Excitonic resonances control the temporal dynamics of nonlinear optical wave mixing in monolayer semiconductors

Jonas Bauer
University of Regensburg
Lijue Chen
Xiamen University https://orcid.org/0000-0002-8872-2723
Philipp Wilhelm
University of Regensburg
Kenji Watanabe
National Institute for Materials Science https://orcid.org/0000-0003-3701-8119
Takashi Taniguchi
National Institute for Materials Science, Tsukuba, Ibaraki https://orcid.org/0000-0002-1467-3105
Sebastian Bange
University of Regensburg https://orcid.org/0000-0002-5850-264X
John Lupton
University of Regensburg https://orcid.org/0000-0002-7899-7598
Kai-Qiang Lin (kaiqiang.lin@ur.de)
University of Regensburg https://orcid.org/0000-0001-9609-749X

Article

Keywords:

Posted Date: April 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1347888/v1

Version of Record: A version of this preprint was published at Nature Photonics on October 17th, 2022.
See the published version at https://doi.org/10.1038/s41566-022-01080-1.

Excitonic resonances control the temporal dynamics of nonlinear optical wave mixing in monolayer semiconductors

Jonas M. Bauer[1†], Lijue Chen[1,2†], Philipp Wilhelm[1], Kenji Watanabe[3], Takashi Taniguchi[4], Sebastian Bange[1], John M. Lupton[1], Kai-Qiang Lin[1]*

[1]*Department of Physics, University of Regensburg, Regensburg, Germany.*

[2]*College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, China.*

[3]*Research Center for Functional Materials, National Institute for Materials Science, Tsukuba, Japan.*

[4]*International Center for Materials Nanoarchitectonics, National Institute for Materials Science, Tsukuba, Japan.*

Monolayer semiconductors are emerging platforms for strong nonlinear light-matter interaction, which is enhanced by the giant oscillator strength of tightly bound excitons. Little attention has been paid to the impact of excitonic resonances on the temporal dynamics of nonlinear light-matter interaction, since harmonic generation and optical-wave mixing are generally considered instantaneous processes. We find that a significant time difference, ranging from -40 fs to +120 fs, is necessary between two light pulses for optimal sum-frequency generation (SFG) and four-wave mixing (FWM) to occur from monolayer WSe_2 when one of the pulses is in resonance with an excitonic transition. These resonances involve both band-edge A-excitons (AX) as well as high-lying excitons (HX) comprising electrons from conduction bands far above the gap. Numerical simulations of the density-matrix evolution reproduce and explain the distinct dynamics of SFG and FWM. The interpulse delays for maximal SFG and FWM are governed primarily by the lifetime of the one-photon and two-photon resonant states, respectively. The method therefore offers an unconventional probe of excitonic dynamics that are either one-photon or two-photon allowed. Remarkably, the longest delay times occur at the lowest excitation powers, indicating a strong nonlinearity that offers exploration potential for excitonic quantum nonlinear optics.

Nonlinear optical wave mixing processes[1,2] such as second-harmonic generation (SHG), sum-frequency generation (SFG) and four-wave mixing (FWM) play an important role in the design of lasers[3], the characterization of ultrashort laser pulses, in photon detection, and in optical sensing[4]. Since light does not interact appreciably with itself, the source of the nonlinear optical response is a material's electronic polarization that is driven nonlinearly by several incident light fields[1]. Descriptions of nonlinear polarization usually emphasize the perspective that the polarization measured is the product of several incident electric fields with the nonlinear susceptibility at a set of exact frequencies, and as such are only directly applicable under continuous-wave conditions[2]. Wave mixing is thus generally described as a *quasi-instantaneous* parametric process, for which the quantum state of matter does not change. Ultimately, though, this assumption can only be true for the initial and final states of the interaction since, in the quantum picture of the interaction of light and matter, every interaction inevitably changes the quantum state for both interaction partners. In the density-matrix representation of the quantum states of matter, the expectation value for the observable polarization $P(t) = \langle \hat{\mu}\, \hat{\rho}(t) \rangle$ therefore normally contains an integration over all possible quantum pathways and involves all state transitions with non-zero dipole-moment matrix elements μ_{ij} for all possible times or frequencies at which electric fields are present[1,5]. This approach has been applied successfully to multidimensional coherent spectroscopy[6], where excitation of a nonlinear response by multiple pulsed laser sources is common. An important consequence of this extended picture of wave mixing is that parametric interactions progress via diagonal elements ρ_{ii} of the density matrix, which means that they can be sensitive not only to the dephasing of transition dipoles, but also to the lifetime of excited states, state depletion, and transition saturation.

Transition-metal dichalcogenide (TMDC) monolayers have emerged as an intriguing two-dimensional medium to accommodate strong light-matter interactions[7-11]. As a direct-gap

semiconductor with weak dielectric screening, TMDC monolayers host strongly bound excitons near the K-points of the Brillouin zone[12-14]. Due to broken inversion symmetry, the materials exhibit nonlinear responses of both even and odd order, making them a target for pursuing a wide range of applications in nonlinear wave mixing[7,11,15,16]. Among other effects, the giant excitonic enhancement in nonlinearity[7,15,17] near resonance exhibited by these materials has enabled the demonstration of SHG and SFG under continuous-wave irradiation[18,19]. A pronounced splitting of the SHG spectrum due to Rabi flopping of the excited-state population occurs even under comparatively weak pulsed-laser excitation[20], and is associated with light-driven quantum interference between the excitation pathways of the bound band-edge A-exciton (AX) and a dark higher-lying exciton (HX) involving an upper conduction band[19]. Such a ladder-type excitonic multilevel system differs fundamentally from conventional biexcitonic systems[21], where two band-edge excitons are excited and form a bound state that contributes to the nonlinear interaction[22,23]. Besides SHG and SFG, single-pass optical parametric amplification has also been demonstrated in TMDC monolayers[24]. Although excitonic enhancement of the nonlinear light-matter interaction is now well established, it is not at all clear whether excitonic transitions affect the *temporal dynamics* of the nonlinear interactions.

