Supplementary Information

Correlated Excitonic Signatures of Individual van der Waals NiPS3 Antiferromagnet Nanoflakes

Vigneshwaran Chandrasekaran,[*a] Christopher R. DeLaney,[a] Cong Tai Trinh,[a] David Parobek,[a] Christopher A. Lane,[b] Jian-Xin Zhu,[a,b] Xiangzhi Li,[a] Huan Zhao,[a] Marshall A. Campbell,[a] Laura Martin,[c] Edward F. Wyckoff,[c] Andrew C. Jones,[a] Matthew M. Schneider,[d] John Watt,[a] Michael T. Pettes,[a] Sergei A. Ivanov,[a] Andrei Piryatinski,[b] David H. Dunlap[e] and Han Htoon[*a]

[a.] *Center for Integrated Nanotechnologies, Materials Physics and Applications Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States.*
[b.] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States.*
[c.] *Sandia National Laboratories, 11515 Eubank SE, Albuquerque, New Mexico 87185, United States.*
[d.] *Materials Science in Radiation and Dynamics Extremes, Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States.*
[e.] *Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, United States.*
Supplementary Information available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

Contents

S1. HR-TEM images of nanoflakes

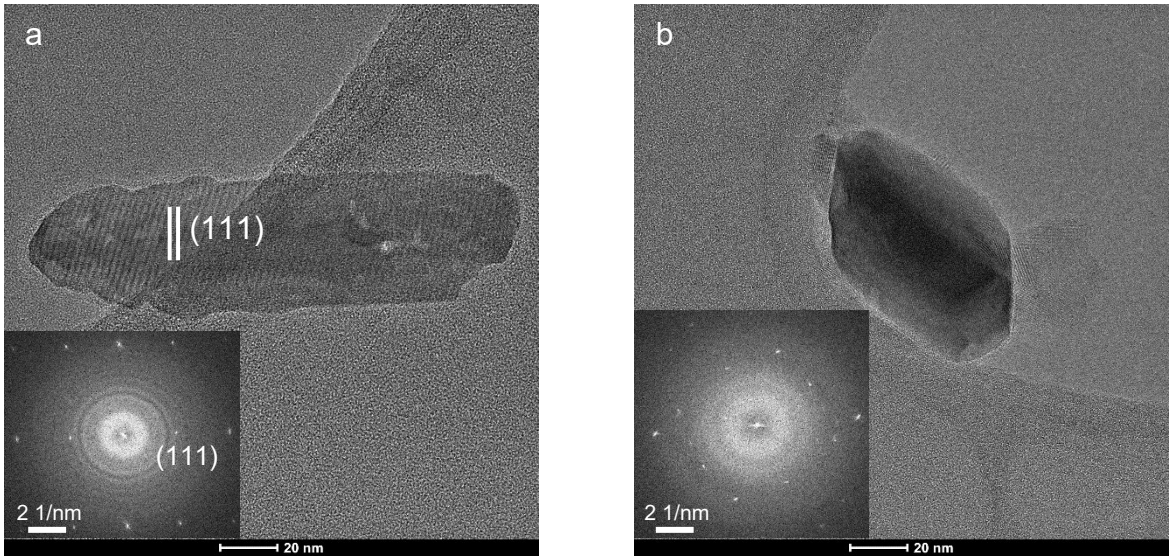

Figure S1. HR-TEM of crystalline nanoflakes. (a,b) A flake displaying a (111) lattice plane, and the other composing of a large central grain with smaller randomly oriented grains attached mostly on the side of the central grains.

S2. Characterization of chemically synthesized $NiPS_3$ nanoflakes

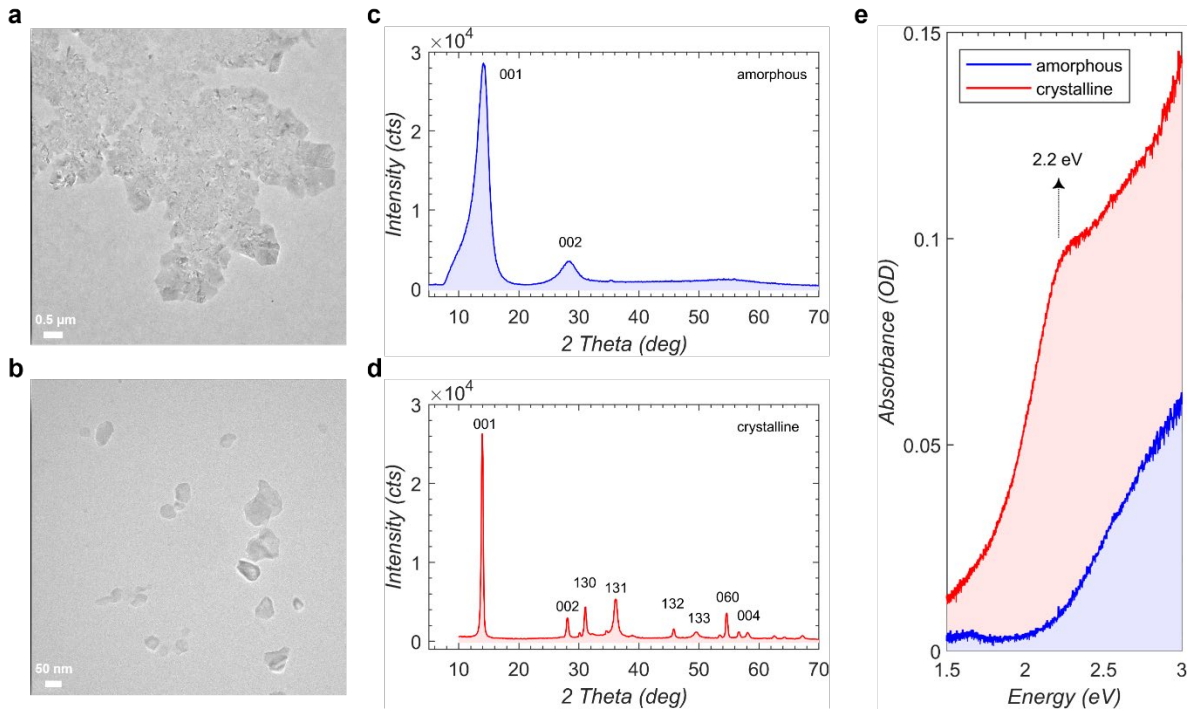

Figure S2. TEM, XRD and absorption spectra of amorphous and crystalline nanoflakes. (a,b) TEM images of nanoflakes in amorphous and crystalline phases respectively. Size range of crystalline nanoflakes is 20-100 nm. (c,d) XRD obtained from amorphous and crystalline nanoflakes of $NiPS_3$. (e) Absorption spectra of amorphous and crystalline nanoflakes of $NiPS_3$. Crystalline sample shows a pronounced peak around 2.2 eV.

