Low-temperature annihilation rate for quasilocalized excitons in monolayer MoS$_2$

Eric Chatterjee,[1,*] Daniel B. S. Soh,[1,2] Christopher Rogers,[1] Dodd J. Gray,[1] and Hideo Mabuchi[1,†]

[1]*Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA*
[2]*Sandia National Laboratories, Livermore, California 94550, USA*

The strong Coulomb forces in monolayer transition metal dichalcogenides ensure that optical excitation of band electrons gives rise to Wannier-Mott excitonic states, each of which can be conceptualized as a composite of a wave packet corresponding to center-of-mass motion and an orbital state corresponding to the motion of the electron and hole about the center-of-mass. Here we show that at low temperature in monolayer MoS$_2$, given quasilocalized excitons and consequently a significant interexciton spacing, the excitons undergo dipole-dipole interaction and annihilate one another in a manner analogous to Auger recombination. To design our model, we assume that each exciton is localized in a region whose length is on the same scale as the excitonic diameter, thus causing the exciton to behave in a fermionlike manner, while the distance between neighboring excitons is much larger than the exciton diameter. We construct the orbital ladder operators for each exciton and apply Fermi's golden rule to derive the overall recombination rate as a function of exciton density.

I. INTRODUCTION

Recent advances in nanophotonic signal processing have focused increasing attention on novel materials that could provide strong and/or novel nonlinear optical properties in near-planar structures [1–7]. Monolayer transition-metal dichalcogenides (TMDs) are of particular interest in this regard because of their direct band gaps [8–13] and substantial optical nonlinearity relative to conventional bulk materials [14–18]. The primary nonlinear susceptibilities of monolayer TMDs are excitonic in nature, including Kerr-type nonlinearity associated with multiphoton transitions among exciton internal states [18] and saturable absorption-type nonlinearity associated with exciton-exciton interactions at high density. Here we present a theoretical study connected to the latter effect, which can in principle provide a basis for optical bistability and thus all-optical switching [19].

Electron-hole binding energies in monolayer TMDs range from 0.3 to 1 eV [20–24], resulting in the formation of Wannier-Mott excitons in such materials. One important topic of research involves the calculation of the excitonic decay rates through various channels, which play an essential role in determining the dynamic optical response of a monolayer TMD near the exciton resonance. To this end, recent theoretical and phenomenological studies have derived the radiative loss rate for excitons in MoS$_2$ [18,25] and for other TMDs [26,27], as well as for other low-dimensional systems such as plasmons in graphene nanoribbons [28]. In addition, the nonradiative decay rate due to exciton-phonon scattering has been numerically and phenomenologically derived as a function of temperature for WS$_2$ and MoSe$_2$ [27].

At sufficiently low temperatures in high quality crystals, the nonradiative decay consisting of exciton-phonon interactions is minimal (due to lack of significant phonon population) compared to the exciton-exciton interaction. This interaction can potentially take the form of either annihilation (via a process analogous to Auger recombination, in which Coulomb interaction between neighboring excitons causes one to be annihilated and the other to be excited) or scattering (direct or exchange). Given two bright ground-state excitons spaced significantly apart, only the recombination process will occur (the reasoning for this will be discussed later). Intuitively, the rate of this decay process should vary with the excitonic spatial density, which can be modulated via the intensity of the incident electromagnetic field. On the theoretical front, Auger recombination of excitons has been previously analyzed in tightly confined one-dimensional (1D) systems such as carbon nanotubes [29]. Experimentally, recent measurements have demonstrated the presence of rapid exciton-exciton annihilation in monolayer MoS$_2$ [30], as well as in other monolayer TMDs such as WSe$_2$ [31] and MoSe$_2$ [32]. Here we will analytically derive the corresponding rate in monolayer MoS$_2$ as a function of the exciton density, in the quasilocalized regime. Although most recent experiments with monolayer TMDs would seem to take place outside the quasilocalized regime, we note that our results nonetheless appear to be consistent with measured exciton annihilation rates. We discuss interpretations of this apparent agreement in the Discussion.

This paper is organized as follows: In Sec. II we derive the exciton-exciton coupling energy as a function of the relative electron to hole positions for a pair of excitons, under the assumption that the spacing between neighboring excitons is much greater than their diameters. This supposition is supported by the fact that for MoS$_2$, the ground state excitonic diameter is about 0.67 nm [18], whereas the experimental data shows rapid Auger recombination even at a

*echatter@stanford.edu
†hmabuchi@stanford.edu

nearest-neighbor exciton spacing greater than 10 nm [30], while measurements on $MoSe_2$ (a TMD with properties similar to MoS_2) demonstrate a saturation of the excitonic density at a nearest-neighbor spacing of 4 nm [32]. In Sec. III we calculate the matrix elements of the hole-to-electron vectors in the basis of the respective ladders of orbital states and thus obtain the exciton-exciton interaction Hamiltonian in terms of the ladder operators for the two excitons. In Sec. IV we apply Fermi's golden rule to derive the annihilation rate as a function of excitonic density. Finally, in Sec. V we evaluate these expressions and discuss the results. Specifically, we focus on the effect of the localization of the excitonic centers-of-mass and the saturation density of excitons on the annihilation rate.

II. EXCITON-EXCITON INTERACTION ENERGY

We set up this problem by separately considering the coupling of the electron and the hole, respectively, with another exciton. As shown in Fig. 1, we apply the assumption that the spacing between neighboring exciton centers is much larger than the hole-electron distance for each exciton. Since the hole-electron distance resembles the monolayer MoS_2 film thickness of 0.65 nm [33,34], the Rytova-Keldysh potential energy at the range of the exciton-exciton spacing can be considered to approximately equal the three-dimensional (3D) Coulomb potential energy [35,36]. Labeling the constant factor in the Coulomb energy as C, the electron-exciton potential energy is written in terms of spatial coordinates as follows:

$$V_{e\text{-}ex} = C\left(\frac{1}{|\boldsymbol{r_e} - \boldsymbol{r_{e'}}|} - \frac{1}{|\boldsymbol{r_e} - \boldsymbol{r_{h'}}|}\right). \tag{1}$$

The hole-exciton potential energy is similar, except with the electron coordinate $\boldsymbol{r_e}$ replaced by hole coordinate $\boldsymbol{r_h}$ and the sign of the energy flipped since the hole carries a positive charge:

$$V_{h\text{-}ex} = -C\left(\frac{1}{|\boldsymbol{r_h} - \boldsymbol{r_{e'}}|} - \frac{1}{|\boldsymbol{r_h} - \boldsymbol{r_{h'}}|}\right). \tag{2}$$

Since the spacing between electron and hole for each exciton is minimal compared to the distance between the excitons, we can use the well-known dipole-dipole approximation to determine the sum of the above two potential energies [37]. We label the exciton-exciton vector as $\boldsymbol{\Delta r} = \boldsymbol{r_{CM}} - \boldsymbol{r_{CM'}}$ (where $\boldsymbol{r_{CM}}$ and $\boldsymbol{r_{CM'}}$ represent the center-of-mass positions of the excitons) and the hole-to-electron vector for the unprimed (primed) exciton as $\boldsymbol{d} = \boldsymbol{r_e} - \boldsymbol{r_h}$ ($\boldsymbol{d'} = \boldsymbol{r_{e'}} - \boldsymbol{r_{h'}}$). We find that the interaction energy varies linearly with both $\boldsymbol{d}$ and $\boldsymbol{d'}$, as

FIG. 1. Diagram of the two interacting excitons. The green arrows depict the Coulomb coupling between the two excitons that generates the interaction energy, and the purple arrows depict the hole-to-electron vectors for the excitons ($\boldsymbol{d}$ and $\boldsymbol{d'}$, respectively). Note that the interexciton spacing is much greater than the size of each exciton.

desired:

$$V_{ex\text{-}ex} = -\frac{C}{(\Delta r)^3}\left(\frac{3\boldsymbol{\Delta r} \cdot \boldsymbol{d}}{(\Delta r)^2}\boldsymbol{\Delta r} - \boldsymbol{d}\right) \cdot \boldsymbol{d'}. \tag{3}$$

Labeling the ratio of electron mass to total mass as $C_e = \frac{m_e}{m_e + m_h}$ and the corresponding ratio for the hole as $C_h = \frac{m_h}{m_e + m_h}$ (where m_e and m_h are positive constants denoting the electron and hole masses, respectively), we note that $\boldsymbol{r_{CM}} = C_e \boldsymbol{r_e} + C_h \boldsymbol{r_h}$ and $\boldsymbol{r_{CM'}} = C_e \boldsymbol{r_{e'}} + C_h \boldsymbol{r_{h'}}$.