Figure 1a illustrates the setup for a two-color excitation experiment using ultrafast laser pulses of 130-140 fs length and 80 MHz repetition rate at 840 nm ($\hbar\omega_{\mathrm{pump}} = 1.739$ eV) and 713 nm ($\hbar\omega_{\mathrm{probe}} = 1.476$ eV), impinging on an exfoliated WSe_2 monolayer with a time difference $\Delta t = t_{\mathrm{probe}} - t_{\mathrm{pump}}$. For notational convenience, we refer to the pulse resonant with the A-exciton as the "pump" pulse and the pulse detuned from any excitonic resonance as the "probe" pulse, although we do not limit Δt to positive values nor do we restrict the probe beam to lower intensities than the pump beam. The sample is cooled to a temperature of 5 K in a microscope cryostat and excited through a long-working-distance microscope objective with a numerical aperture (NA) of 0.6. Further details of the setup are shown in Supplementary Fig. 1. Figure 1b

plots a representative spectrum of the anti-Stokes radiation detected in back-reflected geometry through the same objective, for a time delay of $\Delta t = 0$ and using a 680 nm (1.82 eV) short-pass filter. Five characteristic spectral contributions are identified from their frequencies: the SHG of the pump beam at $2\omega_{\mathrm{pump}}$, the SFG of the pump and probe beams at $\omega_{\mathrm{pump}} + \omega_{\mathrm{probe}}$, the SHG of the probe beam at $2\,\omega_{\mathrm{probe}}$, a FWM contribution at $2\omega_{\mathrm{pump}} - \omega_{\mathrm{probe}}$, and the upconverted photoluminescence (UPL) of the so-called "B-exciton" close to the FWM peak[19]. As monolayer TMDCs are characterized by giant second- and third-order optical susceptibilities[7], both second-order nonlinear processes, such as SFG, and third-order nonlinear processes, such as FWM, are efficient[15,24,25]. Excitation and detection through the high-NA objective bypasses most of the constraints on beam direction due to conservation of momentum that are usually necessary for efficient nonlinear wave mixing[2].

Figure 1c shows a 2D map of the anti-Stokes range as a function of the pump-probe delay time Δt. As expected, neither the SHG of either pump or probe beam nor the UPL show any dependence on the temporal overlap between pump and probe impulses. On the other hand, both SFG and FWM are sensitive to the delay, as would be expected. However, the maximum efficiency of these two processes does not occur at time delays of zero, which is surprising given that parametric, instantaneous nonlinear optical wave mixing should be most effective for optical pulse overlap at zero delay. As the time delay between the two laser impulses is sensitive to group delays accumulated in the optical setup, the zero delay is referenced to the SFG response of a 100-µm thick beta barium borate (BBO) crystal cut at 29.2°, which is placed next to the monolayer WSe_2 sample as detailed in the Supplementary Fig. 1. Figure 1d shows the SFG intensity from the BBO crystal (gray spheres) as well as the spectrally integrated SFG (blue spheres) and FWM (orange spheres) intensities of the WSe_2 sample as a function of the delay time, along with Gaussian fits (lines). The dashed lines mark the center position for each

Gaussian, with the FWM peaking at a delay time Δt_{FWM} = 80 fs (orange) and the SFG peaking at Δt_{SFG} = 120 fs (blue).

To examine the role of excitons in the retardation of these nonlinear processes, we perform the pump-probe measurements as a function of photon energy. Figure 2a shows the dependence of the time delay for maximal SFG, Δt_{SFG}, (blue spheres) on the energy of the pump laser photons for a fixed $\hbar\omega_{\mathrm{probe}} = 1.476$ eV. Comparison with the PL spectrum (black curve) reveals that the largest Δt_{SFG} of 120 fs is reached precisely when the pump laser is in resonance with the A-exciton. Under detuning from the resonance, Δt_{SFG} decreases to zero. Since both pump and probe pulses have a short pulse duration and therefore a broad spectrum, the peak in Δt_{SFG} on resonance appears broader than the A-exciton PL peak.

If the delay is caused by an excitonic resonance, it should depend on excitation density. Figure 2b shows Δt_{SFG} for monolayer WSe_2 (blue spheres) as a function of pump power for $\hbar\omega_{\mathrm{pump}} = 1.739$ eV, $\hbar\omega_{\mathrm{probe}} = 1.476$ eV, and a probe-beam power of 400 µW. Upon increasing the pump power by three orders of magnitude, Δt_{SFG} decreases from values exceeding 120 fs down to near zero. One may be tempted to assign such a decrease to the decreasing lifetime of the A-exciton with increasing excitation density as observed in transient intra and interband spectroscopy[26] and optical two-dimensional Fourier-transform spectroscopy[27]. However, as we demonstrate below, this decrease can also be rationalized in terms of the dynamics of Rabi flopping[28]. In contrast to the effect of a variation in pump power, Supplementary Fig. 2a shows that comparable changes of the probe power do not have any detectable influence on Δt_{SFG} for monolayer WSe_2. To exclude any possibility of optical artifacts and establish a further reference for the measurements, the experiments were repeated for a monolayer of WS_2 in place of the WSe_2. Since the band gap of monolayer WS_2 is approximately 0.3 eV larger than that of monolayer WSe_2, the lasers will not be in resonance with the excitonic transitions. Indeed,

Δt_{SFG} from monolayer WS_2 (purple spheres in Fig. 2b) is clearly close to zero over the entire range of pump powers.

The FWM signals detected simultaneously in the experiments above are shown in Fig. 2c-d (orange spheres). In contrast to Δt_{SFG}, the delay time for maximum FWM, Δt_{FWM}, does not peak at the A-exciton resonance (Fig. 2d), but rather at a pump-photon energy that also gives rise to quantum interference in the corresponding SHG as shown in Supplementary Fig. 3[19,20]. This unusual high-lying excitonic state (HX) appears at 3.44 eV and contributes to the delay Δt_{FWM} through a two-photon resonance with the pump laser as illustrated in the inset of Fig. 2c. As found above for Δt_{SFG}, Δt_{FWM} decreases with increasing pump power (orange circles in Fig. 2d) and is independent of probe power (Supplementary Fig. 2b). Unexpectedly, however, Δt_{FWM} actually drops below 0 in the regime of high pump powers: at powers exceeding 1 mW, the center of the "probe" pulse must reach the sample 40 fs *before* the center of the "pump" pulse for maximum efficiency of FWM. As reported in Supplementary Fig. 4a, the maximal values of $|\Delta t_{\mathrm{SFG}}|$ and $|\Delta t_{\mathrm{FWM}}|$ become even larger for WSe_2 monolayer samples with hBN encapsulation. This change may conceivably arise from a suppression of local sample inhomogeneities[29,30] and the modification of the radiative decay rate due to a cavity effect[31]. The fact that the maximal values of Δt_{SFG} and Δt_{FWM} appear at the *lowest* powers of both pump and probe beams (Supplementary Fig. 4b) demonstrates that strong nonlinear interactions extend into the regime of the low excitation densities, opening up avenues for further exploration of quantum nonlinear optics[32,33].