S3. Raman spectrum of nanoflakes

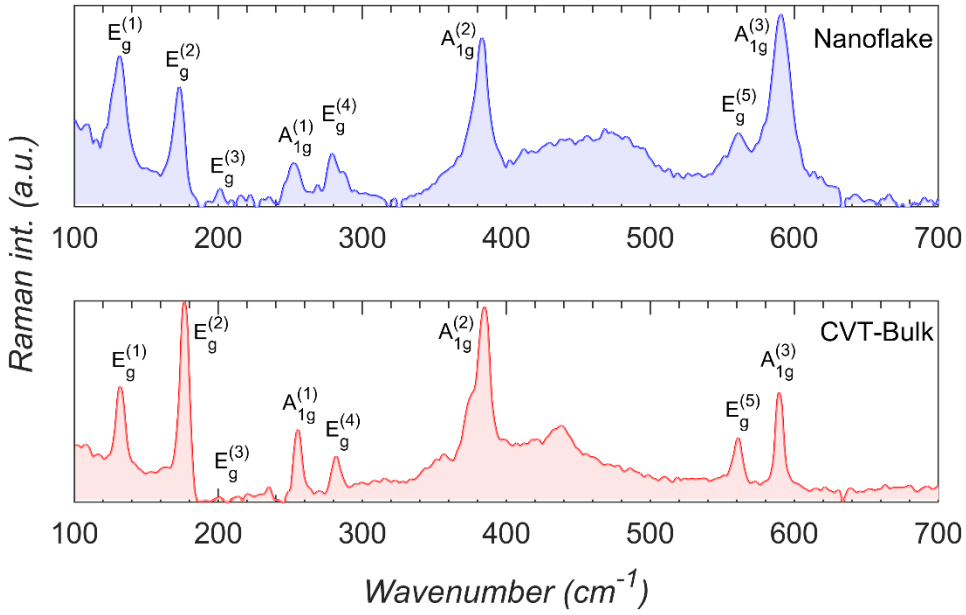

Figure S3. Raman spectra of nanoflakes and CVT-Bulk. Raman spectra of nanoflakes and CVT-Bulk display the same characteristic Raman phonon modes of bulk $NiPS_3$ (*Sci Rep* **6**, 20904 (2016)) providing evidence that chemically synthesized nanoflakes and CVT-Bulk sample share identical chemical composition and crystalline structure. Data is collected with an excitation energy of 2.33 eV at 298 K.

S4. Excitation power dependence of PL in a nanoflake

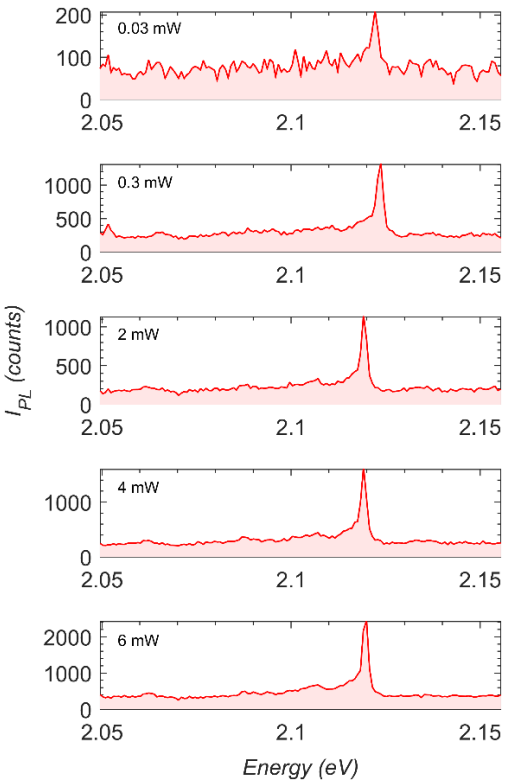

Figure S4. Excitation power dependence of PL in a nanoflake. PL from a nanoflake under different excitation power density for the dataset shown in main Figure 2b.

