

Fast exciton annihilation by capture of electrons or holes by defects via Auger scattering in monolayer metal dichalcogenides

Haining Wang, Jared H. Strait, Changjian Zhang, Weimin Chan, Christina Manolatou, Sandip Tiwari, and Farhan Rana*

School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA

(Received 8 December 2014; revised manuscript received 10 March 2015; published 13 April 2015)

The strong Coulomb interactions and the small exciton radii in two-dimensional metal dichalcogenides can result in very fast capture of electrons and holes of excitons by mid-gap defects from Auger processes. In the Auger processes considered here, an exciton is annihilated at a defect site with the capture of the electron (or the hole) by the defect and the hole (or the electron) is scattered to a high energy. In the case of excitons, the probability of finding an electron and a hole near each other is enhanced many folds compared to the case of free uncorrelated electrons and holes. Consequently, the rate of carrier capture by defects from Auger scattering for excitons in metal dichalcogenides can be 100–1000 times larger than for uncorrelated electrons and holes for carrier densities in the 10^{11} – 10^{12} cm $^{-2}$ range. We calculate the capture times of electrons and holes by defects and show that the capture times can be in the subpicosecond to a few picoseconds range. The capture rates exhibit linear as well as quadratic dependence on the exciton density. These fast time scales agree well with the recent experimental observations and point to the importance of controlling defects in metal dichalcogenides for optoelectronic applications.

DOI: [10.1103/PhysRevB.91.165411](https://doi.org/10.1103/PhysRevB.91.165411)

PACS number(s): 72.20.Jv, 71.35.–y, 79.20.Fv, 81.07.–b

I. INTRODUCTION

Many body interactions play an important role in determining the electronic and optoelectronic properties of two-dimensional (2D) transition metal dichalcogenides (TMDs). The exciton binding energies in 2D chalcogenides are almost an order of magnitude larger compared to other bulk semiconductors [1–5]. The strong Coulomb interactions and small exciton radii in 2D-TMDs result in large optical oscillator strengths [3,6,7] and short radiative lifetimes [8]. In this paper we show that the same factors also result in very fast capture of electrons and holes of excitons by defects from Auger processes leading to fast nonradiative recombination rates. The basic idea can be understood as follows. Consider the Auger process in which a hole (in the valence band) scatters off an electron (in the conduction band) and is captured by a mid-gap defect level and the electron (in the conduction band) takes the energy released in the hole capture process. In the case of uncorrelated electrons and holes, the rate for this process is proportional to the product of the hole density p and the probability of finding an electron near the hole, which is proportional to the electron density n . But in the case of tightly bound excitons, an electron is present near the hole with a very high probability proportional to $|\phi(\vec{r}=0)|^2$, where $\phi(\vec{r})$ is related to the exciton wave function (see the discussion below). Therefore, the rate for a hole (or an electron) in a tightly bound exciton to get captured by a defect is proportional to the exciton density times $|\phi(\vec{r}=0)|^2$. Generally speaking, Auger rates in semiconductors are considered to be important only at large carrier densities [9]. But given the small exciton radii in 2D-TMDs (in the 7–10 Å range), $|\phi(\vec{r}=0)|^2$, which is inversely proportional to the square of the exciton radius, can be extremely large and, consequently, Auger capture rates in 2D-TMDs can be very fast. Compared to the rates for direct

electron-hole recombination via interband Auger scattering, which can be limited by the orthogonality of the conduction and valence band Bloch states, the rates for the capture of electrons and holes of excitons by defects can be very fast when the defect states have a good overlap with the conduction or valence band Bloch states.

Quantum efficiencies of TMD light emitters and detectors that have been reported are extremely poor; in the 0.0001–0.01 range [10–14]. Similar quantum efficiencies for TMDs have been observed in photoluminescence experiments [1,15,16]. Therefore, most of the electrons and holes injected electrically or optically in TMDs recombine nonradiatively. Given that the average radiative lifetimes of excitons in TMDs are in the range of hundreds of picoseconds to a few nanoseconds [8], the nonradiative recombination or capture times in TMDs are expected to be of the order of a few picoseconds. Several experimental results on the ultrafast carrier dynamics in photoexcited monolayer MoS $_2$ do indeed point to nonradiative recombination and/or capture times in the few picoseconds range [15,17–19]. The mechanisms by which electrons and holes recombine nonradiatively and/or are captured by defects, and the associated time scales, remain to be clarified. The results in this paper show that electrons and holes of excitons in TMDs can get captured by defects on very short time scales that are in the subpicosecond to a few picoseconds range resulting in exciton annihilation. The capture rates exhibit linear as well as quadratic dependence on the exciton density (defect-assisted exciton-exciton annihilation). The quadratic dependence of the exciton annihilation rate on the exciton density is generally considered to be an exclusive characteristic of exciton-exciton annihilation processes via direct interband Auger scattering. Although the discussion in this paper focuses on monolayer MoS $_2$, the analysis and the results presented here are expected to be relevant to all 2D-TMDs, and are expected to be useful in designing metal dichalcogenide optoelectronic devices as well as in helping to understand and interpret experimental data [15,17–19].

*fr37@cornell.edu

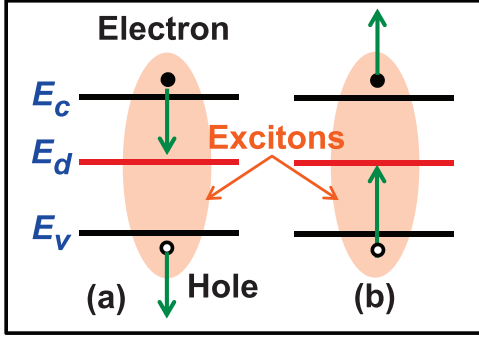


FIG. 1. (Color online) Two basic Auger processes for the capture of an electron (a) or a hole (b) of an exciton by a defect state are depicted [9,20].

II. THEORETICAL MODEL

A. Introduction

The two basic Auger processes for the capture of an electron (a) or a hole (b) of an exciton by a defect state are depicted in Fig. 1. Proper partitioning of the Hamiltonian is important in order to compute the rates of these processes. We discuss the terms in the Hamiltonian describing various processes below.

B. The noninteracting Hamiltonian

The crystal structure of a monolayer of group-VI dichalcogenides MX_2 (e.g., $M = \text{Mo, W}$ and $X = \text{S, Se}$) consist of $X-M-X$ layers, and within each layer the M atoms (or the X atoms) form a 2D hexagonal lattice. Each M atom is surrounded by six nearest neighbor X atoms in a trigonal prismatic geometry with D_{3h}^1 symmetry. The valence band maxima and conduction band minima occur at the K and K' points in the Brillouin zone. Most of the weight in the conduction and valence band Bloch states near the K and K' points resides on the d orbitals of M atoms [4,21,22]. The spin-up and -down valence bands are split near the K and K' points by 0.1–0.2 eV due to the spin-orbit coupling [4,21–23]. In comparison, the spin-orbit-coupling effects in the conduction band are much smaller [23]. Assuming only d orbitals for the conduction and valence band states, and including spin-orbit coupling, one obtains the following simple spin-dependent tight-binding Hamiltonian (in matrix form) near the $K(K')$ points [21],

$$\begin{bmatrix} \Delta/2 & \hbar v k_- \\ \hbar v k_+ & -\Delta/2 + \lambda \tau \sigma \end{bmatrix}. \quad (1)$$

Here, Δ is related to the material bandgap, $\sigma = \pm 1$ stands for the electron spin, $\tau = \pm 1$ stands for the K and K' valleys, 2λ is the splitting of the valence band due to spin-orbit coupling, $k_{\pm} = \tau k_x \pm i k_y$, and the velocity parameter v is related to the coupling between the orbitals on neighboring M atoms. From density functional theories [22,24], $v \approx 5 - 6 \times 10^5$ m/s. The wave vectors are measured from the $K(K')$ points. The d orbital basis used in writing the above Hamiltonian are $|d_{z^2}\rangle$ and $(|d_{x^2-y^2}\rangle + i\tau|d_{xy}\rangle)/\sqrt{2}$ [21]. We will use the symbol s for the combined valley (τ) and spin (σ) degrees of freedom. Defining Δ_s as $\Delta - \lambda \tau \sigma$, the energies and eigenvectors of the

conduction and valence bands are [21,25]

$$E_{c,s}(\vec{k}) = \frac{\lambda \tau \sigma}{2} + \gamma \sqrt{(\Delta_s/2)^2 + (\hbar v k)^2}, \quad (2)$$

$$|v_{c,\vec{k},s}\rangle = \begin{bmatrix} \cos(\theta_{\gamma,\vec{k},s}/2) e^{-i\tau\phi_{\vec{k}}/2} \\ \tau \gamma \sin(\theta_{\gamma,\vec{k},s}/2) e^{i\tau\phi_{\vec{k}}/2} \end{bmatrix}. \quad (3)$$

Here, $\gamma = 1$ (or -1) stands for the conduction (or the valence) band, $\phi_{\vec{k}}$ is the phase of the wave vector \vec{k} , and

$$\cos(\theta_{\gamma,\vec{k},s}) = \gamma \frac{\Delta_s}{2\sqrt{(\Delta_s/2)^2 + (\hbar v k)^2}}. \quad (4)$$

Near the conduction band minima and valence band maxima, the band energy dispersion is parabolic with well-defined effective masses, m_e and m_h , for electrons and holes, respectively.

