Journal of the Korean Physical Society, Vol. 73, No. 11, December 2018, pp. 1735∼1743

Theoretical Study of Auger Recombination of Excitons in Monolayer Transition-metal Dichalcogenides

Hyun Cheol Lee*
Department of Physics, Sogang University, Seoul 04107, Korea

(Received 26 October 2018)

Excitons are the most prominent features of the optical properties of monolayer transition-metal dichalcogenides(TMDC). In view of optoelectronics it is very important to understand the decay mechanisms of the excitons of these materials. Auger recombination of excitons are regarded as one of the dominant decay processes. In this paper the Auger constant of recombination is computed based on the approach proposed by Kavoulakis and Baym. We obtain both temperature dependent (from type A, A' processes) and temperature independent (from type B, B' processes) contributions, and a numerical estimate of theoretical result yields the value of constant in the order of 10^{-2} cm^2s^{-1}, being consistent with existing experimental data. This implies that Auger decay processes severely limit the photoluminescence yield of TMDC-based optoelectronic devices.

PACS numbers: 78.20.Bh, 78.40.−q, 78.47.−p
Keywords: Transition-metal dichalcogenides,Exciton, Auger recombination
DOI: 10.3938/jkps.73.1735

I. INTRODUCTION

Atomically thin two-dimensional crystals [1,2] is a very unique class of materials which are currently under intensive study, and they also provide a new arena for the next generation electronic devices [3,4].

Layered transition metal dichalcogenides (TMDC) is a class of semiconductors which can be tailored into monolayer with *direct* band gap semiconductors [1]. In monolayer TMDC the inversion symmetry is broken, while the time reversal symmetry is kept intact.

TMDC has the same hexagonal Bravais lattice as that of graphene, therefore, they inherit the same valley degrees of freedom with the sublattice degrees of freedom of graphene replaced by the species of d-orbitals of TMDC. Also, the strong spin-orbit coupling stemming from d-orbital of transition metal, being combined with the broken inversion symmetry of monolayer, leads to a significant spin-orbit splitting in valence band [5]. The valley degrees of freedom is an internal quantum number and it can play a role similar to that of real spin [6].

The direct gap (in visible range) nature of monolayer TMDC is very crucial in understanding its optical properties. The optical response is dominated by excitons with very large binding energy of the order of a few hundred meV [7, 8], which are consistent with theoretical calculations [9,10]. These large binding energies are attributed to the insufficient screening in 2-dimensional monolayer compared to bulk 3-dimensional material [11]. Thus it is essential to take the *long range* nature of Coulomb interaction into account in our problem.

In view of optoelectronics application, the monolayer TMDC has been proposed as ideal platform for quantum light emitters based on exciton emissions [7]. For the implementation of this proposal, it is very important to understand and control the exciton decay mechanisms. The Auger recombination of excitons (also referred to as biexciton nonradiative recombination or exciton-exciton annihilation) are regarded as one of the dominant decay mechanisms of excitons [12–14].

The Auger recombination of excitons consists of two processes: the recombination of an electron and a hole constituting excitons and the ionization of an exciton into an electron and a hole. No absorption or emission of radiation is involved, so it is a nonradiative process.

A basic framework for the theoretical understanding of Auger recombination of excitons has been proposed by Kavoulakis and Baym [15] in the context of Bose-Einstein condensation of excitons in Cu_2O. Their approach is best illustrated by the following set of Feynman diagrams, Fig. 1 (this is almost identical with Fig. 1 of Ref. 15 which has been redrawn here by author for reader's convenience).

In Type A and A' processes of Fig. 1 the electron and the hole of a single exciton recombine, while for Type B and B' process, the electron of one exciton and the hole of the other exciton recombine. This difference in recombination channel leads to a significant difference in momentum transfer of Coulomb interaction, which will

*E-mail: hyunlee@sogang.ac.kr

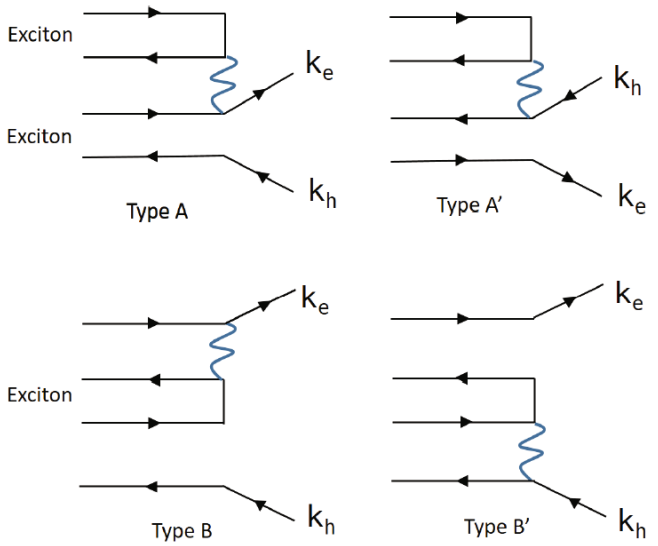

Fig. 1. (Color online) 4 types of Feynman diagrams contributing to the Auger recombination of two excitons. Time flows from left to right. The wavy line indicates Coulomb interaction. $\mathbf{k}_e$ and $\mathbf{k}_h$ is the momentum of ionized electron and hole, respectively.

result in the different temperature dependence of Auger decay constant. Diagrammatically, Type A and A' involves Coulomb interaction in *direct* channel while Type B and B' corresponds to exchange channel. The long range nature of Coulomb interaction (weak screening) is most manifest in direct channel. As is evident from Fig. 1, the recombination processes (in the second quantization formulation) require 3 conduction band electron operators and 1 valence band electron operator or 3 valence band electron operators or 1 conduction band electron operator. We note that the exciton formation is due to the Coulomb interaction with 2 conduction band electron operators and 2 valence band electron operators [16]. Thus we need to sort out those parts of Coulomb interaction responsible for Auger recombinations.

In the original work of [15] the phonon assisted Auger processes have also been considered because the band overlap integral turned out to be too small due to inversion symmetry. In the case of monolayer TMDC the inversion symmetry is absent, so we don't need to take the phonon assisted processes into account at least in the leading order.

The goal of this paper is rather modest. We simply apply the approach by Kavoulakis and Baym [15] of Auger recombination of two excitons to the monolayer TMDC materials and compute the exciton decay constant. The electronic band structure of the monolayer TMDC is described by $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian [5](see section II).

The Auger decay rate of excitons is computed via the Fermi golden rule:

$$\Gamma = \frac{2\pi}{\hbar}\sum_{\mathsf{i}}\rho_{\mathsf{i}}\sum_{\mathsf{f}}\left|\langle \mathsf{f}|\widehat{\mathsf{H}}_{\mathrm{C}}|\mathsf{i}\rangle\right|^2\delta(E_{\mathsf{i}}-E_{\mathsf{f}}), \tag{1}$$

$|\mathsf{i}\rangle$ is the initial state of two incoming excitons and $|\mathsf{f}\rangle$ is the final state of *ionized* electron and hole. $\widehat{\mathsf{H}}_{\mathrm{C}}$ is the part of Coulomb interaction responsible for Auger recombination as mentioned above. $E_{\mathsf{i,f}}$ is the initial and final state energy, respectively. $\sum_{\mathsf{i}}\rho_{\mathsf{i}}$ indicates an average over initial excitons states with probability ρ_{i}. We will assume that the density excitons is not high, so that the Maxwell-Boltzmann distribution can be used for the average. This average will put the exciton kinetic energy to be thermal energy and the center of mass momentum of exciton to be thermal momentum, which often can be neglected when compared to the generic electronic momentum scale. We have adopted the second quantization formalism for easier bookkeeping, since all the relevant band overlap integrals are naturally incorporated into density operators. For the purpose of the numerical estimate of our theoretical results, the fundamental band gap and the binding energy of exciton is taken to be 2.5 eV and 0.4 eV, respectively [17]. The typical size of the exciton (the Bohr radius of the internal exciton wavefunction) is known to be in the range of $0.6\sim 1$nm [17,18], and we will take 0.6 nm for our estimate. We also compare our results with those of 1-dimensional system where screened *short* range Coulomb interaction has been considered [19].

