi Q^2}{3} & \text{if } Q \ll 1, \\
\frac{\pi^2 Q^4}{4} & \text{if } Q \gg 1.
\end{cases} \quad (9)
\]

The full system polarization is obtained by doubling Eq. (4) to account for the valley degeneracy and summing over \(l = \pm 1\) to take care of both spins. Dividing the Coulomb term \(2\pi e^2/q\) by the static \(\epsilon_{\text{RPA}}\) gives the screened interaction potential. Taking the inverse Fourier transform of \(V_{\text{RPA}}(q)\) results in:
\[
V_{\text{RPA}}(r) = \frac{\hbar v_\alpha}{\rho a} \int_0^\infty dx \frac{J_0(x)}{1 + \frac{x^2}{2} \sum_{l=\pm 1} \frac{F(x/(S_l^*) \rho)}{x/(S_l^*) \rho}}, \quad (10)
\]
\[
\alpha = \frac{e^2}{\hbar v_\alpha}, \quad S_l = \sqrt{m_l / \tau}, \quad (11)
\]
where \(\hbar v = t a\). For very small \(r\), we can use the large-\(Q\) approximation for \(F(Q)\) in Eq. (10). This removes all \(x\)-dependence from the denominator in (10) and results in a simple renormalization of the coupling constant \(\alpha \to \alpha/(1 + \alpha \pi/2)\) which is identical to the static screening in single layer graphene\[{13,14}\]. In the opposite limit of large \(r\), there is no such renormalization and the Coulomb interaction remains unscreened. We are considering excitons where the distance \(r\) between an electron and a hole is much larger than the lattice constant \(a\) as the mass gap \(m_l\) is much smaller than the electron mass, and hence the exciton radius much larger than the Bohr radius.

Having established the nature of the Coulomb interaction in the system, we now turn our attention to the EQ. In order to keep the notation as simple as possible, we will dispense with the details of the SOC and simply denote the size of the gap by \(2m_l\), keeping in mind that this quantity depends on the spin and the valley. It is convenient to perform a particle-hole transformation so that the electron and hole free Hamiltonians become:
\[
H_{\epsilon/h} = \begin{pmatrix} & & \pm m_l & & \\
& & \pm \hbar v e^{\pm i \theta} & & & \mp m_l &
\end{pmatrix}. \quad (12)
\]
when dealing with excitons, one works in the center-of-
mass frame and \( \omega = E = \pm \sqrt{m_i^2 + \hbar^2 v_e^2 k^2} \). Positive energies correspond to the states in the electron and hole bands while the negative ones represent the states in the electron and hole “seas”. The total Hamiltonian for two particles without interactions is given by the tensor product \( H_2 = H_e \otimes I + I \otimes H_h \).

\[
H_2 = \begin{pmatrix}
0 & -\hbar v_e e^{i\theta_k} & \hbar v_e e^{-i\theta_r} & 0 \\
-\hbar v_e e^{i\theta_r} & 2m_l & 0 & \hbar v_e e^{-i\theta_p} \\
0 & -2m_l & -\hbar v_e e^{i\theta_k} & 0 \\
0 & \hbar v_e e^{i\theta_p} & -\hbar v_e e^{-i\theta_k} & 0 \\
\end{pmatrix}
\]

Diagonalizing the Hamiltonian gives two branches with \( \mathcal{E} = \pm \sqrt{m_i^2 + \hbar^2 v_e^2 k^2} \). Positive energies correspond to the states in the electron and hole bands while the negative ones represent the states in the electron and hole “seas”. The total Hamiltonian for two particles without interactions is given by the tensor product \( H_2 = H_e \otimes I + I \otimes H_h \).

When dealing with excitons, one works in the center-of-
mass frame of reference. Since in the model the masses of
electrons and holes are the same, the momentum must be
opposite. Setting \( p = -k \) results in four eigenvalues:
\( \pm 2\sqrt{m_i^2 + \hbar^2 v_e^2 k^2} \) and a doubly degenerate 0. The zero energy eigenstates arise from the cases when the system has a single electron or a hole and its complementary particle is in its sea. This way, since they have the same momentum, they give equal and opposite contributions to the total energy. The negative eigenvalue corresponds to the situation where both the electron and the hole are in their respective seas. Finally, the positive eigenvalue is what we are interested in. There, an electron is found in the electron band and the hole is the hole band. This can be regarded as an excitonic state with a vanishingly weak interaction. Therefore, the states of interest constitute a subspace of a full Hilbert space describing a two-particle system.

If we go back to the laboratory frame and investigate the kinetic energy of the exciton, \( k \to K/2 - q \) and \( p \to K/2 + q \). Here, \( q \) is the motion in the center of mass frame and \( K \) is the momentum of the center of mass. Diagonalizing Eq. (13), one cannot separate the two momenta. However, assuming that \( q \ll K \), it is possible to estimate the kinetic energy as \( \mathcal{E}_K \approx 2\sqrt{m_i^2 + \hbar^2 v_e^2 K^2/4} \).