S5. Linear polarized PL in nanoflakes

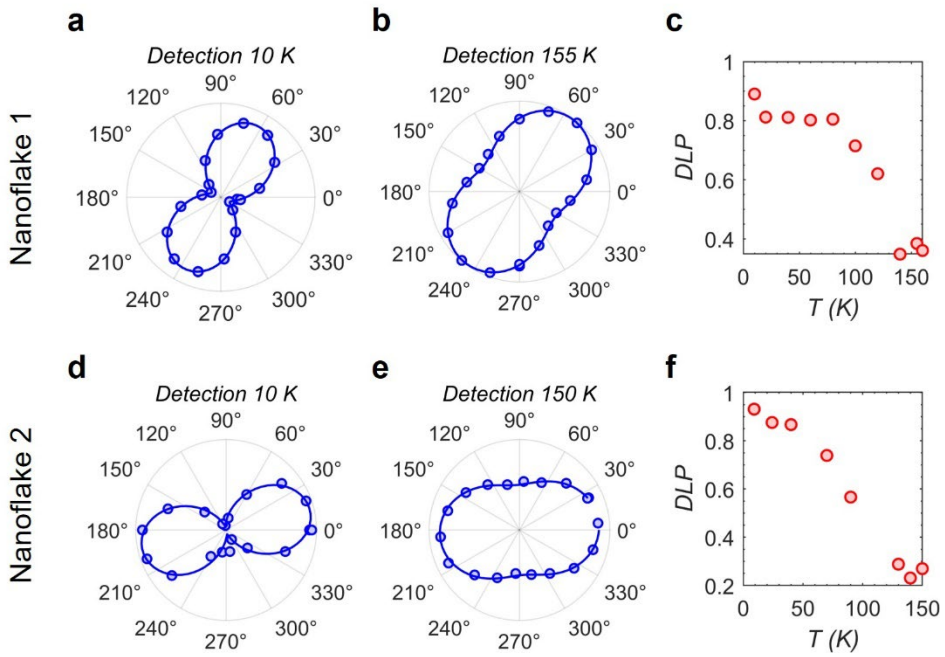

Figure S5. Spin-correlated exciton in nanoflakes. (a,d) Strong degree of linear polarization (~90%) for 2 different nanoflakes at 10 K. (b,e) At >150K, degree of linear polarization decreased to about 38% and 22% for 2 different nanoflakes respectively. (c,f) Degree of linear polarization for different temperatures shows a drop from a near-unity degree of linear polarization to about 38% & 22% at around $T_N \approx 140$ K. A residual degree of linear polarization is still present beyond the T_N.

S6. Magnetization curve of nanoflakes

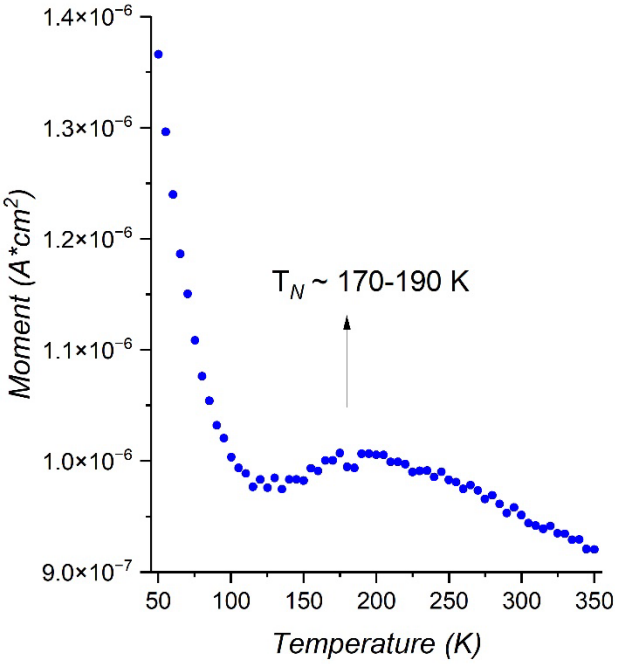

Figure S6. Magnetization curve of nanoflakes. Magnetization curve of crystalline $NiPS_3$ nanoflakes, showing Néel transition temperature T_N around 170-190 K. The sample was measured at higher magnetic field ($1x10^4$ Oe) in the range 50-350 K. At each temperature point, the sample was allowed to equilibrate for 10 min. It is believed that the magnetic susceptibility comprises antiferromagnetic contribution from the interior of the NFs and the paramagnetic contribution from the uncompensated spins arising across the edge of the NFs. This attribution is further supported by the inflection point at 110 K, which is most likely the consequence of AFM and paramagnetic contributions overlapping with each other. The apparent T_N dependence on the size of nanoparticles was expected and currently being investigated separately. The diamagnetic contribution of the sample holder was measured to be nearly constant in the range between $-1.3x10^{-7}$ and $-1.4x10^{-7}$ A*cm^2, which would not alter the profile of the magnetization curve in any significant extent and hence wasn't subtracted from the overall signal.