It is worth analyzing $\boldsymbol{\Delta r}$, $\boldsymbol{d}$, and $\boldsymbol{d'}$, since we will promote these variables to operator form later. In general, the Hilbert space spanned by the plane wave states of the two constituent charge carriers (electron and hole) of an exciton is also spanned by a tensor Hilbert space of the center-of-mass degree of freedom and the orbital degree of freedom. This is mathematically demonstrated by the fact that we can express the Schrödinger equation for a single exciton in both of these position bases. The following is the Schrödinger equation in the electron-hole position basis:

$$H_{ex} = -\frac{\hbar^2}{2m_e}\nabla_e^2 - \frac{\hbar^2}{2m_h}\nabla_h^2 + V_{int}(\boldsymbol{r_e} - \boldsymbol{r_h}). \tag{4}$$

Here ∇_e and ∇_h represent the spatial gradients with respect to $\boldsymbol{r_e}$ and $\boldsymbol{r_h}$, respectively. The convenience of converting the electron-hole basis to the basis consisting of the center-of-mass position and the hole-to-electron vector derives from the fact that the electron-hole electrostatic potential energy V_{int} is specifically a function of the displacement between the electron and hole positions while being otherwise invariant in the individual positions themselves. In that second basis, the Hamiltonian is therefore separable into a center-of-mass part H_{CM} and an orbital part H_{orb}:

$$H_{ex} = -\frac{\hbar^2}{2M}\nabla_{CM}^2 - \frac{\hbar^2}{2\mu}\nabla_{orb}^2 + V_{int}(d)$$
$$= H_{CM} + H_{orb}. \tag{5}$$

Here ∇_{CM} and ∇_{orb} denote the spatial gradients with respect to $\boldsymbol{r_{CM}}$ and $\boldsymbol{d}$, respectively, and M and μ represent the total electron-hole mass and reduced electron-hole mass, respectively. Note that V_{int} is entirely included in the orbital part H_{orb}, thus ensuring the separability of the Hamiltonian.

Introducing interaction between excitons, we find that the Schrödinger equation for an unprimed exciton and a primed exciton interacting with each other takes the following form:

$$H = H_{ex} + H_{ex'} + H_{ex\text{-}ex}. \tag{6}$$

Here $H_{ex\text{-}ex}$ denotes the perturbation to the total Hamiltonian due to exciton-exciton interaction and is given by the interaction energy from Eq. (3):

$$H_{ex\text{-}ex} = V_{ex\text{-}ex}(\boldsymbol{r_{CM}} - \boldsymbol{r_{CM'}}, \boldsymbol{d}, \boldsymbol{d'}). \tag{7}$$

For the Auger process, the fact that we start with two bright excitons from the lowest excitonic state (since both were optically excited to that level) implies that we finish with a higher-energy bright exciton. As a result, we only consider cases for which the center-of-mass state is centered at zero momentum. If the excitonic center-of-mass were fully delocalized, this would reduce the composite wave function to only the orbital part, multiplied by the normalization factor
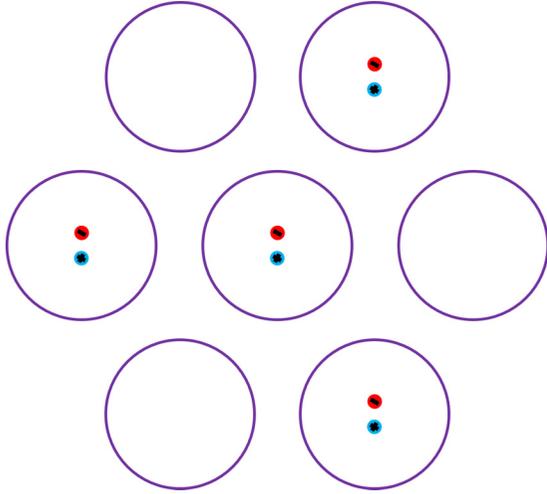
FIG. 2. Visualization of the layout of the excitonic regions in position space given partial filling. Each circle forms the boundary of a localization region, which contains either 0 or 1 exciton. The area of each region will be labeled as A, the density of regions as σ_r, and the excitonic density as σ. Note that if $\sigma < \sigma_r$, then the occupation probability for a given region is $\frac{\sigma}{\sigma_r}$.

$1/\sqrt{A_{\text{beam}}}$, where A_{beam} represents the cross-sectional area of the beam generating the excitons. In reality, defects will constrain the diffusion range [38,39], and quantum dots can also be artificially synthesized in order to generate the same effect (as will be discussed in Sec. V). We will solve the annihilation rate specifically for the case in which each exciton is localized to a unique region (the valid range of localization areas will be analyzed in Sec. V).

Figure 2 depicts the layout of these regions. We will label the effective area of each region as A, the density of the localization regions as σ_r, and the actual excitonic density as σ. The fact that each region contains either 0 or 1 exciton implies that $\sigma \leqslant \sigma_r$ and that the occupation probability for any region is $\frac{\sigma}{\sigma_r}$. For an exciton in the orbital state β localized to a region centered at a generic position $\boldsymbol{R''}$, we will use a stationary wave packet $f_{\boldsymbol{R''}}$ centered at $\boldsymbol{R''}$ to model the zero-point momentum excitonic center-of-mass, leading to the following composite wave function:

$$\Psi_{\beta,\mathbf{0}}(\boldsymbol{r}_{\text{CM}}, \boldsymbol{d}) = f_{\boldsymbol{R''}}(\boldsymbol{r}_{\text{CM}})\psi_\beta(\boldsymbol{d}). \tag{8}$$

We will label the center of the wave packet for the unprimed (primed) exciton as $\boldsymbol{R}$ ($\boldsymbol{R'}$). The lack of overlap between neighboring wave packets, as well as the separability of the center-of-mass part and the orbital part, imply that the matrix elements for $\boldsymbol{\Delta r}$, $\boldsymbol{d}$, and $\boldsymbol{d'}$ can be determined by expanding these variables in different Hilbert subspaces, i.e., the center-of-mass position basis for the two excitons for $\boldsymbol{\Delta r}$, the orbital states of the unprimed exciton for $\boldsymbol{d}$, and the orbital states of the primed exciton for $\boldsymbol{d'}$.

III. INTERACTION HAMILTONIAN MATRIX ELEMENTS

In deriving the matrix elements, we will restrict the range of our summation to bright excitonic states centered at zero center-of-mass momentum. It is worth explaining how such

a reduction is physically valid. Given two generic excitons, there are three possible interaction processes that can occur: annihilation (which preserves CM momentum), direct scattering (which alters CM momentum), or exchange. Though direct scattering has recently been theoretically analyzed [40], it is forestalled if both of the excitons are initially bright and in the ground states of their respective orbital ladders. This is because the scattering process must lower the energy of one exciton while increasing the energy of the other, and since both excitons are at the minimum possible energy level, neither can decrease in energy while still maintaining coherence.

In addition, the large spacing between the excitons ensures that the decay rate is unaffected by exchange. Single-particle exchange (electron or hole) is hindered by the fact that the ground-state orbital restricts the electron-hole separation to a distance far closer than the separation between the localization region centers. Similarly, two-particle (full-exciton) exchange is hindered by the negligible overlap between the wave packets corresponding even to neighboring localization regions.

The only relevant exciton-exciton decay process is thus Auger recombination, for which center-of-mass momentum conservation requires that the nonannihilated exciton be excited to another bright state. Due to the large distance between neighboring wave packet centers relative to the exciton size and the lack of overlap between the wave packets, the $\boldsymbol{\Delta r}$ operator simplifies to the corresponding C number, i.e., the displacement between the coordinate centers of the respective excitons. Henceforth, we will use $\boldsymbol{\Delta R}$ to label the constant $\boldsymbol{R} - \boldsymbol{R'}$.

