Carrier Plasmon Induced Nonlinear Band Gap Renormalization in Two-Dimensional Semiconductors

Yufeng Liang, Li Yang

Department of Physics, Washington University in St. Louis, St. Louis, MO 63130, USA

(Dated: August 26, 2014)

In reduced-dimensional semiconductors, doping-induced carrier plasmons can strongly couple with quasiparticle excitations, leading to a significant band gap renormalization. This effect has been long known and is essential for understanding transport and optical properties. However, the physical origin of this generic effect remains obscure. We develop a new plasmon-pole theory that efficiently and accurately capture this coupling. Using monolayer MoS$_2$ as a prototype two-dimensional (2D) semiconductor, we reveal a striking band gap renormalization around 400 meV and an unusual nonlinear evolution of its band gap with doping. This 2D prediction significantly differs from the linear behaviors that are common to one-dimensional structures. Our developed approach allows for a quantitative understanding of many-body interactions in general doped 2D semiconductors and paves the way for novel band gap engineering techniques.

PACS numbers:

The band gap is a defining property of semiconductors and it is typically not strongly influenced by extrinsic factors, such as doping. However, because of dramatically enhanced many-electron interactions, [1 2], doped one-dimensional (1D) structures exhibit an unexpectedly large band gap normalization (BGR) around several hundreds meV [3 4]. Recently, graphene-inspired two-dimensional (2D) semiconductors and their excited-state properties have garnered enormous interest [5 11]. Since doping is a common occurrence [11 14], understanding the effects of BGR is essential for interpreting experimental measurements, such as angle-resolved photoemission spectroscopy (ARPES) [13 16] and extracting exciton and trion [12 14] binding energies.

Beyond immediate practical applications, obtaining accurate quasiparticle (QP) energies and the corresponding band gap in doped reduced-dimensional semiconductors stands as a fundamental challenge. A particular difficulty is capturing the screening that are dominated by a unique low-energy acoustic carrier plasmon, [15 18]. Unlike in an undoped semiconductor, here the carrier plasmon strongly couples with QP excitations. This results in an enhanced BGR and can even lead to a satellite structure in the QP spectral function [13 21]. The QP-plasmon coupling also results in a nonlinear resonance profile in the self-energy [4 21] that complicates the solutions to the Dyson equation. In short, this subtle but important carrier plasmon calls for a special dynamical treatment of the dielectric screening that is beyond the scope of the widely used general plasmon pole (GPP) model [22], but is crucial for understanding electronic structure of general 2D semiconductors.

In this work, we focus on a material of broad interest, monolayer MoS$_2$ and study its BGR over a wide range of doping densities ($n_{2D}$) using the $GW$ approximation. We propose and implement a generic plasmon-pole model (PPM) approach that captures the essential screening effect and markedly improves the efficiency of many-body calculations. Our study reveals that the QP band gap of MoS$_2$ exhibits a strongly nonlinear evolution when varying the doping density; it drops sharply from 2.7 eV across low doping densities but nearly saturates at 2.3 eV for high densities. This is a consequence of the delicate interplay between carrier occupation and dielectric screening. Ultimately, beyond the one-shot $G_0W_0$ approximation [23], we show that inclusion of self-consistency is crucial for producing reliable QP energies in the presence of strong QP-plasmon coupling.

The quasiparticle self-energy can be obtained using the $GW$ approximation [24], i.e., $\Sigma = iGW$, where $G$ is the single-particle propagator and $W$ is the screened Coulomb interaction. For doped materials, the self-energy can be decomposed into four terms [4]

$$\Sigma = i(G_{\text{int}}W_{\text{int}} + \delta GW_{\text{int}} + G_{\text{int}}\delta W + \delta G\delta W)$$

(1)

where the subscript “int” denotes the operator of the intrinsic (undoped) system and the $\delta$ terms capture the full effects of the doping. The primary goal, then, is to find the variation in dielectric screening $\delta\epsilon^{-1}$ and hence $\delta W = \delta e^{-1}v$.

**Dielectric function.** For a suspended 2D crystal structure, the dielectric function $\epsilon$ can be calculated using the plane-wave representation [22]:

$$\epsilon_{\text{GPP}}(q, \omega) = \delta_{\text{GPP}} - v_{2D}(q + G)\chi_{\text{GPP}}(q, \omega),$$

where $v_{2D}(q)$ is the 2D-truncated Coulomb interaction [24 26] and the polarizability $\chi$ is obtained using the random phase approximation. For a semiconductor with a sizeable band gap, the dynamical matrix $\epsilon_{\text{GPP}}(q, \omega)$ is often described via the generalized plasmon-pole (GPP) model with a simple-pole function that is based on the static limit and $f$-sum rule [22].