S7. Exciton-phonon bound states

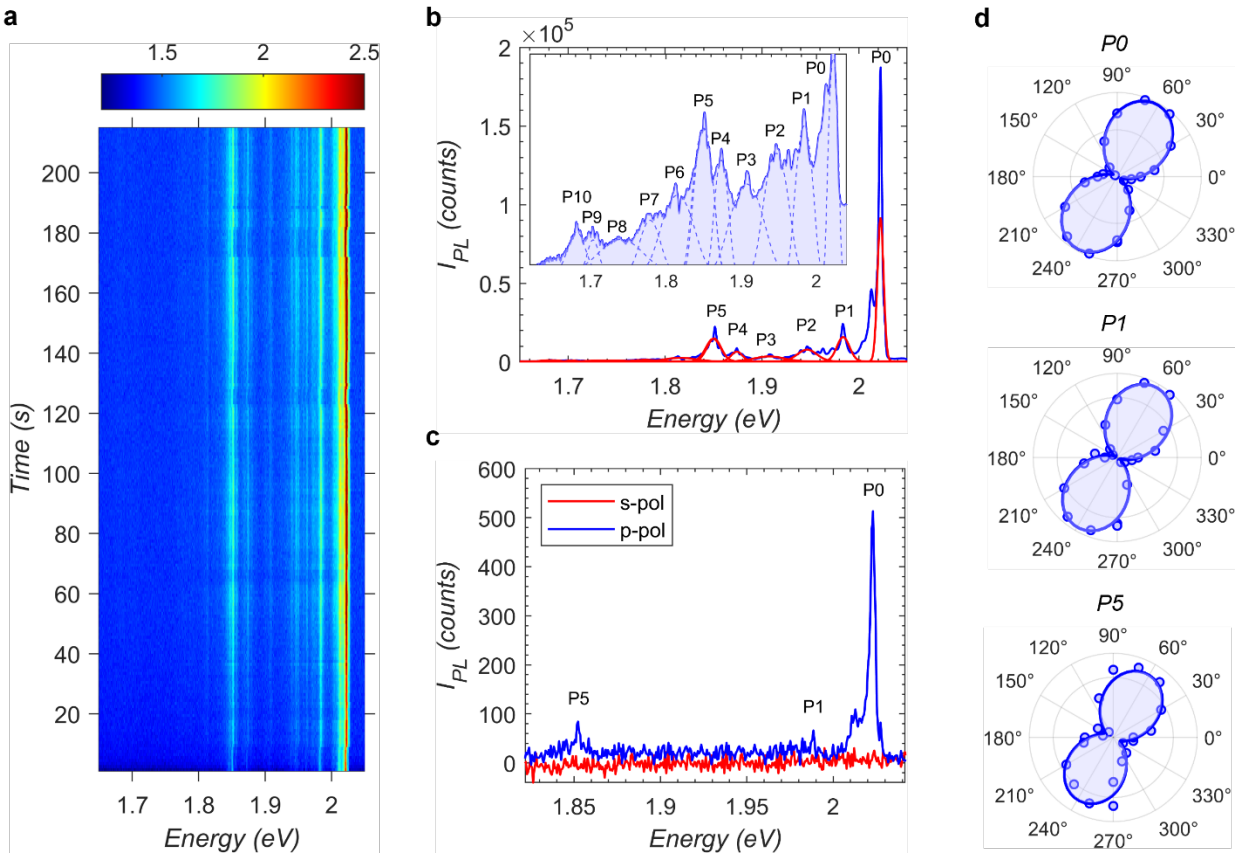

Figure S7. Spectral diffusion of phonon replicas. (a) A time dependent spectral image sequence of PL collected from a nanoflake with 1s each acquisition. This shows the phonon replica of the low energy tail following the spectral diffusion of the zero-phonon peak. (b) Integrated spectrum of the same nanoflake (blue line) showing the main peak P0 and 10 low energy tail peaks labelled from P1 to P10. Red lines are individual Gaussian fits to different peaks. Inset shown in log plot. (c) Polarization resolved spectra of the same nanoflake with 3 peaks P0, P1 and P5 visible. (d) Polarplot of the respective peaks displaying a strong linear polarization oriented in the same direction. All data are collected at 10 K.

S8. Replica feature

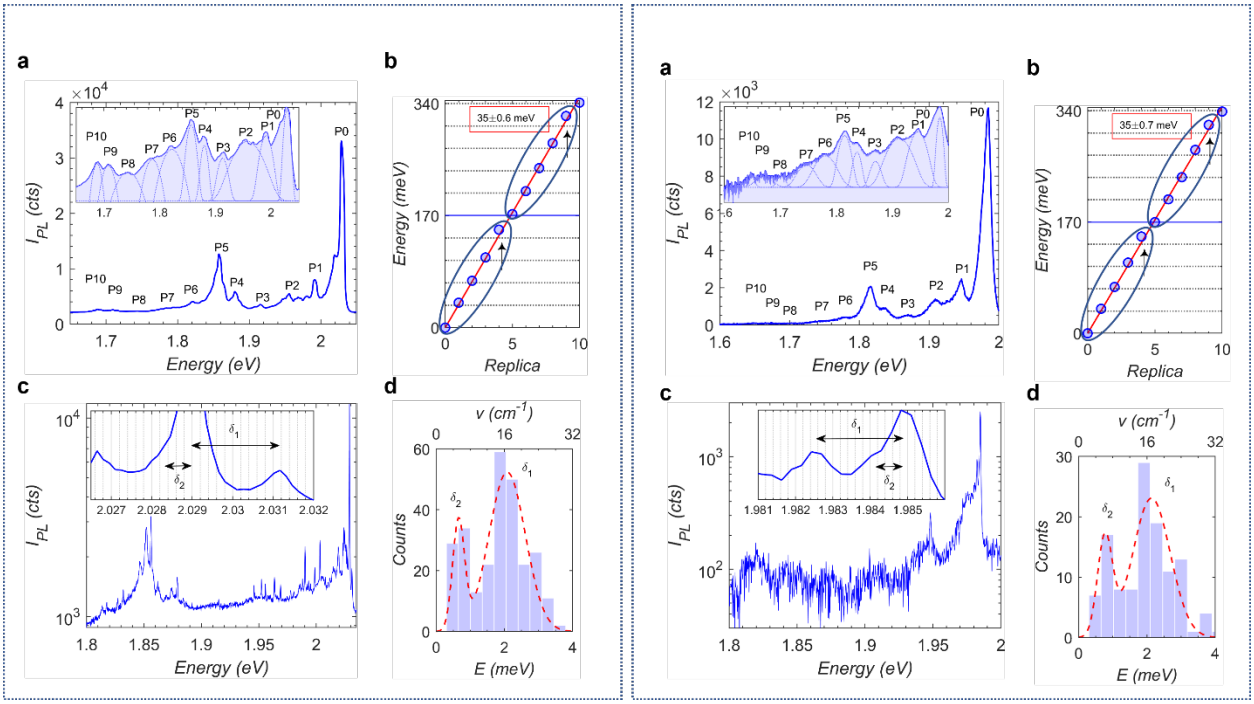

Figure S8. Low- and high-resolution PL spectra of 2 different nanoflakes. (a) Low-resolution spectrum of a nanoflake showing the main peak P0 and 10 low energy tail peaks labelled from P1 to P10. Inset shown in log plot. (b) Tail peak positions from the central emission energy of P0 fitted with a linear function giving a slope of ~35 meV. Arrows show that P4 and P9 slightly deviate from the linear function. Ellipsoid shows the pattern of P0-P4 repeats at P5-P9. (c) High-resolution spectrum of the same nanoflake. Inset shows a zoomed-in portion with noticeable equidistant consecutive peak separations marked as δ1 and δ2. (d) Histogram of consecutive peak difference gives a value of ~2 meV for δ1 and ~0.7 meV for δ2. All data are collected at 10 K.