Our main results are Eq. (65) and Eq. (66) for the Auger decay rate and decay constant. Type A and A' processes produce temperature dependent (proportional to temperature) contribution, while type B and B' processes yields temperature independent contribution. Numerical estimates show that these two contributions can be comparable to each other at room temperature scale, and their magnitudes are consistent with the existing experimental data [12].

This paper is organized as follows: we set up the electron band Hamiltonians in Sec. II. The Coulomb interaction in the second quantized form is expressed in the basis of electronic band states in Sec. III. The matrix elements and the Auger decay rate are computed in Secs. IV and V, respectively. We conclude this paper with discussions and summary in Sec. VI.

II. SET UP

We employ the Hamiltonian which is obtained by $\mathbf{k}\cdot\mathbf{p}$ approximation [5] for the electronic structure of monolayer TMDC. The space group of a monolayer TMDC is D^1_{3h} which is one of the space group of hexagonal lattice. The group of wavevector at two valley points $\mathsf{K},\mathsf{K}'=-\mathsf{K}$ is C_{3h}. The minimal two-band Hamiltonian compatible with the symmetry group of wavevector for a given spin $s=\pm1=\uparrow,\downarrow$ and a valley $\tau=\pm1=\mathsf{K},\mathsf{K}'$ (mostly $\hbar=1$ convention and C.G.S unit will be used in this paper) is

$$\widehat{\mathsf{H}} = t_{\mathrm{hop}}a(k_x\tau\hat{\sigma}_x+k_y\hat{\sigma}_y)+\frac{\Delta}{2}\hat{\sigma}_z-\lambda s\tau\left(\frac{\hat{\sigma}_z-1}{2}\right), \tag{2}$$

where $\hat{\sigma}_i$ $(i = x, y, z)$ are Pauli matrices acting on the orbital space $|1\rangle, |2\rangle$. The symmetry adapted bases at $\mathsf{K}_{\tau=1}$ and $\mathsf{K}'_{\tau=-1}$ are given by

$$|1\rangle = |d_{z^2}\rangle, \quad |2_\tau\rangle = \frac{1}{\sqrt{2}}\Big(|d_{x^2-y^2}\rangle + i\tau|d_{xy}\rangle\Big). \quad (3)$$

The orbital space is spanned by linear combinations d-orbitals of transition metal element of TMDC. The values of the parameters for MoS_2 obtained from the first principle band structure calculations are $a = 3.193$ Å (lattice spacing), $t_{\rm hop} = 1.10$ eV (hopping amplitude), $\Delta = 1.66$ eV (band gap), and $2\lambda = 0.15$ eV (spin-orbit splitting of valence band) [5]. The spin-orbit splitting of conduction band is neglected.

The energy eigenvalues of the Hamiltonian Eq. (2) are (+:conduction band, −:valence band)

$$E^{(0)}_{\pm}(\mathbf{k}) = \pm\sqrt{(\Delta'_{s\tau}/2)^2 + v^2k^2} + \frac{s\tau\lambda}{2}, \quad (4)$$

where $k^2 = k_x^2 + k_y^2$, $v \equiv t_{\rm hop}a$, and $\Delta'_{\tau s} = \Delta - s\tau\lambda$.

The first principle band structure calculations suggest that the cutoff momentum scale of $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, $k_{\rm max}$, is about 10 percent of the first Brillouin zone size.

Since the energy gap Δ is larger than the band width ($\approx vk_{\rm max}$) (which is, in turn, larger than the spin-orbit splitting), Eq. (4) can be approximated by

$$E_+(\mathbf{k}) \approx \frac{\Delta}{2} + \frac{v^2k^2}{\Delta}, \quad E_-(\mathbf{k}) \approx -\frac{\Delta}{2} - \frac{v^2k^2}{\Delta} + s\tau\lambda. \quad (5)$$

The eigenvectors can be found exactly (for a given spin s and valley τ), but for our purpose the perturbative results up to the second order in hopping is sufficient (including wave function renormalization).

The eigenket for the conduction band is

$$|\mathrm{c}, \mathbf{k}\rangle_{\tau s} = \Big(|1\rangle + \frac{v(k_x\tau + ik_y)}{\Delta'_{\tau s}}|2\rangle - \frac{1}{2}\frac{v^2k^2}{(\Delta'_{\tau s})^2}|1\rangle\Big)\otimes\chi_s, \quad (6)$$

where χ_s is the two-component spinor for spin s.

The eigenket for the valence band is

$$|\mathrm{v}, \mathbf{k}\rangle_{\tau s} = \Big(|2\rangle + \frac{v(k_x\tau - ik_y)}{(-\Delta'_{\tau s})}|1\rangle - \frac{1}{2}\frac{v^2k^2}{(\Delta'_{\tau s})^2}|2\rangle\Big)\otimes\chi_s. \quad (7)$$

The matrix element between valence band the conduction at different momenta can be computed from Eqs. (6) and (7). Up to the second order in momentum,

$$\langle \mathrm{v}, \mathbf{k}|\mathrm{c}, \mathbf{k}'\rangle \approx \frac{v(k'_x - k_x)\tau + i(k'_y - k_y)}{\Delta'_{\tau s}}. \quad (8)$$

$$\langle \mathrm{c}, \mathbf{k}|\mathrm{c}, \mathbf{k}'\rangle \approx 1 - \frac{v^2((\mathbf{k}-\mathbf{k}')^2 + 2i\tau(\mathbf{k}'\times\mathbf{k})_z)}{2(\Delta'_{\tau s})^2}. \quad (9)$$

$$\langle \mathrm{v}, \mathbf{k}|\mathrm{v}, \mathbf{k}'\rangle \approx 1 - \frac{v^2((\mathbf{k}-\mathbf{k}')^2 - 2i\tau(\mathbf{k}'\times\mathbf{k})_z)}{2(\Delta'_{\tau s})^2}. \quad (10)$$

The *underscreened* Coulomb interaction in 2-dimension is chosen to be ($r = \sqrt{x^2+y^2}$)

$$v(\mathbf{r}) = \frac{e^2}{\epsilon r}, \quad v_c(\mathbf{q}) = \int \mathrm{d}^2\mathbf{r}\, e^{i\mathbf{q}\cdot\mathbf{r}} v(\mathbf{r}) = \frac{2\pi e^2}{\epsilon q}. \quad (11)$$

The internal state of exciton is taken to be the ground state of two-dimensional hydrogen atom problem. The ground state wavefunction is given by

$$\psi_0(\mathbf{r}) = \sqrt{\frac{2}{\pi}}\frac{1}{a_{\rm B}}e^{-r/a_{\rm B}}, \quad a_{\rm B} = 2\mu e^2/\hbar^2\epsilon, \quad (12)$$

where $\mu = m_{\rm e}m_{\rm h}/(m_{\rm e}+m_{\rm h})$ is the reduced mass of exciton and $a_{\rm B}$ is the Bohr radius of exciton (for our numerical estimate $a_{\rm B} \sim 0.6$nm). From [18], we choose $m_{\rm e} \sim 0.37m_0$ and $m_{\rm h} \sim 0.21m_0$ (m_0 is electron mass), so that $\mu \sim 0.13m_0$ and $M = m_{\rm e} + m_{\rm h} \sim 0.58m_0$.

The exciton binding energy in term of Bohr radius is given by

$$E_{\rm b} = \frac{\hbar^2}{2\mu a_{\rm B}^2}. \quad (13)$$

The Fourier transform of the internal wave function is

$$\phi(\mathbf{q}) = \int \mathrm{d}^2\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}}\psi_0(\mathbf{r}) = \frac{\sqrt{8\pi}a_{\rm B}}{(1+(qa_{\rm B})^2)^{3/2}}. \quad (14)$$

III. THE REPRESENTATION OF COULOMB INTERACTION IN BLOCH ENERGY BAND WAVEFUNCTION

The second quantized electron operator expanded in the basis of Bloch energy band wavefunction is

$$\psi_s(\mathbf{r}) = \sum_n \sum_{\mathbf{k}\in 1\text{st BZ}} \mathsf{c}_{n\mathbf{k}s}\,\psi_{n\mathbf{k}s}(\mathbf{r}), \quad (15)$$

where $\mathsf{c}_{n\mathbf{k}s}$ is the electron destruction operator with momentum $\mathbf{k}$ in the n-th band with spin s. The Bloch wavefunction is given by

$$\psi_{n\mathbf{k}s}(\mathbf{r}) = \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{\mathcal{V}}}u_{n\mathbf{k}}(\mathbf{r})\chi_s, \quad (16)$$

$u_{n\mathbf{k}}(\mathbf{r})$ is periodic over the cells of lattice and $u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}+\mathbf{G}}(\mathbf{r})$ ($\mathbf{G}$ is a reciprocal vector). $\mathcal{V}$ is the system volume (an area in our case).