While the change of interpulse delay for SFG can, in principle, be rationalized by a pump-power dependence of a sufficiently long-lived excitonic resonance[27], the *negative* delay for the maximal FWM signal is not straightforward to explain. To account for the excitonic resonances, instead of pursuing the conventional analysis of the parametric light-matter interaction up to a fixed nonlinear order, we employ a full time-dependent solution of the density-matrix dynamics

in the presence of the laser field. This approach uses a simplified state description borrowed from atomic optics and has been successfully employed before to describe the effect of Rabi flopping in the unusual SHG spectra observed from monolayer WSe_2[20]. The polarization $P(t) \sim \sum_{ij} \mu_{ij} \rho_{ij}(t)$ is understood as the source term for scattered radiation. Specific information about the SFG and FWM spectra is obtained by analyzing $P(t)$ in the Fourier domain. We solve the time dynamics of the density matrix $\hat{\rho}$ in the framework of a Lindblad-type equation $\frac{d}{dt}\hat{\rho} = \frac{1}{i\hbar}\left[\hat{H}, \hat{\rho}\right] - \frac{1}{2}\left\{\hat{\Gamma}, \hat{\rho}\right\} + \hat{\Lambda}$. Here, the density matrix is defined in the conventional way as $\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, with p_i the probability for the system to be in state $|\psi_i\rangle$. The Hamiltonian is given by $H_{ii} = E_i$ and $H_{ij} = -\mathcal{E}(t)\mu_{ij}$ $(i \neq j)$, where E_i are the eigenstate energies, $\mathcal{E}(t)$ is the electric-field component of the incident light field, and μ_{ij} are the transition dipole moments linked to the ρ_{ij} coherence. All vectorial properties of the field and the transition dipoles are neglected for sake of simplicity. The decay matrix $\hat{\Gamma}$ is diagonal with non-zero diagonal elements $\Gamma_{ii} = \Gamma_i$ $(i > 1)$, and repopulation of the ground state is described by $\hat{\Lambda}$ with a single non-zero matrix element $\Lambda_{11} = \sum_{i>1} \Gamma_i \, \rho_{ii}$. For realistic excitation conditions, the laser fields are modelled as Gaussian pulses with parameters matching the experimental ones as illustrated in Fig. 3a. Based on our earlier success of correctly reproducing experimental SHG spectra from WSe_2[20], we again describe the allowed eigenstates by a simple ladder-type three-level system shown in Fig. 3a, consisting of a ground state $|1\rangle$, the AX state $|2\rangle$, and the HX state $|3\rangle$. The detailed parameters used for the simulation are based on the experimental conditions and are listed in the Supplementary Fig. 5 and Supplementary Table 1.

Figure 3b shows a typical optical wave mixing spectrum calculated from the time-dependent polarization for the same spectral range as probed experimentally (Fig. 1b), at zero time delay between the pump pulse at $\hbar\omega_{\mathrm{pump}} = 1.739$ eV and the probe pulse at $\hbar\omega_{\mathrm{probe}} = 1.476$ eV. Five prominent peaks appear, including FWM ($2\omega_{\mathrm{pump}} - \omega_{\mathrm{probe}}$), six-wave mixing

($3\omega_{\mathrm{pump}} - 2\omega_{\mathrm{probe}}$, SWM), SHG (both $2\omega_{\mathrm{pump}}$ and $2\omega_{\mathrm{probe}}$), and SFG ($\omega_{\mathrm{pump}} + \omega_{\mathrm{probe}}$). The SWM feature is not as prominent in the experimental spectrum (Fig. 1b) but can be identified when the powers of both pump and probe pulses are sufficiently high (see Supplementary Fig. 6). Figure 3c shows the result of calculating spectrally-integrated SFG and FWM signal intensities from the numerical simulation as a function of the delay time between pump and probe pulses. Here, the pump laser is set to be resonant with state $|2\rangle$ and the pump power is set to a low value of 1 μW. The simulated maximal SFG intensity appears for a positive delay time of approximately 105 fs, close to the corresponding experimental value of 120 fs found in Fig. 1d. The maximal FWM intensity appears at a delay time of 50 fs, which is somewhat smaller than the experimental result, but confirms the experimental trend that Δt_{FWM} tends to be smaller than Δt_{SFG}.

Next, we investigate the influence of the pump power and photon energy on Δt_{SFG} and Δt_{FWM} in the simulations. Figure 4a,c shows the dependence on pump-photon energy of Δt_{SFG} and Δt_{FWM} at a low pump power of 1 μW. Maximum SFG intensity arises for a pump beam that is resonant with the $|1\rangle \rightarrow |2\rangle$ transition (the AX), while the maximum FWM intensity is found when the pump laser is two-photon resonant with the $|1\rangle \rightarrow |3\rangle$ transition, i.e. with the HX. These resonant enhancements coincide with the experimental observations in Fig. 2a,d. Figure 4b,d plots the simulated pump-power dependence of Δt_{SFG} and Δt_{FWM}. Both Δt_{SFG} and Δt_{FWM} drop with increasing pump power, qualitatively matching the experiment. The simulation, however, shows that both Δt_{SFG} and Δt_{FWM} become negative at high pump powers, although only Δt_{FWM} is found to do so in experiment. This divergence is further discussed below.