S9. Holstein model for exciton-polaron quasiparticle and associated PL lineshape

Let us introduce the Holstein model assuming on site interaction of the exciton and phonon states described by Bose {creation, annihilation} operators $\{\hat{a}_n,\hat{a}_n^\dagger\}$ and $\{\hat{b}_q,\ \hat{b}_q^\dagger\}$, respectively. Notice that the exciton operators are presented in the real space whereas the phonon operators are in momentum space q. The model Hamiltonian in such a representation reads

$$\widehat{H} = \hbar\ \sum_n\left[\Delta_0 + \sum_q g_q\omega_q e^{iqn}\left(\hat{b}_q + \hat{b}_{-q}^\dagger\right)\right]\hat{a}_n^\dagger\,\hat{a}_n + \hbar\sum_q \omega_q\hat{b}_q^\dagger\hat{b}_q\ , \tag{1.1}$$

with index n running over the exciton delocalization range, $\hbar\Delta_0$ denoting bare exciton transition energy, $\hbar\omega_q$ standing for the phonon energy, and g_q being dimensionless exciton-phonon coupling parameter represented in the momentum space. Following standard approach, we perform the Lang-Firsov (displacement) transformation

$$\widehat{D} = exp\left\{\sum_{nq} g_q e^{iqn}\left(\hat{b}_q - \hat{b}_{-q}^\dagger\right)\hat{a}_n^\dagger\,\hat{a}_n\right\} \tag{1.2}$$

of the Hamiltonian (1.1), i.e., $\widehat{H}' = \widehat{D}\widehat{H}\widehat{D}^{-1}$. The transformation eliminates the exciton-phonon coupling term resulting in the Hamiltonian for the uncoupled dressed exciton and phonon states

$$\widehat{H}' = \hbar\sum_n\left(\Delta_0 - \sum_{q,n'} g_q^2\omega_q e^{iqn'}\hat{A}_{n'}^\dagger\hat{A}_{n'}\right)\hat{A}_n^\dagger\hat{A}_n + \hbar\sum_{\mathrm{q}}\omega_q\hat{B}_q^\dagger\hat{B}_q. \tag{1.3}$$

Here, the transformed exciton and phonon operators are

$$\hat{A}_n = \widehat{D}\hat{a}_n\widehat{D}^{-1} = \hat{a}_n e^{-\sum_q g_q e^{iqn}\left(\hat{b}_q - \hat{b}_{-q}^\dagger\right)}\ , \tag{1.4}$$

$$\hat{B}_q = \widehat{D}\hat{b}_q\widehat{D}^{-1} = \hat{b}_q + \sum_n g_q e^{iqn}\hat{a}_n^\dagger\hat{a}_n, \tag{1.5}$$

respectively. The inverse transformation gives the following relationship between bare exciton and phonon operators in term of the new operators

$$\hat{a}_n = \hat{A}_n e^{\sum_q g_q e^{iqn}\left(\hat{B}_q - \hat{B}_{-q}^\dagger\right)}\ , \tag{1.6}$$

$$\hat{b}_q = \hat{B}_q - \sum_n g_q e^{iqn}\hat{A}_n^\dagger\hat{A}_n\ . \tag{1.7}$$

To simplify the quartic term in the transformed Hamiltonian (1.3), we apply the mean-field approximation by introducing the on-site exciton density $\langle\hat{A}_n^\dagger\hat{A}_n\rangle$. This approximation results in the diagonalized Hamiltonian

$$\widehat{H}_{MF} = \hbar\Delta\sum_n\hat{A}_n^\dagger\hat{A}_n + \hbar\sum_q\omega_q\hat{B}_q^\dagger\hat{B}_q, \tag{1.8}$$

where the exciton polaron energy is $\Delta = \left(\Delta_0 - 2\sum_{qk} g_q^2\omega_q\,\langle\hat{A}_k^\dagger\hat{A}_{k+q}\rangle\right)$ including the polaron shift $-2\sum_{qk} g_q^2\omega_q\,\langle\hat{A}_k^\dagger\hat{A}_{k+q}\rangle$ proportional to the zero-time exciton correlation function, $\langle\hat{A}_k^\dagger\hat{A}_{k+q}\rangle$, in the momentum space. With the help of Eq. (1.8), we find the following representation of the bare exciton operator Eq. (1.6) at time t in terms of the dressed exciton and phonon operators

$$\hat{a}_n(t) = e^{\frac{i}{\hbar}\widehat{H}_{MF}t}\hat{A}_n e^{\sum_q g_q\,e^{iqn}\left(\hat{B}_q - \hat{B}_{-q}^\dagger\right)}e^{-\frac{i}{\hbar}\widehat{H}_{MF}t} = \hat{A}_n e^{-i\Delta t}\ e^{\sum_q g_{qn}e^{iqn}\left(\hat{B}_q e^{-i\omega_q t} - \hat{B}_{-q}^\dagger e^{i\omega_{-q}t}\right)}. \tag{1.9}$$

Power spectrum for the spontaneous photon emission power can be defined as the Fourier transform

$$S(\omega) = \mathrm{Re}\int_0^\infty dt\, e^{i\omega t}C_{k=0}(t) \tag{1.10}$$

of the exciton auto-correlation function

$$C_k(t) = \langle\hat{a}_k^\dagger(0)\hat{a}_k(t)\rangle \tag{1.11}$$

represented in the momentum space. By setting the exciton momentum $k = 0$, we account for the negligibly small momentum of the emitted photons. The angular breakers in Eq. (1.10) denote the average with the reduced density operator for the excitons and phonons. The time-evolution of the

exciton operator is governed by the Hamiltonian (1.8). By performing the Fourier transformation of Eq. (1.11), we find the desired time dependance for the bare exciton operator

$$\hat{a}_k(t) = \sum_n \hat{A}_n e^{ikn}\, e^{\sum_q g_q\left(\hat{B}_q e^{-i\omega_q t} - \hat{B}_{-q}^{\dagger} e^{i\omega_{-q} t}\right)} e^{iqn}. \tag{1.12}$$

