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

Monolayer semiconductors are an emerging platform for strong nonlinear light–matter interactions that are 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 such nonlinearities, since harmonic generation and optical wave mixing are generally considered instantaneous processes. We find that a significant time difference, ranging from −40 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 and high-lying excitons that comprise electrons from conduction bands far above the bandgap. Numerical simulations in the density-matrix formalism rationalize 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 the dynamics of excitonic states accessible with either one-photon or two-photon transitions. 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 such as second-harmonic generation (SHG), SFG and FWM are important in the design of lasers, the characterization of ultrashort laser pulses, in photon detection and in optical sensing. Since light does not interact appreciably with itself, the source of an optical nonlinearity is the electronic polarization of a material that is driven by several incident light fields. Conventional descriptions of nonlinear polarization 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. Such descriptions are only strictly applicable under continuous-wave conditions. Wave mixing is thus generally described as a quasi-instantaneous parametric process, for which the quantum state of matter does not change. Ultimately, this assumption can only be true for the initial and final states of the interaction since, in the
quantum picture of light–matter interactions, every interaction process inevitably changes the quantum state for both photons and electrons. In the density-matrix representation of the electronic quantum states, the expectation value for the observable polarization $\langle P(t) \rangle = \langle \hat{\mu} \rangle = \hat{\rho} (t)$ (where $\hat{\mu}$ is the dipole-moment matrix and $\hat{\rho}$ is the density matrix) therefore generally contains an integration over all possible quantum pathways, involving all transitions with non-zero dipole-moment matrix elements $\mu_{ij}$ for all possible times and frequencies at which electric fields are present. This approach has proved powerful for multidimensional coherent spectroscopy, where a nonlinear response is excited by multiple laser pulses. An important consequence of this extended picture of wave mixing is that parametric interactions progress via diagonal elements $\rho_{ii}$ of the density matrix, which means that they can be sensitive not only to dephasing of transition dipoles but also to excited-state lifetimes as well as to state depletion and transition saturation.

Transition metal dichalcogenide (TMDC) monolayers have emerged as an intriguing two-dimensional medium to accommodate strong light–matter interactions. As direct-gap semiconductors with weak dielectric screening, TMDC monolayers host strongly bound excitons near the K-points of the Brillouin zone. Owing to broken inversion symmetry, the materials exhibit nonlinear responses of both even and odd order, suggesting a wide range of application potential in nonlinear wave mixing near resonance exhibited by these materials has enabled SHG and SFG under continuous-wave irradiation. A pronounced splitting of the SHG spectrum due to Rabi flopping of the excited-state population occurs even under comparatively weak pulsed-laser excitation, and is associated with light-driven quantum interference between the excitation pathways of the bound band-edge A exciton (AX) and a two-photon resonant high-lying exciton (HX) involving upper conduction bands at the same ±K-points. As shown via experiment and GW plus Bethe–Salpeter equation theory, the energy of the HX species happens to be close to twice that of the AX, forming an almost evenly spaced three-level system. Such a ladder-type excitonic structure offers a new all-solid-state platform for coherent nonlinear optical processes such as electromagnetically induced transparency. This three-level system is intrinsic to the unusual band structure of the monolayer and differs fundamentally from conventional biexcitonic systems, where a high density of excitons is needed for two excitons to meet in space to form a bound state that contributes to the nonlinear interaction. Besides SHG and SFG, single-pass optical parametric amplification has also been demonstrated in TMDC monolayers. Although excitonic enhancement of the nonlinear light–matter interaction appears to be well established, it is not at all clear whether and how excitonic transitions affect the temporal dynamics of nonlinear interactions.

Figure 1a illustrates the setup for a two-colour excitation experiment using ultrafast laser pulses of 130–140 fs length and 80 MHz repetition rate at 713 nm (with an energy $\hbar \omega_{\text{pump}} = 1.739$ 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, UPL, SHG and SFG appear for a time delay $\Delta t = 0$ between the position of the peak intensities of the pump and probe laser impulses as defined in a. The pump beam is in resonance with the AX. c. Two-colour optical wave-mixing spectrum as a function of $\Delta t$. Maximal SF and FWM intensities appear at non-zero delay times. d. Spectrally integrated intensity of SF (blue) and FWM (orange) as a function of $\Delta t$. The SF intensity of a 100-µm-thick BBO crystal (grey) in place of the WSe$_2$ sample is shown as a function of $\Delta t$ to provide a zero-time-delay reference.