The simulations can be rationalized by studying the corresponding dynamics of the density-matrix elements as shown in Fig. 4e,f for both low (1 μW) and high (1 mW) pump powers. We propose that the largest SFG and FWM intensities are obtained when the center of the probe

pulse, for which the field strength is highest, overlaps in time with the largest transient population of the excitonic state that resonantly enhances the nonlinear wave mixing. Since both the AX state $|2\rangle$ and the HX state $|3\rangle$ must have short but finite lifetimes, the population of these states accumulates over the short timescale of the pump-probe pulse sequence. In the regime of low pump powers, the maximum state occupancy then appears at positive delay times as shown in Fig. 4e. The delay times for the maximal population density of states $|2\rangle$ and $|3\rangle$ match well with Δt_{SFG} and Δt_{FWM}, respectively. As the pump power increases, however, the laser field becomes sufficiently strong, so that Rabi flopping in the state populations becomes relevant on the timescale of the interpulse delay. The signatures of the onset of this Rabi flopping have been studied in detail in the context of SHG, when the two-photon resonance of the $|1\rangle \rightarrow |3\rangle$ HX transition is driven optically[20]. Since the early depopulation of the excited states shifts the maximum occupancy to earlier times, Δt_{SFG} and Δt_{FWM} decrease with increasing pump power. Unlike in the experiment, however, in the simulation Δt_{SFG} does become negative. As discussed previously in Ref. 20, these density-matrix simulations do not include effects of many-body interactions such as enhanced Auger recombination at high pump fluences and large exciton densities. We speculate that the efficient Auger-like exciton-exciton annihilation of band-edge excitons[26,34] in monolayer WSe_2 significantly raises the decay rate of state $|2\rangle$ at high exciton densities, and therefore prevents Δt_{SFG} from becoming negative. As shown in Fig. 5, the simulated delay time Δt_{SFG} for maximum SFG intensity is very sensitive to the decay rate of state $|2\rangle$ and even approaches zero as this rate tends to infinity. In stark contrast to this, Fig. 5 shows that Δt_{FWM} is mainly sensitive to the decay rate of state $|3\rangle$ and rather insensitive to that of state $|2\rangle$. This difference in the dependence of SFG and FWM on the excitonic states found in the simulations suggests an explanation for the experimental observation that the experimental Δt_{FWM} becomes negative whereas Δt_{SFG} does not, considering that high fluences will limit the lifetime of the AX. More importantly, this

difference implies that SFG and FWM can be employed to probe the dynamics of one-photon allowed and two-photon allowed excitonic states simultaneously but independently of each other. The excitonic dynamics of one-photon allowed transitions are straightforward to measure by different optical means such as time-resolved PL, transient intra- and interband spectroscopy[26] or optical two-dimensional Fourier-transform spectroscopy[27]. The dynamics of two-photon allowed dark excitonic states remain much more challenging to access.

In summary, we studied the influence of the delay between two laser pulses on the second- and third-order optical wave mixing processes in monolayer WSe_2. We found that significant delays between the two pulses are necessary to maximize the SFG and FWM signal intensities. Understanding these time delays poses challenges to the description of these spectroscopic observables in terms of the conventional picture of low-order perturbative nonlinear interactions[2]. The observations can, however, be rationalized in the framework of coherent exciton population dynamics under optical pumping as described within the density-matrix formalism that computes the interaction of light and matter to arbitrary perturbative order. Further insight into these processes will necessitate a disentangling of the multitude of Liouville-space pathways embodied in the density-matrix time dynamics. The qualitative agreement between simulation and experiment found here suggests that the simplified ladder-type excitonic three-level system indeed captures the essential physical contributions to the delay effects observed in the optical wave mixing. Our work demonstrates that nonlinear pump-probe spectroscopy offers new opportunities for measuring the population dynamics of optically pumped excitonic states that are either one- or two-photon allowed, whereas the dynamics of the latter have proven challenging to access so far. The strong nonlinear interactions that enable the time delays to persist into the low-excitation regime promise to open up an as-yet unexplored regime of exciton-based quantum nonlinear optics in TMDC monolayers.