The density operator can be expressed as

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_s \psi_s^\dagger(\mathbf{r})\psi_s(\mathbf{r}) \\ &= \sum_s \sum_{nn'\mathbf{k}\mathbf{k}'} \mathsf{c}^\dagger_{n'\mathbf{k}'s}\,\mathsf{c}_{n\mathbf{k}s}\Big[\psi^*_{n'\mathbf{k}'s}(\mathbf{r})\psi_{n\mathbf{k}s}(\mathbf{r})\Big].\end{aligned} \quad (17)$$

The density operator in momentum space is given by

$$\rho(\mathbf{q}) = \int \mathrm{d}^2\mathbf{r}\, e^{i\mathbf{q}\cdot\mathbf{r}} \sum_s \psi_s^\dagger(\mathbf{r})\psi_s(\mathbf{r}) \tag{18}$$
$$= \sum_s \sum_{nn'\mathbf{k}\mathbf{k}'} \mathsf{c}^\dagger_{n'\mathbf{k}'s}\,\mathsf{c}_{n\mathbf{k}s} \int \mathrm{d}^3\mathbf{r}\, e^{i\mathbf{q}\cdot\mathbf{r}} \Big[\psi^*_{n'\mathbf{k}'s}(\mathbf{r})\psi_{n\mathbf{k}s}(\mathbf{r})\Big].$$

The Coulomb interaction in momentum space is

$$\widehat{\mathsf{H}}_{\mathrm{C}} = \frac{1}{2\mathcal{V}} \sum_{\mathbf{q}} v(\mathbf{q}) \;:\rho(\mathbf{q})\rho(-\mathbf{q}):, \tag{19}$$

where :: indicates normal ordering, implying no self-contraction of density operator is allowed. By writing the spatial integral of Eq. (18) as the sum over lattice points and an integral over unit cell, $\rho(\mathbf{q})$ can be expressed as

$$\rho(\mathbf{q}) = \sum_s \sum_{nn'} \sum_{\mathbf{k},\mathbf{G}} \mathsf{c}^\dagger_{n'\mathbf{k}+\mathbf{q}+\mathbf{G}s}\mathsf{c}_{n\mathbf{k}s}\langle n'\mathbf{k}+\mathbf{q}|n\mathbf{k}\rangle_{\mathbf{G}}, \tag{20}$$

where v_{cell} is the volume of unit cell, and the band overlap integral is

$$\langle n'\mathbf{k}'|n\mathbf{k}\rangle_{\mathbf{G}} = \frac{1}{v_{\mathrm{cell}}} \int_{\mathrm{cell}} \mathrm{d}^2\mathbf{r} e^{-i\mathbf{r}\cdot\mathbf{G}} u_{n'\mathbf{k}'}(\mathbf{r})^\dagger u_{n\mathbf{k}}(\mathbf{r}). \tag{21}$$

Now let us project the Hilbert space to the truncated subspace of one valence band (whose electron operator denoted by $\mathsf{a}_{\mathbf{k}s}$) and one conduction band (whose electron operator denoted by $\mathsf{b}_{\mathbf{k}s}$). Then the band sum of Eq. (20) can be written as

$$\rho(\mathbf{q}) = \rho_{\mathsf{aa}}(\mathbf{q}) + \rho_{\mathsf{bb}}(\mathbf{q}) + \rho_{\mathsf{ab}}(\mathbf{q}) + \rho_{\mathsf{ba}}(\mathbf{q}), \tag{22}$$

where (the meaning of other density operators are self-evident)

$$\rho_{\mathsf{ab}}(\mathbf{q}) = \sum_{\mathbf{k},\mathbf{G},s} \mathsf{a}^\dagger_{\mathbf{k}+\mathbf{q}+\mathbf{G}s}\mathsf{b}_{\mathbf{k}s}\langle \mathrm{v},\mathbf{k}+\mathbf{q}|\mathrm{c},\mathbf{k}\rangle_{\mathbf{G}}, \quad \textit{etc.} \tag{23}$$

Expanding the momentum $\mathbf{k}$ of Eq. (23) in the vicinity of two valley points $\mathsf{K},\mathsf{K}' = -\mathsf{K}$, we can decompose the density operator $\rho_{ab}(\mathbf{q})$ into two parts coming from each valley.

$$\rho_{\mathsf{ab}}(\mathbf{q}) \approx \rho_{+\mathsf{ab}}(\mathbf{q}) + \rho_{-\mathsf{ab}}(\mathbf{q}), \quad + = \mathsf{K}, \quad - = \mathsf{K}'. \tag{24}$$

The same is true of $\rho_{\mathsf{aa}}(\mathbf{q}), \rho_{\mathsf{bb}}(\mathbf{q}), \rho_{\mathsf{ba}}(\mathbf{q})$, completing the specification of Coulomb interaction for monolayer TMDC.

IV. COMPUTATION OF MATRIX ELEMENTS

In this section we compute the transition matrix element $\langle \mathsf{f}|\widehat{\mathsf{H}}_{\mathrm{C}}|\mathsf{i}\rangle$. The initial two-exciton state is represented by

$$|\mathsf{i}\rangle = A^\dagger_{\tau s_{a1} s_{b2}}(\mathbf{P}_1) A^\dagger_{\tau' s_{a2} s_{b2}}(\mathbf{P}_2)|\mathrm{vac}\rangle. \tag{25}$$

$A^\dagger_{\tau s_a s_b}(\mathbf{P})$ is the creation operator for the exciton coming from valley τ with hole spin s_a and electron spin s_b, and the center of mass momentum $\mathbf{P}$. The vacuum state is the state where the valence band is completely filled and the conduction band is empty. The explicit form of the exciton operator is

$$A^\dagger_{\tau s_a s_b}(\mathbf{P}) = \sum_{\mathbf{p}} \frac{\phi^*(\mathbf{p})}{\sqrt{\mathcal{V}}}\, \mathsf{b}^\dagger_{\tau s_b}(\mathbf{p}+\frac{m_{\mathrm{e}}}{M}\mathbf{P})\, \mathsf{a}_{\tau \bar{s}_a}(\mathbf{p}-\frac{m_{\mathrm{h}}}{M}\mathbf{P}), \tag{26}$$

where $\mathbf{p}$ is the momentum measured from the valley points. In Eq. (26) the hole creation operator has been expressed in terms of the destruction operator of valence electron with opposite momentum and spin ($\bar{s}_a$ is the spin opposite to s_a). $\phi(\mathbf{p})$ is the internal wavefunction of exciton in momentum space (see Eq. (14)). M is the total mass of exciton which we simply took as $M = m_{\mathrm{e}} + m_{\mathrm{h}}$.

The final state is the ionized electron (with momentum $\mathbf{k}_{\mathrm{e}}$ and spin s_{e}) and the hole state (with momentum $\mathbf{k}_{\mathrm{h}}$ and spin s_{h}):

$$|\mathsf{f}\rangle = b^\dagger_{\tau \mathbf{k}_{\mathrm{e}} s_{\mathrm{e}}} a_{\tau', -\mathbf{k}_{\mathrm{h}} \bar{s}_{\mathrm{h}}}|\mathrm{vac}\rangle. \tag{27}$$

It is convenient to consider two separate cases: (1) two excitons coming from the same valley or (2) two opposite valleys.

1. Recombination of the excitons from the same valley

We have $\tau = \tau'$ in Eq. (25). As is clear from the structure of Feynman diagram of Fig. 1, the ionized electron and hole in the final state should stay in the same valley.

