Supplementary Information:

Hybridized exciton–photon–phonon states in a transition-metal-dichalcogenide van-der-Waals heterostructure microcavity

Donghai Li[1,2], Hangyong Shan[3], Christoph Rupprecht[4], Heiko Knopf[5,6,7], Kenji Watanabe[8], Takashi Taniguchi[9], Ying Qin[10], Sefaattin Tongay[10], Matthias Nuß[1], Sven Schröder[6], Falk Eilenberger[5,6,7], Sven Höfling[4], Christian Schneider[3,4], and Tobias Brixner[1,11*]*

[1] Institut für Physikalische und Theoretische Chemie, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

[2] University of Science and Technology of China, 230026 Hefei, China

[3] Institute of Physics, University of Oldenburg, D-26129 Oldenburg, Germany

[4] Technische Physik and Wilhelm Conrad Röntgen Research Center for Complex Material Systems, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

[5] Institute of Applied Physics, Abbe Center of Photonics, Friedrich Schiller University, Albert-Einstein-Straße 15, 07745 Jena, Germany

[6] Fraunhofer-Institute for Applied Optics and Precision Engineering IOF, Albert-Einstein-Straße 7, 07745 Jena

[7] Max Planck School of Photonics, Albert-Einstein-Straße 7, 07745 Jena, Germany

[8] Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan

[9] International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan

[10] Materials Science and Engineering, School of Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ 85287, USA

[11] Center for Nanosystems Chemistry (CNC), Universität Würzburg, Theodor-Boveri-Weg, 97074 Würzburg, Germany

[*]E-mails: christian.schneider@uol.de, brixner@uni-wuerzburg.de

Supplementary Note 1: Vibronic polariton Hamiltonian

To simulate our experimental observation, we adopted a vibronic polariton Hamiltonian [1],

$$H = \begin{pmatrix} E_X & 0 & 0 & g_{00} & g_{01} & g_{02} \\ 0 & E_X + E_V & 0 & g_{10} & g_{11} & g_{12} \\ 0 & 0 & E_X + 2E_V & g_{20} & g_{21} & g_{22} \\ g_{00} & g_{10} & g_{20} & E_C & 0 & 0 \\ g_{01} & g_{11} & g_{21} & 0 & E_C + E_V & 0 \\ g_{02} & g_{12} & g_{22} & 0 & 0 & E_C + 2E_V \end{pmatrix}. \quad \text{(S1)}$$

Here E_X, E_C, and E_V are the energies of the pure exciton, the cavity photon, and the phonon, respectively. In the matrix of H, we assume three excitonic states, $|\mathrm{G_C; X}; v\rangle$, with phonon quantum number v = 0, 1, 2, one pure cavity photon state, $|\mathrm{E_C; G}; 0\rangle$, and two other phonon-dressed photon states, $|\mathrm{E_C; G}; 1\rangle$ and $|\mathrm{E_C; G}; 2\rangle$, which possess energies of $E_C + v\,E_V$ (v = 1, 2). The phonon-dressed photon states consist of one photon plus v phonons of a particular mode. We truncated the Hamiltonian at the phonon quantum number v = 2 (hence the 6×6 matrix), because this is the lowest number that allows us to simulate the three polariton branches in the energy-momentum-resolved PL map [Fig. 3(a) of the main text]. Here, the energy of the pure exciton E_X = 1.656 eV is known from the absorption spectrum of the DML region before capping the structure with the top DBR [2].

Off-diagonal terms g_{ij} (i, j = 0, 1, 2) are the interaction potentials between the cavity photon and the excitonic states, which can be calculated by [1]

$$\mathrm{g_{ij}} = \langle \mathrm{i}|\mathrm{j}\rangle \frac{\hbar\Omega}{2}. \quad \text{(S2)}$$

Here, Ω is the Rabi frequency and $\langle \mathrm{i}|\mathrm{j}\rangle$ represents the vibrational overlap integral (Franck–Condon factor) that depends on the Huang–Rhys factor S [3],

$$\langle \mathrm{i}|\mathrm{j}\rangle = \mu_X \exp(-S/2),\ (\mathrm{i} = 0, \mathrm{j} = 0) \quad \text{(S3)}$$

$$\langle \mathrm{i}|\mathrm{j}\rangle = \mu_X \sqrt{S/\mathrm{i}} \times \langle \mathrm{i} - 1|\mathrm{j}\rangle,\ (\mathrm{i} \neq 0, \mathrm{j} = 0) \quad \text{(S4)}$$

$$\langle \mathrm{i}|\mathrm{j}\rangle = -\mu_X\sqrt{S/\mathrm{j}} \times \langle \mathrm{i}|\mathrm{j}-1\rangle, (\mathrm{i}=0, \mathrm{j}\neq 0) \quad \text{(S5)}$$

$$\langle \mathrm{i}|\mathrm{j}\rangle = -\mu_X\sqrt{S/\mathrm{j}} \times \langle \mathrm{i}|\mathrm{j}-1\rangle + \mu_X\sqrt{\mathrm{i}/\mathrm{j}} \times \langle \mathrm{i}-1|\mathrm{j}-1\rangle. (\mathrm{i}\neq 0, \mathrm{j}\neq 0) \quad \text{(S6)}$$

Here, μ_X is an "overall" transition dipole moment of the exciton, which is set as 1 in this work.

Diagonalization of Supplementary Eq. (S1) allows us to obtain eigenstates $|\mathrm{P_k}\rangle$, with k = 0, 1, …, 5. The resulting eigenstates are linear combinations of the collective states $|\mathrm{L}\rangle = |\mathrm{G_C}; \mathrm{X}; \mathrm{v}\rangle$ or $|\mathrm{E_C}; \mathrm{G}; \mathrm{v}\rangle$ with v = 0, 1, 2, 3,

$$|\mathrm{P_k}\rangle = \sum_{\mathrm{L}} c_{\mathrm{kL}}|\mathrm{L}\rangle, \quad \text{(S7)}$$

with appropriate mixing coefficients c_{kL}.

