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

Yufeng Liang and Li Yang

Phys. Rev. Lett. 114, 063001 — Published 10 February 2015
DOI: 10.1103/PhysRevLett.114.063001
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: January 16, 2015)

In reduced-dimensional semiconductors, doping-induced carrier plasmons can strongly couple with quasiparticle excitations, leading to a significant band gap renormalization. However, the physical origin of this generic effect remains obscure. We develop a new plasmon-pole theory that efficiently and accurately captures this coupling. Using monolayer MoS$_2$ and MoSe$_2$ as prototype two-dimensional (2D) semiconductors, we reveal a striking band gap renormalization above 400 meV and an unusual nonlinear evolution of their band gaps with doping. This prediction significantly differs from the linear behavior that is observed in one-dimensional structures. Notably, our predicted band gap renormalization for MoSe$_2$ is in excellent agreement with recent experimental results. 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], doping in one-dimensional (1D) structures can induce an unexpectedly large band gap normalization (BGR) of around several hundreds meV [3, 4]. Recently, graphene-inspired two-dimensional (2D) semiconductors and their excited-state properties have garnered enormous interest [5–10]. 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) [15–17] 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 [18, 19]. Unlike in an undoped semiconductor, here the carrier plasmon strongly couples with QP excitations. This can result in a nonlinear resonance profile in the self-energy [4, 22] that complicates the solutions to the Dyson equation. In other words, 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 [23] but is crucial for understanding electronic structure of general 2D semiconductors.

In this work, we focus on two materials of broad interest, monolayer MoS$_2$ and MoSe$_2$ and study their 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 by virtue of pure many body effects the QP band gap of MoS$_2$ exhibits a strongly nonlinear evolution when varying the doping density: it drops sharply from 2.66 eV across low doping densities but nearly saturates at 2.18 eV for high densities. This is a consequence of the delicate interplay between carrier occupation and dielectric screening. For doped MoSe$_2$, our $ab$ initio calculation gives a band gap in good agreement with recent measurements.

The QP self-energy can be obtained using the GW approximation [25], 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]

$$\Sigma = i(G_{\text{int}}W_{\text{int}} + \delta G_{\text{int}} + G_{\text{int}}\delta W + \delta G\delta W) = i\left(\Sigma_{\text{int}} + \Sigma_1 + \Sigma_2 + \Sigma_3\right)$$

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 $\delta W = \delta\epsilon^{-1}v$.

Dielectric function. For a 2D crystal, the dielectric function $\epsilon$ can be calculated using the plane-wave representation [23]:

$$\epsilon_{\text{G}}(q, \omega) = \delta_{\text{G}} - \nu_{2D}(q+G)\chi\epsilon_{\text{G}}(q, \omega)$$

where $\nu_{2D}(q)$ is the 2D-truncated Coulomb interaction [26, 27] and the polarizability $\chi$ is obtained using the random phase approximation. For a semiconductor with a sizeable band gap, the dynamical matrix $\epsilon_{\text{G}}^{-1}(q, \omega)$ 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 [23].

A full-frequency calculation of the dielectric function may give the accurate band gap of doped 2D semiconductors. However, mimicking experimentally-accessible doping densities requires one to use an ultra fine sampling
scheme over the $k$-$\omega$ space to capture the carrier screening, making the simulation formidable for converged results. We are thus motivated to develop an efficient and accurate $ab$ initio model.

Our model begins with the static $\varepsilon^{-1}_{GG'}(q, \omega = 0)$ of doped monolayer MoS$_2$. Fig. 1 (a) compares several representative dielectric matrix elements of the undoped system and the doped one at $n_{2D} = 3.4 \times 10^{13}$ cm$^{-2}$. Given the isotropy about small $q$ [28], the elements are only plotted along a single reciprocal primitive vector $b_1$. Remarkably, all of the matrix elements remain nearly unaffected (difference $< 0.01$) even at this high doping level, except the “head” matrix element $\varepsilon^{-1}_{00}(q, 0)$. In the undoped case, $\varepsilon^{-1}_{00}(q, 0)$ approaches 1 [27, 29] as $q \to 0$, reflecting the absence of screening in a 2D semiconductor at long-wavelengths. However, including doping causes $\varepsilon^{-1}_{00}(q, 0)$ immediately drops to 0 [29], due to metallic screening. This result leads us to focus primarily on the variation of the head matrix element $\delta\varepsilon^{-1}_{00}$.