Let us consider the type A and type A' diagrams first. Since one of the exciton should be annihilated (meaning recombination), we need density operator $\rho_{\tau,\mathsf{ab}}(\mathbf{q})$. Following the electron and the hole line we can see that the type A and A' diagram is associated with $\rho_{\tau,\mathsf{bb}}(\mathbf{q})$ and $\rho_{\tau,\mathsf{aa}}(\mathbf{q})$, respectively. Then the matrix element for type A process can be expressed as

$$\begin{aligned} M_{\mathrm{A},\tau,11} = {} & \langle \mathrm{vac}| a^\dagger_{\tau,-\mathbf{k}_{\mathrm{h}}\bar{s}_{\mathrm{h}}} b_{\tau,\mathbf{k}_{\mathrm{e}}s_{\mathrm{e}}} \\ & \times \Big(\frac{1}{\mathcal{V}} \sum_{\mathbf{q}} v_c(\mathbf{q})\rho_{\tau\mathsf{bb}}(\mathbf{q})\rho_{\tau\mathsf{ab}}(-\mathbf{q})\Big) \\ & \times \sum_{\mathbf{p}_1} \frac{\phi^*(\mathbf{p}_1)}{\sqrt{\mathcal{V}}} \mathsf{b}^\dagger_{\tau s_{b1}}(\mathbf{p}_1 + \frac{m_{\mathrm{e}}}{M}\mathbf{P}_1)\mathsf{a}_{\tau\bar{s}_{a1}}(\mathbf{p}_1 - \frac{m_{\mathrm{h}}}{M}\mathbf{P}_1) \\ & \times \sum_{\mathbf{p}_2} \frac{\phi^*(\mathbf{p}_2)}{\sqrt{\mathcal{V}}} \mathsf{b}^\dagger_{\tau s_{b2}}(\mathbf{p}_2 + \frac{m_{\mathrm{e}}}{M}\mathbf{P}_2)\mathsf{a}_{\tau\bar{s}_{a2}}(\mathbf{p}_2 - \frac{m_{\mathrm{h}}}{M}\mathbf{P}_2)|\mathrm{vac}\rangle. \end{aligned} \tag{28}$$

The subscript 11 of $M_{\mathrm{A},\tau,11}$ is the reminder for the contraction for exciton 1 only. The final result must be

symmetrized under the interchange of 1 and 2 indices of the initial state. Similarly, the matrix element for type A' process is

$$\begin{aligned} M_{\mathrm{A}',\tau,11} &= \langle \mathrm{vac}| a^{\dagger}_{\tau,-\mathbf{k}_\mathrm{h}\bar{s}_\mathrm{h}} b_{\tau,\mathbf{k}_\mathrm{e}s_\mathrm{e}} \\ &\times \Big(\frac{1}{\mathcal{V}}\sum_{\mathbf{q}} v_c(\mathbf{q})\rho_{\tau\mathsf{aa}}(\mathbf{q})\rho_{\tau\mathsf{ab}}(-\mathbf{q})\Big) \\ &\times \sum_{\mathbf{p}_1} \frac{\phi^*(\mathbf{p}_1)}{\sqrt{\mathcal{V}}} \mathsf{b}^{\dagger}_{\tau s_{b1}}(\mathbf{p}_1+\frac{m_\mathrm{e}}{M}\mathbf{P}_1)\mathsf{a}_{\tau\bar{s}_{a1}}(\mathbf{p}_1-\frac{m_\mathrm{h}}{M}\mathbf{P}_1) \\ &\times \sum_{\mathbf{p}_2} \frac{\phi^*(\mathbf{p}_2)}{\sqrt{\mathcal{V}}} \mathsf{b}^{\dagger}_{\tau s_{b2}}(\mathbf{p}_2+\frac{m_\mathrm{e}}{M}\mathbf{P}_2)\mathsf{a}_{\tau\bar{s}_{a2}}(\mathbf{p}_2-\frac{m_\mathrm{h}}{M}\mathbf{P}_2)|\mathrm{vac}\rangle. \end{aligned} \tag{29}$$

Now all we have to do is to perform the normal orderings (which is of course just Wick contraction of many-body perturbation theory) while remembering the explicit form of density operators Eqs. (20), (23), and (24).

The annihilating exciton 1 with $\rho_{\tau\mathsf{ab}}(-\mathbf{q})$ yields (for both type A and type A')

$$\mathbf{q} = \mathbf{P}_1 + \mathbf{G}, \quad s_{b1} = \bar{s}_{a1}. \tag{30}$$

The spin condition means that the spin of the exciton 1 must be singlet. Recalling that $v_c(\mathbf{q}) \sim 1/|\mathbf{q}|$ and that $\mathbf{G}$ is a reciprocal vector and $\mathbf{P}_1$ is of the magnitude of thermal momentum, we can see that the umklapp process of nonzero $\mathbf{G}$ is severely suppressed. We mention that the annihilation of spin triplet exciton is also possible via contraction to ρ_{aa} or ρ_{bb}, but in such case the momentum transfer becomes too large, so that those processes are neglected here.

The overlap integral of the operator $\rho_{\tau\mathsf{ab}}(-\mathbf{q})$ is given by

$$\langle \mathrm{v},\mathbf{k}-\mathbf{q}|\mathrm{c},\mathbf{k}\rangle_{\tau,\mathbf{G}} = \langle \mathrm{v},\mathbf{p}_1-\frac{m_\mathrm{h}}{M}\mathbf{P}_1|\mathrm{c},\mathbf{p}_1+\frac{m_\mathrm{e}}{M}\mathbf{P}_1\rangle_{\tau,\mathbf{G}}. \tag{31}$$

For $\mathbf{G}=0$ (normal process), Eq. (31) can be computed with the aid of Eq. (8).

The remaining (unique) Wick contractions connect the exciton 2 with the final state of ionized electron and hole through the density operator $\rho_{\tau\mathsf{bb}}(\mathbf{q})$ (type A) and $\rho_{\tau\mathsf{aa}}(\mathbf{q})$(type A'). Two processes contribute to the matrix element with opposite sign due to the anti-commutativity fermion operators.

For Type A with obtain

$$\begin{aligned} &s_\mathrm{e} = s_{b2}, \quad s_\mathrm{h} = s_{a2}, \quad \mathbf{p}_2 = \frac{m_\mathrm{h}}{M}\mathbf{P}_2 - \mathbf{k}_\mathrm{h}, \\ &\langle \mathrm{c},\mathbf{k}'+\mathbf{q}|\mathrm{c},\mathbf{k}',\rangle_{\tau,\mathbf{G}} = \langle \mathrm{c},\mathbf{k}_\mathrm{e}|\mathrm{c},\mathbf{k}_\mathrm{e}-\mathbf{P}_1\rangle_{\tau,\mathbf{G}}, \quad \text{type A}. \end{aligned} \tag{32}$$

For Type A',

$$\begin{aligned} &s_\mathrm{e} = s_{b2}, \quad s_\mathrm{h} = s_{a2}, \quad \mathbf{p}_2 = \mathbf{k}_\mathrm{e} - \frac{m_\mathrm{e}}{M}\mathbf{P}_2, \\ &\langle \mathrm{v},\mathbf{k}'+\mathbf{q}|\mathrm{v},\mathbf{k}'\rangle_{\tau,\mathbf{G}} = \langle \mathrm{v},-\mathbf{k}_\mathrm{h}+\mathbf{P}_1|\mathrm{v},-\mathbf{k}_\mathrm{h},\rangle_{\tau,\mathbf{G}}, \text{type A'}. \end{aligned} \tag{33}$$

The spin of the electron and the hole in the final state are identical with those of exciton 2, just as expected. Combining the momentum relations, we obtain the conservation of total momentum

$$\mathbf{k}_\mathrm{e}+\mathbf{k}_\mathrm{h} = \mathbf{P}_1+\mathbf{P}_2+\mathbf{G}+\mathbf{G}'. \tag{34}$$

Since $\mathbf{k}_\mathrm{e}+\mathbf{k}_\mathrm{h}$ must reside in the vicinity of valley (remember that $\mathbf{k}_\mathrm{e},\mathbf{k}_\mathrm{h}$ are measured from the valley points) and $\mathbf{P}_{1,2}$ are thermal, Eq. (34) can be satisfied only for

$$\mathbf{G}+\mathbf{G}' = 0. \tag{35}$$

As has been argued above, the normal process with with $\mathbf{G}=0$ is dominant one, so that

$$\mathbf{G} = \mathbf{G}' = 0, \quad \text{for Type A and A' processes.} \tag{36}$$

Then this leads to the restricted momentum conservation

$$\mathbf{k}_\mathrm{e}+\mathbf{k}_\mathrm{h} = \mathbf{P}_1+\mathbf{P}_2. \tag{37}$$

Furthermore, since $\mathbf{P}_1$ and $\mathbf{P}_2$ are small thermal momentum compared to generic electronic momentum, Eq. (37) implies that the electron and the hole momentum in the final state should be almost opposite to each other.