Making substitution of Eq. (1.12) and its Hermitian conjugate at time zero into Eq. (1.11), we find the exciton auto-correlation function

$$C_k(t) = \sum_{nn'} \langle \hat{A}_n^{\dagger} \hat{A}_{n'} \rangle e^{-i\Delta t + ik(n-n')} \left\langle 0 \left| e^{-\sum_q g_q^*\left(\hat{B}_{-q} - \hat{B}_q^{\dagger}\right) e^{-iqn}}\, e^{\sum_q g_q\left(\hat{B}_q e^{-i\omega_q t} - \hat{B}_{-q}^{\dagger} e^{i\omega_{-q} t}\right) e^{iqn'}} \right| 0 \right\rangle, \tag{1.13}$$

where we used a reasonable assumption that at zero time the reduced density operator partitions into the product of the phonon vacuum density $|0\rangle\langle 0|$ and the exciton density matrix $\langle \hat{A}_n^{\dagger} \hat{A}_{n'} \rangle$. Further applying the Baker-Hausdorff formula $e^{\hat{F}+\hat{G}} = e^{\frac{1}{2}[\hat{F},\hat{G}]} e^{\hat{G}} e^{\hat{F}}$ and the relationship $g_q^* = g_{-q}$, we evaluate the phonon matrix element in Eq. (1.13)

$$\begin{aligned}
&\left\langle 0 \left| e^{-\sum_q g_q^*\left(\hat{B}_{-q} - \hat{B}_q^{\dagger}\right) e^{-iqn}}\, e^{\sum_q g_q\left(\hat{B}_q e^{-i\omega_q t} - \hat{B}_{-q}^{\dagger} e^{i\omega_{-q} t}\right) e^{iqn'}} \right| 0 \right\rangle \\
&= e^{i\sum_q |g_q|^2 \sin\left(\omega_q t + q(n-n')\right)} \\
&\times \left\langle 0 \left| e^{\sum_q g_q \hat{B}_q\left(e^{-i\omega_q t + iqn'} - e^{iqn}\right) - \sum_q g_q \hat{B}_q^{\dagger}\left(e^{i\omega_q t - iqn'} - e^{-iqn}\right)} \right| 0 \right\rangle \\
&= e^{\sum_q |g_q|^2 \{i \sin\left(\omega_q t + q(n-n')\right) + \cos\left(\omega_q t + q(n-n')\right) - 1\}} \\
&\times \left\langle 0 \left| e^{-\sum_q g_q \hat{B}_q^{\dagger}\left(e^{i\omega_q t - iqn'} - e^{-iqn}\right)} e^{\sum_q g_q \hat{B}_q\left(e^{-i\omega_q t + iqn'} - e^{iqn}\right)} \right| 0 \right\rangle \\
&= e^{\sum_q |g_q|^2 \left(e^{i\omega_q t + iq(n-n')} - 1\right)}.
\end{aligned} \tag{1.14}$$

This calculation simplifies the exciton auto-correlation function (1.14), to the form

$$C_k(t) = e^{-\sum_q |g_q|^2} \sum_{nn'} \langle \hat{A}_n^{\dagger} \hat{A}_{n'} \rangle e^{-i\Delta t + ik(n-n')}\, e^{\sum_q |g_q|^2 e^{i\omega_q t + iq(n-n')}}. \tag{1.15}$$

A final touch in our derivation is to express the exciton density matrix via the exciton Wigner distribution function

$$\langle \hat{A}_n^{\dagger} \hat{A}_{n'} \rangle = \sum_{\kappa} W\left(\frac{n+n'}{2}, \kappa\right) e^{i\kappa(n'-n)}, \tag{1.16}$$

of momentum κ and coordinate $(n + n')/2$. Introducing this result in Eq. (1.15), we obtain

$$C_k(t) = e^{-\sum_q |g_q|^2} \sum_{n\kappa} f_{\kappa}\, e^{-i\Delta t + i(k-\kappa)n}\, e^{\sum_q |g_q|^2 e^{i\omega_q t + iqn}}, \tag{1.17}$$

as a function of the momentum space exciton distribution function $f_{\kappa} \equiv \sum_{n+n'} W\left(\frac{n+n'}{2}, \kappa\right)$. To complete the calculation, we expand the last exponential in Eq. (1.17) into the Taylor series and perform the summation over *n* and κ to obtain

$$C_k(t) = e^{-\sum_q |g_q|^2} \sum_{m=0}^{\infty} \sum_{q_1} \dots \sum_{q_m} \frac{f_{k+q_1+\dots+q_m}}{m!} |g_{q_1}|^2 \dots |g_{q_m}|^2\, e^{-i(\Delta - (\omega_{q_1} + \dots + \omega_{q_m}))t}. \tag{1.18}$$

In this expression, we sum over all possible *m*-phonon scattering events contributing to the exciton auto-correlation function. It is important to notice that the distribution function for the exciton momenta, $f_{k+q_1+\dots+q_m}$, accounts for the phonon recoil effect as required by the momentum conservation, i.e., explicitly depends on $k + q_1 + \dots + q_m$. Therefore, $f_{k+q_1+\dots+q_m}$ is an important parameter defining relative contribution of each multi-phonon process.

Making the substitution of Eq. (1.18) into Eq. (1.10) and evaluating the Fourier integral, we find desired expression for the photon emission lineshape

$$S(\omega) = e^{-\sum_q |g_q|^2} \sum_{m=0}^{\infty} \sum_{q_1} \dots \sum_{q_m} \frac{f_{q_1+\dots+q_m}}{m!} \left|g_{q_1}\right|^2 \dots \left|g_{q_m}\right|^2 \frac{\gamma/\pi}{\left(\omega-\Delta+\omega_{q_1}+\dots+\omega_{q_m}\right)^2+\gamma^2}, \quad (1.19)$$

where the exciton dephasing rate γ is introduced. Notice that in accords with the discussion above, the exciton momentum distribution function $f_{q_1+\dots+q_m}$ defines spectral weight of each m-photon spectral replica. Eq. (1.19) significantly simplifies under the assumption that the exciton-phonon coupling, g_q, and optical phonon frequency, ω_q, have flat dispersion curve, i.e., independent of the momentum q. Respectively, denoting them as g and ω_{ph} we recast Eq. (1.19) to the form

$$S(\omega) = e^{-N|g|^2} \sum_{m=0}^{\infty} \frac{|g|^{2m}}{m!} \frac{A_m \gamma/\pi}{\left(\omega-\Delta+m\omega_{ph}\right)^2+\gamma^2}, \quad (1.20)$$

where N is the total number of sites in the lattice. In Eq. (1.20) each m-phonon lineshape function is weighted with the factor

$$A_m = \sum_{q_1} \dots \sum_{q_m} f_{q_1+\dots+q_m}. \quad (1.21)$$

A simple estimate for A_m can be performed assuming that the exciton momentum distribution is sharply peaked around zero, i.e., the exciton is spatially delocalized over all N sites of the lattice. In this case the distribution function can be approximated as $f_q = \left(\frac{2\pi}{a}\right)\delta(q)$, where a is the lattice period. Performing intergradation over the Brillouin zone in Eq. (1.21), we find that $A_m = N^m$. This result shows that the exciton-phonon coupling scales as $\sqrt{N}g$ allowing for the thermodynamic limit making application of our theory scalable finite size nano-structures to the bulk limits.