The Hamiltonian describing electron states in the conduction band, valence band, and a mid-gap defect state is

$$H_o = \sum_{\vec{k},s} E_{c,s}(\vec{k}) c_{\vec{k},s}^\dagger c_{\vec{k},s} + \sum_{\vec{k},s} E_{v,s}(\vec{k}) b_{\vec{k},s}^\dagger b_{\vec{k},s} + \sum_{\sigma} E_d d_{\sigma}^\dagger d_{\sigma}. \quad (5)$$

Here, $c_{\vec{k},s}$, $b_{\vec{k},s}$, and d_{σ} are the destruction operators for the conduction band, valence band, and defect states, respectively. The bandgap is $E_{g,s,s'} = E_{c,s}(\vec{k}=0) - E_{v,s'}(\vec{k}=0)$. Since only the smallest bandgap will be relevant in the discussion that follows, we will drop the spin/valley indices from $E_{g,s,s'}$ for simplicity.

C. Electron-hole interaction and exciton states

The Coulomb interaction between the electrons and holes can be included by adding the following term to the Hamiltonian,

$$H_{eh} = \frac{1}{A} \sum_{\vec{k},\vec{k}',\vec{q},s,s'} V(\vec{q}) F_{s,s'}(\vec{k},\vec{k}',\vec{q}) c_{\vec{k}+\vec{q},s}^\dagger c_{\vec{k}-\vec{q},s'}^\dagger b_{\vec{k}',s'} b_{\vec{k},s}. \quad (6)$$

$V(\vec{q})$ is the 2D Coulomb potential and equals $e^2/2\epsilon_o\epsilon(\vec{q})q$. The wave-vector-dependent dielectric constant $\epsilon(\vec{q})$ for monolayer MoS_2 is given by Zhang *et al.* [3] and Berkelbach *et al.* [4]. $F_{s,s'}(\vec{k},\vec{k}',\vec{q})$ is [25]

$$F_{s,s'}(\vec{k},\vec{k}',\vec{q}) = \langle v_{c,\vec{k}+\vec{q},s} | v_{c,\vec{k},s} \rangle \langle v_{v,\vec{k}'-\vec{q},s'} | v_{v,\vec{k}',s'} \rangle. \quad (7)$$

Near the conduction band minima, where $\hbar v k \ll \Delta_s$, $\cos(\theta_{\gamma,\vec{k},s}) \approx 1$, and $\sin(\theta_{\gamma,\vec{k},s}) \ll 1$. Similarly, near the valence band maxima, $\sin(\theta_{\gamma,\vec{k},s}) \approx 1$ and $\cos(\theta_{\gamma,\vec{k},s}) \ll 1$. Therefore, for wave vectors near the band extrema one can make the approximation [25],

$$F_{s,s'}(\vec{k},\vec{k}',\vec{q}) = e^{i(\tau\phi_{\vec{k}+\vec{q}} - \tau\phi_{\vec{k}} + \tau'\phi_{\vec{k}'} - \tau'\phi_{\vec{k}'-\vec{q}})/2}. \quad (8)$$

Exciton states are approximate eigenstates of the Hamiltonian $H_o + H_{eh}$. Assuming that the ground state of the semiconductor is $|\psi_o\rangle$, which consists of a filled valence band and an empty conduction band, an exciton state with in-plane momentum \vec{Q}

can be constructed from the ground state as follows [3,25],

$$|\psi_{s,s',\alpha}(\vec{Q})\rangle = \frac{1}{\sqrt{A}} \sum_{\vec{k}} \psi_{\alpha,\vec{Q}}(\vec{k}) c_{\vec{k}+\frac{m_e}{m_{ex}}\vec{Q},s}^\dagger b_{\vec{k}-\frac{m_h}{m_{ex}}\vec{Q},s'} |\psi_o\rangle. \quad (9)$$

The exciton wave function is $\psi_{\alpha,\vec{Q}}(\vec{k})$. The electron and hole effective masses are m_e and m_h , respectively. The exciton mass is $m_{ex} = m_e + m_h$, and the reduced electron-hole mass is m_r . If one writes the exciton wave function as

$$\psi_{\alpha,\vec{Q}}(\vec{k}) = \phi_{\alpha}(\vec{k}) e^{i(\tau\phi_{\vec{k}+(m_e/m_{ex})\vec{Q}} + \tau'\phi_{\vec{k}'-(m_h/m_{ex})\vec{Q}})/2}, \quad (10)$$

then the exciton wave function $\phi_{\alpha}(\vec{k})$ satisfies the standard exciton eigenvalue equation [3,4],

$$[\bar{E}_c(\vec{k}) - \bar{E}_v(\vec{k})]\phi_{\alpha}(\vec{k}) - \frac{1}{A} \sum_{\vec{q}} V(\vec{q})\phi_{\alpha}(\vec{k} - \vec{q}) = E_{\alpha}(\vec{Q})\phi_{\alpha}(\vec{k}), \quad (11)$$

with an eigenvalue $E_{\alpha}(\vec{Q})$ given by

$$E_{\alpha}(\vec{Q}) = E_g - E_{\alpha} + \frac{\hbar^2 Q^2}{2m_{ex}}, \quad (12)$$

where E_{α} is the exciton binding energy. The energy $E_{\alpha}(\vec{Q})$ is measured with respect to the energy of the ground state $|\psi_o\rangle$. Note that the phase factors cancel out and do not appear in the exciton eigenvalue equation. The exciton wave functions are orthonormal and complete in the sense [26],

$$\int \frac{d^2\vec{k}}{(2\pi)^2} \phi_{\alpha}^*(\vec{k})\phi_{\beta}(\vec{k}) = \delta_{\alpha,\beta}, \quad (13)$$

$$\sum_{\alpha} \phi_{\alpha}(\vec{k})\phi_{\alpha}^*(\vec{k}') = (2\pi)^2 \delta^2(\vec{k} - \vec{k}'). \quad (14)$$

The sum over α above includes all the discrete bound exciton states as well as the continuum of ionized exciton states.

Finally, the probability of finding an electron and a hole at a distance \vec{r} in the exciton state $|\psi_{\alpha,\vec{Q}}(\vec{k})\rangle$ can be computed by destroying an electron and a hole using the real-space field destruction operators and then taking the overlap of the resulting state with the ground state $|\psi_o\rangle$. The result is $|\phi_{\alpha}(\vec{r})|^2$ where $\phi_{\alpha}(\vec{r})$ is the Fourier transform of $\phi_{\alpha}(\vec{k})$. Note that $\phi_{\alpha}(\vec{r})$ is not the Fourier transform of $\psi_{\alpha,\vec{Q}}(\vec{k})$, which also includes extra phase factors [see (10)].

D. Exciton basis

In what follows, we will use the exciton basis. The exciton creation operator $B_{s,s',\alpha}^{\dagger}(\vec{Q})$ can be defined as

$$B_{s,s',\alpha}^{\dagger}(\vec{Q}) = \frac{1}{\sqrt{A}} \sum_{\vec{k}} \psi_{\alpha,\vec{Q}}(\vec{k}) c_{\vec{k}+\frac{m_e}{m_{ex}}\vec{Q},s}^\dagger b_{\vec{k}-\frac{m_h}{m_{ex}}\vec{Q},s'}. \quad (15)$$

Using the completeness and the orthogonality of the exciton wave functions given in (14) and (13), we get

$$c_{\vec{k},s}^\dagger b_{\vec{k}',s'} = \frac{1}{\sqrt{A}} \sum_{\alpha} \psi_{\alpha,\vec{Q}}^*(\vec{k}_r) B_{s,s',\alpha}^{\dagger}(\vec{Q}). \quad (16)$$

Here, \vec{k}_r and \vec{Q} equal $(m_h/m_{ex})\vec{k} + (m_e/m_{ex})\vec{k}'$ and $\vec{k} - \vec{k}'$ on the left-hand side, respectively. Products of electron and hole

creation and destruction operators can thus be expressed in terms of the exciton operators.