In a first step, let us assume that the phonon-dressed photon states do not couple to the excitonic states, hence, g_{ij} = 0 (for all terms with $j \neq 0$). Then the 6×6-sized Hamiltonian can be reduced to 4×4 entries,

$$H = \begin{pmatrix} E_X & 0 & 0 & g_{00} \\ 0 & E_X + E_V & 0 & g_{10} \\ 0 & 0 & E_X + 2E_V & g_{20} \\ g_{00} & g_{10} & g_{20} & E_C \end{pmatrix}. \quad \text{(S8)}$$

Under this premise, we consider a varying magnitude of *S*. If *S* = 0, then $g_{00} = \Omega/2$ and $g_{10} = g_{20} = 0$, the Hamiltonian reduces to the widely employed two-coupled-oscillators model [4], leading to two polariton branches, i.e., the upper (green) and lower (black) polariton branches as depicted in Fig. 1(b) of the main text . However, the simulated results obviously cannot explain the observed multiple polariton dispersion curves in the energy-momentum-resolved PL map. Next, we set *S* to be 0.3, and show the simulated dispersion curves of the resulting eigenstates in Fig. 1(c) of the main text. As *S* increases, new polariton branches (with curved dispersion) are found. These new polariton states are formed due to a threefold coupling between the exciton, the photon, and the phonon. However, from the nature of the mathematical

model, it follows that the first excited state in Fig. 1(c) of the main text, i.e., the second simulated curve as counted from the bottom, will always appear above the line of E = 1.656 eV, whereas experimentally we observe polariton dispersion curves emerging from 1.605, 1.630 and 1.643 eV [Fig. 3(a) of the main text] for the combined system of heterostructure and cavity. Thus, independent of the particular choice of parameters in the simplified Hamiltonian, the experimentally observed energy structure cannot be explained by the model without including the coupling of phonon-dressed photon states to excitonic states.

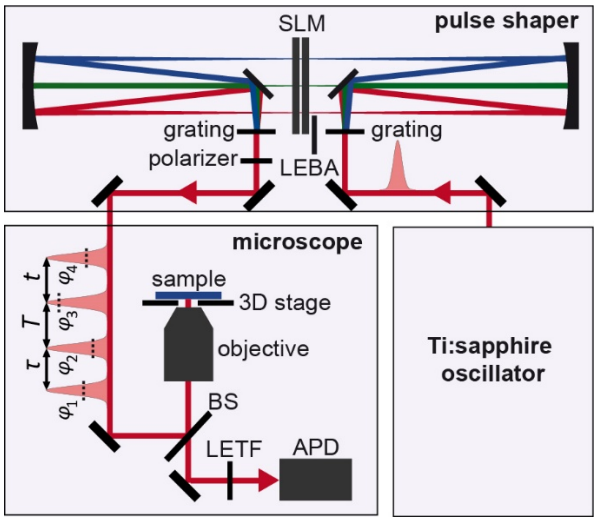
Thus, as the next step, we activate the coupling terms between the phonon-dressed photon states and the excitonic states and calculate the dispersion curves of the resulting eigenstates [Fig. 1(d) of the main text]. The inclusion of coupled phonon-dressed photon states introduces additional polariton states after diagonalizing the Hamiltonian. Most importantly, the appearance of the polariton branches emerging from 1.630 eV and 1.643 eV at $k_{\|}$ =0 in the energy-momentum-resolved PL map shown in Fig. 3(a) of the main text can be reproduced by the simulation.

Supplementary Note 2: Sample fabrication

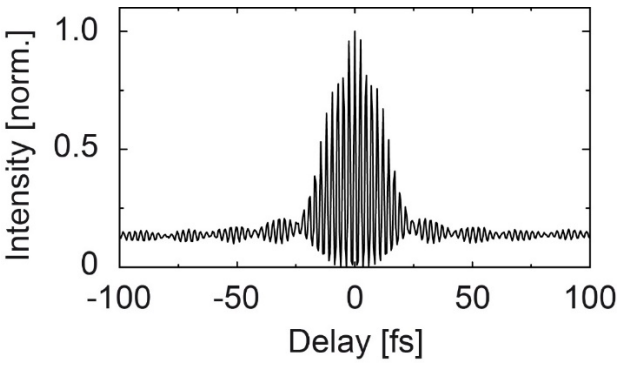
Our microcavity is composed of a bottom distributed Bragg reflector (DBR), which consists of 10 pairs of SiO_2/TiO_2 with a central Bragg wavelength of 750 nm. The embedded van-der-Waals heterostructure was prepared by the Scotch tape method combined with a dry PDMS transfer. We utilize high-quality, vapor-phase-grown WSe_2 crystals to implement a layer sequence of WSe_2/hBN/WSe_2 on top of the bottom DBR. Our top DBR is composed of 9 repetitions of SiO_2/TiO_2, which is deposited on top of the structure utilizing a gentle low-temperature technique (more details can be found in Ref. [5]). The quality factor of the sample is $Q = 4760 \pm 370$.

Supplementary Note 3: 2D micro-spectroscopy setup

Supplementary Figure 1 schematically illustrates the setup of the 2D micro-spectroscopy. Four laser pulses in one collinear beam were generated by a 4*f*-based pulse shaper and, then, focused on the double-monolayer region by a high-numerical-aperture (NA = 1.4) objective. The PL signal was collected by the same objective and detected by an avalanche photodiode (APD). A low-energy-blocking aperture (LEBA, Supplementary Figure 1) was inserted into the light path in front of the spatial light modulator (SLM, Supplementary Figure 1) in the pulse shaper, removing the low-energy part (below 1.615 eV) of the laser spectrum [Fig. 2(c) of the main text]. This ensures that the low-energy transmission filter in the detection path (LETF, Supplementary Figure 1) blocks any scattered or reflected excitation light and leaves only PL to be detected by the APD. The pulse duration was compressed down to ~20 fs at the sample position by a PRISM procedure [6,7], as indicated by the second-order interferometric autocorrelation curve (Supplementary Figure 2). Additionally, the positon of the LEBA was adjusted for a second 2D measurement with the excitation laser spectrum below 1.685 eV blocked, such that the 2D spectrum in Fig.3(c) of the main text was measured.



Supplementary Figure 1. Experimental setup. Four collinear laser pulses are generated by a 4*f*-based pulse shaper and focused by a high-numerical-aperture objective. PL is collected by the same objective and its intensity detected as a function of the time delays (τ, T, and t) between pulses and as a function of their phases (φ_1, φ_2, φ_3, φ_4).



Supplementary Figure 2. second-order interferometric autocorrelation measured by a two-photon photodiode at the sample position.

Supplementary Note 4: Data acquisition

The data acquisition of coherent 2D micro-spectroscopy has been detailed previously [8]. In this work, we obtained 2D spectra by scanning the coherence times τ and t in steps of 15 fs each from 0 to 210 fs, for a given waiting time T, using the spectral modulation function imposed on the pulse shaper,

$$M(\omega) = \exp\left[i\left(\omega - \omega_0(1-\gamma)\right)(-\tau - T)\right] + \exp\left[i\left(\omega - \omega_0(1-\gamma)\right)(-T) + i\varphi_{12}\right] + \exp[i\varphi_{13}] + \exp\left[i\left(\omega - \omega_0(1-\gamma)\right)t + i\varphi_{14}\right], \quad \text{(S10)}$$

at a center frequency of $\omega_0 = 2.513 \times 10^{15}$ s^{-1}. We employed a rotating frame by setting the parameter $\gamma = 0$. The third pulse was fixed at time 0, so that when 2D spectra were measured at a certain waiting time T, only the first and fourth pulses were delayed.