Next, we consider the expansion of the hole-to-electron vectors $\boldsymbol{d}$ and $\boldsymbol{d'}$ in the basis of the ladder of orbitals for their respective excitons. Knowing that each exciton is initially in the ground state $|\nu\rangle$ and summing over all possible final bound states ($|\nu_f\rangle$), unbound states ($|\phi_q\rangle$), and the vacuum state ($|fs\rangle$), $\boldsymbol{d}|\nu\rangle\langle\nu|$ takes the following form:

$$\boldsymbol{d}|\nu\rangle\langle\nu|$$
$$= \left(\sum_{\nu_f}|\nu_f\rangle\langle\nu_f| + \sum_{q}|\phi_q\rangle\langle\phi_q| + |fs\rangle\langle fs|\right)\boldsymbol{d}|\nu\rangle\langle\nu|$$
$$= \langle fs|\boldsymbol{d}|\nu\rangle B_\nu + \sum_{\nu_f}\langle\nu_f|\boldsymbol{d}|\nu\rangle B_{\nu_f}^\dagger B_\nu$$
$$+ \sum_{q}\langle\phi_q|\boldsymbol{d}|\nu\rangle D_q^\dagger B_\nu. \tag{9}$$

Here $B_{\nu''}$ denotes the annihilation operator corresponding to the bound state $|\nu''\rangle$, whereas $D_{q''}$ represents the annihilation operator for the unbound state $|\phi_{q''}\rangle$. We consider the first matrix element in this expression, which couples the ground excitonic state with the vacuum state, by decomposing the excitonic state into the electronic band states of the lattice, per the Keldysh formalism. Specifically, a bright excitonic state $|\nu''\rangle$ can be conceptualized as a superposition across wave vectors $\boldsymbol{k''}$ of composite electron-hole states, with the electron deriving from the lowest conduction band and the hole from the same wave vector at the highest valence band. The wave vectors are weighted by the Fourier envelope function

$\psi_{v''}(\boldsymbol{k''})$, which has been previously derived for MoS$_2$ [41]:

$$|v''\rangle = \sum_{\boldsymbol{k''}} \psi_{v''}(\boldsymbol{k''}) |c(\boldsymbol{k''})v(\boldsymbol{k''})\rangle. \qquad (10)$$

Using this representation, we calculate the matrix element representing the transition from the initial excitonic state $|v\rangle$ to the vacuum state $|fs\rangle$:

$$\langle fs|\boldsymbol{d}|v\rangle = \sum_{\boldsymbol{k}} \psi_v(\boldsymbol{k}) \langle fs|\boldsymbol{d}|c(\boldsymbol{k})v(\boldsymbol{k})\rangle. \qquad (11)$$

Physically, the process described by the right-hand-side inner product corresponds to the electron at $|c(\boldsymbol{k})\rangle$ dropping to $|v(\boldsymbol{k})\rangle$ and annihilating the hole there [18]:

$$\langle fs|\boldsymbol{d}|v\rangle = \sum_{\boldsymbol{k}} \psi_v(\boldsymbol{k}) \langle v(\boldsymbol{k})|\boldsymbol{r}_e|c(\boldsymbol{k})\rangle. \qquad (12)$$

Next, we examine the matrix element connecting the ground state to another bound excitonic state $|v_f\rangle$. The most convenient and physically intuitive method for solving this is by expanding in the spatial basis $\boldsymbol{d}$ instead of in the band basis:

$$\langle v_f|\boldsymbol{d}|v\rangle = \int_{\text{lat}} d^2 d\,\psi_{v_f}^*(d)\boldsymbol{d}\psi_v(d). \qquad (13)$$

Finally, we analyze the matrix element connecting $|v\rangle$ to an unbound state denoted by $|\phi_{\boldsymbol{q}}\rangle$. This state can be considered as a composite of the free electron plane wave with momentum $\hbar\boldsymbol{q}$ and the free hole plane wave with momentum $-\hbar\boldsymbol{q}$ (the opposite values of the electron and hole momenta derive from the fact that the center-of-mass momentum must equal 0):

$$\phi_{\boldsymbol{q}}(\boldsymbol{d}) = \frac{1}{\sqrt{A}}e^{i\boldsymbol{q}\cdot\boldsymbol{r}_e}e^{-i\boldsymbol{q}\cdot\boldsymbol{r}_h} = \frac{1}{\sqrt{A}}e^{i\boldsymbol{q}\cdot\boldsymbol{d}}. \qquad (14)$$

Similarly, we calculate the matrix element representing the transition from the ground state to a generic unbound state by expanding in $\boldsymbol{d}$:

$$\langle \phi_{\boldsymbol{q}}|\boldsymbol{d}|v\rangle = \int_{\text{lat}} d^2 d\,\phi_{\boldsymbol{q}}^*(\boldsymbol{d})\boldsymbol{d}\psi_v(d)$$
$$= \frac{1}{\sqrt{A}} \int_{\text{lat}} d^2 d\, e^{-i\boldsymbol{q}\cdot\boldsymbol{d}}\boldsymbol{d}\psi_v(d). \qquad (15)$$

Since one of the excitons drops to the Fermi sea from the lowest excitonic state in the annihilation process, the other exciton must be excited by an energy equaling the gap between the lowest excitonic state and vacuum. Based on our tight-binding calculations, we know that the gap between conduction and valence band at the K/K' points (which equals the maximum bound exciton energy) is approximately 2.2 eV, and it has been shown that the electron-hole binding energy reduces the ground state exciton energy to approximately 1.9–2.1 eV [42,43]. Therefore, if one ground state exciton is annihilated, then the other must rise to an energy of 3.8–4.2 eV, thus exceeding the maximum bound exciton energy and creating an unbound exciton. Applying the exciton-exciton interaction energy derived in Eq. (3) to a pair of excitons in the lowest excitonic state and substituting the operator representations of $\boldsymbol{d}$ and $\boldsymbol{d'}$ derived above, we obtain the following column of matrix elements for the perturbation $H_{\text{ex-ex}}$ to the Hamiltonian due to exciton-exciton interaction:

$$H_{\text{ex-ex}}|v,v\rangle\langle v,v| = -\frac{C}{(\Delta R)^3} \sum_{\boldsymbol{q}} \left[\left(\frac{3\Delta\boldsymbol{R}\cdot\langle fs|\boldsymbol{d}|v\rangle}{(\Delta R)^2}\Delta\boldsymbol{R} - \langle fs|\boldsymbol{d}|v\rangle \right) \cdot \langle \phi_{\boldsymbol{q}}|\boldsymbol{d}|v\rangle B_v D_{\boldsymbol{q}}^{\prime\dagger}B_v' \right. $$
$$\left. + \left(\frac{3\Delta\boldsymbol{R}\cdot\langle \phi_{\boldsymbol{q}}|\boldsymbol{d}|v\rangle}{(\Delta R)^2}\Delta\boldsymbol{R} - \langle \phi_{\boldsymbol{q}}|\boldsymbol{d}|v\rangle \right) \cdot \langle fs|\boldsymbol{d}|v\rangle D_{\boldsymbol{q}}^{\dagger}B_v B_v' \right]\delta_{\omega_q,2\omega_v}. \qquad (16)$$

Here $\hbar\omega_v$ denotes the energy of the lowest excitonic state $|v\rangle$, while $\hbar\omega_{\boldsymbol{q}}$ denotes the energy of the generic unbound state $|\phi_{\boldsymbol{q}}\rangle$. Note that this summation consists of two parts: one corresponding to the annihilation of the unprimed exciton and excitation of the primed exciton to an unbound state, and the other corresponding to the converse process. Henceforth, we will reduce the summation over $\boldsymbol{q}$ to only include the wave vectors which satisfy the condition $\omega_q = 2\omega_v$ as required by the energy conservation condition encapsulated by the Kronecker delta function. Also note that we have replaced $\boldsymbol{d'}$ with $\boldsymbol{d}$ in the inner products, since the hole-to-electron position operator acts on the states of the corresponding exciton in the same manner.