**Figure 1** Delayed SFG and FWM from monolayer WSe$_2$. a. Schematic illustration of the nonlinear optical wave-mixing 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, UPL, SHG and SFG appear for a time delay $\Delta t = 0$ between the position of the peak intensities of the pump and probe laser impulses as defined in a. The pump beam is in resonance...
Comparison of the PL spectrum of monolayer WSe₂ (black curve) with the delay PL spectrum of monolayer WSe₂ (black curve) compared with the delay time for and the pump beam is set to 3 µW (1.32 µJ cm⁻²). The blue line is a guide for (blue spheres) and monolayer WS₂ (purple spheres) using the eye. The dashed line marks the two-photon resonance with an HX, which is congruent with the position of the dip in the SHG spectrum (see Supplementary Fig. 4). Dependence of Δₚₗ₉m on the pump power for monolayer WSe₂ (orange spheres) and monolayer WS₂ (purple spheres) using ħωₚₗ₉m = 1.739 eV in resonance with the AX, ħωₚₗ₉m = 1.476 eV and a probe-beam power of 400 µW. The red and dark-red wave packets indicate the temporal relation of the pump and probe pulses. The grey dashed line marks the zero time delay. Throughout, the error bars are the standard deviation of the results of three independent measurements on the same sample position.

Fig. 2 | Excitonic effects on the delay time for maximal SFG and FWM. a. The measurement conditions are the same as in a. The orange line is a guide for the eye. The dashed line marks the two-photon resonance with an HX, which is congruent with the position of the dip in the SHG spectrum (see Supplementary Fig. 4). d. Dependence of Δₚₗ₉m on the pump power for monolayer WSe₂ (orange spheres) and monolayer WS₂ (purple spheres) using ħωₚₗ₉m = 1.739 eV, ħωₚₗ₉m = 1.476 eV and a probe-beam power of 400 µW. The red and dark-red wave packets indicate the temporal relation of the pump and probe pulses. The grey dashed line marks the zero time delay. Throughout, the error bars are the standard deviation of the results of three independent measurements on the same sample position.

‘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 using a microscope cryostat and excited through a long-working-distance microscope objective with a numerical aperture (NA) of 0.6. Details of the setup are shown in Supplementary Fig. 1. Figure 1b plots a representative spectrum of the upconverted radiation detected in back-reflected geometry through the same objective, for a time delay of Δₜ = 0 and using a 680 nm (1.82 eV) short-pass filter. Five characteristic spectral contributions are identified by their frequencies: SHG of the pump beam at 2ωₚₗ₉m, SFG of the pump and probe beams at ωₚₗ₉m + ωₚₗ₉m, SHG of the probe beam at 2ωₚ₉m, FWM at 2ωₚₗ₉m − ωₚₗ₉m, and upconverted photoluminescence (UPL) of the so-called ‘B exciton’ close to the FWM peak. As monolayer TMDCs have giant second- and third-order optical susceptibilities, both second-order nonlinear processes, such as SFG, and third-order nonlinear processes, such as FWM, are very efficient. Monolayer TMDCs therefore constitute broadband optical frequency mixers. Excitation and detection through the high-NA objective bypasses most of the constraints on beam direction due to the conservation of momentum that is usually necessary for efficient nonlinear wave mixing.

Figure 1c shows a two-dimensional map of the upconverted radiation as a function of the pump–probe delay time Δₜ. As expected, neither the SHG of either the pump or probe beam nor the UPL show any dependence on the coincidence of pump and probe impulses. On the other hand, both SFG and FWM are sensitive to Δₜ. However, the maximum efficiency of these two processes does not occur at zero delay, which is surprising given that parametric, instantaneous nonlinear optical wave mixing should be most effective for an optimal temporal pulse overlap. 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 WSe₂ sample as detailed in Supplementary Fig. 1. Figure 1d shows the SFG intensity from the BBO crystal (grey spheres) as well as the spectrally integrated SFG (blue spheres) and FWM (orange spheres) intensities of the WSe₂ sample as a function of the delay time, along with Gaussian fits (lines). The dashed lines mark the centre of each Gaussian, with the FWM peaking at the delay time Δₕ₉m = 80 fs (orange) and the SFG peaking at Δₚₗ₉m = 120 fs (blue).