E. Defect states

TMDs (MX_2), and in particular monolayer MoS_2 , are known to have several different kinds of point defects, such as M and X vacancies and interstitials, impurity atoms, in addition to grain boundaries and dislocations [27–35]. The goal in this section is not to give a detailed description of different defect states in TMDs, something well beyond the scope of this paper, but to capture the essential physics in a way that would enable us to obtain capture rates for electrons and holes and present the main ideas associated with the capture processes.

Since the Bloch states form a complete set, the wave function $\psi_d(\vec{r})$ of the electron in the defect state can be expanded in terms of the Bloch states from all the bands [9]. In most cases of practical interest, only Bloch states in the vicinity of certain points \vec{K}_s in the Brillouin zone, such as Γ , M , K , and K' in the case of 2D-TMDs, need to be included in the expansion and therefore one may write

$$\psi_d(\vec{r}) = \frac{1}{\sqrt{A}} \sum_{n,\vec{k},s} c_{n,s}(\vec{k}) \frac{e^{i(\vec{K}_s+\vec{k})\cdot\vec{r}}}{\sqrt{A}} u_{n,\vec{k},s}(\vec{r}). \quad (17)$$

In the expression above, $u_{n,\vec{k},s}(\vec{r})$ are the periodic parts of the Bloch functions. The sum over n runs over all the energy bands. Whereas shallow defect levels can usually be described well by limiting the summation above to a single band, deep mid-gap defect levels generally have contributions from multiple bands [9,20]. The above expression can usually be cast in much simpler forms for specific defect states.

As an example, we consider the case of the deep point defect in MoS_2 due to a sulfur atom vacancy. A sulfur atom vacancy is a common defect in MoS_2 monolayers and can have a small formation energy [31–33]. The three states within the bandgap associated with a sulfur vacancy have been obtained previously using *ab initio* techniques [31–33]. These defect states consist of (i) a single A_1 state, made up of mostly the d_{xz} and d_{yz} orbitals of the Mo atoms adjacent to the missing S atom, with an energy a few tenths of an eV above the valence band maxima, and (ii) two degenerate E states, made up of mostly the d_{z^2} , $d_{x^2-y^2}$, and d_{xy} orbitals of the Mo atoms adjacent to the missing S atom, with an energy 1.4–1.6 eV above the valence band maxima. All the defect states are spin degenerate and correspond to the one- (A_1) and two-dimensional (E) representations of the trigonal symmetry group C_{3v} . The computed orbitals of these states are shown in Fig. 2 (from Noh *et al.* [32]). A defect state can be an efficient center for nonradiative recombination due to Auger scattering only if it has good overlaps with the Bloch states of both the conduction and the valence bands. The E states fit this criterion. The E states can be described well by limiting the summation in the expression above to the Bloch states of the conduction and the valence band extrema at the K and K' points. Since all the orbitals forming the E states have weights almost entirely on the Mo atoms adjacent to the missing S atom, one may write $c_{v,s}(\vec{k}) \approx \chi_d(\vec{k}) e^{i\gamma\tau\phi_{\vec{k}}/2} b_{v,s}$. Since $e^{i\gamma\tau\phi_{\vec{k}}/2} u_{c,\vec{k},s}(\vec{r})$ does not vary much with \vec{k} near the

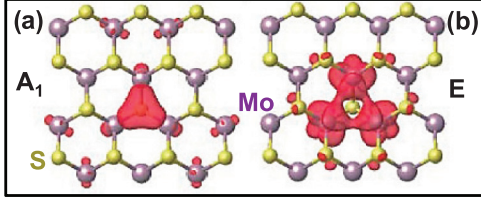


FIG. 2. (Color online) The computed orbitals of the defect states in a MoS₂ monolayer corresponding to a sulfur vacancy are shown (from Noh *et al.* [32]). (a) and (b) The A₁ state and the two degenerate E states, respectively.

band extrema, the sum in (17) can be rearranged to give

$$\psi_d(\vec{r}) = \chi_d(\vec{r}) \sum_{n=c,v} b_{n,s} e^{i\vec{k}_s \cdot \vec{r}} e^{i\gamma\tau\phi_{\vec{k}}/2} u_{n,\vec{k},s}(\vec{r}). \quad (18)$$

Here, the line under \vec{k} means that any wave vector near the band extrema can be chosen. The function $\chi_d(\vec{r})$ is expected to be localized at the defect, becoming very small at the second nearest Mo atom near the defect site.

F. Hamiltonian for the capture of holes and electrons

Consider process (b) in Fig. 1 in which a hole scatters off an electron and is captured by a defect and the electron is scattered to a higher energy. The relevant term in the Coulomb interaction Hamiltonian that describes the hole capture process in Fig. 1(b) can be written as

$$H_{hc} = \frac{1}{A} \sum_{\vec{k}, \vec{k}', \vec{q}, s, s'} V(\vec{q}) M_{s,s'}(\vec{k}, \vec{k}', \vec{q}) c_{\vec{k}+\vec{q}, s}^\dagger b_{\vec{k}', s'}^\dagger d_{\sigma'} c_{\vec{k}, s} + \text{H.c.} \quad (19)$$

The overlap factor $M_{s,s'}(\vec{k}, \vec{k}', \vec{q})$ equals

$$\begin{aligned} M_{s,s'}(\vec{k}, \vec{k}', \vec{q}) &= \langle v_{c, \vec{k}+\vec{q}, s} | v_{c, \vec{k}, s} \rangle \\ &\times \sum_{n=c,v} b_{n,s'} e^{i\gamma\tau\phi_{\vec{k}}/2} \langle v_{v, \vec{k}', s'} | v_{n, \vec{k}, s} \rangle \\ &\times \frac{1}{\sqrt{A}} \int d^2\vec{r} \chi_d(\vec{r}) e^{-i(\vec{k}+\vec{q}) \cdot \vec{r}} \\ &\approx e^{i(\tau\phi_{\vec{k}+\vec{q}} - \tau\phi_{\vec{k}} - \tau'\phi_{\vec{k}'})/2} \frac{b_{v,s'}}{\sqrt{A}} \chi_d(\vec{k}' + \vec{q}). \end{aligned} \quad (20)$$

Similarly, the electron capture process [Fig. 1(a)] is described by the Hamiltonian,

$$H_{ec} = \frac{1}{A} \sum_{\vec{k}, \vec{k}', \vec{q}, s, s'} V(\vec{q}) L_{s,s'}(\vec{k}, \vec{k}', \vec{q}) d_{\sigma}^\dagger b_{\vec{k}', s'}^\dagger b_{\vec{k}+\vec{q}, s'} c_{\vec{k}, s} + \text{H.c.}, \quad (21)$$

where the overlap factor $L_{s,s'}(\vec{k}, \vec{k}', \vec{q})$ equals

$$L_{s,s'}(\vec{k}, \vec{k}', \vec{q}) \approx e^{i(\tau'\phi_{\vec{k}'+\vec{q}} - \tau'\phi_{\vec{k}'} - \tau\phi_{\vec{k}})/2} \frac{b_{c,s}^*}{\sqrt{A}} \chi_d^*(\vec{k} + \vec{q}). \quad (22)$$

The potential of the defect does not appear in the Hamiltonian above. The reason for this is that it has already been taken into account in defining the noninteracting Hamiltonian, and its eigenstates, in Sec. II B.