A full-frequency calculation of the dielectric function may give the accurate band gap of doped 2D semiconduc-
Interestingly, the turning point for each density is located at several high-symmetry points of the first Brillouin zone (BZ). It can be seen that the intrinsic and doped cases are almost identical (<0.1%) except for some spectral features at low frequencies (inset of Fig. 1(c)). A refined calculation (Fig. 1(d)) provides clear evidence of a dispersive carrier plasmon, as manifested by the peak in the loss function.

The above calculation reveals a key fact about the dielectric function of a doped 2D system: the doping effects are exclusively concentrated in the head element $\epsilon^{-1}_{00}(q, \omega)$ at long wavelengths and low frequencies; this is where the carrier plasmon dominates. An overall picture is illustrated with schematics in Fig. 2. For an undoped semiconductor, only the optical plasmons arising from interband transitions are present, which can be represented by the single-pole function (Fig. 2(a)). For a doped 2D semiconductor, a branch of low-energy carrier acoustic plasmon emerges while the high-energy optical plasmon remains intact (Fig. 2(b)). The GPP model no longer accurately describes the dynamical effects in the system when this new plasmon emerges; it exaggerates the doping effects across a broader frequency region (Fig. 2(c)).

Motivated by the simple plasmon structure in Fig. 2(d), we model the variation of the head matrix element caused by doping as

$$\delta\epsilon^{-1}_{00}(q, \omega) = \frac{\Omega_d^2(q)}{\omega^2 - \omega_0^2(q)}$$  \hspace{1cm} (2)$$

where the parameter $\Omega_d(q)$ and $\omega_0(q)$ are the plasmon-pole strength and frequency, respectively. These are determined by the following two constraints. First, the plasmon energy $\omega_d(q)$ can be extracted from a frequency-dependent calculation. We find that $\omega_d(q)$ converges quickly with cutoff energy $E_G$ and occupied band number $N_c$. In our case, $E_G = 2 \text{ Ry}$ and $N_c = 4$ are sufficient for producing a converged $\omega_d(q)$ [27]. Second, the plasmon-

FIG. 1: (color online) (a) Static $\epsilon^{-1}_{GG'}(q, \omega = 0)$ with different combination of G-vectors. (b) Static $\epsilon^{-1}_{G0'}$ with at various doping density $n_{2D}$. (c) Loss function $\text{Im}[\epsilon^{-1}_{G0'}(q, \omega)]$ for wave vector $q$ some high-symmetry points. (d) Loss function at low energies featured by a shifting carrier-plasmon peak.

FIG. 2: (color online). Schematics of $\text{Re}[\epsilon^{00}_{G0}](q, \omega)$ produced by simple pole functions, as indicated by colormap plot on the $q - \omega$ plane. (a) Intrinsic case, (b) doped case, and (c) GPP model for doped case.
pole strength is given by $\Omega_{\text{CH}}^2(q) = -\delta e_{\text{CH}}(q, \omega = 0)\omega_{\text{CH}}^2(q)$, where $\delta^{-1}$ can be extracted from static calculations for the intrinsic and doped systems (Fig. 1(a)). Although a dense $\mathbf{k}$-grid ($120 \times 120 \times 1$) and a typical $E_{\mathbf{G}}$ (10Ry) are needed, only a few conduction bands ($N_c = 13$) are needed for convergence for calculations involving only small $q$.

As shown in Fig. 3, this PPM satisfactorily reproduces the frequency-dependence of the head $e_{\text{q,GG}}$ that are obtained from ab initio simulations, for various wave vector $q$. Our proposed calculation scheme focuses on the head dielectric function at small wave vectors and low frequencies, which circumvents the process of inverting the dielectric matrix $\varepsilon_{\text{GG}}$. Ultimately, this model is a far more efficient than the full-frequency scheme.