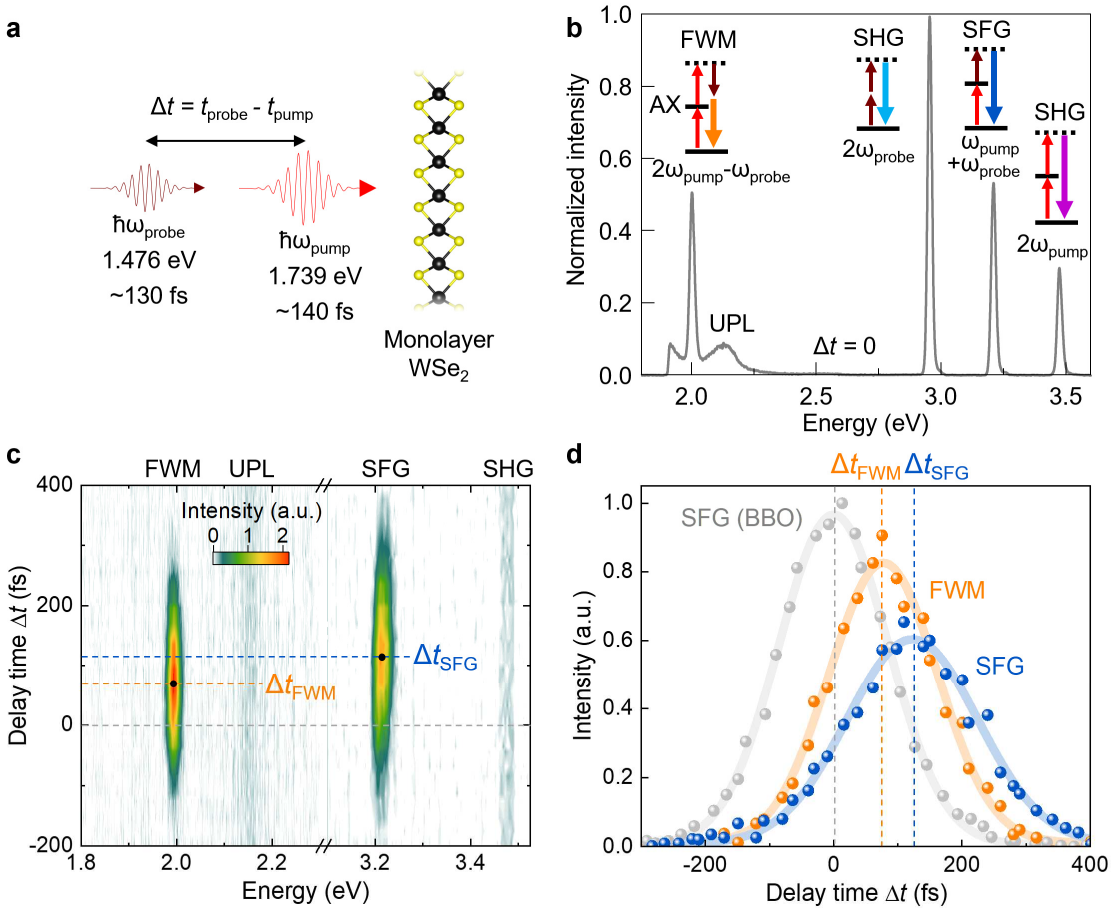

Fig. 1 | Delayed sum-frequency generation (SFG) and four-wave mixing (FWM) from monolayer WSe_2. a, Schematic illustration of the pump-probe experiment with two laser pulses at 1.739 eV and 1.476 eV focused onto a monolayer WSe_2 at a temperature of 5 K. **b**, Spectrum of the light generated, detected in reflection geometry. FWM, upconverted photoluminescence (UPL), second-harmonic generation (SHG) and SFG appear for a time delay $\Delta t = 0$ between the position of peak intensities of pump and probe laser impulses as defined in panel **a**. The pump beam is in resonance with the band-edge A-exciton (AX). **c**, Two-color optical wave mixing spectrum as a function of Δt. Maximal SFG and FWM intensities appear at non-zero delay times. **d**, Spectrally integrated intensity of SFG (blue) and FWM (orange) as a function of Δt. The SFG intensity of a 100-µm-thick beta barium borate (BBO) crystal (grey) in place of the WSe_2 sample is shown as a function of Δt to provide a zero-time-delay reference.

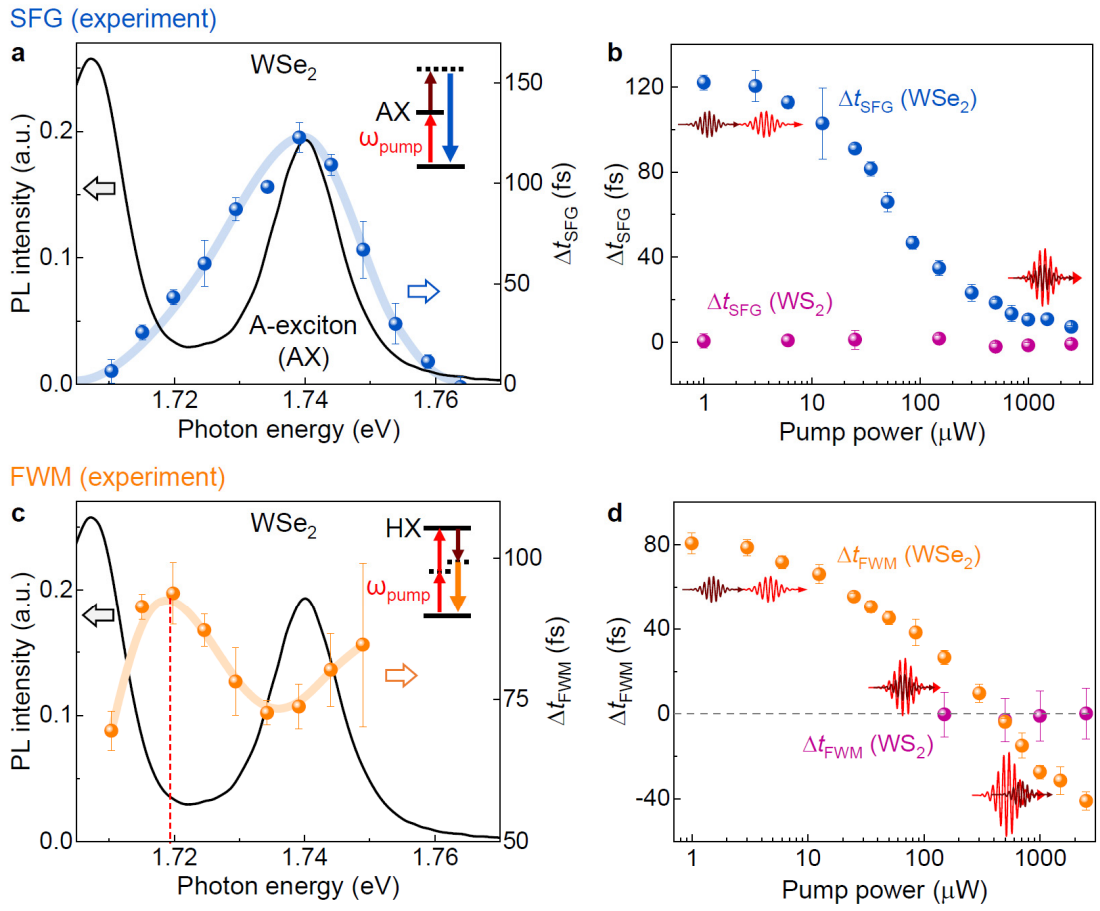

Fig. 2 | Excitonic effects on the delay time for maximal SFG and FWM. **a**, Comparison of the PL spectrum of monolayer WSe_2 (black curve) to the delay time for maximal SFG, Δt_{SFG}, as a function of the pump photon energy (blue spheres). The power of the probe beam is set to 400 μW at 1.476 eV and the pump beam is set to 3 μW. The blue line is a guide to the eye. **b**, Dependence of Δt_{SFG} on pump beam power for monolayer WSe_2 (blue spheres) and monolayer WS_2 (purple spheres) using $\hbar\omega_{\mathrm{pump}} = 1.739$ eV in resonance with the A-exciton, $\hbar\omega_{\mathrm{probe}} = 1.476$ eV, and a probe-beam power of 400 μW. **c,** Delay time for maximal FWM, Δt_{FWM}, as a function of the pump photon energy (orange spheres) for monolayer WSe_2. The measurement conditions are the same as in panel **a**. The orange line is a guide to the eye. The dashed line marks the two-photon resonance with a high-lying exciton (HX). **d,** Dependence of Δt_{FWM} on probe-beam power for monolayer WSe_2 (orange) and monolayer WS_2 (purple) using $\hbar\omega_{\mathrm{pump}} = 1.739$ eV, $\hbar\omega_{\mathrm{probe}} = 1.476$ eV, and a probe-beam power of 400 μW. The red and dark-red wave packets indicate the temporal relation of pump and probe pulses.