IV. CALCULATING THE ANNIHILATION RATE

We calculate the overall annihilation rate for a single exciton as a function of spatial excitonic density using Fermi's golden rule [44]. We assume that the excitonic localization regions form a closely packed (triangular) lattice resembling the layout of the TMD itself, though it is worth noting that the overall annihilation rate for a grid (square) layout would differ by less than 15 percent. The spacing l_0 between the centers of nearest-neighbor regions relates to the area density of regions (i.e., the saturation excitonic density) σ_r as follows:

$$\sigma_r = \frac{2}{l_0^2\sqrt{3}}. \qquad (17)$$

In order to determine the total annihilation rate for a given exciton, we will use series summation over the annihilation rates of the exciton upon interaction with all possible primed excitons (i.e., all other excitonic locations). We start by deriving the transition amplitude for a single exciton-exciton axis $\hat{l}$, where $\Delta\boldsymbol{r} = l\hat{l}$, using the Hamiltonian column from Eq. (16):

$$\langle fs, \phi_{\boldsymbol{q}}|H_{\text{ex-ex}}|v,v\rangle = -\frac{C}{l^3}(3\langle fs|\hat{l}\cdot\boldsymbol{d}|v\rangle\langle \phi_{\boldsymbol{q}}|\hat{l}\cdot\boldsymbol{d}|v\rangle$$
$$- \langle fs|\boldsymbol{d}|v\rangle\cdot\langle \phi_{\boldsymbol{q}}|\boldsymbol{d}|v\rangle). \qquad (18)$$

The rate $\Gamma_{\text{single,drop}}(l,\boldsymbol{q})$ at which the unprimed exciton drops from the lowest excitonic state $|v\rangle$ to the Fermi sea upon interacting with a single primed exciton that jumps to a state $|\phi_{\boldsymbol{q}}\rangle$ is calculated through Fermi's golden rule:

$$\Gamma_{\text{single,drop}}(l,\boldsymbol{q}) = \frac{2\pi}{\hbar}\rho(\boldsymbol{q})|\langle fs, \phi_{\boldsymbol{q}}|H_{\text{ex-ex}}|v,v\rangle|^2. \qquad (19)$$
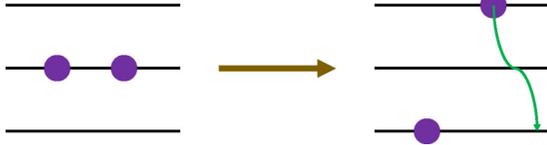
FIG. 3. Simplified three-level diagram representing the exciton-exciton annihilation process. Note that the excited exciton decays to the vacuum state upon interacting with phonons, as shown by the curved green arrow.

Here $\rho(\boldsymbol{q})$ denotes the unbound excitonic density of states at the orbital wave vector $\boldsymbol{q}$. Note that all unbound states $|\phi_q\rangle$ for a given magnitude q are degenerate. In order to find the total rate $\Gamma_{\text{single,drop}}(l)$ at which an exciton drops from $|v\rangle$ to the Fermi sea due to the interaction with another exciton at a distance l, we therefore need to sum over the transition rates to all possible final wave vectors $\boldsymbol{q}$ on a ring of radius q. This is equivalent to replacing the density of states at a single value $\boldsymbol{q}$ with the total density of states for the ring, which we label $\rho(E_q)$, and averaging the amplitude-squared term over all possible directions $\hat{\boldsymbol{q}}$:

$$
\begin{aligned}
\Gamma_{\text{single,drop}}(l) &= \frac{2\pi}{\hbar}\rho(E_q)\langle|\langle fs, \phi_q|H_{\text{ex-ex}}|v, v\rangle|^2\rangle_{\hat{\boldsymbol{q}}} \\
&= \frac{2\pi}{\hbar}\rho(E_q)\frac{C^2}{l^6}\langle|3\langle fs|\hat{\boldsymbol{l}}\cdot\boldsymbol{d}|v\rangle\langle\phi_q|\hat{\boldsymbol{l}}\cdot\boldsymbol{d}|v\rangle \\
&\quad + \langle fs|\boldsymbol{d}|v\rangle\cdot\langle\phi_q|\boldsymbol{d}|v\rangle|^2\rangle_{\hat{\boldsymbol{q}}}.
\end{aligned}
\tag{20}
$$

In addition, it is important to consider how the newly generated unbound exciton relaxes. Since the constituent electron and hole feature the same wave vector in reciprocal space, the electron can radiatively decay into the hole. However, due to the weak oscillator strength corresponding to the transition between conduction and valence bands at a particular wave vector (relative to the exciton-vacuum oscillator strength), this process is inefficient compared to the radiative decay of an exciton from the lowest excitonic state. A faster decay channel is by multiphonon relaxation, as previously analyzed in linear molecular aggregates [45,46]. Due to the very small value of the transition dipole moment connecting the lowest excitonic state to a generic unbound state [18], the unbound exciton is eventually annihilated to vacuum by such a process (as depicted in Fig. 3) instead of returning to the original lowest excitonic state. As such, the actual annihilation rate for a single exciton upon interaction with another exciton $\Gamma_{\text{single}}(l)$ is double that of $\Gamma_{\text{single,drop}}(l)$:

$$
\begin{aligned}
\Gamma_{\text{single}}(l) &= \frac{4\pi}{\hbar}\rho(E_q)\frac{C^2}{l^6}\langle|3\langle fs|\hat{\boldsymbol{l}}\cdot\boldsymbol{d}|v\rangle\langle\phi_q|\hat{\boldsymbol{l}}\cdot\boldsymbol{d}|v\rangle \\
&\quad - \langle fs|\boldsymbol{d}|v\rangle\cdot\langle\phi_q|\boldsymbol{d}|v\rangle|^2\rangle_{\hat{\boldsymbol{q}}}.
\end{aligned}
\tag{21}
$$

The density of states $\rho(E_q)$ can be derived by envisioning a circular ring of free-particle states in a two-dimensional reciprocal space. In this space, each state occupies a reciprocal area of $\frac{(2\pi)^2}{A}$. The fact that each of the two charge carriers (electron and hole) is propagating at a wave vector of magnitude q results in the following relationship between the unbound

state energy and q:

$$
E_q = E_{\text{gap}} + \frac{\hbar^2 q^2}{2}\left(\frac{1}{m_e} + \frac{1}{m_h}\right).
\tag{22}
$$

Here E_{gap} represents the lattice band gap energy, equaling the sum of the ground state excitonic energy and the exciton binding energy. Defining the region of the reciprocal space enclosed by a circle of radius q as $a_q = \pi q^2$, we calculate the density of states as follows:

$$
\begin{aligned}
\rho(E_q) &= \frac{dN}{da_q}\frac{da_q}{dq}\frac{dq}{dE_q} \\
&= \frac{A}{(2\pi)^2}2\pi q\frac{1}{\hbar^2 q}\left(\frac{1}{m_e} + \frac{1}{m_h}\right)^{-1} \\
&= \frac{m_e m_h}{2\pi\hbar^2(m_e + m_h)}A.
\end{aligned}
\tag{23}
$$

Next, we determine the total annihilation rate for a given exciton. Note that the probability of interaction between two excitons separated by a distance l is proportional to l^{-6}, and therefore for a given exciton, the interaction with the nearest-neighbor excitons should dominate. Establishing the unprimed exciton as the origin of a two-dimensional (2D) coordinate system, we label the reduced coordinates for a generic localization region as (n, m), where n and m are integers and the horizontal and vertical coordinates (h and v, respectively) equal the following:

$$
\begin{aligned}
h_{n,m} &= n + m\cos\frac{\pi}{3} = n + \frac{1}{2}m, \\
v_{n,m} &= m\sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}m.
\end{aligned}
\tag{24}
$$

For the case in which every localization region is occupied by an exciton, the distance between nearest-neighbor excitons equals l_0. We aim to obtain the ratio between the total annihilation rate for the unprimed exciton to the annihilation rate with a single nearest-neighbor primed exciton. To do so, we use the following summation over the localization regions indexed by (n, m):

$$
\begin{aligned}
\sum_{n,m}\left(\frac{l_0}{l_{n,m}}\right)^6 &= \sum_{n,m}\frac{1}{[(n + \frac{1}{2}m)^2 + (\frac{\sqrt{3}}{2}m)^2]^3} \\
&= \sum_{n,m}\frac{1}{(n^2 + nm + m^2)^3}.
\end{aligned}
\tag{25}
$$