The sum of the matrix element for Type A and A' processes is

$$\begin{aligned} &M_{\mathrm{A},\tau} + M_{\mathrm{A}'\tau} \\ &= \frac{1}{\mathcal{V}} v_c(\mathbf{P}_1)\sum_{\mathbf{p}_1}\frac{\phi(\mathbf{p}_1)}{\mathcal{V}}\langle \mathrm{v},\mathbf{p}_1-\frac{m_\mathrm{h}}{M}\mathbf{P}_1|\mathrm{c},\mathbf{p}_1+\frac{m_\mathrm{e}}{M}\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0} \\ &\quad\times\Big(\phi(\frac{m_\mathrm{h}}{M}\mathbf{P}_2-\mathbf{k}_\mathrm{h})\langle \mathrm{c},\mathbf{k}_\mathrm{e}|\mathrm{c},\mathbf{k}_\mathrm{e}-\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0} \\ &\qquad-\phi(\mathbf{k}_\mathrm{e}-\frac{m_\mathrm{e}}{M}\mathbf{P}_2)\langle \mathrm{v},-\mathbf{k}_\mathrm{h}+\mathbf{P}_1|\mathrm{v},-\mathbf{k}_\mathrm{h},\rangle_{\tau,\mathbf{G}=0}\Big) \\ &\quad+\big(1\leftrightarrow 2\big) \end{aligned} \tag{38}$$

Since $\langle \mathbf{p}_1-\frac{m_\mathrm{h}}{M}\mathbf{P}_1, v|\mathbf{p}_1+\frac{m_\mathrm{e}}{M}\mathbf{P}_1, c\rangle_{\tau,\mathbf{G}=0}$ and the quantity in parenthesis of Eq. (38) vanish in the limit $\mathbf{P}_1,\mathbf{P}_2\to 0$, and taking the fact $v_c(\mathbf{P}_1)\propto 1/|\mathbf{P}_1|$ into account we can see that $M_{\mathrm{A},\tau}+M_{\mathrm{A}'\tau}$ proportional to linear combinations of $\mathbf{P}_1$ and $\mathbf{P}_2$. From Eq. (8) we find

$$\langle \mathrm{v},\mathbf{p}_1-\frac{m_\mathrm{h}}{M}\mathbf{P}_1|\mathrm{c},\mathbf{p}_1+\frac{m_\mathrm{e}}{M}\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0} \approx \frac{vP_{1x}\tau+viP_{1y}}{\Delta'_{\tau s}}. \tag{39}$$

$\mathbf{p}_1$ integral can be easily performed $\frac{1}{\mathcal{V}}\sum_{\mathbf{p}_1}\phi(\mathbf{p}_1) = \sqrt{\frac{2}{\pi}}\frac{1}{a_\mathrm{B}}$. As can be seen from Eq. (9), $\langle \mathrm{c},\mathbf{k}_\mathrm{e}|\mathrm{c},\mathbf{k}_\mathrm{e}-\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0} \sim 1+\frac{v^2Pk_\mathrm{e}}{\Delta'^2}$, while the correction coming from the internal wavefunction is larger by a factor of $\Delta^2/v^2k_\mathrm{e}^2$. Thus in our approximation we can set the conduction and valence band overlap integral in the parenthesis of Eq. (38) to 1. Typical electron and hole

momenta is greater than $a_{\rm B}^{-1}$, so that the quantity in the parenthesis of Eq. (38) can be approximated to $3\sqrt{8\pi}\mathbf{P}_2\cdot\mathbf{k}_{\rm e}/k_{\rm e}^5a_{\rm B}^2$. With these approximations taken into account, Eq. (38) becomes

$$M_{{\rm A},\tau}+M_{{\rm A}'\tau}=12\frac{v_c(\mathbf{P}_1)}{\mathcal{V}}\frac{\mathbf{P}_2\cdot\mathbf{k}_{\rm e}}{k_{\rm e}^5a_{\rm B}^3}\frac{vP_{1x}\tau+ivP_{1y}}{\Delta'_{\tau s}}+(1\leftrightarrow 2). \tag{40}$$

In the above approximation, $\mathbf{k}_{\rm e}\sim-\mathbf{k}_{\rm h}$ has been used. So there is only one momentum integration in the final state.

Next turn to type B and B' processes. In these processes, the operator $\rho_{\tau\mathsf{ab}}$ annihilate electron and hole operators of *different* excitons.

$$\begin{aligned}M_{{\rm B},\tau,12}&=\langle{\rm vac}|a^\dagger_{\tau,-\mathbf{k}_{\rm h}\bar{s}_{\rm h}}b_{\tau,\mathbf{k}_{\rm e}s_{\rm e}}\\&\times\Big(\frac{1}{\mathcal{V}}\sum_{\mathbf{q}}v_c(\mathbf{q})\rho_{\tau\mathsf{bb}}(\mathbf{q})\rho_{\tau\mathsf{ab}}(-\mathbf{q})\Big)\\&\times\sum_{\mathbf{p}_1}\frac{\phi^*(\mathbf{p}_1)}{\sqrt{\mathcal{V}}}\mathsf{b}^\dagger_{\tau s_{b1}}(\mathbf{p}_1+\frac{m_{\rm e}}{M}\mathbf{P}_1)\mathsf{a}_{\tau\bar{s}_{a1}}(\mathbf{p}_1-\frac{m_{\rm h}}{M}\mathbf{P}_1)\\&\times\sum_{\mathbf{p}_2}\frac{\phi^*(\mathbf{p}_2)}{\sqrt{\mathcal{V}}}\mathsf{b}^\dagger_{\tau s_{b2}}(\mathbf{p}_2+\frac{m_{\rm e}}{M}\mathbf{P}_2)\mathsf{a}_{\tau\bar{s}_{a2}}(\mathbf{p}_2-\frac{m_{\rm h}}{M}\mathbf{P}_2)|{\rm vac}\rangle.\end{aligned} \tag{41}$$

The subscript 12 of $M_{{\rm B},\tau,12}$ is the reminder for the contraction of electron and hole from *different* excitons. The final result again must be symmetrized under the interchange of 1 and 2 indices of the initial state.

Annihilating the electron of exciton 1 and the hole of exciton 2 using the operator $\rho_{\tau\mathsf{ab}}$, we obtain

$$\begin{aligned}&s_{b1}=\bar{s}_{a2},\ \ \mathbf{p}_1=-\mathbf{k}_{\rm h}+\frac{m_{\rm h}}{M}\mathbf{P}_1,\\&\mathbf{q}=\mathbf{G}-\mathbf{k}_{\rm h}-\mathbf{p}_2+\mathbf{P}_1+(m_{\rm h}/M)\mathbf{P}_2.\end{aligned} \tag{42}$$

An internal momentum $\mathbf{p}_1$ is fixed by the exciton center of mass momentum and the momentum of the hole in the final state, while $\mathbf{p}_2$ is unconstrained. As is clear from Eq. (14), the effective cutoff for the momentum $\mathbf{p}_2$ is $a_{\rm B}^{-1}$. Both $a_{\rm B}^{-1}$ and $|\mathbf{k}_{\rm h}|$ is much smaller than the reciprocal vector $\mathbf{G}$. Recalling that $\mathbf{P}_1,\mathbf{P}_2$ are thermal momenta, the Coulomb matrix element $v_c(\mathbf{q})$ with the momentum transfer $\mathbf{q}$ for the umklapp process with nonzero $\mathbf{G}$ is greatly suppressed compared to normal process, being similar to type A, A' cases. However, there is no low momentum enhancement of Coulomb interaction as in type A, A' processes. The momentum transfer is set by the electronic momentum scale or $1/a_{\rm B}$. The momentum conservation Eq. (34) also holds, and for the normal processes Eq. (37) is valid. Then the overlap integrals for type B process are given by

$$\begin{aligned}&\langle{\rm v},\mathbf{k}-\mathbf{q}|{\rm c},\mathbf{k}\rangle_{\tau,\mathbf{G}=0}\\&=\langle{\rm v},\mathbf{p}_2-\frac{m_{\rm h}}{M}\mathbf{P}_2|{\rm c},-\mathbf{k}_{\rm h}+\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0},\\&\langle{\rm c},\mathbf{k}'+\mathbf{q}|{\rm c},\mathbf{k}'\rangle_{\tau,\mathbf{G}'=0}\\&=\langle{\rm c},\mathbf{k}_{\rm e}|{\rm c},\mathbf{p}_2+\frac{m_{\rm e}}{M}\mathbf{P}_2\rangle_{\tau,\mathbf{G}'=0},\quad\text{type B.}\end{aligned} \tag{43}$$