Diagonalizing the Hamiltonian gives two branches with \( \mathcal{E} = \pm \sqrt{m_i^2 + \hbar^2 v_e^2 k^2} \). Positive energies correspond to the states in the electron and hole bands while the negative ones represent the states in the electron and hole “seas”. The total Hamiltonian for two particles without interactions is given by the tensor product \( H_2 = H_e \otimes I + I \otimes H_h \).

When dealing with excitons, one works in the center-of-
mass frame of reference. Since in the model the masses of
electrons and holes are the same, the momentum must be
opposite. Setting \( p = -k \) results in four eigenvalues:
\( \pm 2\sqrt{m_i^2 + \hbar^2 v_e^2 k^2} \) and a doubly degenerate 0. The zero energy eigenstates arise from the cases when the system has a single electron or a hole and its complementary particle is in its sea. This way, since they have the same momentum, they give equal and opposite contributions to the total energy. The negative eigenvalue corresponds to the situation where both the electron and the hole are in their respective seas. Finally, the positive eigenvalue is what we are interested in. There, an electron is found in the electron band and the hole is the hole band. This can be regarded as an excitonic state with a vanishingly weak interaction. Therefore, the states of interest constitute a subspace of a full Hilbert space describing a two-particle system.

If we go back to the laboratory frame and investigate the kinetic energy of the exciton, \( k \to K/2 - q \) and \( p \to K/2 + q \). Here, \( q \) is the motion in the center of mass frame and \( K \) is the momentum of the center of mass. Diagonalizing Eq. (13), one cannot separate the two momenta. However, assuming that \( q \ll K \), it is possible to estimate the kinetic energy as \( \mathcal{E}_K \approx 2\sqrt{m_i^2 + \hbar^2 v_e^2 K^2/4} \).

It is possible to keep the masses of electrons and holes different. One particular case of interest is when one of these masses goes to infinity. In this case, the problem reduced to a particle moving in the field of a charged impurity. Careful expansion of the two-particle eigenvalues gives \( \sqrt{m_i^2 + \hbar^2 v_e^2 k^2} \) for the excitonic energy, where \( m_i \) is finite. The collapsing states of this system were treated in Ref. 14.

In the case of equal masses, the minimum energy of exciting an electron to the conduction band from the valence band is \( 2m_i \). We set up the bands in such a way that zero energy is located half-way between their extrema. This results in a reduced excitonic Hamiltonian:

\[
H_E = \begin{pmatrix}
m_l & 2qe^{-i\theta} -m_l \\
2qe^{i\theta} & -m_l \\
\end{pmatrix} - \frac{\hbar v_\alpha}{r},
\]

where we have included the interaction term. Note that compared to the impurity problem [15], the momentum is doubled since the electron and hole are moving with the same momentum. We can perform a variable transformation \( 2q \to p, r \to 2x \) to preserve the commutation relation, yielding:

\[
\tilde{H}_E = \begin{pmatrix}
m_l & pe^{-i\theta} -m_l \\
p e^{i\theta} & -m_l \\
\end{pmatrix} - \frac{\hbar v_\alpha}{x},
\]

with \( \alpha = \alpha/2 \). Notice that for the exciton problem the effective fine structure constant \( \tilde{\alpha} \) is half of the impurity case because of the doubling in the momentum. Equation [15] describes a massive Dirac particle in an attractive Coulomb potential of a single charge with coupling that is \( 2e/(\sqrt{\varepsilon}c) \) times stronger than the vacuum fine structure constant. The energies of the bound states are given by:

\[
\mathcal{E}_{jn} = m_l \frac{n + \sqrt{j^2 - \tilde{\alpha}^2}}{\sqrt{j^2 + \sqrt{n^2 + j^2 - \tilde{\alpha}^2^2}}},
\]

where \( n \) is the principal quantum number and \( j = \pm 1/2, \pm 3/2 \ldots \) is the angular momentum quantum number. These determine the binding energies for the EQ and can be measured, say, by optical means. The optical absorption energies are given by the sum of \( m_l \) and Eq. (16) since \( n \to \infty \), this sum approaches the size of the gap. Notice that for \( \tilde{\alpha} > 1/2 \), some energies become imaginary. This is caused by the breakdown of the point-charge treatment of the center of the Coulomb well and is usually remedied by a regularization procedure, as we discuss below. For now, we avoid this problem by choosing \( \tilde{\alpha} < 1/2 \) to illustrate a particular case. As an example, we pick MoS2 (gap of 1.66 eV) on a SiC substrate \( (\varepsilon_0 \approx 5.5 \text{ and } \tilde{\alpha} \approx 0.42) \). Fig. 2 shows the exciton energies in this case.