To examine the role of excitons in the retardation of these nonlinear processes, we perform the measurements as a function of the photon energy. Figure 2a shows the dependence of the time delay for maximal SFG (blue spheres) on the energy of the pump laser for a fixed probe energy ħωₚ₉m = 1.476 eV. Comparison with the photoluminescence (PL) spectrum (black curve) reveals that the largest Δₚₗ₉m of 120 fs is reached precisely when the pump laser is in resonance with the AX. Under off-resonance detuning, Δₚₗ₉m decreases to zero. Since both
the pump and probe pulses are short and therefore spectrally broad, the peak in the energy dependence of $\Delta_{\text{SG}}$, appears to be broader than the PL peak of the AX.

If the delay is caused by an excitonic resonance, it should depend on the excitation density. Figure 2b shows $\Delta_{\text{SG}}$. For monolayer WSe$_2$, (blue spheres) as a function of the pump power for $\hbar\omega_{\text{pump}} = 1.739$ eV, $\hbar\omega_{\text{probe}} = 1.476$ eV and a probe-beam power of 400 $\mu$W (0.18 mJ cm$^{-2}$). On increasing the pump power by three orders of magnitude (below the Mott transition, see Supplementary Note 2), $\Delta_{\text{SG}}$ decreases from $\sim$120 fs to close to zero. Experimentally, this can lead to the counterintuitive phenomenon of an increasing SFG intensity with a decreasing pump power, as shown in Supplementary Fig. 2. It may be tempting to assign such a decrease in $\Delta_{\text{SG}}$ to the decrease of the AX lifetime with increasing excitation density, as observed in transient intra- and interband spectroscopy and optical two-dimensional Fourier-transform spectroscopy. However, as we demonstrate below, this decrease can also be rationalized in terms of the dynamics of Rabi flopping. In contrast to the effect of varying the pump power, Supplementary Fig. 3a shows that comparable changes in the probe power do not affect $\Delta_{\text{SG}}$. To eliminate the possibility of optical artefacts and establish a further reference for the measurements, the experiments were repeated for monolayer WS$_2$ in place of the WSe$_2$ sample. Since the bandgap of monolayer WS$_2$, is approximately 0.3 eV larger than that of monolayer WSe$_2$, the lasers will not be in resonance with excitonic transitions. Indeed, $\Delta_{\text{SG}}$ from monolayer WS$_2$ (purple spheres in Fig. 2b) is close to zero over the entire range of pump power.

The amplitudes of the FWM signals detected simultaneously are shown in Fig. 2c, d (orange spheres). In contrast to $\Delta_{\text{SG}}$, the delay time for maximum FWM, $\Delta_{\text{FWM}}$, does not peak at the pump photon energy of the AX resonance (Fig. 2c), but rather at an energy that also gives rise to quantum interference in the SHG as shown in Supplementary Fig. 4 (refs. 12,13). This quantum interference arises due to the high-lying excitonic state HX, which lies at 3.44 eV and contributes to the delay $\Delta_{\text{FWM}}$ through a two-photon resonance with the pump laser, as illustrated in the inset of Fig. 2c. As found for $\Delta_{\text{FSG}}, \Delta_{\text{FWM}}$ decreases with increasing pump power (orange circles in Fig. 2d) and is independent of the probe power (Supplementary Fig. 3b) and the probe photon energy (Supplementary Fig. 5c). Unexpectedly, however, $\Delta_{\text{FWM}}$ drops below zero in the high-pump-power regime: above 1 mW (0.44 mJ cm$^{-2}$), the centre of the probe pulse must reach the sample 40 fs before the centre of the pump pulse for maximum efficiency of FWM. As reported in Supplementary Fig. 5a, the maximum values of $|\Delta_{\text{FSG}}|$ and $|\Delta_{\text{FWM}}|$ become even larger for WSe$_2$ monolayer samples with hexagonal boron nitride (hBN) encapsulation. This change may conceivably arise from a suppression of local sample inhomogeneities and a modification of the radiative decay rate due to a cavity effect. The fact that the maximal values of $\Delta_{\text{FSG}}$ and $\Delta_{\text{FWM}}$ appear at the lowest powers of both the pump and probe beams (Supplementary Fig. 5b) demonstrates that strong nonlinear interactions extend into the regime of the low excitation densities, opening up avenues for further exploration of quantum nonlinear optics.