$\varepsilon^{-1}_{00}(q, 0)$ is shown in Fig. 1 (b) at various doping densities $n_{2D}$. For increasing $q$, all of the dielectric functions $\varepsilon^{-1}_{00}(q, 0)$ at finite $n_{2D}$ first grow up linearly at the same rate for small $q$, but then individually turn up towards, and merge with, the intrinsic $\varepsilon^{-1}_{00}(q, 0)$. This behavior is in accord with the 2D free electron gas (FEG) polarizability [30]. $\delta\chi_{00}(q) = -\frac{\varepsilon_{0}^{-1}}{2\pi i} [1 - \theta(q - 2k_F) \sqrt{1 - 4k_F^2/q^2}],$ for which the static doping effect only dominates at small $q$ and damps away rapidly beyond $2k_F$. The rapid damping profile suggests that the 2D doping effects are predominantly localized at the long-wavelength limit, due to the low-doping densities that commonly occur and the resulting large screening length.

Next, we turn to dynamical screening effects to treat the carrier plasmon. Fig. 1 (c) presents the frequency-dependence of the loss function $\text{Im}[\varepsilon^{-1}_{00}(q, \omega)]$ for wave vectors of several high-symmetry $q$ (see details in [28]) with a 0.4-eV smearing. (d) Loss function at low energies featured by a shifting carrier-plasmon peak with a 0.03-eV smearing.

FIG. 1: (color online) (a) Static $\varepsilon^{-1}_{GG'}$ with different combination of $G$-vectors. (b) Static $\varepsilon^{-1}_{00}$ with at various doping density $n_{2D}$. $b_1$ is the reciprocal vector along the $\Gamma - M$ direction of the first Brillouin zone. (c) Loss function $\text{Im}[\varepsilon^{-1}_{00}(q, \omega)]$ for wave vectors of several high-symmetry $q$ (see details in [28]) with a 0.4-eV smearing. (d) Loss function at low energies featured by a shifting carrier-plasmon peak with a 0.03-eV smearing.

FIG. 2: (color online). Schematics of $\text{Re}[\varepsilon_{00}^2(q, \omega)]$ 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.

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 $\varepsilon^{-1}_{00}(q, \omega)$ 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 in low-frequency region across a broader frequency region due to the sum rule [23] (Fig. 2(c)).

Motivated by the simple plasmon structure in Fig. 1 (d), we model the variation of the head matrix element caused by doping as

$$\delta\varepsilon^{-1}_{00}(q, \omega) = \frac{\Omega_{d}^2(q)}{\omega^2 - \omega_{d}^2(q)}$$

where the parameter $\Omega_d(q)$ and $\omega_d(q)$ 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
where the SX and CH kernels are defined as obtained from full \textit{ab initio} simulations, and our proposed PPM.

\(N_c\). In our case, \(E_G = 2\) Ry and \(N_c = 4\) are sufficient for producing a converged \(\omega_d(q)\) [28]. The second, the plasmon-pole strength is given by \(\Omega^2_d(q) = -\delta\epsilon_{00}^{-1}(q, \omega = 0)\omega_d^2(q)\), where \(\delta\epsilon^{-1}\) can be extracted from static calculations for the intrinsic and doped systems (Fig. 1(a)). Although a dense \(k\)-grid \((120 \times 120 \times 1)\) and a typical Eq. (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 \(\epsilon^{-1}\) that are obtained from \textit{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 [31]. Ultimately, this model is a far more efficient than the full-frequency scheme.

\textbf{Self-Energy:} We will calculate the self-energy of doped 2D semiconductors using the developed PPM. Following the COHSEX approximation [23, 25], the self-energy \(\Sigma^k\) can be split into a screened-exchange (SX) term and a Coulomb-hole (CH) term

\[
\Sigma^k_{\text{SX/CH}}(E) = \sum_{n', q, G, G'} \xi_{n', G', G} n_{n'}(q, k, q, G) M_{n'n}(q, k, G') \delta W_{n'n}(E),
\]

where the SX and CH kernels are defined as \(K_{\text{SX}} = -f_{n'k'-q}W_{G'G}(q, E - \epsilon_{n'k'-q})\) and \(K_{\text{CH}} = W_{GG'}(q, E - \epsilon_{n'k'-q})\), with \(W^\pm(E) = \pm \frac{1}{2} P \int_0^{\pm \infty} dE' \frac{\Im W(E)}{E - E'},\) which only encompasses the positive (negative) poles in \(W\). \(\epsilon_{n'n'}_{G'G'}(k, q, G)\) represents the band structure effect, where \(M_{n'n}(k, q, G)\) is the plane-wave matrix element[23, 29]. \(\epsilon_{n'k'-q} f_{n'k-q}\) is the single-particle energy and occupation number of the state on band \(n'\) at \(k\), respectively.