By setting the phase of the first pulse to 0, we scanned three relative phases φ_{12}, φ_{13}, and φ_{14}, where $\varphi_{1j} = \varphi_j - \varphi_1$ for individual pulse phases φ_j, in a 64-step phase-cycling (PC) scheme so that each relative phase took values of 0, $\frac{\pi}{2}$, π, and $\frac{3\pi}{2}$. One particular phase combination can thus be presented as $[\varphi_{12}, \varphi_{13}, \varphi_{14}]$. Supplementary Figure 3 shows four 2D maps with PL intensity plotting against the coherence times τ and t for phase combinations of [0, 0, 0], [$\frac{\pi}{2}$, 0, 0], [π, 0, 0], and [$\frac{3\pi}{2}$, 0, 0], as examples. The data were normalized at the pixel locating at τ = 210 fs and t = 210 fs, that is, we divided the values of all pixels by the value of

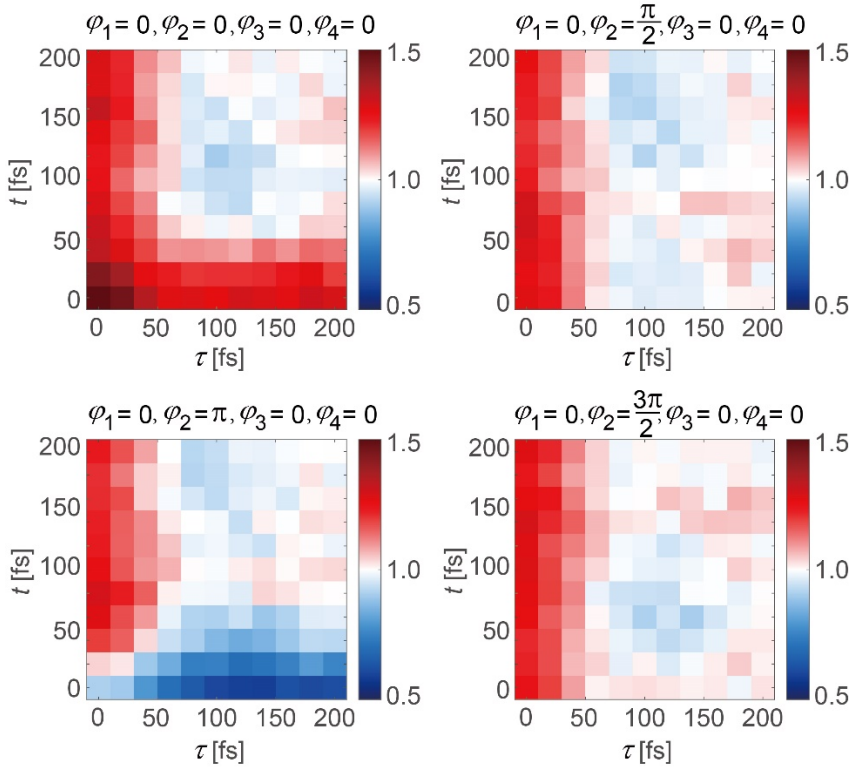
the chosen pixel. The reason is that due to the short dephasing time of the studied sample, the fourth-order nonlinear contributions at long coherence time averaged out, leaving only linear contribution at these positions. Therefore, the adopted normalization procedure can help us to eliminate the scaling effect of the linear signal. In addition, owing to the employment of rotating frame, no high-frequency oscillation occurs when plot a line cut along either the τ axis or the t axis, such that we can use relatively large scan step of 15 fs to monitor the envelope evolution. At every T, we measured 64 2D maps with different phase combinations. Then, for a certain signal, e.g., the rephasing or the nonrephasing signal, a weighted summing of the 64 2D maps has to be conducted using

$$\tilde{p}^{(n)}(\beta,\gamma,\delta) = \frac{1}{LMN}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sum_{l=0}^{L-1} p(\varphi_{12},\varphi_{13},\varphi_{14}) \times \exp[-i(\beta\varphi_{12}+\gamma\varphi_{13}+\delta\varphi_{14})] \tag{S11}$$

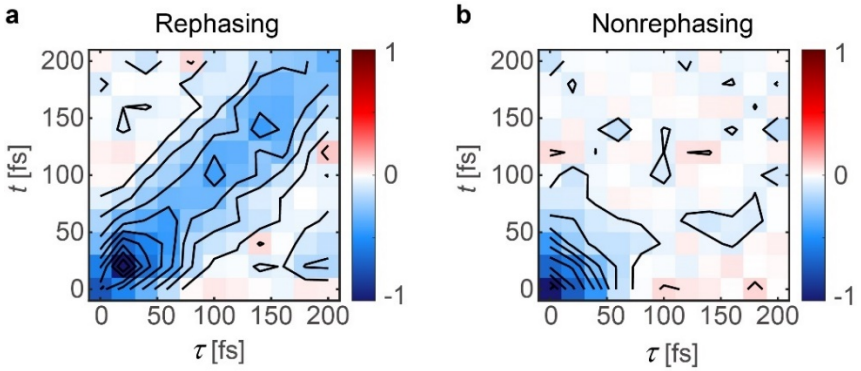
where L, M, N are the numbers of steps we scan within a 2π range for the phase of each pulse, β,γ,δ are integers used to define one specific nonlinear signal. For the rephasing signal, $\beta = 1$, $\gamma = 1$, and $\delta = -1$, whereas for the nonrephasing signal, $\beta = -1$, $\gamma = 1$, and $\delta = -1$. After the weighted summation, time-domain 2D maps can be obtained. E.g., Supplementary Figures 4a and 4b show the time-domain 2D maps of the rephasing and nonrephasing signals for T = 50 fs, respectively. Subsequent 2D Fourier transformations of these two figures against τ and t result in the 2D spectra shown in Figs. 4(b,c) of the main text. The 64-step PC ensures that there is no sixth-order nonlinear signal mixing with the fourth-order rephasing signal observed in the 2D spectra. This is crucial for our TMD microcavity, for which the amplitudes of sixth-order nonlinear signals are comparable to those of the rephasing signal and could seriously affect its 2D lineshape.

Due to the finite response time of the liquid crystals of the employed pulse shaper, we waited for 500 ms after changing the phase mask before taking data. The laser spectrum was confined by the LEBA, which acted as a short-pass filter at 768 nm wavelength (1.615 eV). The LETF

with a cutoff wavelength of 773 nm (1.604 eV) was placed before the APD. Photoluminescence was averaged over 1 ms for each acquisition period of the APD. Including additional averaging (4000 times for each pulse shape), the total measurement time for one 2D map was ~10 h. The group delay dispersion at the sample position was compensated by adding an additional phase to the modulation function.