Self-Energy: We will calculate the self-energy of doped 2D semiconductors using the developed PPM. Following the COHSEX approximation, the self-energy $\Sigma_{\text{K}}$ can be split into a screened-exchange (SX) term and a Coulomb-hole (CH) term

$$\Sigma_{\text{SX/CH}}(E) = \sum_{n'<q,q',\text{GG}} \varepsilon_{n',G}^n \mathcal{G}_{\text{CH}}(k-q) K_{\text{SX/CH}}$$

where the SX and CH kernels are defined as $K_{\text{SX}} = -f_{n'k-q} W_{\mathbf{G}}(\mathbf{q}, E - \varepsilon_{n'k-q})$ and $K_{\text{CH}} = W_{\mathbf{G}}^+(\mathbf{q}, E - \varepsilon_{n'k-q})$, with $W^\pm(E) = \pm \frac{1}{2} \mathcal{P} \int_0^{\pm \infty} dE' \frac{\text{Im}W(E')}{E-E'}$, which only encompasses the positive (negative) poles in $W$. $\varepsilon_{n'G}^n(\mathbf{k}, \mathbf{q}) = M_{nn'}(\mathbf{k}, \mathbf{q}, G) M_{nn'}(\mathbf{k}, \mathbf{q}, G')$ represents the band structure effect, where $M_{nn'}(\mathbf{k}, \mathbf{q}, G)$ is the plane-wave matrix element. $\varepsilon_{nk}$ and $f_{nk}$ are the single-particle energy and occupation number of the state on band $n$ at $k$, respectively.

According to Eq. (4), we can obtain the total self-energy term by term. The calculation of $\Sigma_{\text{int}}$ and $\Sigma_3$ is straightforward. The next two terms, $\Sigma_2$ and $\Sigma_3$, are closely tied to the carrier screening $\delta W$, which can be reproduced using our proposed PPM in Eq. (4).

Finally, the energy dependence of the self-energy plays a crucial role in determining the QP energies. For a state $(n, k)$, both $\Sigma_{\text{int}}$ and $\Sigma_3$ vary slowly near the single-particle energy $\varepsilon_{nk}$ because the optical-plasmon feature occurs at high-energies. However, this is not the case for $\Sigma_2$ and $\Sigma_3$, on account of the emergence of the low-energy carrier plasmon. Instead, they exhibit strongly nonlinear behaviors near the $\varepsilon_{nk}$. Fig. 4 displays $\Sigma_2$ and $\Sigma_3$ for the valence band maximum (VBM) and conduction band minimum (CBM) of n-doped monolayer MoS$_2$ with $\omega = E - \varepsilon_{\text{VBM}}$ and $\omega = E - \varepsilon_{\text{CBM}}$, respectively. Thus $\omega = 0$ defines the on-shell energy. For simplicity, spin-orbital coupling is not considered here. In Fig. 4(a), $\Sigma_{\text{VBM}}$ exhibits typical feature of Fano resonance, resulting from the coupling of a quasi-electron with the dispersive carrier plasmon. With increasing carrier density $n_{2D}$, the resonance peak position systematically shifts left due to a blue shift in the plasmon energy. The case of $\Sigma_{\text{VBM}}$ is the reverse (Fig. 4(b)) because it corresponds to a quasi-hole state. Fig. 4(c) shows the energy dependence of $\Sigma_3$. For electron doping, while $\Sigma_{\text{VBM}}$ is $\approx 0$, $\Sigma_{\text{CBM}}$ rises as an asymmetric finite-width plateau. Finally, the total contribution of $\Sigma_{\text{CBM}} + \Sigma_{\text{CBM}}$ is shown in Fig. 4(d), which features a transition from a hole-like resonance to an electron-like resonance.

With all of the self-energy contributions taken under consideration, the QP band energies $E$ can be obtained by solving the Dyson equation $\Sigma_{\text{int}}(E_{\text{QP}}) = E_{\text{QP}} - V_{\text{xc}}^\text{PPM} + \text{Re}[\Sigma_{\text{K}}(E_{\text{QP}})]$ where $V_{\text{xc}}$ is the exchange-correlation potential from the DFT. Although the $G_0W_0$ approximation works well for intrinsic semiconductors, it is no longer appropriate for treating a doped 2D system, for which the self-energy becomes strongly nonlinear. This
FIG. 4: (color online) Self-energy contributions \( \Sigma_2 = iG_{\text{int}} \delta W \) and \( \Sigma_3 = i\delta G \delta W \) for both VBM and CBM at different doping levels. \( \omega = 0 \) is set at the on-shell energy. The arrows in (a) and (b) are used for guiding readers’ eyes for the evolution of those resonant peaks under higher doping densities.

is manifested by \( \Sigma_2 \) and \( \Sigma_3 \). This nonlinear profile causes the QP solutions to depend sensitively on the preliminary DFT calculation. Even worse, now there can be multiple solutions. One has to determine which is the most important one by considering the spectral weight, which also relies on the DFT starting-point. This has exacerbates the inherent deficiency of the \( G_0 W_0 \) approximation. One needs to proceed with a self-consistent GW calculation [33, 34]. Fortunately, we find that the carrier-plasmon resonance profile depends weakly on the band curvature [27]. This suggests that the self-consistency can be achieved by rigidly shifting [35] the whole resonance profile along energy-axis such that the on-shell energy of \( \Sigma \) coincides with the QP solution. This procedure is equivalent to performing the \( GW_0 \) approximation [26, 36].