We simplify this sum by only applying it to a single $\frac{\pi}{3}$ slice. Due to the sixfold rotational symmetry of the excitonic layout, the result for a single slice will apply to every other quadrant and axis, respectively. For the first slice (starting from the positive horizontal axis, inclusive, and ending just short of the $\frac{\pi}{3}$ axis), we obtain $\sum_{n=1}^{\infty}\sum_{m=0}^{\infty}\frac{1}{(n^2+nm+m^2)^3} \approx 1.0626$. The result of Eq. (25) is calculated by multiplying these results by 6:

$$
\sum_{n,m}\frac{1}{(n^2 + nm + m^2)^3} \approx 6 \times 1.0626 \approx 6.4.
\tag{26}
$$

We determine the total decay rate for a single exciton *given full filling of the localization regions* $\Gamma_{\text{total,full}}$ by multiplying the expression from Eq. (21) by the ratio from Eq. (26) and substituting the density of states from Eq. (23), while

averaging over all possible exciton-exciton axis orientations $\hat{l}$:

$$
\begin{aligned}
\Gamma_{\text{total,full}} &= \frac{12.8 m_e m_h}{\hbar^3 (m_e + m_h)} \frac{C^2 A}{l_0^6} \langle |3 \langle fs|\hat{l} \cdot \boldsymbol{d}|\nu\rangle \langle \phi_q|\hat{l} \cdot \boldsymbol{d}|\nu\rangle - \langle fs|\boldsymbol{d}|\nu\rangle \cdot \langle \phi_q|\boldsymbol{d}|\nu\rangle |^2 \rangle_{\hat{l},\hat{q}} \\
&= \frac{8.3 m_e m_h}{\hbar^3 (m_e + m_h)} C^2 A \sigma_r^3 \langle |3 \langle fs|\hat{l} \cdot \boldsymbol{d}|\nu\rangle \langle \phi_q|\hat{l} \cdot \boldsymbol{d}|\nu\rangle - \langle fs|\boldsymbol{d}|\nu\rangle \cdot \langle \phi_q|\boldsymbol{d}|\nu\rangle |^2 \rangle_{\hat{l},\hat{q}}.
\end{aligned} \quad (27)
$$

For the general case including full or partial filling, i.e., $\sigma \leqslant \sigma_r$, Eq. (27) must be weighted by the occupation probability of any given localization region, i.e., $\frac{\sigma}{\sigma_r}$. This yields an annihilation rate Γ_{total} for a single exciton that varies linearly with the excitonic density σ:

$$
\begin{aligned}
\Gamma_{\text{total}}(\sigma) &= \frac{\sigma}{\sigma_r} \Gamma_{\text{total,full}} \\
&= \frac{8.3 m_e m_h}{\hbar^3 (m_e + m_h)} C^2 A \sigma_r^2 \sigma \langle |3 \langle fs|\hat{l} \cdot \boldsymbol{d}|\nu\rangle \langle \phi_q|\hat{l} \cdot \boldsymbol{d}|\nu\rangle \\
&\quad - \langle fs|\boldsymbol{d}|\nu\rangle \cdot \langle \phi_q|\boldsymbol{d}|\nu\rangle |^2 \rangle_{\hat{l},\hat{q}}.
\end{aligned} \quad (28)
$$

V. RESULTS AND DISCUSSION

The final step in the remaining process of obtaining the exciton-exciton annihilation rate is to substitute numerical values into the matrix elements and coefficients in Eq. (28). The matrix element $\langle fs|\boldsymbol{d}|\nu\rangle$, corresponding to the annihilation of an exciton at the ground state $|\nu\rangle$, was previously determined (via decomposition into the band basis [18]) as proportional to the square root of the localization region area A:

$$
\langle fs|\boldsymbol{d}|\nu\rangle \approx \hat{\boldsymbol{x}}(0.1425i)\sqrt{A}. \quad (29)
$$

Here we have defined $\hat{\boldsymbol{x}}$ as the direction of the vector $\langle fs|\boldsymbol{d}|\nu\rangle$. Another element used in Eq. (28) is $\langle fs|\hat{l} \cdot \boldsymbol{d}|\nu\rangle$, i.e., the $\hat{l}$ component of $\langle fs|\boldsymbol{d}|\nu\rangle$. Labeling the angle between $\hat{\boldsymbol{x}}$ and $\hat{l}$ as ϕ_l, we obtain the following expression for that inner product:

$$
\begin{aligned}
\langle fs|\hat{l} \cdot \boldsymbol{d}|\nu\rangle &= \hat{l} \cdot \langle fs|\boldsymbol{d}|\nu\rangle \\
&\approx (0.1425i)\sqrt{A} \cos \phi_l.
\end{aligned} \quad (30)
$$

We calculate the inner product $\langle \phi_q|\boldsymbol{d}|\nu\rangle$, corresponding to the excitation of a ground state exciton to the unbound state exhibiting an energy twice that of the ground state, by expanding in the hole-to-electron vector position basis and integrating:

$$
\begin{aligned}
\langle \phi_q|\boldsymbol{d}|\nu\rangle &= \frac{1}{\sqrt{A}} \int_{\text{lat}} d^2 d\, e^{-i\boldsymbol{q}\cdot\boldsymbol{d}} \boldsymbol{d}\, \psi_\nu(d) \\
&= -\hat{\boldsymbol{q}} \frac{8i\sqrt{2\pi} a_0^3 q}{(a_0^2 q^2 + 4)^2 \sqrt{A}}.
\end{aligned} \quad (31)
$$

Here $\frac{a_0}{2}$ represents the expectation value of the electron-hole distance, which approximately equals 0.67 nm for MoS_2 [18]. Labeling the angle between $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{q}}$ as ϕ_q, the $\hat{l}$ component of the inner product is solved as follows:

$$
\begin{aligned}
\langle \phi_q|\hat{l} \cdot \boldsymbol{d}|\nu\rangle &= \hat{l} \cdot \langle \phi_q|\boldsymbol{d}|\nu\rangle \\
&= -i \frac{8\sqrt{2\pi} a_0^3 q}{(a_0^2 q^2 + 4)^2 \sqrt{A}} \cos(\phi_l - \phi_q).
\end{aligned} \quad (32)
$$

Since $\hat{l}$ and $\hat{\boldsymbol{q}}$ can take any direction relative to $\hat{\boldsymbol{x}}$, we need to average over ϕ_l and ϕ_q. Substituting the matrix elements into Eq. (28), and averaging over the two angle variables, we obtain the following expression for the total annihilation rate as a function of the excitonic density σ:

$$
\begin{aligned}
\Gamma_{\text{total}}(\sigma) &= \frac{8.3 m_e m_h}{\hbar^3 (m_e + m_h)} C^2 A \sigma_r^2 \sigma \left| \frac{(-8i)(0.1425i)\sqrt{2\pi} a_0^3 q}{(a_0^2 q^2 + 4)^2} \right|^2 \\
&\quad \times \langle \langle |3 \cos \phi_l \cos(\phi_l - \phi_q) - \cos \phi_q|^2 \rangle_{\phi_l} \rangle_{\phi_q} \\
&= \frac{27\pi a_0^6 q^2}{(a_0^2 q^2 + 4)^4} \frac{\mu}{\hbar^3} C^2 A \sigma_r^2 \sigma.
\end{aligned} \quad (33)
$$