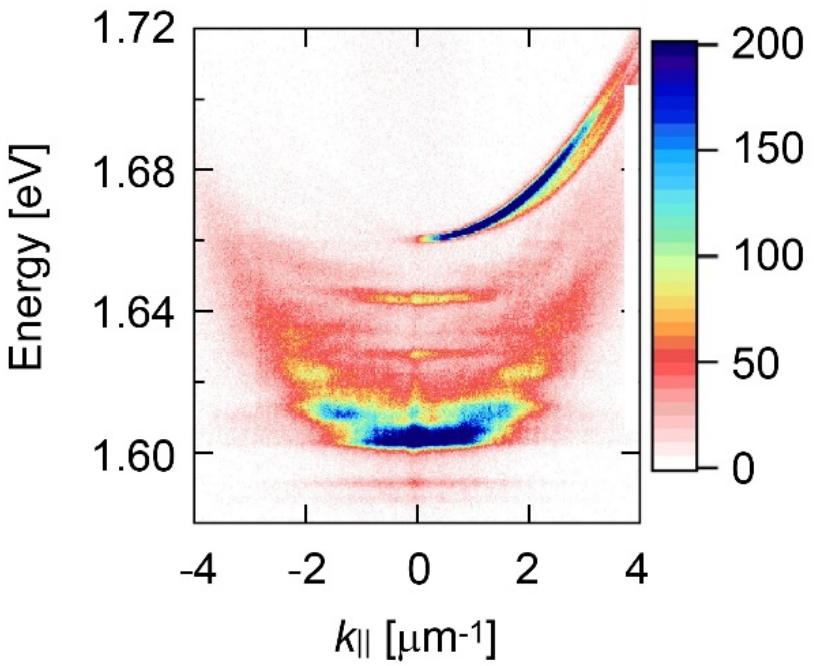
Supplementary Figure 3. Four 2D maps with normalized PL intensity plotting against the coherence times τ and t for phase combinations of [0, 0, 0, 0], [0, $\frac{\pi}{2}$, 0, 0], [0, π, 0, 0], and [0, $\frac{3\pi}{2}$, 0, 0].



Supplementary Figure. 4. Time-domain 2D maps of the rephasing (a) and nonrephasing (b) signals for T = 50 fs.

Supplementary Note 5: Full set of energy-momentum-resolved PL map

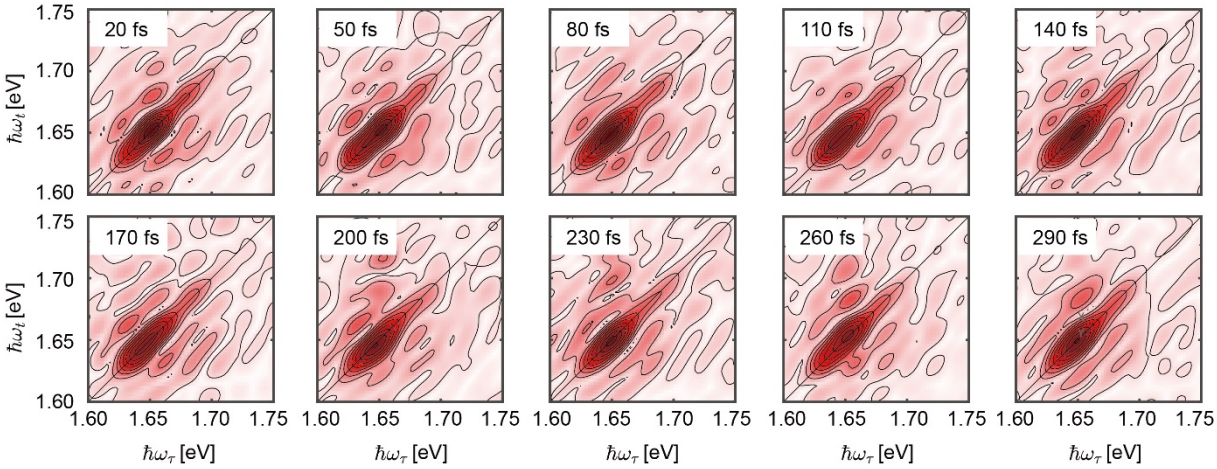
For the momentum-resolved PL measurements, a standard back-Fourier-plane imaging configuration is used, as schematically shown in Fig. 2(c) of the main text. The first lens on top of the sample is Thorlabs C105TMD-B (NA=0.6), and its back-focal plane carries the momentum-dependent information of the sample. A second lens collects the back-focal-plane information of the first lens, and a third lens projects the information into the focal plane (slit) of a spectrometer (model A-SR-500i-B2-SIL). The CCD camera (model iKon-M DU934P-BEX2-DD-9FL) attached to the spectrometer is operated in the regime for highest sensitivity (50 kHz A/D rate, pre-amplifier setting x4, and sensor cooling temperature of -80°C). Supplementary Figure 5 shows the energy-momentum-resolved PL map of a double-monolayer (DML) region.



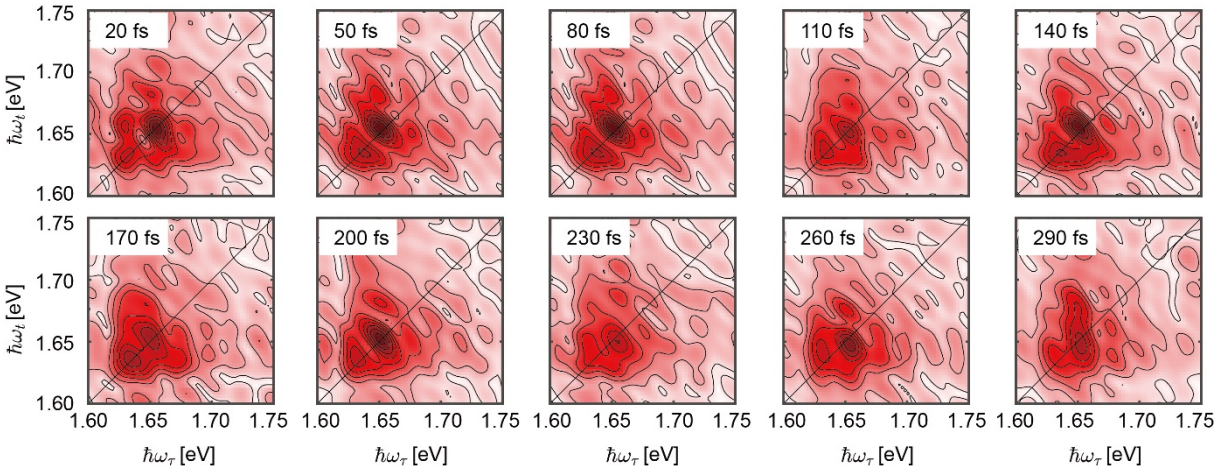
Supplementary Figure 5. Energy-momentum-resolved PL map of a double-monolayer (DML) region.