III. ELECTRON AND HOLE CAPTURE RATES FOR EXCITONS

We assume an initial state described by the density operator ρ_i in which the exciton occupation $n_{s,s',\alpha}(\vec{Q})$, defect occupation f_d , and conduction and valence band occupations are given by

$$\begin{aligned} \langle d_{\sigma}^\dagger d_{\sigma'} \rangle &= f_d \delta_{\sigma,\sigma'}, \\ \langle c_{\vec{k},s}^\dagger c_{\vec{k}',s'} \rangle &= f_{c,s}(\vec{k}) \delta_{s,s'} \delta_{\vec{k},\vec{k}'}, \\ \langle b_{\vec{k},s}^\dagger b_{\vec{k}',s'} \rangle &= f_{v,s}(\vec{k}) \delta_{s,s'} \delta_{\vec{k},\vec{k}'}, \\ \langle B_{s,s',\alpha}^\dagger(\vec{Q}) B_{s,s',\alpha}(\vec{Q}) \rangle &= n_{s,s',\alpha}(\vec{Q}) \\ &+ \frac{1}{A} \sum_{\vec{k}} |\phi_{\alpha}(\vec{k})|^2 f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \\ &\times \left[1 - f_{v,s'} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right]. \end{aligned} \quad (23)$$

The angled brackets stand for ensemble averaging with respect to the density operator ρ_i . Since the excitons are not exact bosons, the value of $\langle B_{s,s',\alpha}^\dagger(\vec{Q}) B_{s,s',\alpha}(\vec{Q}) \rangle$ is not just equal to the exciton occupation $n_{s,s',\alpha}(\vec{Q})$. Using the cluster expansion to evaluate $\langle B_{s,s',\alpha}^\dagger(\vec{Q}) B_{s,s',\alpha}(\vec{Q}) \rangle$ results in the additional Hartree-Fock term shown above [36,37]. The same extra term also shows up in the luminescence spectra of excitons [26], and, as discussed below, this term results in a quadratic dependence of the capture rate on the exciton density at large exciton densities.

We assume that the electron and hole densities for different spins/valleys (including both free carriers and bound excitons) are n_s and $p_{s'}$, respectively, and the defect density is n_d . The initial ensemble consists of states that are approximate eigenstates of $H_o + H_{eh}$ but not of $H_o + H_{eh} + H_{hc} + H_{ec}$. Therefore, we consider H_{hc} and H_{ec} as perturbations.

A. Electron capture rate

We first consider process Fig. 1(a) in which the electron is captured by a defect. The average electron capture rate R_{ec} (units, per unit area per second) can be calculated from the first-order perturbation theory using the exciton basis described in Sec. II D and the average values given in (23). The details of the calculations are given in the appendix. The final result is

$$\begin{aligned} R_{ec} &\approx \frac{2\pi}{\hbar} n_d (1 - f_d) \sum_{s,s',\alpha} D_{v,s'}(q_{\alpha}) |\chi_d(q_{\alpha})|^2 |b_{c,s}|^2 \\ &\times \left| \frac{1}{A} \sum_{\vec{k}_r} V(q_{\alpha} \hat{x} - \vec{k}_r) \phi_{\alpha}(\vec{k}_r) \right|^2 \\ &\times \left[n_{s,s',\alpha} + \frac{1}{A^2} \sum_{\vec{k}, \vec{Q}} |\phi_{\alpha}(\vec{k})|^2 f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \right. \\ &\times \left. \left[1 - f_{v,s'} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right] \right]. \end{aligned} \quad (24)$$

Here, $D_{v,s'}$ is the valence band density of states (per valley per spin) evaluated at the energy of the scattered hole whose wave vector is q_α . q_α is approximately given by the relation, $E_{v,s'}(0) - E_{v,s'}(q_\alpha) = E_g - E_\alpha - E_d$. Note that none of the phase factors appear in the above result. The exciton density $n_{s,s',\alpha}$ is

$$n_{s,s',\alpha} = \int \frac{d^2\vec{Q}}{(2\pi)^2} n_{s,s',\alpha}(\vec{Q}). \quad (25)$$

If $q_\alpha \gg k_r$ for all values of k_r for which $\phi_\alpha(\vec{k}_r)$ is significant, then the above expression reduces to

$$R_{ec} = \frac{2\pi}{\hbar} n_d (1 - f_d) \sum_{s,s',\alpha} D_{v,s'}(q_\alpha) |V(q_\alpha)|^2 |\chi_d(q_\alpha)|^2 \times |b_{c,s}|^2 [|\phi_\alpha(\vec{r} = 0)|^2 n_{s,s',\alpha} + G_\alpha n_s p_{s'}]. \quad (26)$$

The expression for G_α is given in the appendix. G_α is significant for only the lowest few exciton states.

B. Hole capture rate

The rate for process [Fig. 1(b)] in which the hole is captured by a defect can be calculated in the same way. The result is

$$R_{hc} \approx \frac{2\pi}{\hbar} n_d f_d \sum_{s,s',\alpha} D_{c,s}(q_\alpha) |\chi_d(q_\alpha)|^2 |b_{v,s'}|^2 \times \left| \frac{1}{A} \sum_{\vec{k}_r} V(q_\alpha \hat{x} - \vec{k}_r) \phi_\alpha(\vec{k}_r) \right|^2 \times \left[n_{s,s',\alpha} + \frac{1}{A^2} \sum_{\vec{k}, \vec{Q}} |\phi_\alpha(\vec{k})|^2 f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \times \left[1 - f_{v,s'}(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q}) \right] \right], \quad (27)$$

where now q_α is approximately given by the relation $E_{c,s}(q_\alpha) - E_{c,s}(0) = E_d - E_\alpha$. And, as before, if $q_\alpha \gg k_r$ for all values of k_r for which $\phi_\alpha(\vec{k}_r)$ is significant, then the above expression reduces to

$$R_{hc} = \frac{2\pi}{\hbar} n_d f_d \sum_{s,s',\alpha} D_{c,s}(q_\alpha) |V(q_\alpha)|^2 |\chi_d(q_\alpha)|^2 \times |b_{v,s'}|^2 [|\phi_\alpha(\vec{r} = 0)|^2 n_{s,s',\alpha} + G_\alpha n_s p_{s'}]. \quad (28)$$

C. Coulomb correlations and enhancement of the Auger capture rates

Equation (26) for the electron capture rate can also be written as

$$R_{ec} = \frac{2\pi}{\hbar} n_d (1 - f_d) \sum_{s,s',\alpha} D_{v,s'}(q_\alpha) |V(q_\alpha)|^2 |\chi_d(q_\alpha)|^2 \times |b_{c,s}|^2 n_s p_{s'} [G_\alpha + g_{s,s',\alpha}(\vec{r} = 0)], \quad (29)$$

where $g_{s,s',\alpha}(\vec{r} = 0) = |\phi_\alpha(\vec{r} = 0)|^2 n_{s,s',\alpha} / (n_s p_{s'})$. The quantity inside the square brackets in (29), $G_\alpha + g_{s,s',\alpha}(\vec{r} = 0)$, describes the enhancement in the probability of finding an electron and a hole close to each other as a result of the attractive Coulomb interactions. It is interesting to compare the

electron capture rate in (29) with the result obtained assuming no electron-hole attractive interaction (i.e., $H_{eh} = 0$),

$$R_{ec} = \frac{2\pi}{\hbar} n_d (1 - f_d) \times \sum_{s,s'} D_{v,s'}(q_o) |V(q_o)|^2 |\chi_d(q_o)|^2 |b_{c,s}|^2 n_s p_{s'}, \quad (30)$$

where q_o is approximately given by the relation $E_{v,s'}(0) - E_{v,s'}(q_o) = E_g - E_d$. It can be seen that the capture rate in (29) is larger by the same enhancement factor. Assuming all the electrons and holes are in the lowest ($\alpha = 1$) bound exciton state, values of $D_{v,s'}$ and $|b_{c,s}|$ are independent of the valley/spin indices, and the exciton density is $n_{ex} = \sum_{s,s'} n_{s,s',\alpha=1}$, the comparison between (29) and (30) shows that the enhancement of the electron capture rate in the case of excitons is roughly proportional to $G_{\alpha=1} + |\phi_{\alpha=1}(\vec{r} = 0)|^2 / n_{ex}$. Given that the radius of the lowest exciton state in monolayer MoS₂ is in the 7–10 Å range [3], the enhancement, assuming an exciton density of 10^{12} cm^{-2} , is in the 72–138 range, and in the 644–1308 range if the exciton density is assumed to be 10^{11} cm^{-2} . Therefore, the correlations in the positions of the electrons and the holes as a result of the attractive Coulomb interaction make electrons and holes in tightly bound excitons in TMDs far more susceptible to capture by defects compared to uncorrelated free carriers. Interestingly, even when the exciton density $n_{s,s',\alpha}$ is zero the capture rate in (29) is enhanced by the factors G_α compared to the rate in (30) for uncorrelated electrons and holes. Therefore, Coulomb correlations in the positions of electrons and holes due to the attractive interaction between them enhances the Auger scattering rates even at the Hartree-Fock level.