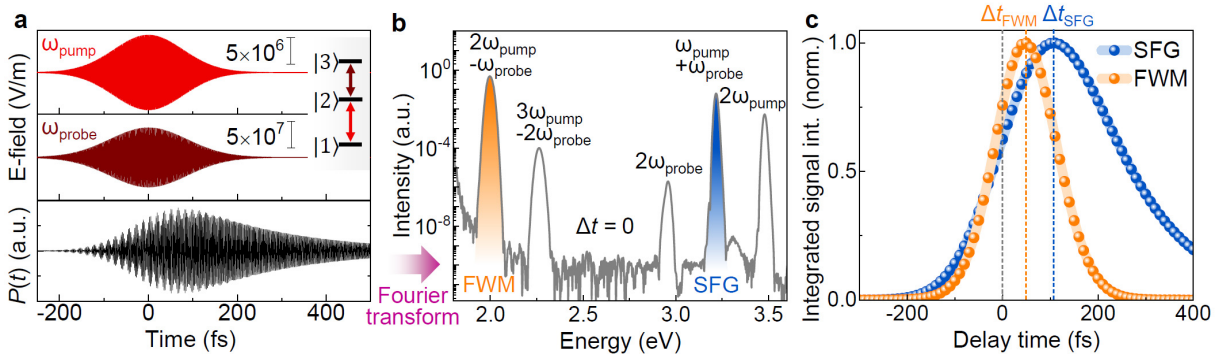

Fig. 3 | Simulation of nonlinear optical wave mixing spectroscopy. **a,** Illustration of electric fields and polarization in the numerical simulation model used to describe the population evolution in a ladder-type three-level excitonic system. **b,** Representative calculated optical wave mixing spectrum at delay time zero. The FWM contribution is highlighted in orange and SFG is marked blue. **c,** Spectrally integrated intensity of SFG and FWM signals as a function of delay time between pump and probe pulses. Δt_{SFG} and Δt_{FWM} mark the delay times for maximal intensity of SFG and FWM, respectively.

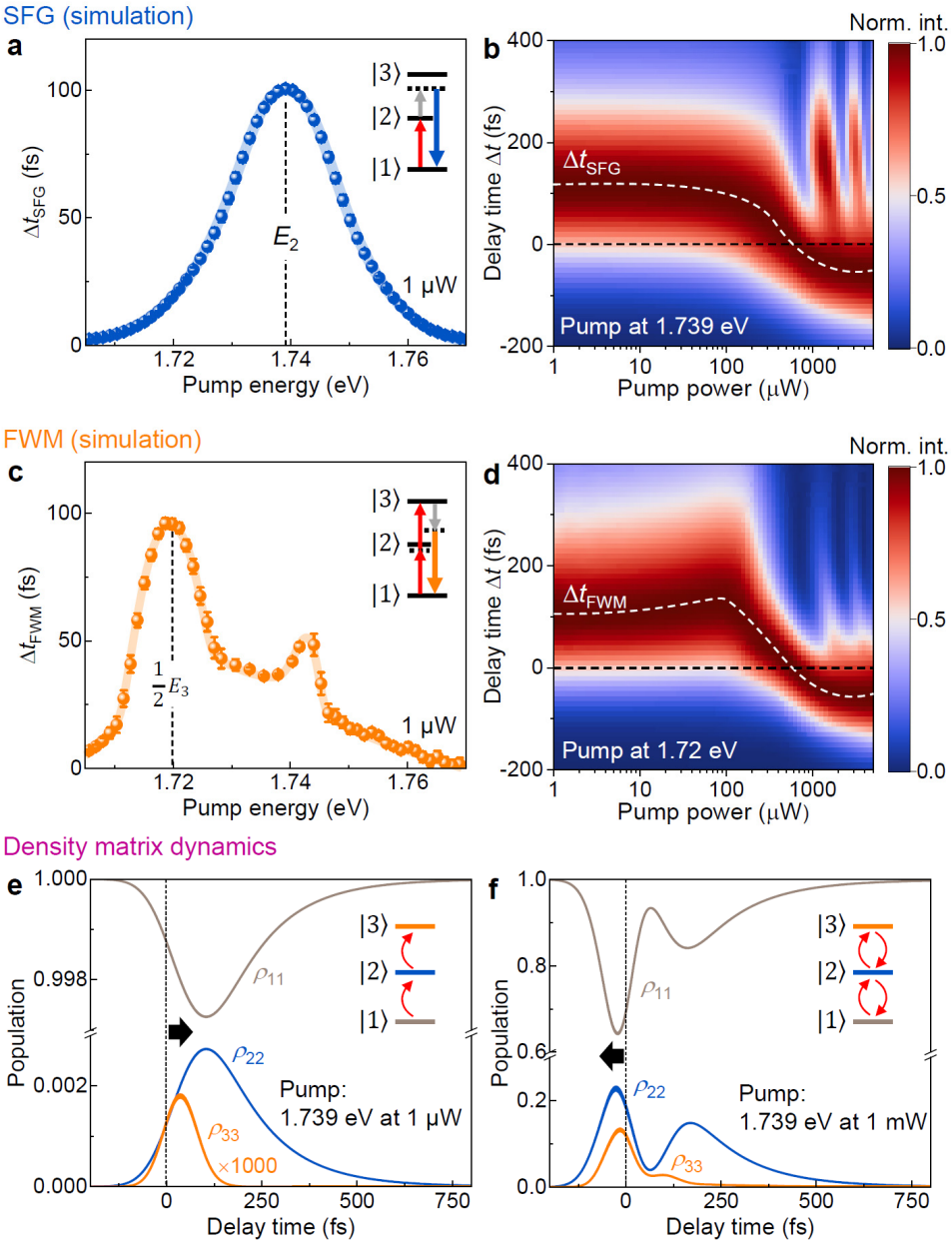

Fig. 4 | Delay time for maximal SFG and FWM intensity computed in the density-matrix formalism. a, Delay times Δt_{SFG} for maximal simulated SFG as a function of the pump-photon energy. The pump at 1.739 eV is set to a power of 1 μW while the probe at 1.476 eV is set to 200 μW. **b,** Corresponding dependence of the intensity of SFG on delay time as a function of pump power, with the pump in resonance with the $|1\rangle \rightarrow |2\rangle$ excitonic transition (AX) at 1.739 eV. **c,** Delay time for maximal FWM intensity, Δt_{FWM}, as a function of the pump-photon energy. The pump power and the parameters for the probe beam are chosen as for panel a. **d,** Corresponding dependence of the simulated FWM intensity on delay time as a function of pump power, with the pump in two-photon resonance with the $|1\rangle \rightarrow |3\rangle$ (HX) transition. **e,** Density-matrix dynamics of the excitonic three-level system for a pump

power of 1 μW, without the probe beam. State |2⟩ reaches maximum occupancy at a delay time of 105 fs with respect to the center of the pump pulse. **f**, Density-matrix dynamics for a pump power of 1 mW. State |3⟩ reaches maximum occupancy before the arrival of the peak of the pump pulse, and all three states show pronounced Rabi oscillations.