Supplementary Note 6: 2D spectra for different *T*

Supplementary Figures 6 and 7 show the rephasing and the nonrephasing 2D spectra at different *T*, respectively.



Supplementary Figure 6. Rephasing 2D spectra at various *T*.



Supplementary Figure 7. Nonrephasing 2D spectra at various *T*.

Supplementary Note 7: Determination of the diagonal peak positions

We fit the diagonal slices from rephasing and nonrephasing 2D spectra with a sum of Gaussians. The fitting results for different *T* are shown in Supplementary Figures 8 and 9, respectively. Specifically, the positions and the amplitudes of five Gaussian peaks are scanned in certain ranges to search for the minimum deviation between the sum of the five Gaussian functions and the experimental data. Retrieved energy positions of the five components are listed in

Supplementary Tables 1 and 2 for the rephasing and the nonrephasing signals, respectively. Owing to the nature of rephasing and nonrephasing 2D spectra, the nonrephasing 2D spectra show narrower peaks along the diagonal direction such that the two bands with lowest energies are spectrally separated (Supplementary Figure 8), while the two bands are merged into one peak in the rephasing spectra (Supplementary Figure 9).

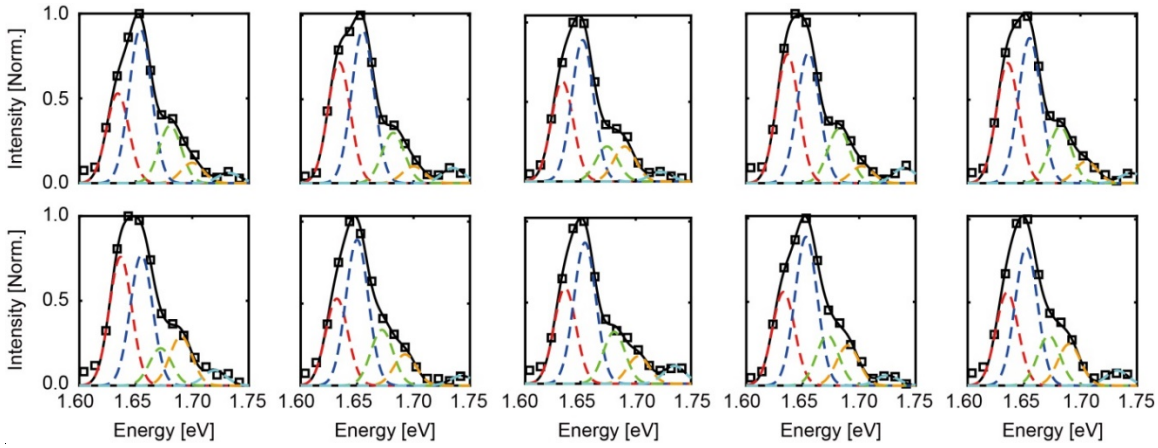
Supplementary Table 1. The energy positions of the five components of the rephasing signal retrieved from the fitting procedure.

T	20 fs	50 fs	80 fs	110 fs	140 fs	170 fs	200 fs	230 fs	260 fs	290 fs
Com. 1	1.6340 eV	1.6340 eV	1.6335 eV	1.6365 eV	1.6355 eV	1.6365 eV	1.6320 eV	1.6360 eV	1.6330 eV	1.6350 eV
Com. 2	1.6540 eV	1.6550 eV	1.6520 eV	1.6550 eV	1.6550 eV	1.6550 eV	1.6500 eV	1.6540 eV	1.6530 eV	1.6520 eV
Com. 3	1.6805 eV	1.6820 eV	1.6740 eV	1.6820 eV	1.6820 eV	1.6720 eV	1.6720 eV	1.6805 eV	1.6720 eV	1.6720 eV
Com. 4	1.7000 eV	1.7000 eV	1.6900 eV	1.7025 eV	1.7050 eV	1.6900 eV	1.6925 eV	1.7025 eV	1.6900 eV	1.6900 eV
Com. 5	1.7325 eV	1.7350 eV	1.7225 eV	1.7400 eV	1.7450 eV	1.7200 eV	1.7425 eV	1.7350 eV	1.7250 eV	1.7325 eV

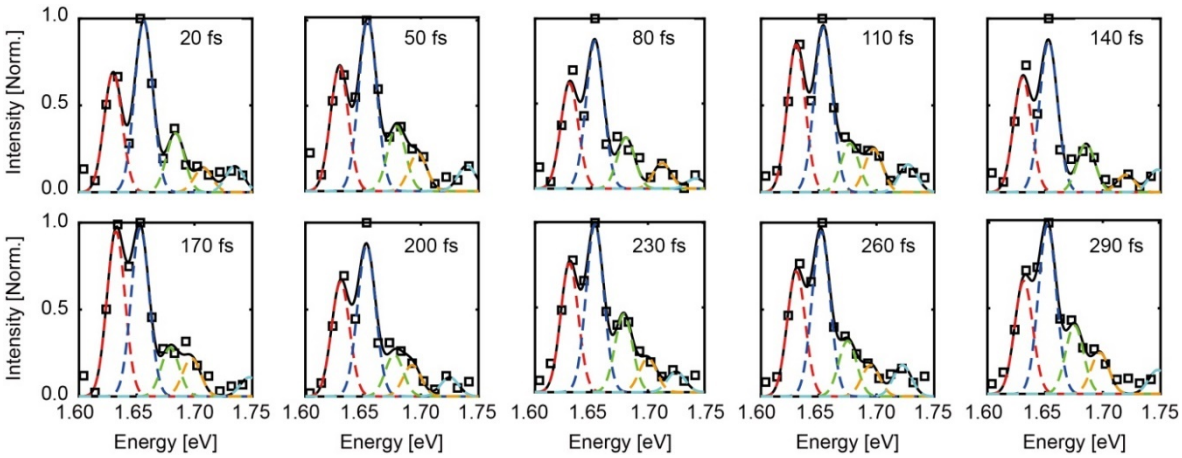
Supplementary Table 2. The energy positions of the five components of the nonrephasing signal retrieved from the fitting procedure.