Fitting PL spectra using exciton-polaron Holstein model

To fit the experimental data, we generalize Eq. (1.21) with $A_m = N^m$ to the case of three independent vibration modes. The result is

$$S(\omega) = S_o e^{-g_1^2-g_2^2-g_3^2} \sum_{m_1=0}^{M_{max}} \sum_{m_2=0}^{M_{max}} \sum_{m_3=0}^{M_{max}} \frac{g_1^{2m_1} g_2^{2m_2} g_3^{2m_3}}{m_1! m_2! m_3!} \times \frac{\left(m_1\gamma_{ph1}+m_2\gamma_{ph2}+m_3\gamma_{ph3}+\gamma\right)}{\left(\omega-\Delta+m_1\omega_{ph1}+m_2\omega_{ph2}+m_3\omega_{ph3}\right)^2+\left(m_1\gamma_{ph1}+m_2\gamma_{ph2}+m_3\gamma_{ph3}+\gamma\right)^2}, \quad (1.23)$$

where for the sake of brevity we replaced $\sqrt{N}|g| \to g$ for the exciton-phonon coupling parameters and added the phonon line broadening γ_{ph}.

Results of the data fit with the three-mode model (1.23) are presented in Figure S9 and Table S1. The model reproduces well the main features of the experimental plot. Extracted values for the exciton-phonon coupling parameter, g, are all less than one suggesting relatively weak coupling regime which is consistent with the observed Poissonian as opposed to the Gaussian type (blue and gray lines) of the phonon replica progression.

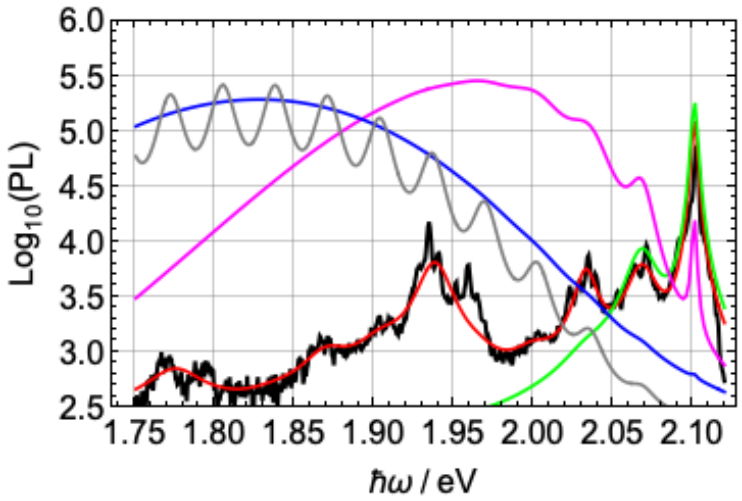

Figure S9. Holstein exciton-polaron model fits. Log plot of experimental PL spectra (solid black) from Figure 3a in the main text overlaid with line-shape fits (red line) using Eq. (1.23) derived using the adopted Holstein exciton-polaron model. Fit parameters are summarized in Table S1. Using the fit parameter for Phonon mode 1 (Tables S1), we calculated PL spectrum (green line) for the single phonon mode model. An increase of the exciton phonon coupling to $g_1 = 2.3$ and $g_1 = \sqrt{9.9} = 3.15$ in this model results in the PL lineshapes shown by the magenta and blue lines, respectively. Further changing the line broadening parameters to $\hbar\boldsymbol{\gamma_{ph}}$ = 0 meV and $\hbar\gamma = 10$ meV in the single mode calculation but retaining $g_1 = \sqrt{9.9} = 3.15$ allowed us to resolve the phonon mode replica (gray line) washed out in the blue line.

Table S1. Fit parameters. Parameters obtained from fitting PL shown on Figure S9 using Eq. (1.23) with $M_{max} = 10$. Zero phonon line is characterized by $\hbar\Delta = 2.1 \pm 2 \times 10^{-4}$ eV and $\hbar\gamma = 2.4 \pm 0.2$ meV. Obtained normalization perfector is $S_o = 1293 \pm 36$.

	$\hbar\boldsymbol{\omega_{ph}}$ / meV	$\hbar\boldsymbol{\gamma_{ph}}$ / meV	$\boldsymbol{g}$
Phonon mode 1	32.8 ± 0.7	7.4 ± 1.3	0.49 ± 0.03
Phonon mode 2	68.5 ± 0.7	4.8 ± 1.0	0.39 ± 0.02
Phonon mode 3	163.0 ± 0.6	8.8 ± 0.9	0.58 ± 0.01

S10. Charging model for spectral diffusion

The model has two essential parts. First, the inhomogeneous broadening of the fluorescence spectrum is assumed to be caused by fluctuations in the exciton binding energy. These fluctuations are proportional to fluctuations in the intralayer electric field through the first-order (linear) Stark shift. Second, electric field fluctuations are caused by sudden shifts of the charge density in the nanoflake, which arise from interlayer charge transfer due to hopping of hot electrons and holes just after laser excitation and before thermal relaxation.

Linear Stark Shift

Let us assume that the stochastic jumping of the exciton emission spectrum, with periodic frequency shifts in multiples of ~1 meV, is due to fluctuations in the Stark shift of the excitons confined to two-dimensional layers due to charge transfer between adjacent layers. In first order perturbation, the energy shift

$$\delta = -\vec{\mu} \cdot \vec{E}\,, \tag{2.1}$$

associated with exciton emission due to an electric field E depends on the electric dipole operator μ. For an electric dipole moment $\mu \sim 10D$, a shift of 1 meV requires fields as high as 10^5 V/cm.