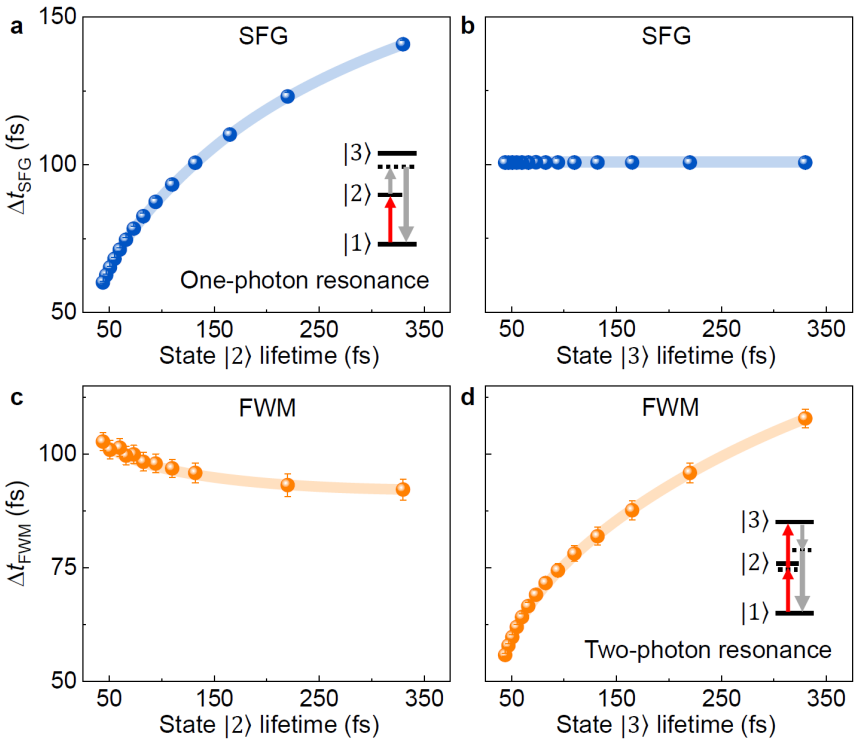

Fig. 5 | Dependence of the delay time for maximal SFG and FWM on the lifetime of the excitonic states, from the density-matrix simulations. Delay time for maximum SFG intensity as a function of the lifetime of state |2⟩ (**a**) and state |3⟩ (**b**) at a pump-photon energy of 1.739 eV. Delay time for maximum FWM intensity as a function of the lifetime of state |2⟩ (**c**) and state |3⟩ (**d**) at a pump-photon energy of 1.72 eV. The probe-photon energy is fixed at 1.476 eV. The lines are guides to the eye. The delay time Δt_{SFG} is determined mainly by the lifetime of the one-photon resonant state |2⟩, while the delay time Δt_{FWM} depends more strongly on the lifetime of the two-photon resonant state |3⟩.

Data availability

The raw data that support the plots within this paper and the other findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgments

Financial support is gratefully acknowledged from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) SFB 1277 (Project-ID 314695032) projects B03 and B11, SPP 2244 (Project-ID LI 3725/1-1, 443378379). Growth of hexagonal boron nitride crystals was supported by the Elemental Strategy Initiative conducted by the MEXT, Japan (Grant Number JPMXP0112101001) and JSPS KAKENHI (Grant Numbers 19H05790 and JP20H00354).

Author contributions

K.-Q.L. conceived and supervised the project. J.M.B., L.C., P.W., K.-Q.L, and S.B. carried out the experiments and simulations. S.B. wrote the simulation code. K.W. and T.T. provided the hBN crystals. All authors analyzed the data, discussed the results, and contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Author information

†These authors contributed equally.

*Correspondence and requests for materials should be addressed to K.-Q.L. (kaiqiang.lin@ur.de).

References

1 Shen, Y. R. *The principles of nonlinear optics*. (1984).

2 Boyd, R. W. *Nonlinear optics*. 3rd edn, (Academic, 2008).

3 Keller, U. Recent developments in compact ultrafast lasers. *Nature* **424**, 831-838 (2003).

4 Garmire, E. Nonlinear optics in daily life. *Opt Express* **21**, 30532-30544 (2013).

5 Mukamel, S. *Principles of nonlinear optical spectroscopy*. (Oxford University Press, 1999).

6 Cundiff, S. T. & Mukamel, S. Optical multidimensional coherent spectroscopy. *Physics Today* **66**, 44-49 (2013).

7 Autere, A. *et al.* Nonlinear optics with 2D layered materials. *Adv. Mater.* **30**, 1705963 (2018).

8 Schneider, C. *et al.* Two-dimensional semiconductors in the regime of strong light-matter coupling. *Nat. Commun.* **9**, 2695 (2018).

9 Langer, F. *et al.* Lightwave valleytronics in a monolayer of tungsten diselenide. *Nature* **557**, 76-80 (2018).

10 Liu, H. *et al.* High-harmonic generation from an atomically thin semiconductor. *Nat. Phys.* **13**, 262-265 (2017).

11 You, J. W., Bongu, S. R., Bao, Q. & Panoiu, N. C. Nonlinear optical properties and applications of 2D materials: theoretical and experimental aspects. *Nanophotonics* **8**, 63-97 (2019).

12 Wang, G. *et al.* Colloquium: Excitons in atomically thin transition metal dichalcogenides. *Rev. Mod. Phys.* **90**, 021001 (2018).

13 Liu, G.-B. *et al.* Electronic structures and theoretical modelling of two-dimensional group-VIB transition metal dichalcogenides. *Chem. Soc. Rev.* **44**, 2643-2663 (2015).

14 Mak, K. F. & Shan, J. Photonics and optoelectronics of 2D semiconductor transition metal dichalcogenides. *Nat. Photonics* **10**, 216-226 (2016).

15 Lafeta, L. *et al.* Second- and third-order optical susceptibilities across excitons states in 2D monolayer transition metal dichalcogenides. *2D Mater.* **8**, 035010 (2021).

16 Hernandez-Rueda, J., Noordam, M. L., Komen, I. & Kuipers, L. Nonlinear optical response of a WS2 monolayer at room temperature upon multicolor laser excitation. *ACS Photonics* **8**, 550-556 (2021).