While the change in the interpulse delay for SFG can, in principle, be rationalized through a pump-power dependence of a sufficiently long-lived excitonic resonance, the negative $\Delta_{\text{FWM}}$ is not straightforward to explain. To account for the excitonic resonances, instead of following the conventional analysis of the parametric light–matter interaction up to fixed nonlinear order, we use a full time-dependent solution of the density-matrix dynamics in the presence of the laser field (see Methods for details). 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 usual SHG spectra from monolayer WSe$_2$ (ref. 29). For realistic excitation conditions, the laser fields are modelled as Gaussian pulses with parameters matching the experimental ones as illustrated in Fig. 3a. We describe the allowed eigenstates using a simple ladder-type three-level system shown in Fig. 3a, which consists of a ground state $|1\rangle$, the AX state $|2\rangle$ and the HX state $|3\rangle$. The detailed parameters used for the simulation, listed in the Supplementary Fig. 6 and Supplementary Table 1, are based on experimental conditions.

Figure 3b shows a calculated optical wave-mixing spectrum for the same spectral range as probed experimentally (Fig. 1b), at zero time delay between the pump and probe pulses ($\hbar\omega_{\text{pump}} = 1.739$ eV and $\hbar\omega_{\text{probe}} = 1.476$ eV). Five prominent peaks are identified, including FWM $(2\omega_{\text{pump}} - \omega_{\text{probe}})$, SHG $(2\omega_{\text{pump}}$ and $2\omega_{\text{probe}})$, and SFG ($\omega_{\text{pump}} + \omega_{\text{probe}}$). A six-wave-mixing feature $(3\omega_{\text{pump}} - 2\omega_{\text{probe}})$ appears at a higher pump fluence (see Supplementary Fig. 8). Figure 3c shows the result of calculating the spectrally integrated SFG and FWM intensities as a function of the delay time between the pump and probe pulses. Here, the pump laser is set to be resonant with state $|2\rangle$ (with a low power of 1 mW (0.44 mJ cm$^{-2}$). The simulated maximal SFG intensity appears for a positive delay time of approximately 105 fs, which is close to the experimental value of 120 fs in Fig. 1d. The maximal FWM intensity appears at a delay time of 50 fs, which is somewhat smaller than the experimental result although it confirms the experimental trend that $\Delta_{\text{FWM}}$ tends to be smaller than $\Delta_{\text{SG}}$. 
Next, we investigate the influence of the pump power and photon energy on $\Delta t_{\text{SFG}}$ and $\Delta t_{\text{FWM}}$. Figure 4a,c shows the dependence of the pump photon energy on $\Delta t_{\text{SFG}}$ and $\Delta t_{\text{FWM}}$ at a low pump power of 1 µW. The maximum SFG arises for a pump beam resonant with the $|1\rangle \rightarrow |2\rangle$ transition (the AX), whereas the maximum FWM is found when the pump laser is two-photon resonant with the $|1\rangle \rightarrow |3\rangle$ transition, that is, the HX. These resonant enhancements coincide with the experimental observations (Fig. 2a,c). Figure 4b,d plots the simulated pump-power dependence of $\Delta t_{\text{SFG}}$ and $\Delta t_{\text{FWM}}$. Both $\Delta t_{\text{SFG}}$ and $\Delta t_{\text{FWM}}$ drop with increasing power, qualitatively matching with experiments. The simulation, however, shows that both $\Delta t_{\text{SFG}}$ and $\Delta t_{\text{FWM}}$ become negative at high pump powers, although only $\Delta t_{\text{FWM}}$ 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, 0.44 µJ cm$^{-2}$) and high (1 mW, 0.44 mJ cm$^{-2}$) pump powers. We