T	20 fs	50 fs	80 fs	110 fs	140 fs	170 fs	200 fs	230 fs	260 fs	290 fs
Com. 1	1.6295 eV	1.6295 eV	1.6305 eV	1.6305 eV	1.6305 eV	1.6320 eV	1.6305 eV	1.6305 eV	1.6305 eV	1.6315 eV
Com. 2	1.6560 eV	1.6540 eV	1.6530 eV	1.6540 eV	1.6530 eV	1.6530 eV	1.6530 eV	1.6530 eV	1.6520 eV	1.6520 eV
Com. 3	1.6835 eV	1.6790 eV	1.6805 eV	1.6770 eV	1.6855 eV	1.6790 eV	1.6770 eV	1.6785 eV	1.6750 eV	1.6755 eV
Com. 4	1.7075 eV	1.6975 eV	1.7125 eV	1.6975 eV	1.7200 eV	1.6975 eV	1.6925 eV	1.7000 eV	1.6950 eV	1.6975 eV
Com. 5	1.7350 eV	1.7400 eV	1.7425 eV	1.7275 eV	1.7475 eV	1.7475 eV	1.7250 eV	1.7250 eV	1.7225 eV	1.7475 eV

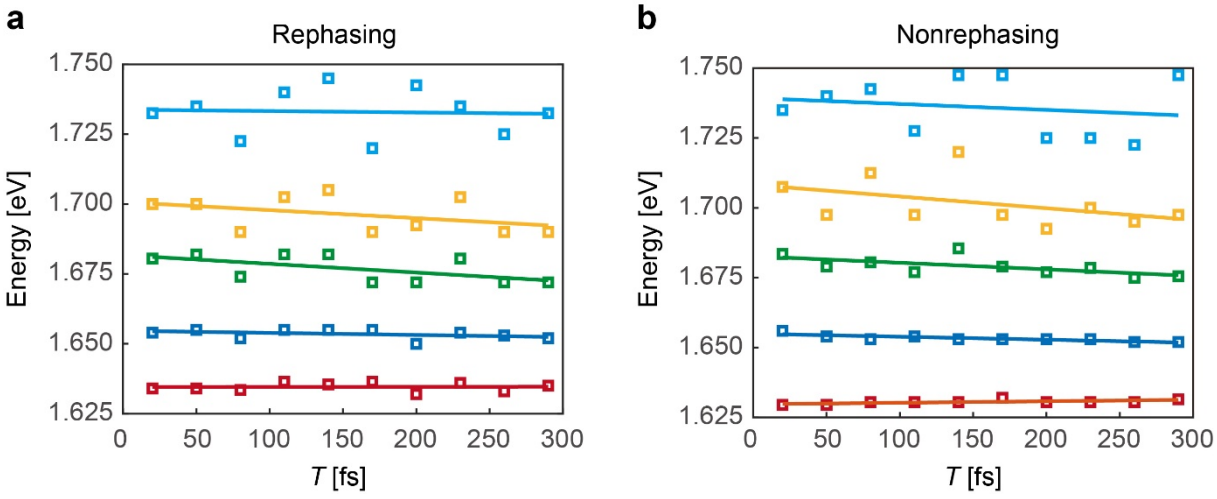
Supplementary Figure 10 shows the fitted positions of the five diagonal peaks from the rephasing and the nonrephasing 2D spectra for T = 20, 50, 80, …, 290 fs. Firstly, we found that the rephasing and the nonrephasing results are in line with each other, both clearly showing a five-peak band structure. Therefore, to reduce the errors, we combine the rephasing and the nonrephasing results by calculating the mean value of the retrieved peak positions from the rephasing and the nonrephasing 2D spectra for every individual T, with the combined results shown in Fig. 3(c) of the main text. From that, we found that the four high-energy bands show red-shift behavior as T increases, as indicated by the linear fitting lines with negative slopes. The error bar of each band can be determined by the standard deviation of the fluctuations calculated from the differences between the measured values (marked by squares in Fig. 3c of the main text) and the linear fit results (solid lines in Fig. 3c of the main text) at each T. The fitting slopes of the five bands are plotted in Fig. 3(d) of the main text, with the error bars depicting 95% confidence bounds from the fitting. It can be seen that for bands 2 and 3 the whole 95% regions are below the line of zero, for band 4 most part of the 95% region is below the line of zero, whereas for band 5 the probability of the slope to be negative is reduced because of the large uncertainty.



Supplementary Figure 8. Fitting (black solid) of the diagonal slices (black squares) from rephasing 2D spectra using a sum of Gaussians (red, blue, green, orange, and cyan dashed).



Supplementary Figure 9. Fitting (black solid) of the diagonal slices (black squares) from nonrephasing 2D spectra using a sum of Gaussians (red, blue, green, orange, and cyan dashed).

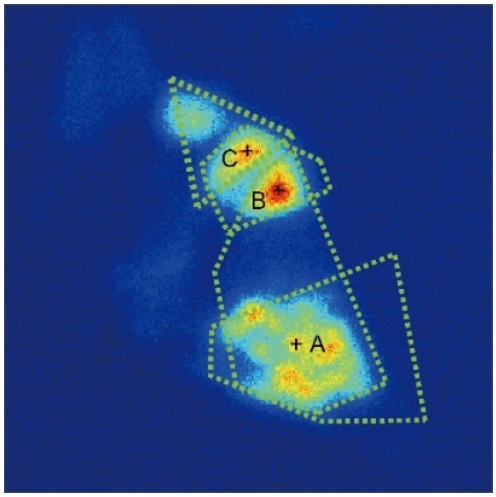


Supplementary Figure 10. Fitted positions of the five diagonal peaks of the rephasing (a) and the nonrephasing (b) 2D spectra at different *T*.