Fig. 5 displays the evolution of the BGR and the different contributing mechanisms versus doping density \( n_{2D} \) in monolayer MoS\(_2\). Remarkably, the QP band gap is strongly renormalized in the light-doping regime; it drops dramatically from 2.73 eV to around 2.25 eV as \( n_{2D} \) increases from 0 to \( 10^{13} \text{cm}^{-2} \) (Fig. 5(a)). The origin of this enhanced band-gap shrinkage is due to the following mechanisms. First, a major contribution comes from the carrier-occupation energy, \( \Sigma_1 \), is particularly large. Within the largest \( n_{2D} \) in our simulation, it can reach up to \( \sim 500 \text{meV} \). Similar magnitudes of carrier-occupation energy have also been reported in doped semiconducting carbon nanotubes (CNTs) [3, 4]. Another important band-gap shrinkage mechanism comes from the carrier screening \( \Sigma_2 = iG_{\text{int}} \delta W \). It weakens the electron-electron interaction and hence leads to a significant reduction in QP band gap correction up to a few hundred meV (see Fig. 5(b) and (c)).

As the doping density \( n_{2D} \) increases, the QP band gap saturates at 2.25eV eventually. This results from several factors. When \( n_{2D} \) is high enough, the extra carrier screening can in turn reduce the carrier occupation energy. At high \( n_{2D} \), the double-difference term \( \Sigma_3 = i\delta G \delta W \) becomes a prominent contribution (Fig. 5(b)). For n-doping, it scales as \( -\delta \epsilon^{-1} k_F \). Therefore, \( \Sigma_3 \) varies in an opposite trend to \( \Sigma_1 \) due to this minus sign and is responsible for enlarging the band gap. Furthermore, the carrier screening effect \( \Sigma_2 \) itself also exhibits a saturation behavior (Fig. 5(b) and (c)). With increasing doping density, the carrier plasmon blueshifts significantly and its effect on the QP state becomes weaker, as evidenced from the departing Fano feature that is indicated by arrows in Fig 5(a) and (b). Finally, dimensionality effects also plays a role in the band gap evolution. Our case differs substantially from that in CNTs, for which the band gap shrinks almost linearly with the doping density [3]. In the 1D case, some major self-energy contributions, like \( \Sigma_1 \), still scale as \( -k_F \), but which is proportional to the doping density.

For comparison, we calculate the band gap via the conventional GPP method for \( n_{2D} = 3.4 \times 10^{13} \) and
$5.7 \times 10^{13}$ cm$^{-2}$. As shown in Fig. 3(a), the GPP model overestimates the band gap reduction by $\approx 200$ meV, which is 40% larger than in our refined PPM method.

In conclusion, we have developed and applied a highly efficient generic computing scheme for calculating the dielectric function of doped 2D semiconductors. From it, we obtained the enhanced and nonlinear reduction of the QP band gap over a wide range of doping densities. This has not been observed in one-dimensional and bulk structures. We also found that the band gap drops to a certain limit at sufficiently high doping densities because of a delicate competition between the exchange and correlation energies. This enhanced BGR is crucial for explaining the excitonic effects and trions observed in experiments and it can be directly observed by a combination of the APRES and inverse ARPES spectroscopies.

We acknowledge valuable discussions with Giovanni Vignale and Ryan Soklaski. This work is supported by NSF DMR-1207141. The ground-state calculation is performed with Quantum Espresso [38]. The intrinsic self-energies are obtained by the BerkeleyGW code [28]. The computational resources have been provided by Lonestar of Teragrid at the Texas Advanced Computing Center (TACC).