The spin of electron and hole for final state is

$$s_{\rm e}=s_{b2},\quad s_{\rm h}=s_{a1}. \tag{44}$$

Similarly, the matrix element for type B' process is given by

$$\begin{aligned}M_{{\rm B}',\tau,12}&=\langle{\rm vac}|a^\dagger_{\tau,-\mathbf{k}_{\rm h}\bar{s}_{\rm h}}b_{\tau,\mathbf{k}_{\rm e}s_{\rm e}}\\&\times\Big(\frac{1}{\mathcal{V}}\sum_{\mathbf{q}}v_c(\mathbf{q})\rho_{\tau\mathsf{aa}}(\mathbf{q})\rho_{\tau\mathsf{ab}}(-\mathbf{q})\Big)\\&\times\sum_{\mathbf{p}_1}\frac{\phi^*(\mathbf{p}_1)}{\sqrt{\mathcal{V}}}\mathsf{b}^\dagger_{\tau s_{b1}}(\mathbf{p}_1+\frac{m_{\rm e}}{M}\mathbf{P}_1)\mathsf{a}_{\tau\bar{s}_{a1}}(\mathbf{p}_1-\frac{m_{\rm h}}{M}\mathbf{P}_1)\\&\times\sum_{\mathbf{p}_2}\frac{\phi^*(\mathbf{p}_2)}{\sqrt{\mathcal{V}}}\mathsf{b}^\dagger_{\tau s_{b2}}(\mathbf{p}_2+\frac{m_{\rm e}}{M}\mathbf{P}_2)\mathsf{a}_{\tau\bar{s}_{a2}}(\mathbf{p}_2-\frac{m_{\rm h}}{M}\mathbf{P}_2)|{\rm vac}\rangle.\end{aligned} \tag{45}$$

The momentum conservation Eq. (37) also applies to type B', and we get

$$s_{b1}=\bar{s}_{a2},\ \ \mathbf{p}_2=\mathbf{k}_{\rm e}-\frac{m_{\rm e}}{M}\mathbf{P}_2,\ \ \mathbf{q}=\mathbf{k}_{\rm h}+\mathbf{p}_1-(m_{\rm h}/M)\mathbf{P}_1. \tag{46}$$

For type B' process, $\mathbf{p}_1$ is unconstrained. The overlap integrals are given by

$$\begin{aligned}&\langle{\rm v},\mathbf{k}-\mathbf{q}|{\rm c},\mathbf{k},\rangle_{\tau,\mathbf{G}=0}=\langle{\rm v},\mathbf{k}_{\rm e}-\mathbf{P}_2|{\rm c},\mathbf{p}_1+\frac{m_{\rm e}}{M}\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0}\\&\langle{\rm v},\mathbf{k}'+\mathbf{q}|{\rm v},\mathbf{k}'\rangle_{\tau,\mathbf{G}'=0}=\langle{\rm v},\mathbf{p}_1-\frac{m_{\rm h}}{M}\mathbf{P}_1|{\rm v},-\mathbf{k}_{\rm h}\rangle_{\tau,\mathbf{G}'=0},\end{aligned}\quad\text{Type B'} \tag{47}$$

The explicit form of the matrix elements for type B processes is then

$$\begin{aligned}M_{{\rm B},\tau,12}&=(-1)\phi(-\mathbf{k}_{\rm h}+\frac{m_{\rm h}}{M}\mathbf{P}_1)\\&\times\sum_{\mathbf{p}_2}\phi(\mathbf{p}_2)\frac{1}{\mathcal{V}^2}v_c(-\mathbf{k}_{\rm h}-\mathbf{p}_2+\mathbf{P}_1+\frac{m_{\rm h}}{M}\mathbf{P}_2)\\&\times\langle{\rm v},\mathbf{p}_2-\frac{m_{\rm h}}{M}\mathbf{P}_2|{\rm c},-\mathbf{k}_{\rm h}+\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0}\\&\times\langle{\rm c},\mathbf{k}_{\rm e}|{\rm c},\mathbf{p}_2+\frac{m_{\rm e}}{M}\mathbf{P}_2\rangle_{\tau,\mathbf{G}'=0}+(1\leftrightarrow 2).\end{aligned} \tag{48}$$

The overlap integral between valence and conduction band is given by (using Eq. (8))

$$\begin{aligned}&\langle{\rm v},\mathbf{p}_2-\frac{m_{\rm h}}{M}\mathbf{P}_2|{\rm c},-\mathbf{k}_{\rm h}+\mathbf{P}_1\rangle_{\tau,\mathbf{G}=0}\\&\qquad\approx\frac{-v(k_{{\rm h}x}+p_{2x})\tau-iv(k_{{\rm h}y}+p_{2y})}{\Delta'_{\tau s}}.\end{aligned} \tag{49}$$

In our $\mathbf{k}\cdot\mathbf{p}$ approximation, the overlap integral between conduction band can be taken to be 1. So Eq. (48) simplifies to

$$M_{\mathrm{B},\tau,12} \approx 2\times\phi(-\mathbf{k}_\mathrm{h})\frac{1}{\mathcal{V}^2}\sum_{\mathbf{p}_2}\phi(\mathbf{p}_2)v_c(-\mathbf{k}_\mathrm{h}-\mathbf{p}_2) \times\frac{v(k_{\mathrm{h}x}+p_{2x})\tau+iv(k_{\mathrm{h}y}+p_{2y})}{\Delta'_{\tau s}}. \quad (50)$$

Note that $\mathbf{P}_1$ and $\mathbf{P}_2$ disappeared from the integral in our approximation, so that no temperature dependence is expected for Type B, B' processes just like the result of [19]. Performing similar calculations for type B' processes, we arrive at

$$M_{\mathrm{B}',\tau,12} \approx 2\times\phi(+\mathbf{k}_\mathrm{e})\frac{1}{\mathcal{V}^2}\sum_{\mathbf{p}_1}\phi(\mathbf{p}_1)v_c(+\mathbf{k}_\mathrm{h}+\mathbf{p}_1) \times\frac{v(p_{1x}-k_{\mathrm{e}x})+iv(p_{1y}-k_{\mathrm{e}y})}{\Delta'_{\tau s}}. \quad (51)$$

In the approximation of $\mathbf{k}_\mathrm{e}\approx-\mathbf{k}_\mathrm{h}$, Eq. (50) and Eq. (51) become almost identical.

The momentum integrals of Eqs. (50) and (51) can be approximately done by keeping the leading order term of multipole expansion. Later we will see that $a_\mathrm{B}k_\mathrm{h} > (2\sim 3)$, and in this limit we obtain

$$M_{\mathrm{B}+B',\tau}\sim 2\frac{\phi(\mathbf{k}_\mathrm{e})}{\mathcal{V}}\,\frac{e^2}{\epsilon}\,\frac{1}{k_\mathrm{h}a_\mathrm{B}}\,\frac{(-1)v(k_{\mathrm{e},x}\tau+ik_{\mathrm{e},y})}{\Delta'_{\tau s}}. \quad (52)$$

2. Recombination of excitons from different valleys

Now the initial state is given by

$$|\mathrm{i}\rangle = A^\dagger_{+,\bar{s}_{a1}s_{b1}}(\mathbf{P}_1)A^\dagger_{-,\bar{s}_{a2}s_{b2}}(\mathbf{P}_2)|\mathrm{vac}\rangle. \quad (53)$$

The final state is of the same form as Eq. (27). Since two excitons come from different valley, the coulomb interaction must include the contributions from both valleys:

$$\sum_{\mathbf{q}}\frac{v_c(\mathbf{q})}{\mathcal{V}}\,(\rho_{+\mathsf{aa}}+\rho_{-\mathsf{aa}}+\rho_{+\mathsf{bb}}+\rho_{-\mathsf{bb}})_{\mathbf{q}}(\rho_{+\mathsf{ab}}+\rho_{-\mathsf{ab}})_{-\mathbf{q}}. \quad (54)$$

We can immediately see that type B and B' processes are suppressed since there is no density operators in $\mathbf{k}\cdot\mathbf{p}$ approximation which can annihilate electron and hole of different excitons (The annihilation is possible in the exchange channel but its magnitude will be much smaller).