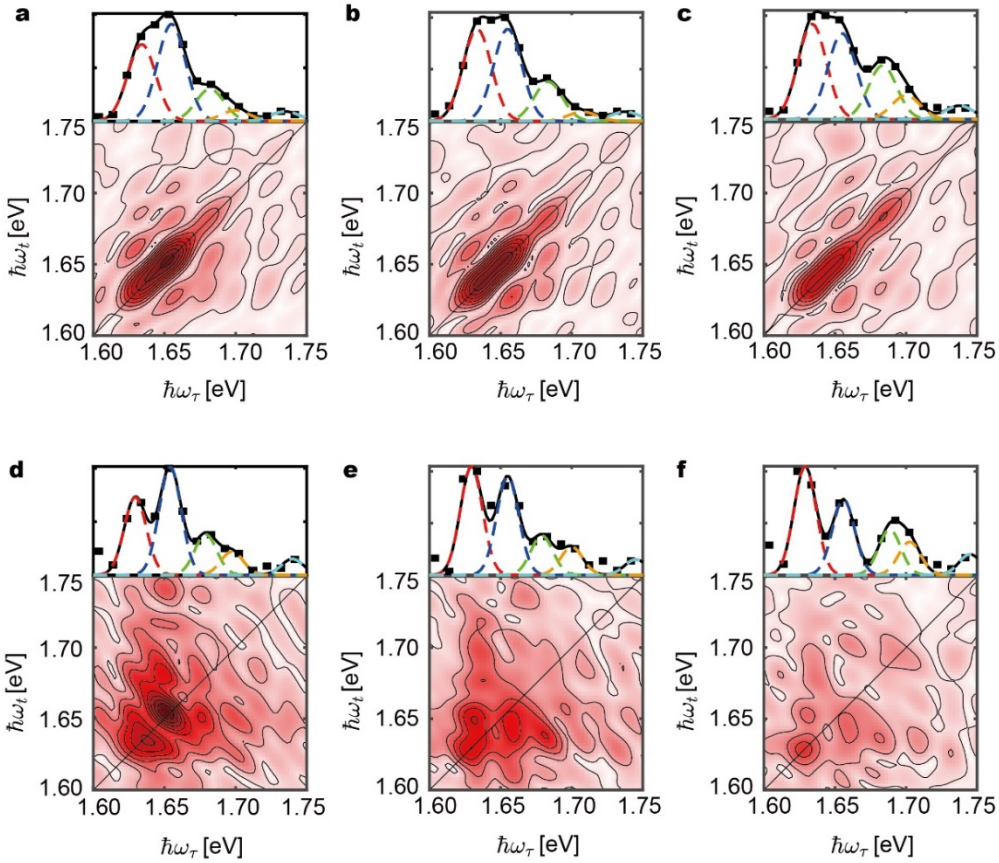
Supplementary Note 8: 2D spectra from different positions

We measured 2D spectra at different DML regions (marked as Pos. A, B, and C in Supplementary Figure 11) and show the rephasing (a,b,c) and the nonrephasing (d,e,f) 2D spectra in Supplementary Figure 12. On top of each 2D spectrum is the slice along the diagonal line of the corresponding 2D spectrum (black squares). The slices are fitted by five Gaussian functions (colored dashed curves), the sum of which shown as the black curve. The retrieved peak positions of the five components are illustrated in Supplementary Figure 13 for different spatial positions within the DML of the sample, where squares are employed for the rephasing

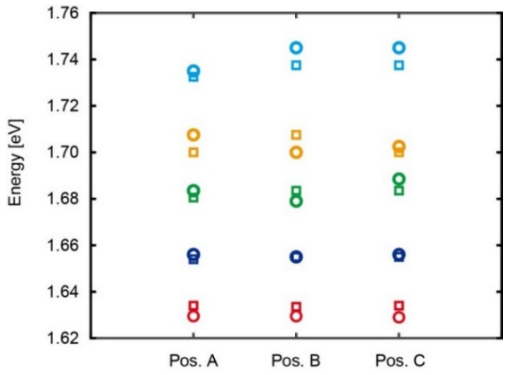
and circles for the nonrephasing signals. We observe the same five-peak energy structure for all three positions, indicating reproducibility of our results throughout the DML regions.



Supplementary Figure 11. Three different positions (Pos. A, B, C marked by black crosses) where 2D spectra were collected.



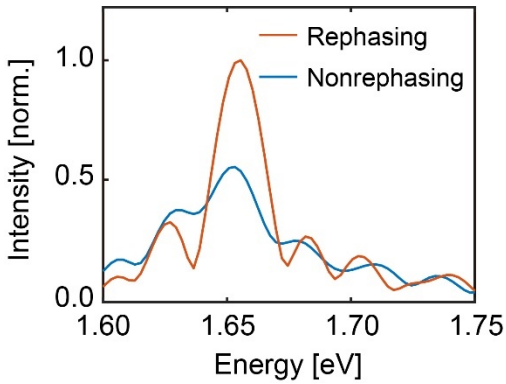
Supplementary Figure 12. Rephasing (a,b,c) and nonrephasing (d,e,f) 2D spectra collected at position A (a,d), position B (b,e), and position C (c,f). Top: Slices through the 2D spectra along the diagonal lines (squares) and the fitting results (black solid curves) using five Gaussian functions (colored dashed curves).



Supplementary Figure 13. Retrieved peak positions of the five components for different positions, where squares indicate rephasing and circles nonrephasing data.

Supplementary Note 9: Positions of off-diagonal peaks

We plotted the line cuts at ω_τ = 1.655 eV for the rephasing (Fig. 4b) and the nonrephasing (Fig. 4c) 2D spectrum for T = 50 fs, as shown in Supplementary Figure 14. The peak positions are retrieved from the line cuts by locating the maximum positions and are listed in Supplementary Table 3. The retrieved off-diagonal peak positions all lie within the errors of the experimentally determined diagonal peak positions, indicating the same five-peak energy structure.



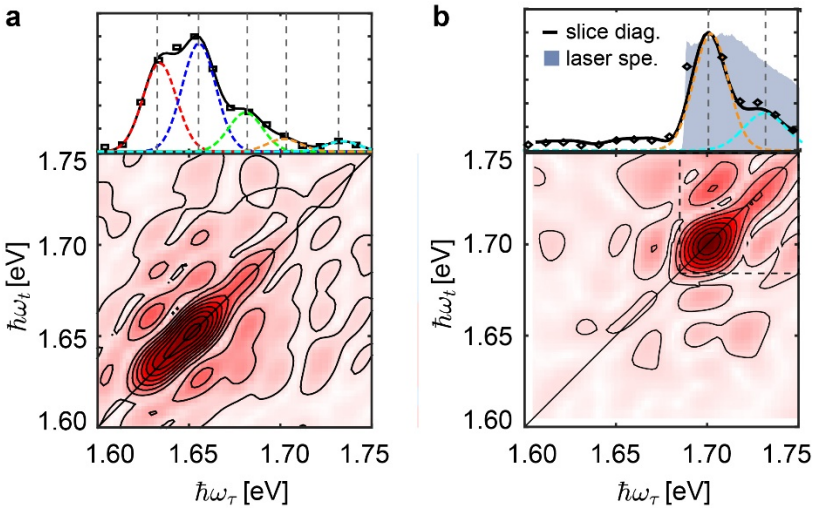
Supplementary Figure 14. Line cuts at ω_τ = 1.655 eV for the rephasing (red curve) and the nonrephasing (blue curve) 2D spectrum for T = 50 fs.

Supplementary Table 3. Retrieved positions of off-diagonal peaks of the line cuts through the rephasing 2D (Row 2) and nonrephasing (Row 3) 2D spectra for T = 50 fs.