Here μ denotes the reduced mass of the electron-hole pair, defined as $\mu = (1/m_e + 1/m_h)^{-1}$. Figure 4 provides a schematic of the relevant excitonic energy levels (Fermi sea, ground state, and unbound states) and the gaps between those levels. Recall that the excited exciton gains an energy of E_ν, equaling that of the ground state exciton relative to the Fermi sea. Some of this energy is spent in overcoming the exciton binding energy E_{BE} in order to reach the conduction band energy, while the remainder is used in reaching the unbound state $|\phi_q\rangle$, which exhibits an energy (relative to the conduction band) corresponding to a free electron and free hole carrying wave vectors of magnitude q. In quantitative terms, the binding energy, wave vector magnitude, and ground state energy are related by the following expression:

$$
E_{\text{BE}} + \frac{\hbar^2 q^2}{2\mu} = E_\nu. \quad (34)
$$

Note that the annihilation rate varies with the reduced mass μ (which is defined by the effective electron and hole masses
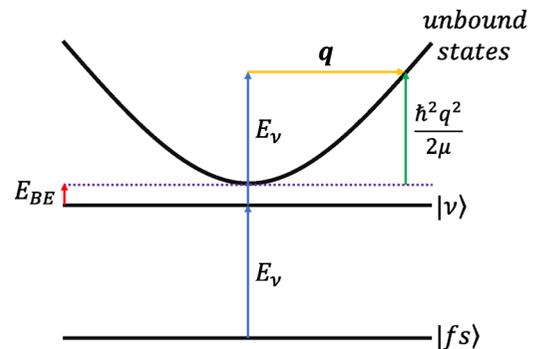
FIG. 4. Depiction of excitonic energy levels. The energy gap between the excitonic ground state $|\nu\rangle$ and the excited unbound state at wave vector $\boldsymbol{q}$ equals the gap E_ν between $|\nu\rangle$ and the vacuum state $|fs\rangle$. Note that μ represents the reduced mass of the electron-hole pair, given by $(\frac{1}{m_e} + \frac{1}{m_h})^{-1} \approx \frac{m}{2}$, where $m \approx m_e \approx m_h$.

and is always smaller than the smallest effective mass of composite particles), as evidenced from Eqs. (33) and (34). In a previous study [18], we determined m_e and m_h each as 0.55 times the actual electron mass m_0, E_{BE} as 0.31 eV, and E_ν as 1.92 eV. It is worth mentioning that values for m_h and m_e equaling $0.65m_0$ and $0.49m_0$, respectively, have been determined in other analysis [47], yielding the same reduced mass as the case in which both carriers have a mass of $0.55m_0$. Substituting these values into the above expression, we find that $q = 3.4$ nm^{-1}. Given this value for q, along with $a_0 = 1.34$ nm and $m_e \approx m_h \approx 5.0 \times 10^{-31}$ kg, our last computational step is to obtain C. This constant simply equals $\frac{e^2}{4\pi\epsilon_0\epsilon_r}$, where ϵ_r represents the ambient dielectric constant for the monolayer MoS$_2$ film. For a free-standing film, the dielectric constant equals 1, yielding $C = 2.3 \times 10^{-28}$ $\frac{\text{kg m}^3}{\text{s}^2}$ and reducing the annihilation rate from Eq. (33) to a function of effective sample area, density of regions, and exciton density:

$$\Gamma_{\text{ex-ex}}(\sigma) \approx \left(1.7 \times 10^{-22} \frac{\text{m}^4}{\text{s}}\right) A\sigma_r^2 \sigma. \qquad (35)$$

Note that this represents the decay rate for a single exciton. From this expression, we derive the rate of change of density due to the annihilation process, labeling the total sample area being measured as A_{total} and the total number of excitons as N:

$$\frac{d\sigma}{dt} = \frac{1}{A_{\text{total}}}\frac{dN}{dt} = -\frac{N}{A_{\text{total}}}\Gamma_{\text{ex-ex}} = -\alpha\sigma^2. \qquad (36)$$

Here α represents the annihilation rate constant, given by the following value:

$$\alpha \approx \left(1.7 \times 10^{-22} \frac{\text{m}^4}{\text{s}}\right) A\sigma_r^2. \qquad (37)$$

It is worth determining a feasible range for α by analyzing A and σ_r. As previously mentioned, A represents the localization area of each charge carrier wave function, while σ_r represents the exciton saturation density. Assuming that the localization regions approximately fill the position space of the sample, we find that $\sigma_r \approx \frac{1}{A}$, yielding the following value for α in terms of A:

$$\alpha \approx \left(1.7 \times 10^{-22} \frac{\text{m}^4}{\text{s}}\right)\frac{1}{A}. \qquad (38)$$

Next, we aim to find a feasible range for A, starting with a consideration of the defects that give rise to exciton and charge carrier localization. Since the grain boundaries of graphene were mapped by Kim et al. [48], the characterization of inhomogeneities in 2D materials has remained an active field. To this end, the high concentration of defects in TMD samples obtained by mechanical exfoliation or grown epitaxially has been well established by experimental measurements. For mechanically exfoliated monolayer MoS$_2$, Wang et al. [49] determined a defect density ranging from 0.3×10^{12} to 2×10^{13} cm^{-2} through a pump-probe measurement, while Vancso et al. [50] used a similar method to obtain a defect density of 5×10^{12} to 5×10^{13} cm^{-2}. Similarly, Liu et al. [51] measured an average defect density of 8×10^{12} cm^{-2} on monolayer MoS$_2$ grown on epitaxial graphene. CVD-grown TMDs exhibit an even greater density of inhomogeneities,

although a recent study by Rogers et al. [52] has shown that laser annealing can remove some of the impurities and reduce the strain gradient.

The range of defect densities gives rise to an array of possible excitonic coherence lengths. For a particular region of a sample, the local value of A can be inferred via the fundamental A dependence of the excitonic radiative decay rate. We use the following well-known expression based on the Einstein coefficients [53] to calculate the radiative loss rate from the excitonic state $|\nu\rangle$ to the vacuum state $|fs\rangle$:

$$\Gamma_{\text{rad}} = \frac{\omega^3}{3\pi\epsilon_0\hbar c^3}|\langle fs|e\boldsymbol{d}|\nu\rangle|^2. \qquad (39)$$

Substituting the excitonic energy $\hbar\omega = 1.9$ eV and the dipole matrix element $|\langle fs|\boldsymbol{d}|\nu\rangle| = 0.1425\sqrt{A}$ [see Eq. (29)] into this expression, we find that the radiative decay rate varies linearly with the localization area A:

$$\Gamma_{\text{rad}} = \left(5.0 \times 10^{25} \frac{1}{\text{m}^2\,\text{s}}\right)A. \qquad (40)$$

Note that the linear relationship between the radiative decay rate and localization area agrees with Wang et al. [25]. For excitons with large localization areas, the radiative lifetime drops to the femtosecond range and will be far shorter than the exciton-exciton annihilation lifetime (even observations of rapid annihilation in molybdenum-based TMDs yield lifetimes in the tens of picoseconds [30,32]), thus rendering the latter process negligibly slow.

Next, we resolve the range of localization areas that fit our assumption that each localization region contains up to one exciton. To this end, it is important to note that although delocalized excitons act in a bosonic manner, the behavior becomes fermionlike given significant localization (on the scale of the excitonic size). For example, Ohtsu [54] shows that for an sufficiently small exciton localization region indexed by n, the commutator between the excitonic annihilation and creation operators takes the following form:

$$[B_n, B_n^\dagger] = 1 - 2B_n^\dagger B_n. \qquad (41)$$

The fact that the formation of more than one exciton in a single narrow localization region is not favored is further supported by experimental evidence of phase space filling, e.g., the absorption saturation measured in GaAs by Hunsche et al. [55]. For 2D materials, the seminal theoretical analysis by Schmitt-Rink et al. [56] showed that the effects of Pauli exclusion result in a blocked area of $8.5\pi a^2$ around each exciton (where a represents the exciton size) [57], indicating a diameter of $5.8a$ for the blocked zone. For molybdenum-based TMDs, the exciton size approximately equals 0.7 nm as previously mentioned, leading to a blocked zone diameter of about 4 nm. This is confirmed by recent measurements on monolayer MoSe$_2$ by Kumar et al. [32], which suggest that the excitonic density saturates at a nearest-neighbor spacing of 4 nm.