For a regular Coulomb potential, the energies in Eq. (16) for the states are always positive. At sufficiently large \( \tilde{\alpha} \), however, the energy “dives” to negative value until it eventually merges with the valence band at \( \mathcal{E} = -m_i \).
Setting $E \rightarrow -m$ and solving Eq. (15) for $x > a/2$ yields the form of the solution to Eq. (15) that we impose is:

$$
\Psi_j(x, \theta) = \frac{1}{\sqrt{\alpha}} \left( e^{-\nu j/\alpha} A(x) + e^{\nu j/\alpha} B(x) \right). \quad (17)
$$

Where $\nu = 2\sqrt{\alpha^2 - j^2}$, $K_\nu$ is a modified Bessel function, and $N$ is the normalization constant. Similarly, for $x < a/2$, the result becomes:

$$
A^<(x) = N K_\nu \left( \frac{\sqrt{8\alpha m_x} x}{\hbar v} \right), \quad (18)
$$

$$
B^<(x) = \frac{N}{\sqrt{\alpha}} \left[ \left( j + \frac{\nu v}{2} \right) K_\nu \left( \frac{\sqrt{8\alpha m_x} x}{\hbar v} \right) + \sqrt{\frac{2\alpha m_x}{\hbar v} K_{\nu-1} \left( \frac{\sqrt{8\alpha m_x} x}{\hbar v} \right)} \right], \quad (19)
$$

where $\nu = 2\sqrt{\alpha^2 - j^2}$, and $m_x$ is a modified Bessel function.

The lowest energy levels, corresponding to $j = 1/2$, are the first to merge with the continuum. We determine the smallest values of $\alpha$ that satisfy Eq. (23) for a given $S_l$ at $j = 1/2$ and plot the result in Fig. 3. We also show where four STMDCs are located in the phase diagram depending on the dielectric screening. Suspended samples with $\epsilon_0 = 1$ are all located in the supercritical regime. On the other hand, samples placed on BN with $\epsilon_0 \approx 2.63$ are located in the subcritical regime. Therefore, in order to possibly observe the excitonic collapse, for $S_l \approx 1$ one needs to work substrates with $1 < \epsilon_0 \leq 1.25$.

In conclusion, we have analyzed a long wavelength model for STMDC's and obtained both the long-range and static polarization functions. We used the latter to determine the screening of Coulomb interactions in the system. This allowed us to solve for the energy levels of exciton quasiparticles. In addition, we obtained an equation which gives the critical coupling $\alpha$ in a system that leads to a so-called excitonic collapse. Our results indicate that this phenomenon should be omnipresent in this class of materials.

We acknowledge DOE grant DE-FG02-08ER46512, ONR grant MURI N00014-09-1-1063, and the NRF-CRP award “Novel 2D materials with tailored properties: beyond graphene” (R-144-000-295-281).
1 K. S. Novoselov et al., Science 306, 666 (2004).
2 K. S. Novoselov et al., Proc. Natl. Acad. Sci. U.S.A. 102, 10451 (2005).
3 A. H. Castro Neto and K. S. Novoselov, Rep. Prog. Phys. 74, 082501 (2011).
4 J. A. Wilson, F. J. DiSalvo, and S. Mahajan, Adv. Phys. 28, 117 (1975); Q.H. Wang et al., Nat. Nanotech. 7, 699 (2012); M. Chhowalla et al., Nat. Chem. 5, 263 (2013).
5 A. H. Castro Neto, Phys. Rev. Lett. 86, 4382 (2001).
6 M. S. Fuhrer and J. Hone, Nat. Nanotech. 8, 146 (2013).
7 K. F. Mak et al., Phys. Rev. Lett. 105, 136805 (2010).
8 Di Xiao et al., Phys. Rev. Lett. 108, 196802 (2012).
9 V. M. Pereira, J. Nilsson, and A. H. Castro Neto, Phys. Rev. Lett. 99, 166802 (2007); A. V. Shytov, M. I. Katsnelson, and L. S. Levitov, Phys. Rev. Lett. 99, 246802 (2007).
10 Y. Wang et al., Science, 2013 (DOI: 10.1126/science.1234320).
11 H. Rostami, A. G. Moghaddam, and R. Asgari, arXiv:1302.5901; A. Kormanyos et al., arXiv:1304.4084.
12 V. N. Kotov et al., Rev. Mod. Phys. 84, 1067 (2012).
13 J. Sabio, F. Sols, and F. Guinea, Phys. Rev. B 81, 045428 (2010).
14 V. M. Pereira, V. N. Kotov, and A. H. Castro Neto, Phys. Rev. B 78, 085101 (2008).