Let us consider type A process first. If we annihilate exciton 1 (with valley K), then the ionized electron and hole will come from exciton 2 (with valley K'). The computations are very similar to those of the identical valley case, except that the overlap integral of conduction electron is that of valley K'.

$$\begin{aligned}[M_\mathrm{A}+M_{\mathrm{A}'}]_{12} &= \frac{1}{\mathcal{V}^2}\,v_c(\mathbf{P}_1)\sum_{\mathbf{p}_1}\phi(\mathbf{p}_1)\\ &\times\langle\mathbf{p}_1-\frac{m_\mathrm{h}}{M}\mathbf{P}_1,v|\mathbf{p}_1+\frac{m_\mathrm{e}}{M}\mathbf{P}_1,c\rangle_{+,\mathbf{G}=0}\\ &\times\Big(\phi(\frac{m_\mathrm{h}}{M}\mathbf{P}_2-\mathbf{k}_\mathrm{h})\langle\mathrm{c},\mathbf{k}_\mathrm{e}|\mathrm{c},\mathbf{k}_\mathrm{e}-\mathbf{P}_1\rangle_{-,\mathbf{G}=0}\\ &\quad-\phi(\mathbf{k}_\mathrm{e}-\frac{m_\mathrm{e}}{M}\mathbf{P}_2)\langle\mathrm{v},-\mathbf{k}_\mathrm{h}+\mathbf{P}_1|\mathrm{v},-\mathbf{k}_\mathrm{h},\rangle_{-,\mathbf{G}=0}\Big).\end{aligned} \quad (55)$$

With two excitons interchanged, the valley is also exchanged, so that the above result is modified to

$$\begin{aligned}[M_\mathrm{A}+M_{\mathrm{A}'}]_{21} &= \frac{1}{\mathcal{V}^2}\,v_c(\mathbf{P}_2)\sum_{\mathbf{p}_1}\phi(\mathbf{p}_1)\\ &\times\langle\mathbf{p}_1-\frac{m_\mathrm{h}}{M}\mathbf{P}_2,v|\mathbf{p}_1+\frac{m_\mathrm{e}}{M}\mathbf{P}_2,c\rangle_{-,\mathbf{G}=0}\\ &\times\Big(\phi(\frac{m_\mathrm{h}}{M}\mathbf{P}_1-\mathbf{k}_\mathrm{h})\langle\mathrm{c},\mathbf{k}_\mathrm{e}|\mathrm{c},\mathbf{k}_\mathrm{e}-\mathbf{P}_2\rangle_{+,\mathbf{G}=0}\\ &\quad-\phi(\mathbf{k}_\mathrm{e}-\frac{m_\mathrm{e}}{M}\mathbf{P}_1)\langle\mathrm{v},-\mathbf{k}_\mathrm{h}+\mathbf{P}_2|\mathrm{v},-\mathbf{k}_\mathrm{h},\rangle_{+,\mathbf{G}=0}\Big).\end{aligned} \quad (56)$$

The full matrix element is the sum of Eqs. (55) and (56). The approximations leading to Eq. (40) can be employed except that the interchange of 1 and 2 indices also interchanges the valley index.

$$\begin{aligned}M_{\mathrm{A},\tau}+M_{\mathrm{A}'\tau} &\approx 12\frac{v_c(\mathbf{P}_1)}{\mathcal{V}}\,\frac{\mathbf{P}_2\cdot\mathbf{k}_\mathrm{e}}{k_\mathrm{e}^5a_\mathrm{B}^3}\,\frac{vP_{1x}+ivP_{1y}}{\Delta'_{+s}}\\ &+12\frac{v_c(\mathbf{P}_2)}{\mathcal{V}}\,\frac{\mathbf{P}_1\cdot\mathbf{k}_\mathrm{e}}{k_\mathrm{e}^5a_\mathrm{B}^3}\,\frac{vP_{2x}(-1)+ivP_{2y}}{\Delta'_{-s}}.\end{aligned} \quad (57)$$

V. COMPUTATION OF AUGER RECOMBINATION RATE

The Auger recombination rate can be computed from Eq. (1) using the matrix elements obtained in section IV The final state momentum sum is constrained by the momentum conservation (of normal processes) $\mathbf{k}_\mathrm{e}+\mathbf{k}_\mathrm{h} = \mathbf{P}_1+\mathbf{P}_2$. Since $\mathbf{P}_1$ and $\mathbf{P}_2$ are thermal, $\mathbf{k}_\mathrm{e}\approx-\mathbf{k}_\mathrm{h}$, effectively there is only one independent momentum in the sum. The final state spin configuration is uniquely determined by the initial state and by the process type. At our level of approximation, it just modifies order one constant, so that we will not elaborate on it.

Next we take the energy conservation into account. The initial state energy is the sum of two exciton energies, while the final state energy is the sum of the energy of ionized electron and hole (recall E_b is the binding en-

ergy and Δ is the (fundamental) energy gap)

$$E_{\rm e}+E_{\rm h}=E_{\mathbf{P}_1}+E_{\mathbf{P}_2},$$
$$\Delta+\frac{\hbar^2k_{\rm e}^2}{2m_{\rm e}}+\frac{\hbar^2k_{\rm h}^2}{2m_{\rm h}}=(\Delta-E_{\rm b}+\frac{P_1^2}{2M})+(\Delta-E_{\rm b}+\frac{P_2^2}{2M}). \tag{58}$$

Neglecting thermal kinetic energy of excitons and employing $\mathbf{k}_{\rm e}\sim-\mathbf{k}_{\rm h}$,

$$\frac{\hbar^2\mathbf{k}_{\rm e}^{*2}}{2\mu}=\Delta-2E_{\rm b}, \tag{59}$$

which *fixes* the energy and the momentum of final state electron (and hole) (see [19] for a similar result).

Since $\Delta-2E_{\rm b}\sim(1\sim1.5)$eV, $k_{\rm e}$ is quite sizable. Using the estimate of various parameters mentioned in the previous sections, we find

$$k_{\rm e}a_{\rm B}\approx(2\sim4). \tag{60}$$

More precisely, we have to insert the $(1-n_{\mathbf{k}_{\rm e}})(1-n_{\mathbf{k}_{\rm h}})$ factor for the final state sum. As we have argued the energy of electron and hole in the final state is high, so that the thermal populations can neglected: $n_{\mathbf{k}_{\rm e}}\approx 0, n_{\mathbf{k}_{\rm h}f}\approx 0$.

Now the Auger recombination rate can be expressed as

$$\begin{aligned}\Gamma&=\frac{2\pi}{\hbar}\sum_{\rm i}\rho_{\rm i}\sum_{\mathbf{k}_{\rm e},s,\tau}|M|^2\delta(\hbar^2\mathbf{k}_{\rm e}^2/2\mu-(E_g-2E_{\rm b}))\\&=\frac{2\pi}{\hbar}\frac{\mathcal{V}}{4\pi}\frac{2\mu}{\hbar^2}\int\frac{{\rm d}\theta}{2\pi}\left\langle|M|^2_{k_{\rm e}=k_{\rm e}^*}\right\rangle,\end{aligned} \tag{61}$$

where $\langle\cdots\rangle$ indicates the implicit average over initial states and the sum over appropriate spin and valley configurations. θ is the azimuth angle of $\mathbf{k}_{\rm e}$.