[1] C. D. Spataru, S. Ismail-Beigi, L. X. Benedict, and S. G. Louie, Physical Review Letters 92, 077402 (2004).
[2] L. Yang, C.-H. Park, Y.-W. Son, M. L. Cohen, and S. G. Louie, Physical Review Letters 99, 186801 (2007).
[3] C. D. Spataru and F. Léonard, Physical review letters 104, 177402 (2010).
[4] C. D. Spataru and F. Léonard, Chemical Physics 413, 81 (2013).
[5] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Physical Review Letters 105, 136805 (2010).
[6] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, Nano letters 10, 1271 (2010).
[7] K. F. Mak, K. He, J. Shan, and T. F. Heinz, Nature nanotechnology 7, 494 (2012).
[8] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nature nanotechnology 7, 699 (2012).
[9] M. Chhowalla, H. S. Shin, G. Eda, L.-J. Li, K. P. Loh, and H. Zhang, Nature chemistry 5, 263 (2013).
[10] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao, Physical Review Letters 108, 196802 (2012).
[11] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacomelli, and A. Kis, Nature nanotechnology 6, 147 (2011).
[12] K. F. Mak, K. He, C. Lee, G. H. Lee, J. Hone, T. F. Heinz, and J. Shan, Nature materials 12, 207 (2013).
[13] B. Radisavljevic and A. Kis, Nature materials 12, 815 (2013).
[14] S. Mouri, Y. Miyauchi, and K. Matsuda, Nano letters 13, 5944 (2013).
[15] W. Jin, P.-C. Yeh, N. Zaki, D. Zhang, J. T. Sadowski, A. Al-Mahboob, A. M. van der Zande, D. A. Chemet, J. I. Dadap, I. P. Herman, et al., Physical Review Letters 111, 106801 (2013).
[16] T. Eknaphakul, P. D. King, M. Asakawa, P. Buaphet, R.-H. He, S.-K. Mo, H. Takagi, K. M. Shen, F. Baumberger, T. Sasagawa, et al., Nano letters 14, 1312 (2014).
[17] T. Ando, A. B. Fowler, and F. Stern, Reviews of Modern Physics 54, 437 (1982).
[18] S. D. Sarma and E. Hwang, Physical Review B 54, 1936 (1996).
[19] M. Guzzo, G. Lani, F. Sottile, P. Romaniello, M. Gatti, J. J. Kas, J. J. Rehr, M. G. Silly, F. Sirotti, and L. Rein- ing, Physical review letters 107, 166401 (2011).
[20] J. Lischner, D. Vigil-Fowler, and S. G. Louie, Physical Review Letters 110, 146801 (2013).
[21] J. Lischner, D. Vigil-Fowler, and S. G. Louie, Physical Review B 89, 125430 (2014).
[22] M. S. Hybertsen and S. G. Louie, Physical Review B 34, 5390 (1986).
[23] A. Steinhoff, M. Rsner, F. Jahnke, T. O. Wehling, and C. Gies, Nano Letters 0, null (2014), http://pubs.acs.org/doi/pdf/10.1021/nl500595u, URL http://pubs.acs.org/doi/abs/10.1021/nl500595u.
[24] L. Hedin, Physical Review 139, A796 (1965).
[25] C. A. Rozzi, D. Varsano, A. Marini, E. K. Gross, and A. Rubio, Physical Review B 73, 205119 (2006).
[26] S. Ismail-Beigi, Physical Review B 73, 235103 (2006).
[27] See Supplemental Material (0).
[28] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, Computer Physics Communications 183, 1269 (2012).
[29] G. Giuliani, Quantum theory of the electron liquid (Cambridge University Press, 2005).
[30] A. Oschlies, R. Godby, and R. Needs, Physical Review B 51, 1527 (1995).
[31] D. Y. Qiu, H. Felipe, and S. G. Louie, Physical review letters 111, 216805 (2013).
[32] Y. Liang, S. Huang, R. Soklaski, and L. Yang, Applied Physics Letters 103, 042106 (2013).
[33] F. Bruneval, N. Vast, and L. Reinig, Physical Review B 74, 045102 (2006).
[34] M. van Schilfgaarde, T. Kotani, and S. Faleev, Physical review letters 96, 226402 (2006).
[35] L. Hedin, Journal of Physics: Condensed Matter 11, R489 (1999).
[36] S. D. Sarma and E. Hwang, Physical review B 87, 045425 (2013).
[37] L. Yang, J. Deslippe, C.-H. Park, M. L. Cohen, and S. G. Louie, Physical review letters 103, 186802 (2009).
[38] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, et al., Journal of Physics: Condensed Matter 2013, 395502 (2009).