Fig. 4 | Simulated delay times for maximal SFG and FWM intensities. a, Delay time $\Delta t_{\text{SFG}}$ for the maximal simulated SFG intensity as a function of the pump photon energy. The pump at 1.739 eV is set to a power of 1 µW (0.44 µJ cm$^{-2}$) while the probe at 1.476 eV is set to 200 µW (88 µJ cm$^{-2}$). b, Corresponding dependence of the intensity of SFG on the 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 the maximal simulated FWM intensity, $\Delta t_{\text{FWM}}$, as a function of the pump photon energy. The pump power and the parameters for the probe beam are as in a. The error bars mark the uncertainty in peak position from fitting the numerical results. The dashed line marks the energy for two-photon excitation of the HX (state $|3\rangle$). d, Corresponding dependence of the simulated FWM intensity on the 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 $|2\rangle$ reaches maximum occupancy at a delay time of 105 fs with respect to the centre of the pump pulse. f, Density-matrix dynamics for a pump power of 1 mW. State $|3\rangle$ reaches maximum occupancy before the arrival of the peak of the pump pulse, and all three states show pronounced Rabi oscillations.
propose that the largest SFG and FWM intensities are obtained when the centre of the probe pulse, that is, where the field strength is highest, coincides with the largest transient population of the excitonic state. Since both the AX state $|2\rangle$ and the HX state $|3\rangle$ have short but finite lifetimes, the state populations accumulate over the short timescale of the pump–probe pulse sequence. In the low-pump-power regime, maximum state occupancy then appears at positive delay times as seen in Fig. 4e. The delay times for maximal population of states $|2\rangle$ and $|3\rangle$ match well with $\Delta t_{SFG}$ and $\Delta t_{FWM}$, 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. A detailed time evolution of the state populations as a function of the pump power is shown in Supplementary Fig. 9. The signatures of the onset of this Rabi flopping have been studied in the context of SHG, when the two-photon resonance of the $|1\rangle \rightarrow |3\rangle$ AX transition is driven\textsuperscript{34}. Rabi flopping is directly observed experimentally in the frequency domain, by examining the fingerprints of quantum interference in the SHG spectra as a function of the pulse length and amplitude—the number of transparency dips in the spectrum relates directly to the number of Rabi cycles undergone by the driven system in a single laser pulse\textsuperscript{35}. Since the early depopulation of the excited states associated with Rabi flopping shifts maximum occupancy to earlier times (Fig. 4f), $\Delta t_{SFG}$ and $\Delta t_{FWM}$ decrease with increasing pump power (Fig. 4b,d).

Unlike in the experiments, however, the simulated $\Delta t_{SFG}$ does turn negative. As we discussed in ref. 35, 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\textsuperscript{36,38} in monolayer WSe\textsubscript{2} substantially raises the decay rate of state $|2\rangle$ at high exciton densities, and therefore prevents $\Delta t_{SFG}$ from becoming negative. As shown in Fig. 5a,b, the simulated delay time $\Delta t_{SFG}$ for maximum SFG intensity is very sensitive to the lifetime of state $|2\rangle$ and even approaches zero as this lifetime tends to zero. In stark contrast to this observation, Fig. 5c,d shows that $\Delta t_{FWM}$ is mainly sensitive to the lifetime of state $|3\rangle$ but is quite insensitive to that of state $|2\rangle$.

The different dependencies of $\Delta t_{SFG}$ and $\Delta t_{FWM}$ on the lifetimes of the two states imply that SFG and FWM can be used to probe the dynamics of one-photon allowed and two-photon allowed excitonic states simultaneously but independently of each other. The largest experimentally observed delay time of $\Delta t_{SFG} = 120$ fs for bare WSe\textsubscript{2} corresponds to an AX lifetime of approximately 200 fs (Fig. 5a), which matches well with the population relaxation and radiative lifetime measured by transient intra- and interband spectroscopy\textsuperscript{28} as well as optical two-dimensional Fourier-transform spectroscopy\textsuperscript{29}. Although excitonic dynamics of one-photon allowed transitions are straightforward to probe using a variety of conventional optical means, the dynamics of two-photon allowed dark excitonic states have remained much more challenging to access, and our approach provides a simple solution to this problem.