When two excitons come from the same valley we have to compute $|M_{\rm A}+M_{\rm B}|^2$. First using Eq. (40), we find

$$\mathcal{V}^2\int_0^{2\pi}\frac{{\rm d}\theta}{2\pi}\,|M_{{\rm A},\tau}+M_{{\rm A}',\tau}|^2\approx 288\left(\frac{2\pi e^2}{\epsilon}\right)^2\frac{P^2v^2}{\Delta^2a_{\rm B}^6(k_{\rm e}^*)^8}. \tag{62}$$

With thermal average $P^2=2Mk_{\rm B}T$ (in 2-dimension), Eq. (62) is proportional to temperature. Next from Eq. (52), it follows that

$$\mathcal{V}^2\int_0^{2\pi}\frac{{\rm d}\theta}{2\pi}\,|M_{{\rm B},\tau}+M_{{\rm B}',\tau}|^2\approx\frac{32\pi}{(a_{\rm B}k_{\rm e})^3}\left(\frac{e^2v}{\epsilon\Delta}\right)^2. \tag{63}$$

Eq. (63) does not depend on P^2, which means that it is temperature *independent*. It turns out that the mixing term

$$\mathcal{V}^2\int_0^{2\pi}\frac{{\rm d}\theta}{2\pi}\,(M_{\rm A}M_{\rm B}^*+M_{\rm A}^*M_{\rm B}) \tag{64}$$

is proportional to $\mathbf{P}_1\cdot\mathbf{P}_2$, which vanishes upon the average over initial states.

Upon including the contribution of the recombination from excitons with different valleys (upon angle integration of the matrix element squared, the valley difference is gone, so essentially the identical contributions as the same valley case are obtained), we arrive at the final result:

$$\Gamma=\frac{1}{\hbar}\frac{1}{\mathcal{V}}\frac{\mu}{\hbar^2}\left(\frac{e^2\hbar v}{\epsilon\Delta}\right)^2\frac{1}{(a_{\rm B}k_{\rm e}^*)^3}\times\left[c_1\frac{k_{\rm B}T}{(\hbar k_{\rm e}^*)^2/2M}\frac{1}{(a_{\rm B}k_{\rm e}^*)^2}+c_2\right], \tag{65}$$

where $c_1=288\times(2\pi)^2$ and $c_2=32\pi$, and these numerical constants should *not* be taken too seriously since many approximations (mostly in estimating averages) have been involved in obtaining this result. Main factor contributing to these numbers is $\sqrt{8\pi}$ of Eq. (14). One can also check that indeed Γ has the correction dimension of inverse time.

The Auger coefficient A is defined to be $\Gamma\mathcal{V}$ which has the dimension of [area/time]. Expressing the Auger coefficient A using the relation $\frac{1}{a_{\rm B}k_{\rm e}^*}=\left(\frac{E_{\rm b}}{\Delta-2E_{\rm b}}\right)^{1/2}$, we get

$$A=\Gamma\mathcal{V}=\frac{1}{\hbar}\,\mu\left(\frac{e^2v}{\epsilon\Delta}\right)^2\left(\frac{E_{\rm b}}{\Delta-2E_{\rm b}}\right)^{3/2}\times\left[c_1\frac{k_{\rm B}T}{(\hbar k_{\rm e}^*)^2/2M}\frac{1}{(a_{\rm B}k_{\rm e}^*)^2}+c_2\right]. \tag{66}$$

Using $v=t_{\rm hop}a\sim3\times10^7$ cm/sec we can estimate the numerical value for A to be the order of 10^{-2} cm^2/s (for the temperature in the range of 100 - 300 K) , which is consistent with the experimental data of [12]. In these temperature ranges, the manitudes of two contributions are comparable.

VI. DISCUSSIONS AND SUMMARY

First, let us compare our result Eq. (66) with Eq. (14) of [19]. Ref. 19 studied the Auger recombination of 1-dimensional system (*e.g.*, carbon nanotubes) with *short* range interaction ($V(r)\propto\delta(r)$). In such case, there is no enhancement of Coulomb matrix element from the low momentum transfer, so that type A and A' contributions are severely suppressed, and only type B and B' processes contribute. Thus no temperature dependence is expected for the result of [19]. It is also interesting to compare the dependence on $E_{\rm b}/\Delta$. The c_2 part of Eq. (66) has the dependence of $\left(\frac{E_{\rm b}}{\Delta-2E_{\rm b}}\right)^{3/2}$, while Eq. (14) of [19] indicates $(E_{\rm b}/\Delta)^3$. Thus the Auger decay rate of TMDC is much more enhanced than that of 1D system with short range Coulomb interaction.

Second, the Auger constant (annihilation rate) of $A = (4.3 \pm 1.1) \times 10^{-2}$ cm$^2/s$ has been reported experimentally in [12]. Even though our calculations are too crude for the quantitative comparison, it is quite encouraging that both theoretical and experimental results are consistent with each other at the level of the order of magnitude estimates.

Such a large A value of the order of 10^{-2} cm$^2/s$ implies two excitons of size 1nm with distance 100 nm away can recombine. Evidently, this is due to the (underscreened) long range nature Coulomb interaction. Also, this large value means that the exciton lifetime is severely limited by nonradiative Auger decay even at a low exciton density ($\sim 10^{10}$ cm^{-2}). Considering a few nanosecond of radiative lifetime of en exciton, this will in turn limit the luminescence quantum yields of a 2D TMDC monolayer in the optoelectronic device architecture [12–14].

Summarizing, we have computed the Auger decay rate of exciton recombinations of monolayer TMDC based the approach proposed by Kavoulakis and Baym. Both temperature dependent (proportional to T) and temperature independent contributions are obtained. The former essentially comes from the long range nature of Coulomb interaction, while the latter is due to the exchange channel momentum transfer where the long range nature is not effective. Also, the numerical estimate of our result is consistent with the existing experimental data.

ACKNOWLEDGMENTS

The author is grateful to Prof. Joon Ik Jang for the suggestion of this problem and for very fruitful discussions and comments on manuscript. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2016R 1D 1A 1B03930125).

REFERENCES

[1] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov and A. K. Geim, Proc. Natl. Aca. Sci.U.S.A. **102**, 10451 (2005).
[2] A. K. Geim and I. V. Grigorieva, Nature **499**, 419 (2013).
[3] Y. Yoon, K. Ganapathi and S. Salahuddin, Nano Lett. **11**, 3768 (2011).
[4] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti and A. Kis, Nat. Nanotech. **6**, 147 (2011).
[5] D. Xiao, G. B. Liu, W. Feng, X. Xu and W. Yao, Phys. Rev. Lett. **108**, 196802 (2012).
[6] X. Xu, W. Yao, D. Xiao and T. F. Heinz, Nat. Phys. **10**, 343 (2014).
[7] K. F. Mak, C. Lee, J. Hone, J. Shan and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).
[8] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C-Y. Chim, G. Galii and F. Wang, Nano Lett. **10**, 1271 (2010).
[9] D. Y. Qiu, F. H. daJornada and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (2013).
[10] A. Ramasubramaniam, Phys. Rev. B **86**, 115409 (2012).
[11] 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).
[12] D. Sun, Y. Rao, G. A. Reider, G. Chen, Y. You, L. Brézin, A. R. Harutyunyan and T. Heinz, Nano Lett. **14**, 5625 (2014).
[13] M. Amani, P. Taheri, T. Addou, G. H. Ahn, D. Kiriya, D-H. Lien, J. W. Ager, R. M. Wallace and A. Javey, Nano Lett. **16**, 2786 (2016).
[14] Y. Lee, G. Ghimire, S. Roy, Y. Kim, C. Seo, A. K. Sood, J. I. Jang and J. Kim, ACS Photonics **5**, 2904 (2018).
[15] G. M. Kavoulakis and G. Baym, Phys. Rev. B **54**, 16625 (1996).
[16] G. Mahan, *Many-Particle Physics* (Plenum, New York, 2000), p. 592.
[17] Y. Yu, Y. Yu, Y. Cai, W. Li, A. Gurarslan, H. Peelasers, D. E. Aspnes, G. G. Van de Walle, N. V. Nguyen, Y-W. Zhang and L. Cao, Sci. Rep. **5**, 16996 (2015).
[18] D. Y. Qiu, F. H. da Jornada and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (2013).
[19] F. Wang, Y. Wu, M. S. Hybertsen and T. F. Heinz, Phys. Rev. B **73**, 245424 (2006).