Given a diameter of less than about 8 nm for each localization region, the formation of more than one exciton in a single region is hindered since a region of such size cannot fit more than one blocked zone. On the other hand, if the localization region diameter is less than the blocked zone diameter of about 4 nm, then the orbital wave function for an exciton
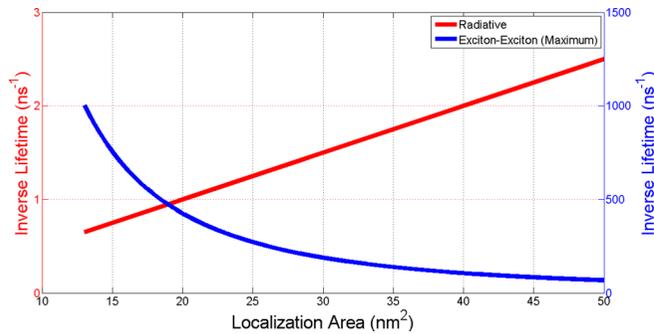
FIG. 5. Plot of the decay rates for a single exciton due to radiative recombination (red) and exciton-exciton annihilation (blue) given maximal filling of the excitonic localization regions, as functions of the localization area A for the range $13 \lesssim A \lesssim 50$ nm^2. Note that the radiative recombination rate is more than an order of magnitude less than the exciton-exciton annihilation rate for all values of A in this range.

formed in that region would start to experience distortion due to the effects of strong localization. As a result, our model of quasilocalized excitons behaving in a fermionlike manner within their respective regions specifically applies for the range of areas given by $13 \lesssim A \lesssim 50$ nm^2. For this range, the density of regions σ_r ($\approx 1/A$) represents the excitonic saturation density and thus serves as the upper bound for σ. As a result, the maximum value of the annihilation rate from Eq. (35) (which applies when the localization regions are saturated with excitons) becomes a function solely of the localization area:

$$\Gamma_{\text{ex-ex,max}} \approx \left(1.7 \times 10^{-22} \, \frac{\text{m}^4}{\text{s}} \right) \frac{1}{A^2}. \quad (42)$$

Figure 5 depicts the radiative recombination and maximum exciton-exciton annihilation rates based on Eqs. (40) and (42), respectively. For the entire range, the annihilation process given maximal excitonic density is more than an order of magnitude faster than the radiative process, with the effect especially pronounced at smaller localization areas. This implies that if a sufficient pump fluence is applied to the sample such that the localization regions become saturated, the initial radiative recombination rate will be negligible compared to the exciton-exciton annihilation rate.

We now return to the general case in which the excitonic density σ is decoupled from the density of localization regions σ_r (i.e., the excitonic density is a variable with an arbitrary value less than or equal to the region density) in order to determine the range for the rate constant α for the relevant range of localization areas. From Eq. (38), it is apparent that α decreases with A. This value is maximized (minimized) when the localization area A is minimized (maximized), which applies when $A \approx 13$ nm^2 (50 nm^2):

$$3.4 \times 10^{-6} \lesssim \alpha \lesssim 1.3 \times 10^{-5} \, \frac{\text{m}^2}{\text{s}}. \quad (43)$$

It is useful to compare this result to experimental findings demonstrating rapid exciton-exciton annihilation. In this respect, the predicted range for the annihilation rate constant for

quasilocalized excitons at low temperatures resembles values measured at room temperature. For example, Sun *et al.* [30] determined a rate constant of 4.3×10^{-6} m^2/s for MoS$_2$, well within the span shown in Eq. (43). Similarly, Kumar *et al.* [32] measured a rate constant of 3.3×10^{-5} m^2/s, slightly exceeding our predicted range. Although it is difficult to glean the localization areas from the radiative decay rate in these experiments since the radiative lifetime measurements were performed on a thermal ensemble consisting of both bright and dark excitons, the similarities between our predictions and the experimental results suggest that the rate constant might be similar for localized and delocalized excitons. A similarity between the annihilation rates of localized and delocalized excitons would imply that the exchange interaction between delocalized excitons is much weaker than the dipole-dipole interaction that gives rise to annihilation in both the localized and delocalized cases.

One possible means of testing our predictions is through an experimental method that exploits the narrow (micron-ranged) spot size of the pumping beam relative to the total area of the 2D sample. Due to the uneven distribution of inhomogeneities across the sample (as evidenced by the wide range of defect densities discussed above), the excitonic localization area should differ from one beam-sized region to the next. The localization area for each region can be gleaned by measuring the radiative decay rate for the region under cryogenic conditions. Of course another method would be to test on samples synthesized using different techniques. As shown in Fig. 5, our calculations will apply specifically when the radiative decay rate falls in the range $6.5 \times 10^8 \lesssim \Gamma_{\text{rad}} \lesssim 2.5 \times 10^9$ s^{-1}. It is worth noting that such long radiative lifetimes have already been observed in superacid-treated TMD samples [58,59]. The best strategy for obtaining a consistent localization area across the sample, however, is likely to break up the lattice into quantum dots. Fabrication techniques for patterning quantum dots into monolayer TMDs have been rapidly advancing in terms of both precision and miniaturization, with Wei *et al.* [60] generating nanodots as small as 15 nm, and more recently Ding *et al.* [61] reducing the size to the single nanometer range.

VI. CONCLUSION

We have derived the annihilation rate for optically generated excitons in monolayer MoS$_2$ as a function of density, given large interexciton spacing (significantly exceeding the excitonic size) and quasilocalization of individual excitons. To start, we derived the exciton-exciton interaction energy as a function of exciton spacing, using the dipole-dipole approximation to show that the energy is linear in the hole-to-electron vector for each exciton. We then promoted the hole-to-electron vectors to orbital ladder operators and demonstrated that the center-to-center vector approximately takes a single value, thus enabling the center-of-mass states to be abstracted out and reducing the dimensionality of the Hilbert space. Finally, we employed Fermi's golden rule to calculate the exciton-exciton annihilation rate, using energy conservation to determine the final state in the orbital ladder reached by the excited exciton. We obtained a rate that varies inversely with the localization area and linearly with the total excitonic

density. The rates measured in experimental observations of rapid annihilation resemble our predicted range, even though the experimental data was obtained at room temperature.

Our findings present both theoretical and experimental applications. On the theoretical side, they can be expanded to other 2D materials by reevaluating the numerical results for the matrix elements as well as the energy levels. On the experimental side, when conducting measurements of nonradiative decay rates on samples with low excitonic densities, the results of this paper can be used to account for the loss rate due to exciton-exciton annihilation. Most importantly, since the annihilation rate varies with exciton density, the process introduces a nonlinearity in the transient behavior of the excitons and knowledge of the rate thus represents a

critical step toward assessing the utility of this material in the context of cavity nonlinear optics.

[1] X. Liu, R. M. Osgood, Y. A. Vlasov, and W. M. Green, Nat. Photon. **4**, 557 (2010).

[2] O. Benson, Nature (London) **480**, 193 (2011).

[3] H. Mabuchi, Phys. Rev. A **85**, 015806 (2012).

[4] A. Majumdar, C. M. Dodson, T. K. Fryett, A. Zhan, S. Buckley, and D. Gerace, ACS Photon. **2**, 1160 (2015).

[5] M. Z. Alam, I. De Leon, and R. W. Boyd, Science **352**, 795 (2016).

[6] D. B. S. Soh, R. Hamerly, and H. Mabuchi, Phys. Rev. A **94**, 023845 (2016).

[7] A. Sipahigil, R. E. Evans, D. D. Sukachev, M. J. Burek, J. Borregaard, M. K. Bhaskar, C. T. Nguyen, J. L. Pacheco, H. A. Atikian, C. Meuwly, R. M. Camacho, F. Jelezko, E. Bielejec, H. Park, M. Loncar, and M. D. Lukin, Science **354**, 847 (2016).

[8] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, Nano Lett. **10**, 1271 (2010).

[9] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).

[10] J. K. Ellis, M. J. Lucero, and G. E. Scuseria, Appl. Phys. Lett. **99**, 261908 (2011).

[11] W. Zhao, Z. Ghorannevis, L. Chu, M. Toh, C. Kloc, P.-H. Tan, and G. Eda, ACS Nano **7**, 791 (2013).

[12] R. S. Sundaram, M. Engel, A. Lombardo, R. Krupke, A. C. Ferrari, Ph. Avouris, and M. Steiner, Nano Lett. **13**, 1416 (2013).

[13] Y. Zhang, T.-R. Chang, B. Zhou, Y.-T. Cui, H. Yan, Z. Liu, F. Schmitt, J. Lee, R. Moore, Y. Chen, H. Lin, H.-T. Jeng, S.-K. Mo, Z. Hussain, A. Bansil, and Z.-X. Shen, Nat. Nanotechnol. **9**, 111 (2013).