17 Wang, G. *et al.* Giant enhancement of the optical second-harmonic emission of WSe2 monolayers by laser excitation at exciton resonances. *Phys. Rev. Lett.* **114**, 097403 (2015).

18 Yao, K. *et al.* Continuous wave sum frequency generation and imaging of monolayer and heterobilayer two-dimensional semiconductors. *ACS Nano* **14**, 708-714 (2020).

19 Lin, K.-Q. *et al.* Narrow-band high-lying excitons with negative-mass electrons in monolayer WSe2. *Nat. Commun.* **12**, 5500 (2021).

20 Lin, K.-Q., Bange, S. & Lupton, J. M. Quantum interference in second-harmonic generation from monolayer WSe2. *Nat. Phys.* **15**, 242-246 (2019).

21 Phillips, M. C. *et al.* Electromagnetically induced transparency in semiconductors via biexciton coherence. *Phys. Rev. Lett.* **91**, 183602 (2003).

22 Kim, D.-S. *et al.* Unusually slow temporal evolution of femtosecond four-wave-mixing signals in intrinsic GaAs quantum wells: Direct evidence for the dominance of interaction effects. *Phys. Rev. Lett.* **69**, 2725-2728 (1992).

23 Saiki, T., Kuwata-Gonokami, M., Matsusue, T. & Sakaki, H. Photon echo induced by two-exciton coherence in a GaAs quantum well. *Phys. Rev. B* **49**, 7817-7820 (1994).

24 Trovatello, C. *et al.* Optical parametric amplification by monolayer transition metal dichalcogenides. *Nat. Photonics* **15**, 6-10 (2021).

25 Dai, Y. *et al.* Electrical control of interband resonant nonlinear optics in monolayer MoS2. *ACS Nano* **14**, 8442-8448 (2020).

26 Poellmann, C. *et al.* Resonant internal quantum transitions and femtosecond radiative decay of excitons in monolayer WSe2. *Nat. Mater.* **14**, 889 (2015).

27 Moody, G. *et al.* Intrinsic homogeneous linewidth and broadening mechanisms of excitons in monolayer transition metal dichalcogenides. *Nat. Commun.* **6**, 8315 (2015).

28 Cundiff, S. T. *et al.* Rabi flopping in semiconductors. *Phys. Rev. Lett.* **73**, 1178-1181 (1994).

29 Cadiz, F. *et al.* Excitonic linewidth approaching the homogeneous limit in MoS2-based van der Waals heterostructures. *Phys. Rev. X* **7**, 021026 (2017).

30 Martin, E. W. *et al.* Encapsulation narrows and preserves the excitonic homogeneous linewidth of exfoliated monolayer MoSe2. *Phys. Rev. Appl.* **14**, 021002 (2020).

31 Fang, H. H. *et al.* Control of the exciton radiative lifetime in van der Waals heterostructures. *Phys. Rev. Lett.* **123**, 067401 (2019).

32 Chang, D. E., Vuletić, V. & Lukin, M. D. Quantum nonlinear optics — photon by photon. *Nat. Photonics* **8**, 685 (2014).

33 Wild, D. S., Shahmoon, E., Yelin, S. F. & Lukin, M. D. Quantum Nonlinear Optics in Atomically Thin Materials. *Phys. Rev. Lett.* **121**, 123606 (2018).

34 Sun, D. *et al.* Observation of rapid exciton–exciton annihilation in monolayer molybdenum disulfide. *Nano Lett.* **14**, 5625-5629 (2014).

Methods

Sample fabrication. The WSe_2 and WS_2 monolayers and thin layers of hBN are mechanically exfoliated from bulk single crystals (WSe_2 and WS_2, HQ Graphene; hBN, NIMS) using Nitto tape and transferred to a silicon substrate with 90 nm silicon dioxide layer through stamping[35]. The hBN encapsulated monolayer WSe_2 sample is fabricated through a pick-up method[36]. The sample is cooled down to 5 K in a microscope cryostat.

Nonlinear pump-probe spectroscopy. A schematic of the experimental setup is depicted in Supplementary Fig. 1. A Ti:sapphire laser beam (Chameleon Ultra II, Coherent, 80 MHz repetition rate) is split with a 20:80 plate beam splitter. The 20% reflected beam is used as the probe beam at the frequency ω_{probe}. The 80% transmitted beam pumps an optical parametric oscillator (OPO) with an intra-cavity frequency doubling (OPO-X fs, APE), which generates the pump beam at a frequency of ω_{pump}. Both beams pass through custom-made prism-based pulse compressors to compensate their chirp, and a power control unit is used to regulate the excitation power. The probe beam is directed through a delay stage and then combined with the pump beam through a 50:50 beam splitter. The reflected beams are focused onto the sample surface inside a cryostat using a microscope objective (Olympus, LUCPLFLN 40×, 0.6 numerical aperture). The transmitted beams are focused into a beta barium borate (BBO) crystal that provides the monitoring of the sum-frequency generation (SFG) signal measured by a photodiode (PD), as detailed in Supplementary Note 1. The signal reflected from the sample passes through a 680 nm short-pass filter, is dispersed by a spectrometer (Acton SP2300, Princeton Instruments), and detected by a cooled CCD camera (PIXIS 100, Princeton Instruments). A 50:50 beam splitter (BS3) is used to separate the excitation and emission beams.

Characterization of excitonic states. To characterize the band-edge A-exciton of monolayer WSe_2, a 488 nm continuous wave laser is focused onto the sample to excite the photoluminescence (PL). To characterize the dark high-lying exciton of monolayer WSe_2, an

80 fs pulsed Ti:sapphire laser (Mai Tai XF, Spectra-Physics, 80 MHz repetition rate) is focused onto the sample to drive the second-harmonic generation (SHG). The reflected PL after a 488 nm long-pass filter and the SHG signal after a 680 nm short-pass filter are measured by the same detection unit described for the nonlinear pump-probe spectroscopy.

References

35 Castellanos-Gomez, A. *et al.* Deterministic transfer of two-dimensional materials by all-dry viscoelastic stamping. *2D Mater.* **1**, 011002 (2014).

36 Kim, K., *et al.* van der Waals heterostructures with high accuracy rotational alignment. *Nano Lett.* **16**, 1989-1995 (2016).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OpticalwavemixingSI.pdf