In summary, we explored the influence of the delay between two laser pulses on second- and third-order optical wave-mixing processes in monolayer WSe\textsubscript{2}. Substantial delays between the two pulses are found to be necessary to maximize the SFG and FWM signal intensities. Understanding these time delays poses challenges to the description...
of 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 population dynamics of the excitons under optical pumping as described within the density-matrix formalism, which computes the interaction of light and matter to arbitrary perturbative order. Further distinction between conventional parametric SF and FWM and processes involving excitons will necessitate disentangling of the multitude of Liouville-space pathways embodied in the density-matrix time dynamics. The qualitative agreement between simulation and experiment 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 therefore demonstrates that nonlinear optical wave-mixing spectroscopy offers new opportunities for measuring the population dynamics of optically pumped excitonic states that are either one- or two-photon allowed, where the dynamics of the latter have proved 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.

**Online content**

Any additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at [https://doi.org/10.1038/s41566-022-01080-1](https://doi.org/10.1038/s41566-022-01080-1).

**References**

1. Shen, Y. R. *The Principles of Nonlinear Optics* (Wiley, 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. Phys. 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 WS\(_2\) 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 WSe\(_2\) 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 WSe\(_2\). Nat. Commun. **12**, 5500 (2021).
20. Lin, K.-Q., Bange, S. & Lupton, J. M. Quantum interference in second-harmonic generation from monolayer WSe\(_2\). Nat. Phys. **15**, 242–246 (2019).
21. Fleischhauer, M., Imamoglu, A. & Marangos, J. P. Electromagnetically induced transparency: optics in coherent media. Rev. Mod. Phys. **77**, 633–673 (2005).
22. Phillips, M. C. et al. Electromagnetically induced transparency in semiconductors via biexciton coherency. Phys. Rev. Lett. **91**, 183602 (2003).
23. 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).
24. Saiki, T., Kuwata-Gonokami, M., Matsuuse, T. & Sakaki, H. Photon echo induced by two-exciton coherence in a GaAs quantum well. Phys. Rev. B **49**, 7817–7820 (1994).
25. Trovatiello, C. et al. Optical parametric amplification by monolayer transition metal dichalcogenides. Nat. Photonics **15**, 6–10 (2021).
26. Dai, Y. et al. Electrical control of interband resonant nonlinear optics in monolayer Mo\(_x\)Se\(_2\). ACS Nano **14**, 8442–8448 (2020).
27. Liu, S. et al. An all-dielectric metasurface as a broadband optical frequency mixer. Nat. Commun. **9**, 2507 (2018).
28. Poellmann, C. et al. Resonant internal quantum transitions and femtosecond radiative decay of excitons in monolayer WSe\(_2\). Nat. Mater. **14**, 889–893 (2015).
29. Moody, G. et al. Intrinsic homogeneous linewidth and broadening mechanisms of excitons in monolayer transition metal dichalcogenides. Nat. Commun. **6**, 8315 (2015).
30. Cundiff, S. T. et al. Rabi flopping in semiconductors. Phys. Rev. Lett. **73**, 1178–1181 (1994).
31. Cadiz, F. et al. Excitonic linewidth approaching the homogeneous limit in Mo\(_x\)S\(_2\)-based van der Waals heterostructures. Phys. Rev. X **7**, 021026 (2017).
32. Martin, E. W. et al. Encapsulation narrows and preserves the excitonic homogeneous linewidth of exfoliated monolayer MoSe\(_2\). Phys. Rev. Appl. **14**, 021002 (2020).
33. Fang, H. H. et al. Control of the exciton radiative lifetime in van der Waals heterostructures. Phys. Rev. Lett. **123**, 067401 (2019).
34. Chang, D. E., Vuletić, V. & Lukin, M. D. Quantum nonlinear optics — photon by photon. Nat. Photonics **8**, 685–694 (2014).
35. Wild, D. S., Shahmoon, E., Yelin, S. F. & Lukin, M. D. Quantum nonlinear optics in atomically thin materials. Phys. Rev. Lett. **121**, 123606 (2018).
36. Sun, D. et al. Observation of rapid exciton–exciton annihilation in monolayer molybdenum disulfide. Nano Lett. **14**, 5625–5629 (2014).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022
Methods
Sample fabrication
The WSe2 and WS2 monolayers and thin layers of hBN were mechanically exfoliated from bulk single crystals (WSe2 and WS2, HQ Graphene; hBN, National Institute for Materials Science) using Nitto tape and transferred to a silicon substrate with a 90 nm silicon dioxide layer through stamping. The hBN encapsulated monolayer WSe2 sample was fabricated using a pick-up method. The sample was cooled to 5 K using a microscope cryostat.