[14] D. J. Clark, V. Senthilkumar, C. T. Le, D. L. Weerawarne, B. Shim, J. I. Jang, J. H. Shim, J. Cho, Y. Sim, M.-J. Seong, S. H. Rhim, A. J. Freeman, K.-H. Chung, and Y. S. Kim, Phys. Rev. B **90**, 121409(R) (2014).

[15] P. Yan, A. Liu, Y. Chen, H. Chen, S. Ruan, C. Guo, S. Chen, I. L. Li, H. Yang, J. Hu, and G. Cao, Opt. Express **23**, 154 (2015).

[16] C.-Y. Wang and G.-Y. Guo, J. Phys. Chem. C **119**, 13268 (2015).

[17] R. I. Woodward, R. T. Murray, C. F. Phelan, R. E. P. de Oliveira, T. H. Runcorn, E. J. R. Kelleher, S. Li, E. C. de Oliveira,

G. J. M. Fechine, G. Eda, and C. J. S. de Matos, 2D Mater. **4**, 011006 (2016).

[18] D. B. S. Soh, C. Rogers, D. J. Gray, E. Chatterjee, and H. Mabuchi, Phys. Rev. B **97**, 165111 (2018).

[19] L. A. Lugiato, Contemp. Phys. **24**, 333 (1983).

[20] M. M. Ugeda, A. J. Bradley, S.-F. Shi, F. H. da Jornada, Y. Zhang, D. Y. Qiu, W. Ruan, S.-K. Mo, Z. Hussain, Z.-X. Shen, F. Wang, S. G. Louie, and M. F. Crommie, Nat. Mater. **13**, 1091 (2014).

[21] A. T. Hanbicki, M. Currie, G. Kioseoglou, A. L. Friedman, and B. T. Jonker, Solid State Commun. **203**, 16 (2015).

[22] B. Zhu, X. Chen, and X. Cui, Sci. Rep. **5**, 9218 (2015).

[23] D. W. Kidd, D. K. Zhang, and K. Varga, Phys. Rev. B **93**, 125423 (2016).

[24] A. F. Rigosi, H. M. Hill, K. T. Rim, G. W. Flynn, and T. F. Heinz, Phys. Rev. B **94**, 075440 (2016).

[25] H. Wang, C. Zhang, W. Chan, C. Manolatou, S. Tiwari, and F. Rana, Phys. Rev. B **93**, 045407 (2016).

[26] M. Palummo, M. Bernardi, and J. C. Grossman, Nano Lett. **15**, 2794 (2015).

[27] M. Selig, G. Berghäuser, A. Raja, P. Nagler, C. Schüller, T. F. Heinz, T. Korn, A. Chernikov, E. Malic, and A. Knorr, Nat. Commun. **7**, 13279 (2016).

[28] N. Zhao, Z. Zhao, I. A. D. Williamson, S. Boutami, B. Zhao, and S. Fan, ACS Photon. **6**, 339 (2019).

[29] F. Wang, Y. Wu, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. B **73**, 245424 (2006).

[30] D. Sun, Y. Rao, G. Reider, G. Chen, Y. You, L. Brezin, A. R. Harutyunyan, and T. F. Heinz, Nano Lett. **14**, 5625 (2014).

[31] C. Poellmann, P. Steinleitner, U. Leierseder, P. Nagler, G. Plechinger, M. Porer, R. Bratschitsch, C. Schuller, T. Korn, and R. Huber, Nat. Mater. **14**, 889 (2015).

[32] N. Kumar, Q. Cui, F. Ceballos, D. He, Y. Wang, and H. Zhao, Phys. Rev. B **89**, 125427 (2014).

[33] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, Nat. Nanotechnol. **6**, 147 (2011).

[34] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nat. Nanotechnol. **7**, 699 (2012).

[35] N. S. Rytova, arXiv:1806.00976.

[36] L. V. Keldysh, JETP Lett. **29**, 658 (1979).

[37] O. L. Berman and R. Y. Kezerashvili, Phys. Rev. B **96**, 094502 (2017).

[38] J. R. Schaibley, H. Yu, G. Clark, P. Rivera, J. S. Ross, K. L. Seyler, W. Yao, and X. Xu, Nat. Rev. Mater. **1**, 16055 (2016).

[39] K. Chen, R. Ghosh, X. Meng, A. Roy, J.-S. Kim, F. He, S. C. Mason, X. Xu, J.-F. Lin, D. Akinwande, S. K. Banerjee, and Y. Wang, NPJ 2D Mater. Appl. **1**, 15 (2017).

[40] V. Shahnazaryan, I. Iorsh, I. A. Shelykh, and O. Kyriienko, Phys. Rev. B **96**, 115409 (2017).

[41] H. Haug and S. W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (World Scientific, Singapore, 2009).

[42] C. Zhang, H. Wang, W. Chan, C. Manolatou, and F. Rana, Phys. Rev. B **89**, 205436 (2014).

[43] A. R. Klots, A. K. M. Newaz, B. Wang, D. Prasai, H. Krzyzanowska, J. Lin, D. Caudel, N. J. Ghimire, J. Yan, B. L. Ivanov, K. A. Velizhanin, A. Burger, D. G. Mandrus, N. H. Tolk, S. T. Pantelides, and K. I. Bolotin, Sci. Rep. **4**, 6608 (2014).

[44] E. Fermi, *Nuclear Physics* (University of Chicago Press, Chicago, 1950), formula VIII.2.

[45] V. A. Malyshev, H. Glaeske, and K.-H. Feller, Chem. Phys. Lett. **305**, 117 (1999).

[46] V. A. Malyshev, G. G. Kozlov, H. Glaeske, and K.-H. Feller, Chem. Phys. **254**, 31 (2000).

[47] A. Kormanyos, V. Zolyomi, N. D. Drummond, and G. Burkard, Phys. Rev. X **4**, 011034 (2014).

[48] K. Kim, Z. Lee, W. Regan, C. Kisielowski, M. F. Crommie, and A. Zettl, ACS Nano **5**, 2142 (2011).

[49] H. Wang, C. Zhang, and F. Rana, Nano Lett. **15**, 1 (2015).

[50] P. Vancso, G. Z. Magda, J. Peto, J.-Y. Noh, Y.-S. Kim, C. Hwang, L. P. Biro, and L. Tapszto, Sci. Rep. **6**, 29726 (2016).

[51] X. Liu, I. Balla, H. Bergeron, and M. C. Hersam, J Phys. Chem. C **120**, 20798 (2016).

[52] C. Rogers, D. Gray, N. Bogdanowicz, and H. Mabuchi, Phys. Rev. Mat. **2**, 094003 (2018).

[53] R. C. Hilborn, Am. J. Phys. **50**, 982 (1982).

[54] M. Ohtsu, *Progress in Nano-Electro Optics III: Industrial Applications and Dynamics of the Nano-Optical System* (Springer, Berlin, 2005).

[55] S. Hunsche, K. Leo, H. Kurz, and K. Kohler, Phys. Rev. B **49**, 16565 (1994).

[56] S. Schmitt-Rink, D. S. Chemla, and D. A. B. Miller, Phys. Rev. B **32**, 6601 (1985).

[57] F. Ceballos, Q. Cui, M. Z. Bellus, and H. Zhao, Nanoscale **8**, 11681 (2016).

[58] M. Amani, D.-H. Lien, D. Kiriya, J. Xiao, A. Azcatl, J. Noh, S. R. Madhvapathy, R. Addou, S. KC, M. Dubey, K. Cho, R. M. Wallace, S.-C. Lee, J.-H. He, J. W. Ager III, X. Zhang, E. Yablonovitch, and A. Javey, Science **350**, 1065 (2015).

[59] M. Amani, P. Taheri, R. Addou, G. H. Ahn, D. Kiriya, D.-H. Lien, J. W. Ager III, R. M. Wallace, and A. Javey, Nano Lett **16**, 2786 (2016).

[60] G. Wei, D. A. Czaplewski, E. J. Lenferink, T. K. Stanev, I. W. Jung, and N. P. Stern, Sci. Rep. **7**, 3324 (2017).

[61] X. Ding, F. Peng, J. Zhou, W. Gong, G. Slaven, K. P. Loh, C. T. Lim, and D. T. Leong, Nat. Commun. **10**, 41 (2019).