According to Eq. (1), we can obtain the total self-energy term by term. The calculation of \(\Sigma_{\text{int}}\) and \(\Sigma_1\) is straightforward [28, 32, 33]. 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. (2). Due to the aforementioned properties of the carrier plasmon, these self-energy contributions can be simplified dramatically by (i) limiting the band summation to \(n' = n\), (ii) retaining only the term with \(G = G' = 0\), and (iii) setting up a cutoff \(q_c\), for the BZ integration.

\[
\Sigma_2 = iG_{\text{int}} \delta W \text{ does not involve partial band filling, it reads}
\]

\[
\Sigma^k_{\text{2}}(E) \approx \pm \int_{q < q_c} \frac{d^2q}{(2\pi)^2} \epsilon_{n0} n(k, q) \delta W_{n0}(q, E - \epsilon_{n-k-q})
\]

where \(\delta W_{n0}(q, \omega) = \frac{\Omega^2_d(q)}{2\omega_d(q)}(\omega + \omega_d(q)) v_{2D}(q)\) and \(\pm\) is for conduction/valence states. For fully filled valence bands, the SX and CH term have been combined via the COHSEX approximation [23, 29]. The energy dependence of the self-energy plays a crucial role in determining the QP energies. For a state \(nk\), both \(\Sigma_{\text{int}}\) and \(\Sigma_1\) vary slowly near the single-particle energy \(\epsilon_{nk}\) because the optical-plasmon feature occurs at high-energies [28]. 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 \(\epsilon_{nk}\). 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 \(\omega = E - \epsilon_{VBM}\) and \(\omega = E - \epsilon_{CBM}\), respectively, where \(\omega = 0\) defines the on-shell energy. In Fig. 4 (a), \(\Sigma^2_{\text{CBM}}\) 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 \(\Sigma^3_{\text{CBM}}\) 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 \(\Sigma^2_{\text{CBM}} \approx 0\), \(\Sigma^3_{\text{CBM}}\) rises as an asymmetric finite-width plateau. Finally, the total contribution of \(\Sigma^2_{\text{CBM}} + \Sigma^3_{\text{CBM}}\) is shown in Fig. 4(d), which features a transition from a hole-like resonance to an electron-like resonance.

Although the \(G_0W_0\) approximation is reliable for many intrinsic semiconductors, it is no longer appropriate for treating a doped 2D system, for which the self-energy becomes strongly nonlinear. This is manifested by \(\Sigma_2\) and \(\Sigma_3\). This nonlinear profile causes the QP solutions to depend sensitively on the entry DFT calculation. One needs to proceed with a self-consistent GW calculation.
The doping density. The sole effects of $\Sigma_2 = iG_{\text{int}} \delta W$ and $\Sigma_3 = iG_{\text{SO}} \delta W$ 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.

As the doping density $n_{2D}$ increases, the QP band gap slowly saturates at 2.18 eV. This phenomenon 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 $\Sigma_3 = iG_{\text{int}} \delta W$ becomes a prominent contribution (Fig. 5 (b)). For $n$-doping, it scales as $-\epsilon_{\text{int}}^{-1} k_F$. 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 4 (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. Meanwhile, the conventional GPP method [23] significantly overestimate the BGR, as shown in Fig. 5 (a).

A recent ARPES experiments [17] that directly accesses the band gap of doped monolayer MoSe$_2$ can serve to justify our ab initio approach. At the experimental...
doping level \((5.3 \times 10^{13} \text{cm}^{-2})\) \([17]\), our calculated QP band gap reduction is 0.45eV. Additionally, this doping density induces a lattice dilation of 0.6% that causes an extra QP band gap reduction of 0.13eV. Finally, by including a SO splitting of \(\Delta_{SO} = 180\text{meV}\) \([17]\) and a band gap reduction of 0.13eV from substrate screening \([40]\), we obtain a band gap of 1.59eV, which is in excellent agreement with the measured value of 1.58eV.

In conclusion, we have developed and applied a highly efficient generic computing scheme for calculating the dielectric function of doped 2D semiconductors. We thereby obtained the enhanced and nonlinear reduction of the QP band gap over a wide range of doping densities unique to 2D systems. 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. Our \textit{ab initio} prediction is in close agreement with recent measurements. This enhanced BGR is crucial for explaining the excitonic effects and trions observed in experiments.

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 \([39]\). The intrinsic self-energies are obtained by the BerkeleyGW code \([29]\). The computational resources have been provided by Lonestar of Teragrid at the Texas Advanced Computing Center (TACC).

---