Nonlinear spectroscopy
A schematic of the experimental setup is depicted in Supplementary Fig. 1. A Ti:sapphire laser beam (Chameleon Ultra II, 80 MHz repetition rate, Coherent) was split using a 20:80 plate beamsplitter. The 20% reflected beam was used as the probe beam at the frequency ωpump. The 80% transmitted beam was used to pump an optical parametric oscillator with intra-cavity frequency doubling (OPO-IXs, APE), which generates the pump beam at the frequency ωpump. Both beams pass through custom-made prism-based pulse compressors to compensate their chirp, and a power-control unit was used to regulate the excitation power. The probe beam was directed through a delay stage and then combined with the pump beam through a 50:50 beamsplitter. The reflected beams were focused onto the sample surface inside a cryostat using a microscope objective (LUCPLFLN40X, 0.6NA, Olympus). The diameter of the laser spot was estimated to be 1.9 μm. The transmitted beams were focused into a BBO crystal that provides monitoring of the SFG signal measured using a photodiode, as detailed in Supplementary Note 1. The signal reflected from the sample passed through a 680 nm short-pass filter, was dispersed by a spectrometer (Acton SP2300, Princeton Instruments) and detected using a cooled CCD camera (PIXIS 100, Princeton Instruments). A 50:50 beamsplitter (BS3) was used to separate the excitation and emission beams.

Characterization of excitonic states
To characterize the AX of monolayer WSe2, a 488 nm continuous-wave laser was focused onto the sample to excite the PL. To characterize the dark HX of monolayer WSe2, an 80 fs pulsed Ti:sapphire laser (Mai Tai XF, 80 MHz repetition rate, Spectra-Physics) was focused onto the sample to drive the SHG. The PL signal after a 488 nm long-pass filter and the SHG signal after a 680 nm short-pass filter were measured using the same detection unit described for the nonlinear spectroscopy.

Density-matrix formalism simulation
The polarization in P(t) ∝ ∑μiμjρi j(t) is understood to be the source term for scattered radiation. Specific information about the SFG and FWM spectra was obtained by analysing P(t) in the Fourier domain. We solved the time dynamics of the density matrix ρ in the framework of a Lindblad-type equation \[ \frac{d}{dt} \rho = -\frac{i}{\hbar} [H, \rho] - \frac{1}{2} \sum_i \{ F^\dagger i, \rho \} + \Lambda. \] Here, the density matrix is defined in the conventional way as \[ \rho = \sum_i p_i |i\rangle \langle i| \] with p_i the probability for the system to be in state |i\rangle. The Hamiltonian \[ H = -\frac{\varepsilon(t)}{\mu} \mu_L (i \neq j) \] where E_i are the eigenstate energies, ε(t) is the electric-field component of the incident light field and \[ \mu_L \] are the transition dipole moments linked to the ground state. The decay matrix \[ \Gamma = \sum_i \Gamma_i \] describes the repopulation of the ground state with a single non-zero matrix element \[ \Lambda_{11} = \sum_i \Gamma_i \rho_{ii}. \] The reduced Planck constant \[ \hbar \] is the reduced Planck constant.

Data availability
Any additional data are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

Code availability
The Mathematica code used for the numerical simulations discussed in this paper is available from the corresponding authors upon reasonable request.

References
37. Castellanos-Gomez, A. et al. Deterministic transfer of two-dimensional materials by all-dry viscoelastic stamping. 2D Mater. 1, 011002 (2014).
38. Kim, K. et al. van der Waals heterostructures with high accuracy rotational alignment. Nano Lett. 16, 1989–1995 (2016).

Acknowledgements
We thank C. Schüller and R. Huber for insightful discussions. Financial support is gratefully acknowledged from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) SFB 1277 (project-ID 314695032) projects B03 (J.M.L. and S.B.) and B11 (K.-Q.L.), SPP 2244 (project-ID LI 3725/1-1, 443378379) (K.-Q.L. and S.B.) and from DFG project number 439215932 (J.M.L.). Growth of the hBN 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 analysed the data, discussed the results and contributed to the writing of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41566-022-01080-1.

Correspondence and requests for materials should be addressed to Kai-Qiang Lin.

Peer review information Nature Photonics thanks Giulio Cerullo and Christoph Lienau for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.