IV. NUMERICAL RESULTS AND DISCUSSION

A. Carrier capture times at low exciton densities

For numerical computations, we consider monolayer MoS₂ on a quartz substrate, as is the case in many experiments. We first assume that the exciton density is small enough ($\leq 10^{12} \text{ cm}^{-2}$) to allow one to ignore phase-space filling effects [3]. We use the wave-vector-dependent dielectric constant $\epsilon(\vec{q})$ for monolayer MoS₂ on quartz given by Zhang *et al.* [3]. The defect state wave function is given in (18). The values of $|b_{c,s}|^2$ and $|b_{v,s}|^2$ are assumed to be independent of the valley/spin indices. This is a good approximation for many important cases. For example, in the case of the sulfur vacancy in MoS₂ discussed earlier, the E states have a total weight of ~ 0.25 on the d_{z^2} orbitals of the Mo atoms adjacent to the missing sulfur atom [38]. Since the conduction band Bloch states of both K and K' valleys are made up of mostly the d_{z^2} orbitals of Mo atoms, $|b_{c,s}|^2$ is the same for both the valleys. We approximate the envelope $\chi_d(\vec{r})$ of the defect state wave function in (18) by a Gaussian, $\chi_d(\vec{r}) = \sqrt{2/(\pi a_d^2)} e^{-r^2/a_d^2}$, where $a_d \approx 3 \text{ Å}$ (see Fig. 2). Note that the in-plane S-Mo bond length in MoS₂ is $\sim 1.83 \text{ Å}$. Figure 3 plots the computed capture times of electrons (τ_{ec}) and holes (τ_{hc}) of excitons assuming that all the excitons are in the lowest state ($\alpha = 1$). In the low exciton density limit considered here these capture times are independent of the exciton density. The defect density n_d is assumed to be $2 \times 10^{11} \text{ cm}^{-2}$. The capture times for electrons

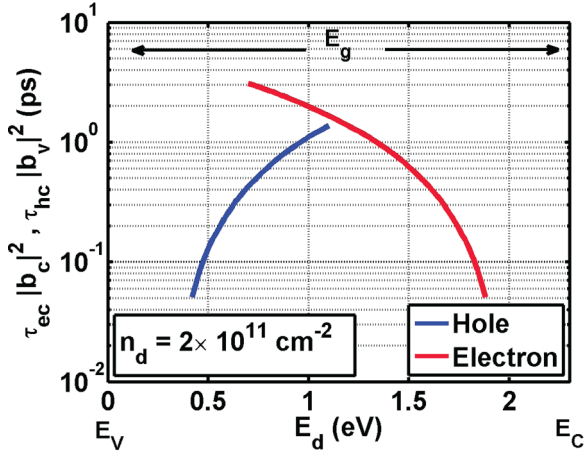


FIG. 3. (Color online) The capture times of electrons (τ_{ec}) and holes (τ_{hc}) of excitons by defects in monolayer MoS₂ on quartz are plotted as a function of the defect energy within the bandgap. The exciton binding energy $E_{\alpha=1}$ is 0.4 eV and the material bandgap is 2.3 eV [3]. The plotted capture times for electrons and holes have been normalized by multiplying them by $|b_c|^2$ and $|b_v|^2$, respectively. The defect density n_d is $2 \times 10^{11} \text{ cm}^{-2}$.

and holes shown in Fig. 3 have been normalized by multiplying them by $|b_c|^2$ and $|b_v|^2$, respectively, given the uncertainty in the exact values of these parameters. In the calculation of the electron capture times the defect state is assumed to be empty ($f_d = 0$), and in the calculation of the hole capture times the defect state is assumed to be full ($f_d = 1$).

The curves shown in Fig. 3 can provide results in different situations. For example, in the case of the E states associated with a sulfur vacancy, if $|b_c|^2$ is assumed to be ~ 0.25 [38], then the electron capture time curve in Fig. 3 would need to be multiplied by 4 in order to get the actual electron capture times. If the E state energy is assumed to ~ 1.5 eV above the valence band edge [32], then the electron capture time comes out to be ~ 2.4 ps. Since the capture times decrease inversely with the defect density n_d , the capture times shown in Fig. 3 can be interpolated for different values of the defect density. For example, a defect density of $8 \times 10^{11} \text{ cm}^{-2}$ would result in an electron capture time of 0.6 ps for the E state of a sulfur vacancy (under the same assumptions as stated above).

Figure 3 shows that shallower traps have much shorter capture times than deeper traps. This can be understood as follows. Energy conservation requires that the scattered electron (hole), in a hole (electron) capture process, takes away most of the energy. The deeper the trap the more the final energy of the scattered particle. Also, momentum conservation requires that the momentum of the scattered particle be provided by the relevant Fourier component of the defect state wave function. Therefore, the deeper the trap the larger the momentum transfer. Since in Fourier space the defect state wave function is $\chi_d(\vec{q}) = \sqrt{2\pi a_d^2} e^{-q^2 a_d^2/4}$, larger momentum transfers result in smaller capture rates. Note that this result is largely independent of the exact assumed form of the defect state wave function. In addition, the Coulomb potential $V(\vec{q})$ also decreases for larger momentum transfers. Although the final density of states available to the scattered particle increases with the particle energy (for nonparabolic

energy band dispersions in 2D), this increase is not enough to offset the reduction in the capture rates due to the factors mentioned above.

Since the energy width of the valence and conduction bands in MoS₂ are less than 1.2 eV and 0.6 eV [24,39], respectively, the limited horizontal extents of the curves in Fig. 3 ensure that the electron (hole) scattered to a high energy in the hole (electron) capture process is scattered within the same band consistent with the assumptions made in this work. It is, however, possible for the scattered particle to go into a different band. For example, slightly away from the K (K') points, the next higher conduction band has Bloch states with a large weight on the d_{z^2} orbitals of Mo atoms and these Bloch states will have large overlap with the Bloch states near the conduction band bottom [40]. It should also be noted that the weights $|b_c|^2$ and $|b_v|^2$ for defects could be very small or zero. For example, in the case of sulfur vacancy A_1 states both $|b_c|^2$ and $|b_v|^2$ are expected to be very small [32,33,38].

B. Carrier capture times at high exciton densities

At large exciton densities (typically larger than 10^{12} cm^{-2} , but smaller than 10^{13} cm^{-2} , for 2D-TMDs [3]), phase-space filling effects cannot be ignored in the description of the exciton states. We use the formalism developed by Kira and Koch [26,36]. When phase-space filling is taken into account, the exciton eigenvalue equation in the relative coordinates becomes non-Hermitian (see the appendix) and its solutions are expressed in terms of the *left* and the *right* eigenfunctions, $\phi_{\alpha,s,s'}^L(\vec{k}, \vec{Q})$ and $\phi_{\alpha,s,s'}^R(\vec{k}, \vec{Q})$, respectively. These eigenfunctions are also a function of the center-of-mass momentum \vec{Q} , and are related as follows [26,36],

$$\phi_{s,s',\alpha}^R(\vec{k}, \vec{Q}) = \phi_{s,s',\alpha}^L(\vec{k}, \vec{Q}) \left[f_{v,s'} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) - f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \right], \quad (31)$$

and obey the orthogonality relation,

$$\int \frac{d^2 \vec{k}}{(2\pi)^2} [\phi_{s,s',\alpha}^L(\vec{k}, \vec{Q})]^* \phi_{s,s',\beta}^R(\vec{k}, \vec{Q}) = \delta_{\alpha,\beta}. \quad (32)$$

In terms of these eigenfunctions, the expression for the electron capture rate becomes

$$\begin{aligned} R_{ec} \approx & \frac{2\pi}{\hbar} n_d (1 - f_d) \frac{1}{A} \sum_{s,s',\alpha,\vec{Q}} D_{v,s'}(q_\alpha) |\chi_d(q_\alpha)|^2 |b_{c,s}|^2 \\ & \times \left| \frac{1}{A} \sum_{\vec{k}_r} V(q_\alpha \hat{x} - \vec{k}_r) \phi_{s,s',\alpha}^R(\vec{k}_r, \vec{Q}) \right|^2 \left[n_{s,s',\alpha}(\vec{Q}) \right. \\ & + \frac{1}{A} \sum_{\vec{k}} |\phi_{s,s',\alpha}^L(\vec{k}, \vec{Q})|^2 f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \\ & \left. \times \left[1 - f_{v,s'} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right] \right]. \end{aligned} \quad (33)$$