	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
Rephasing	1.627 eV	1.655 eV	1.682 eV	1.705 eV	1.741 eV
Nonrephasing	1.629 eV	1.653 eV	1.679 eV	1.708 eV	1.734 eV

Supplementary Note 10: 2D measurements with higher excitation energies

To highlight the relatively weak (i.e., the fourth and the fifth) diagonal peaks located on the high-energy side of 2D spectra, we performed an additional 2D measurement at T = 50 fs with the laser spectrum moved to higher excitation energies and total intensity increased to keep the linear PL signal intensity at the same level. As a result, the peaks inside the black dashed box of Supplementary Figure 15b are strongly enhanced compared to Supplementary Figure 15a, verifying the existence of these two peaks.



Supplementary Figure 15. Bottom: Measured rephasing 2D spectrum (absolute value) at T = 50 fs, with the excitation laser spectrum blocked below 1.615 eV (a) and below 1.685 eV (b). We cut off energies below the quoted values by inserting a low-energy-blocking aperture (LEBA) in front of the spatial light modulator (SLM, Supplementary Figure 1) and adjusting its position correspondingly. Top: Slices through the 2D spectra along the diagonal lines (black solid curves), laser spectrum (blue shaded areas), and energies of the components $|E_{1\text{-}5}\rangle$ (vertical dashed lines) as determined from the fitting of the curves with sums of Gaussians (individual contributions in colored dashed lines).

Supplementary Note 11: Quantifying the vibronic Hamiltonian

From the energy-momentum-resolved PL map, we obtain the energies at $k_{\|}$ = 0 for three polariton branches, which are 1.605, 1.630, and 1.643 eV. The coherent 2D spectra reveal three additional higher branches. Since the 2D micro-spectroscopy measures momentum-integrated signals, we cannot precisely determine the energies at $k_{\|}$ = 0 for these three branches. We thus take the lowest peak positions obtained from the diagonal slices among all the different T (1.672,

1.690 and 1.720 eV) approximately as the zero-momentum values, such that we have the six-point energy structure (listed in the first row of Supplementary Table 4) to deduce the vibronic Hamiltonian. Although such an approximation will induce certain errors, the simulated vibronic polariton dispersion relations can explain the experimental results of both the energy-momentum-resolved PL and spatially resolved 2D spectroscopy, as indicated in Fig. 4 of the main text.

Specifically, to fit the experimentally determined energy positions, we used E_V, E_C, Ω, and S as fitting parameters. For each set of [E_V, E_C, Ω, S], we conducted a diagonalization of Supplementary Eq. (S2) and obtained six eigenenergies E_k (k = 0, …, 5). We calculated the mean-square deviation, $D = \sqrt{\sum_{k=0}^{5}\left(E_k - \tilde{E}_k\right)^2}$, with $\tilde{E}_k$ being the measured energy positions (the first row of Supplementary Table 4) and E_k being the calculated eigenenergies of six polariton states at $k_{\|} = 0$. The minimum deviation can be found with E_V = 0.020 eV, E_C = 1.623 eV, S = 0.02, and Ω = 0.06 eV. The calculated energy positions for this parameter set can be found in Supplementary Table 4 (row 2). The phonon energy is determined to be 20 meV, close to that of the E″ phonon mode (21.8 meV) [9]. The small Huang–Rhys factor signifies relatively weak coupling between the exciton and the phonon mode. This differs from most vibronic polaritons observed in organic microcavities, in which the vibronic coupling between excitons and phonons competes in strength with the exciton–photon coupling. There may be two reasons for this particular phonon mode to contribute significantly to the exciton polariton formation in the studied system. First, because of the low energy of the phonon, many phonons are initially populated on their excited state at room temperature. Second, the likely long lifetime for this phonon mode both on the ground and the excited electronic states facilitates that the vibrational relaxation is slower than coherent Rabi exchange, prompting the 1–1 and the 2–2 transitions to interact strongly with the cavity photons.

Supplementary Table 4. Band positions obtained from experimentally measured energy positions (Row 1) and their fitting results using the vibronic polariton model in Supplementary Note 1 (Row 2).

$\widetilde{E}_k$	**1.605 eV**	**1.630 eV**	**1.643 eV**	**1.672 eV**	**1.690 eV**	**1.720 eV**
E_k	1.605 eV	1.625 eV	1.646 eV	1.674 eV	1.794 eV	1.713 eV

Supplementary References

[1] F. C. Spano, *Optical Microcavities Enhance the Exciton Coherence Length and Eliminate Vibronic Coupling in J-Aggregates*, J. Chem. Phys. **142**, 184707 (2015).

[2] C. Rupprecht, M. Klaas, H. Knopf, H. Knopf, H. Knopf, T. Taniguchi, K. Watanabe, Y. Qin, Y. Qin, S. Tongay, S. Tongay, S. Schröder, F. Eilenberger, F. Eilenberger, F. Eilenberger, S. Höfling, and C. Schneider, *Demonstration of a Polariton Step Potential by Local Variation of Light-Matter Coupling in a van-Der-Waals Heterostructure*, Opt. Express **28**, 18649 (2020).

[3] H. van Amerongen, L. Valkunas, and R. van Grondelle, *Photosynthetic Excitons* (World Scientific Publishing Co. Pte. Ltd., 2000).

[4] A. V. Kavokin, J. J. Baumberg, G. Malpuech, and F. P. Laussy, *Microcavities*, Second Edition (Oxford University Press, Oxford, New York, 2017).

[5] H. Knopf, N. Lundt, T. Bucher, S. Höfling, S. Tongay, T. Taniguchi, K. Watanabe, I. Staude, U. Schulz, C. Schneider, and F. Eilenberger, *Integration of Atomically Thin Layers of Transition Metal Dichalcogenides into High-Q, Monolithic Bragg-Cavities: An Experimental Platform for the Enhancement of the Optical Interaction in 2D-Materials*, Opt. Mater. Express **9**, 598 (2019).

[6] T. Wu, J. Tang, B. Hajj, and M. Cui, *Phase Resolved Interferometric Spectral Modulation (PRISM) for Ultrafast Pulse Measurement and Compression*, Opt. Express **19**, 12961 (2011).

[7] M. Pawł owska, S. Goetz, C. Dreher, M. Wurdack, E. Krauss, G. Razinskas, P. Geisler, B. Hecht, and T. Brixner, *Shaping and Spatiotemporal Characterization of Sub-10-Fs Pulses Focused by a High-NA Objective*, Opt. Express **22**, 31496 (2014).

[8] S. Goetz, D. Li, V. Kolb, J. Pflaum, and T. Brixner, *Coherent Two-Dimensional Fluorescence Micro-Spectroscopy*, Opt. Express **26**, 3915 (2018).

[9] X. Luo, Y. Zhao, J. Zhang, M. Toh, C. Kloc, Q. Xiong, and S. Y. Quek, *Effects of Lower Symmetry and Dimensionality on Raman Spectra in Two-Dimensional WSe_2*, Phys. Rev. B **88**, 195313 (2013).