The expression for the capture rate of holes in the high exciton density case follows similarly from (27). When all electrons

and holes exist as excitons, self-consistency requires that the distribution functions are given by [26]

$$f_{c,s}(\vec{k}) = \frac{1}{A} \sum_{s',\alpha,\vec{Q}} [\phi_{s,s',\alpha}^L(\vec{k},\vec{Q})]^* \phi_{s,s',\alpha}^R(\vec{k},\vec{Q}) n_{s,s',\alpha}(\vec{Q}),$$

$$1 - f_{v,s'}(\vec{k}) = \frac{1}{A} \sum_{s,\alpha,\vec{Q}} [\phi_{s,s',\alpha}^L(\vec{k},\vec{Q})]^* \phi_{s,s',\alpha}^R(\vec{k},\vec{Q}) n_{s,s',\alpha}(\vec{Q}).$$
(34)

Equations (33) and (34) show that the capture rate R_{ec} has terms that go linearly as well as quadratically with the exciton density. The quadratic dependence comes from the Hartree-Fock term in the evaluation of $\langle B_{s,s',\alpha}^\dagger(\vec{Q}) B_{s,s',\alpha}(\vec{Q}) \rangle$ [see Eq. (23)]. It can be understood as coming from the Auger scattering between the electron of one exciton and the hole of another exciton (defect-assisted exciton-exciton annihilation). Recall from the discussion in Sec. III C that even at the Hartree-Fock level Auger scattering between electrons and holes is enhanced due to the Coulomb correlations compared to uncorrelated electrons and holes.

For numerical computations, we again consider monolayer MoS₂ on a quartz substrate, as in Sec. IV A. We solve the exciton eigenvalue equation for different exciton densities and obtain the exciton radii and the exciton binding energies [3]. For simplicity, we consider the case when all the electrons and holes are in the lowest ($\alpha = 1$) bound exciton state. Figure 4 plots the inverse capture time (τ_{ec}^{-1}) of the electron of an exciton in monolayer MoS₂ on quartz as a function of the exciton density. The plotted capture time has been normalized by multiplying it by $|b_c|^2$. The defect density n_d is $2 \times 10^{11} \text{ cm}^{-2}$ and the defect energy E_d is assumed to be 1.5 eV above the valence band edge. The inverse capture time increases

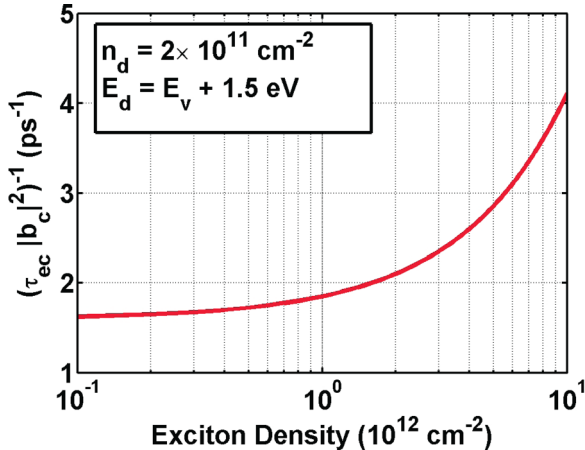


FIG. 4. (Color online) The inverse capture time (τ_{ec}^{-1}) for the electron of an exciton in monolayer MoS₂ on quartz is plotted as a function of the exciton density. The plotted capture time has been normalized by multiplying it by $|b_c|^2$. The defect density n_d is $2 \times 10^{11} \text{ cm}^{-2}$ and the defect energy E_d is assumed to be 1.5 eV above the valence band edge. The inverse capture time increases with the exciton density n_{ex} roughly as $\tau_{ec}^{-1} \sim A + Bn_{ex}$ (A and B are constants).

with the exciton density n_{ex} roughly as, $\tau_{ec}^{-1} \sim A + Bn_{ex}$ (A and B are constants), indicating that the capture rate R_{ec} has both linear and quadratic dependence on the exciton density ($R_{ec} \sim An_{ex} + Bn_{ex}^2$). The term quadratic in the exciton density in R_{ec} becomes significant at exciton densities higher than $\sim 10^{12} \text{ cm}^{-2}$. When interpreting experimental data, this quadratic increase of the carrier capture rate with the exciton density can make exciton annihilation via carrier capture by defects (defect-assisted exciton-exciton annihilation) indistinguishable from exciton-exciton annihilation via direct interband Auger scattering, the rate of which is also expected to go quadratically with the exciton density.

V. COMMENTS AND CONCLUSION

The results presented in this paper show that the capture times for electrons and holes of excitons in TMDs can be very short—from less than a picosecond to a few picoseconds. These numbers agree well with the recently reported experimental results on the ultrafast carrier dynamics in photoexcited monolayer MoS₂ where fast relaxation times in the few picoseconds range were observed [15,17–19]. In addition, the results in Figs. 3 and 4 are largely independent of the carrier temperature which is also consistent with the experimental observations [15,18].

The expressions given in this work could overestimate (underestimate) the capture rates (times). The reasons are as follows. The magnitude of the intraband overlap integrals for Bloch states was assumed to equal unity in Sec. II F and only phase differences were taken into account. At energies much different from the band edge energies, the Bloch states are different from the band edge Bloch states, and consequently the magnitude of the overlap integrals are smaller than unity. For example, the two-band $k \cdot p$ model in Sec. II B shows that at wave vector \vec{k} the conduction (valence) band Bloch states have contributions from the valence (conduction) band Bloch states at $\vec{k} = 0$ with a weight given by $0.5 - 0.5 (\Delta_s/2) / \sqrt{(\Delta_s/2)^2 + (\hbar v k)^2}$. This implies a 15% weight at energies in the band that are ~ 0.5 eV away from the band edge. In addition, both the conduction and valence band Bloch states are expected to get contributions from other lower and higher bands at large wave vectors [22]. However, we don't expect the essential physics to change significantly or the rates to change by more than a factor of unity when these sources of error are removed. We should also point out that the rates for carrier capture by defects in 2D-TMDs can vary from sample to sample as the nature of defects is expected to depend on the method of sample preparation.

ACKNOWLEDGMENTS

The authors would like to acknowledge helpful discussions with Paul L. McEuen and Michael G. Spencer, and support from CCMR under NSF Grant No. DMR-1120296, AFOSR-MURI under Grant No. FA9550-09-1-0705, and ONR under Grant No. N00014-12-1-0072.

APPENDIX

1. Details on the electron capture rate

In this section, we derive the expression for the electron capture rate given in (24). The derivation of the hole capture rate is similar. We assume an initial state described by the density matrix ρ_i in which the exciton occupation is $n_{s,s',\alpha}(\vec{Q})$, the defect density is n_d , the defect occupation is f_d , and the electron and hole densities (including both free carriers and bound excitons) are n_s and $p_{s'}$, respectively. The average values of various operators are as given in (23). The rate of change of the total electron density is

$$\dot{n} = \frac{dn}{dt} = \frac{d}{dt} \left(\frac{1}{A} \sum_{s,\vec{k}} c_{\vec{k},s}^\dagger c_{\vec{k},s} \right). \quad (\text{A1})$$

Defining the interaction representation for the time development of operators as

$$O^I(t) = e^{\frac{i}{\hbar}(H_o+H_{eh})t} O e^{-\frac{i}{\hbar}(H_o+H_{eh})t}, \quad (\text{A2})$$

the rate R_{ec} for the electron capture by the defect can be found by picking the appropriate term from the expression obtained

using the first-order perturbation theory,

$$\left\langle \frac{dn}{dt} \right\rangle = \lim_{\eta \rightarrow 0} \frac{i}{\hbar} A n_d \int_{-\infty}^t dt' e^{\eta t'} \text{Tr} \{ \rho_i [H_{ec}^I(t'), \dot{n}^I(t)] \}. \quad (\text{A3})$$

Since the exciton states are approximate eigenstates of the Hamiltonian $H_o + H_{eh}$ we have

$$e^{\frac{i}{\hbar}(H_o+H_{eh})t} B_{s,s',\alpha}(\vec{Q}) e^{-\frac{i}{\hbar}(H_o+H_{eh})t} \approx B_{s,s',\alpha}(\vec{Q}) e^{-i \frac{E_{\alpha}(\vec{Q})}{\hbar} t}. \quad (\text{A4})$$

It is therefore convenient to express the conduction and valence band creation and destruction operators appearing in H_{ec} using the exciton basis described in Sec. II D. We also point out here that the ensemble average of a product of operators of the form,

$$\left\langle e^{\frac{i}{\hbar}(H_o+H_{eh})t} c_{\vec{k}_1,s_1}^\dagger b_{\vec{k}'_1,s'_1} b_{\vec{k}'_2,s'_2}^\dagger c_{\vec{k}_2,s_2} e^{-\frac{i}{\hbar}(H_o+H_{eh})t} \right\rangle, \quad (\text{A5})$$

needs to be evaluated using the cluster expansion and keeping the correlation terms as well as the Hartree-Fock term [36,37]. The final result is

$$R_{ec} = \frac{2\pi}{\hbar} n_d (1 - f_d) \frac{1}{A^4} \sum_{s,s',\vec{k}_r,\vec{k}'_r} |b_{c,s}|^2 V^*(\vec{q} - \vec{k}'_r) V(\vec{q} - \vec{k}_r) |\chi_d(\vec{q} + (m_e/m_{ex})\vec{Q})|^2 \phi_\alpha^*(\vec{k}'_r) \phi_\alpha(\vec{k}_r) \left[n_{s,s',\alpha}(\vec{Q}) + \frac{1}{A} \sum_{\vec{k}} |\phi_\alpha(\vec{k})|^2 \right. \\ \left. \times f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \left[1 - f_{v,s'} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right] \right] \delta \left(E_g - E_\alpha + \frac{\hbar^2 Q^2}{2m_{ex}} - E_d - E_{v,s'}(0) + E_{v,s'} \left(\vec{q} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right). \quad (\text{A6})$$

Note that all the phase factors have canceled out. The exciton center-of-mass kinetic energy, $\hbar^2 Q^2 / 2m_{ex}$, is expected to be much smaller than the energy difference $E_g - E_\alpha - E_d$. The former is expected to be in the few tens of meV range and the latter in the hundreds of meV range. The energy conserving delta function then enforces q to the value determined by the condition $E_{v,s'}(0) - E_{v,s'}(q_\alpha) = E_g - E_\alpha - E_d$. Once the magnitude of \vec{q} has been fixed in this way, it is easy to see that R_{ec} does not depend on the angle of \vec{q} . So one may assume $q \approx q_\alpha \hat{x}$ and obtain

$$R_{ec} \approx \frac{2\pi}{\hbar} n_d (1 - f_d) \sum_{s,s',\alpha} D_{v,s'}(q_\alpha) |\chi_d(q_\alpha)|^2 |b_{c,s}|^2 \left| \frac{1}{A} \sum_{\vec{k}_r} V(q_\alpha \hat{x} - \vec{k}_r) \phi_\alpha(\vec{k}_r) \right|^2 \left[n_{s,s',\alpha} + \frac{1}{A^2} \sum_{\vec{k},\vec{Q}} |\phi_\alpha(\vec{k})|^2 \right. \\ \left. \times f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \left[1 - f_{v,s'} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right] \right]. \quad (\text{A7})$$

Here, $D_{v,s'}$ is the valence band density of states (per valley per spin) evaluated at the energy of the scattered hole whose wave vector is q_α . The exciton density $n_{s,s',\alpha}$ is

$$n_{s,s',\alpha} = \int \frac{d^2 \vec{Q}}{(2\pi)^2} n_{s,s',\alpha}(\vec{Q}). \quad (\text{A8})$$

If $q_\alpha \gg k_r$ for all values of k_r for which $\phi_\alpha(\vec{k}_r)$ is significant, then the above expression reduces to

$$R_{ec} = \frac{2\pi}{\hbar} n_d (1 - f_d) \sum_{s,s',\alpha} D_{v,s'}(q_\alpha) |V(q_\alpha)|^2 |\chi_d(q_\alpha)|^2 \\ \times |b_{c,s}|^2 [|\phi_\alpha(\vec{r}=0)|^2 n_{s,s',\alpha} + G_\alpha n_s p_{s'}]. \quad (\text{A9})$$

Equation (A9) contains the exciton density $n_{s,s',\alpha}$ as well as the electron and hole densities (including both free carriers and bound excitons) n_s and $p_{s'}$, respectively. The latter appear as a result of the Hartree-Fock term in the cluster expansion [36,37]. G_α is

$$G_\alpha = \frac{|\phi_\alpha(\vec{r}=0)|^2}{n_s p_{s'} A^2} \sum_{\vec{k},\vec{Q}} |\phi_\alpha(\vec{k})|^2 f_{c,s} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \\ \times \left[1 - f_{v,s'} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right]. \quad (\text{A10})$$

G_α is expected to be significant for only the lowest few exciton states.

If all the electrons and holes are assumed to be in the lowest ($\alpha = 1$) bound exciton state then self-consistency requires that the distribution functions are given by [26]

$$\begin{aligned} f_{c,s}(\vec{k}) &= |\phi_{\alpha=1}(\vec{k})|^2 n_s, \\ 1 - f_{v,s'}(\vec{k}) &= |\phi_{\alpha=1}(\vec{k})|^2 p_{s'}. \end{aligned} \quad (\text{A11})$$

Here, $n_s = \sum_{s'} n_{s,s',\alpha=1}$ and $p_{s'} = \sum_s n_{s,s',\alpha=1}$. One then obtains

$$G_1 = |\phi_{\alpha=1}(\vec{r}=0)|^2 \frac{1}{A^2} \sum_{\vec{k}, \vec{Q}} \left| \phi_{\alpha=1} \left(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q} \right) \right|^2 |\phi_{\alpha=1}(\vec{k})|^2 \left| \phi_{\alpha=1} \left(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q} \right) \right|^2. \quad (\text{A12})$$

Assuming the standard 2D exciton wave function [3], G_1 equals $128/(5\pi) \approx 8.15$.

2. Description of excitons states in the high exciton density limit

In the high exciton density case, phase filling effects cannot be ignored in the description of the exciton states [3,26]. The exciton wave functions, $\phi_{\alpha,s,s'}^L(\vec{k}, \vec{Q})$ and $\phi_{\alpha,s,s'}^R(\vec{k}, \vec{Q})$ satisfy the eigenvalue equations [26],

$$\begin{aligned} & [E_{c,s}(\vec{k} + (m_e/m_{ex})\vec{Q}) - E_{v,s'}(\vec{k} - (m_h/m_{ex})\vec{Q})] \phi_{\alpha,s,s'}^L(\vec{k}, \vec{Q}) \\ & - \frac{1}{A} \sum_{\vec{k}'} V(\vec{k} - \vec{k}') \phi_{\alpha,s,s'}^L(\vec{k}', \vec{Q}) \left[f_{v,s'} \left(\vec{k}' - \frac{m_h}{m_{ex}} \vec{Q} \right) - f_{c,s} \left(\vec{k}' + \frac{m_e}{m_{ex}} \vec{Q} \right) \right] \\ & = E_{s,s',\alpha}(\vec{Q}) \phi_{s,s',\alpha}^L(\vec{k}, \vec{Q}). \end{aligned} \quad (\text{A13})$$

$$\begin{aligned} & [E_{c,s}(\vec{k} + (m_e/m_{ex})\vec{Q}) - E_{v,s'}(\vec{k} - (m_h/m_{ex})\vec{Q})] \phi_{\alpha,s,s'}^R(\vec{k}, \vec{Q}) - \left[f_{v,s'}(\vec{k} - \frac{m_h}{m_{ex}} \vec{Q}) - f_{c,s}(\vec{k} + \frac{m_e}{m_{ex}} \vec{Q}) \right] \\ & \times \frac{1}{A} \sum_{\vec{k}'} V(\vec{k} - \vec{k}') \phi_{\alpha,s,s'}^R(\vec{k}', \vec{Q}) \\ & = E_{s,s',\alpha}(\vec{Q}) \phi_{s,s',\alpha}^R(\vec{k}, \vec{Q}). \end{aligned} \quad (\text{A14})$$

The exciton wave functions satisfy the orthogonality and completeness relations,

$$\int \frac{d^2\vec{k}}{(2\pi)^2} [\phi_{s,s',\alpha}^L(\vec{k}, \vec{Q})]^* \phi_{s,s',\beta}^R(\vec{k}, \vec{Q}) = \delta_{\alpha,\beta} \quad (\text{A15})$$

$$\sum_{\alpha} \phi_{s,s',\alpha}^L(\vec{k}, \vec{Q}) [\phi_{s,s',\alpha}^R(\vec{k}', \vec{Q})]^* = (2\pi)^2 \delta^2(\vec{k} - \vec{k}'). \quad (\text{A16})$$

We also define

$$\psi_{s,s',\alpha}^{L/R}(\vec{k}, \vec{Q}) = \phi_{s,s',\alpha}^{L/R}(\vec{k}, \vec{Q}) e^{i(\tau \phi_{\vec{k} + (m_e/m_{ex})\vec{Q}} + \tau' \phi_{\vec{k}' - (m_h/m_{ex})\vec{Q}})/2}. \quad (\text{A17})$$

The exciton creation operator $B_{s,s',\alpha}^\dagger(\vec{Q})$ is defined as

$$B_{s,s',\alpha}^\dagger(\vec{Q}) = \frac{1}{\sqrt{A}} \sum_{\vec{k}} \psi_{s,s',\alpha}^L(\vec{k}, \vec{Q}) c_{\vec{k} + \frac{m_e}{m_{ex}} \vec{Q}, s}^\dagger b_{\vec{k} - \frac{m_h}{m_{ex}} \vec{Q}, s'}. \quad (\text{A18})$$

Using the completeness and the orthogonality of the exciton wave functions given in (A16) and (A15), we get

$$c_{\vec{k}, s}^\dagger b_{\vec{k}', s'} = \frac{1}{\sqrt{A}} \sum_{\alpha} \psi_{s,s',\alpha}^R(\vec{k}, \vec{Q}) B_{s,s',\alpha}^\dagger(\vec{Q}), \quad (\text{A19})$$

where \vec{k}_r and \vec{Q} equal $(m_h/m_{ex})\vec{k} + (m_e/m_{ex})\vec{k}'$ and $\vec{k} - \vec{k}'$ on the left-hand side, respectively.

- [1] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, *Phys. Rev. Lett.* **105**, 136805 (2010).
- [2] J. S. Ross, S. Wu, H. Yu, N. J. Ghimire, A. M. Jones, G. Aivazian, J. Yan, D. G. Mandrus, D. Xiao, W. Yao, and X. Xu, *Nat. Comm.* **4**, 1474 (2013).
- [3] C. Zhang, H. Wang, W. Chan, C. Manolatu, and F. Rana, *Phys. Rev. B* **89**, 205436 (2014).
- [4] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, *Phys. Rev. B* **88**, 045318 (2013).
- [5] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. Li, O. B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, *Phys. Rev. Lett.* **113**, 076802 (2014).
- [6] S. Konabe and S. Okada, *Phys. Rev. B* **90**, 155304 (2014).
- [7] G. Berghauser and E. Malic, *Phys. Rev. B* **89**, 125309 (2014).
- [8] H. Wang, C. Zhang, W. Chan, C. Manolatu, S. Tiwari, and F. Rana, [arXiv:1409.3996](https://arxiv.org/abs/1409.3996).
- [9] P. T. Landsberg, *Recombination in Semiconductors* (Cambridge University Press, Cambridge, 1992).
- [10] O. Lopez-Sanchez, D. Lembke, M. Kayci, A. Radenovic, and A. Kis, *Nat. Nanotechnol.* **8**, 497 (2013).
- [11] J. S. Ross, P. Klement, A. M. Jones, N. J. Ghimire, J. Yan, D. G. Mandrus, T. Taniguchi, K. Watanabe, K. Kitamura, W. Yao *et al.*, *Nat. Nanotechnol.* **9**, 268 (2014).
- [12] Z. Yin, H. Li, H. Li, L. Jiang, Y. Shi, Y. Sun, G. Lu, Q. Zhang, X. Chen, and H. Zhang, *ACS Nano* **6**, 74 (2012).
- [13] R. S. Sundaram, M. Engel, A. Lombardo, R. Krupke, A. C. Ferrari, P. Avouris, and M. Steiner, *Nano Lett.* **13**, 1416 (2013).
- [14] B. W. H. Baugher, H. O. H. Churchill, Y. Yang, and P. Jarillo-Herrero, *Nat. Nanotechnol.* **9**, 262 (2014).
- [15] H. Wang, C. Zhang, and F. Rana, *Nano Lett.* **15**, 339 (2015).
- [16] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, *Nano Lett.* **10**, 1271 (2010).
- [17] H. Shi, R. Yan, S. Bertolazzi, J. Brivio, B. Gao, A. Kis, D. Jena, H. Xing, and L. Huang, *ACS Nano* **7**, 1072 (2013).
- [18] D. Lagarde, L. Bouet, X. Marie, C. R. Zhu, B. L. Liu, T. Amand, P. H. Tan, and B. Urbaszek, *Phys. Rev. Lett.* **112**, 047401 (2014).
- [19] T. Korn, S. Heydrich, M. Hirmer, J. Schmutzler, and C. Schüller, *App. Phys. Lett.* **99**, 102109 (2011).
- [20] D. J. Robbins and P. T. Landsberg, *J. Phys. C* **13**, 2425 (1980).
- [21] D. Xiao, Gui-Bin Liu, W. Feng, X. Xu, and W. Yao, *Phys. Rev. Lett.* **108**, 196802 (2012).
- [22] A. Kormányos, V. Zolyomi, N. D. Drummond, P. Rakytá, G. Burkard, and V. I. Falko, *Phys. Rev. B* **88**, 045416 (2013).
- [23] G.-B. Liu, W.-Y. Shan, Y. Yao, W. Yao, and D. Xiao, *Phys. Rev. B* **88**, 085433 (2013).
- [24] T. Cheiwchanchamnangij and W. R. L. Lambrecht, *Phys. Rev. B* **85**, 205302 (2012).
- [25] D. K. Efimkin and Yu. E. Lozovik, *Phys. Rev. B* **87**, 245416 (2013).
- [26] M. Kira and S. W. Koch, *Semiconductor Quantum Optics* (Cambridge University Press, New York, 2012).
- [27] A. M. van der Zande, P. Y. Huang, D. A. Chenet, T. C. Berkelbach, Y. You, G.-H. Lee, T. F. Heinz, D. R. Reichman, D. A. Muller, and J. C. Hone, *Nat. Mater.* **12**, 554 (2013).
- [28] J. D. Fuhr, A. Saul, and J. O. Sofo, *Phys. Rev. Lett.* **92**, 026802 (2004).
- [29] H. P. Komsa, J. Kotakoski, S. Kurasch, O. Lehtinen, U. Kaiser, and A. V. Krashennnikov, *Phys. Rev. Lett.* **109**, 035503 (2012).
- [30] A. N. Enyashin, M. Bar-Sadan, L. Houben, and G. Seifert, *J. Phys. Chem. C* **117**, 10842 (2013).
- [31] W. Zhou, X. Zou, S. Najmaei, Z. Liu, Y. Shi, J. Kong, J. Lou, P. M. Ajayan, B. I. Yakobson, and J. C. Idrobo, *Nano Lett.* **13**, 2615 (2013).
- [32] J.-Y. Noh, H. Kim, and Y.-S. Kim, *Phys. Rev. B* **89**, 205417 (2014).
- [33] D. Liu, Y. Guo, L. Fang, and J. Robertson, *Appl. Phys. Lett.* **103**, 183113 (2013).
- [34] S. Yuan, R. Roldan, M. I. Katsnelson, and F. Guinea, *Phys. Rev. B* **90**, 041402 (2014).
- [35] H. Qiu, T. Xu, Z. Wang, W. Ren, H. Nan, Z. Ni, Q. Chen, S. Yuan, F. Miao, F. Song *et al.*, *Nat. Commun.* **4**, 2642 (2013).
- [36] M. Kira and S. W. Koch, *Prog. Quant. Electron.* **30**, 155 (2006).
- [37] S. W. Koch, M. Kira, G. Khitrova, and H. M. Gibbs, *Nat. Mater.* **5**, 523 (2006).
- [38] Y.-S. Kim [32] (private communication).
- [39] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, *Phys. Rev. Lett.* **111**, 216805 (2013).
- [40] E. Cappelluti, R. Roldan, J. A. Silva-Guillen, P. Ordejon, and F. Guinea, *Phys. Rev. B* **88**, 075409 (2013).