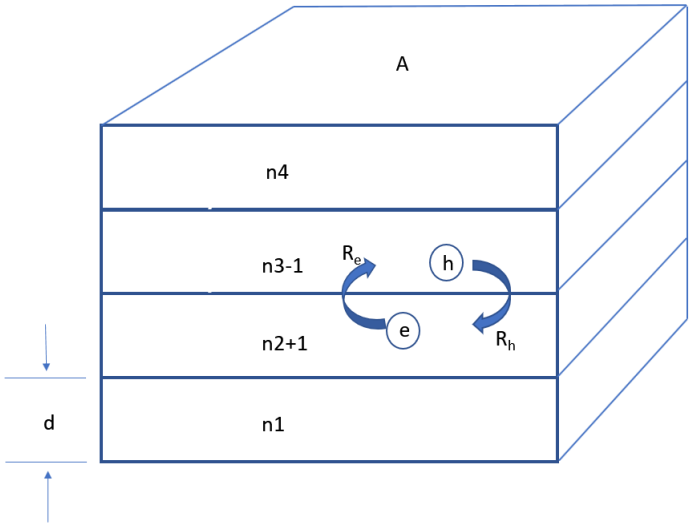

Figure S10. Parallel capacitor model. Shown here is schematic for a 4-layer nanoflake, modeled as a set of 4 parallel conducting layers with area A and layer periodicity d. The i^{th} layer has charge $q_i = n_i e$ which is an integral multiple of the electronic charge, with positive and negative integer n_i corresponding to whether the excess charge is comprised of holes or electrons, respectively. A change in state occurs when either the hole or the electron of an energetic electron-hole pair, created by the laser in any one of the 4 layers, hops to an adjacent layer before complete thermal relaxation can occur. In this illustration, we assume that the transfer occurs between layers 2 and 3, with $n_2 \to n_2 + 1$, and $n_3 \to n_3 - 1$. This can happen with the transfer of an electron from layer 2 to layer 3, or with the transfer of a hole from layer 3 to layer 2.

To first approximation, consider the nanoflake to act as a capacitor, with each layer serving as a uniformly charged plate. In such a case, the electric field,

$$E = \frac{e}{\kappa \varepsilon_0 A} \sim 10^5 \text{V/cm}\,, \tag{2.2}$$

due to single-electron charging is the right order of magnitude for a 1 meV Stark shift if the product $\boldsymbol{\kappa A \simeq 2 \times 10^3 \text{nm}^2}$. For a dielectric constant $\boldsymbol{\kappa \simeq 5}$ in agreement with calculations for bulk $NiPS_3$,[1] this corresponds to an area $\boldsymbol{A \simeq 400 \text{nm}^2}$ suggesting lateral dimensions on the order of 20 nm. This is within reasonable agreement of AFM measurements reported in Figure 1b.

Charge Hopping Rates

Introduction of a rate model for the hopping of electrons and holes sharpens this discussion. Consider a nanoflake consisting of N parallel conducting layers, each with an area A, separated from one another with a thickness d. Figure S10 shows a schematic of a four-layer structure as an illustration. The i^{th} layer has charge $q_i = n_i e$, which is an integral multiple of the electronic charge, with positive and negative integer n_i corresponding to whether excess charge is comprised of holes or electrons, respectively. Thus, the charge state of the nanoflake is determined by the set of integers $\{n_i\}$. A change in state occurs when either the hole or the electron of a laser-produced energetic electron-hole pair, created in any one of the layers, hops to an adjacent layer before thermal relaxation can occur. Figure S10 illustrates the case where charge transfer occurs between layers 2 and 3, with $n_2 \to n_2 + 1$, and $n_3 \to n_3 - 1$. The rate

$$R_{f,i} = R \begin{cases} \exp\left(-\left(U_f - U_i\right)/kT\right) & ;\text{U}_f > U_i \\ 1 & ;\text{U}_f < U_i \end{cases} \tag{2.3}$$

for changing from an initial state $\{n_1, n_2, n_3, n_4\}_i$ to a final state $\{n_1, n_2 + 1, n_3 - 1, n_4\}_f$ is given by the sum $R = R_e + R_h$ of electron and hole hopping rates. The Boltzmann factor $\exp\left(-\left(U_f - U_i\right)/kT\right)$ expresses detailed balance between initial and final states with energies U_i and U_f respectively. After excitation, an electron hole pair will immediately begin thermalizing to a temperature of 10 K (~1 meV), after which hops between layers are frozen out. Thus kT in Eq. (2.3) is an effective thermal energy for rare charge-transfer events, expected to lie somewhere between 1 meV and 200 meV, the difference $h\nu - \varepsilon_g$ between the 2.4 eV laser excitation energy and the 2.2 eV band gap. The initial and final charge-state energies can be determined by the spatial integration of the electrical energy density $\frac{1}{2}\varepsilon E^2$. If the charge is confined to thin conducting planes in each layer, the energy

$$U\left(\{n_i\}\right) = \frac{1}{2}\kappa \varepsilon_0 A d \sum_{j=1}^{N-1} E_j^2 = \frac{e^2}{2C} \sum_{j=1}^{N-1} \left(\sum_{i=1}^{j} n_i \right)^2 , \tag{2.4}$$

can be written as a sum over the square of the electric field penetrating the potential-barrier regions separating each of the planes. For a nanoflake with N layers, there are N-1 such regions. In writing

the second equality in Eq. (2.4) we have used Gauss's law to express the electric field $E_j = e\sum_{i=1}^{j} n_i / \kappa\varepsilon_0 A$ in the j^{th} barrier in terms of the charge $e\sum_{i=1}^{j} n_i$ distributed in the conducting layers preceding it. We have also introduced the capacitance of a single layer,

$$C = \frac{\kappa\varepsilon_0 A}{d} \simeq 4 \times 10^{-17} F \,, \tag{2.5}$$

for a typical interlayer distance $d \sim 0.5\text{nm}$. Thus, the energy change between initial and final states will be quantized in multiples of $e^2/2\,C \simeq 2\text{meV}$.

Kinetic Monte Carlo

Spectral diffusion can be simulated by following the charge state of the nanoflake in time in a Monte Carlo simulation of a continuous-time random walk.[2] For an *N*-layer system, a given charge state,

$$\{n_1, n_2, n_3, n_4, \ldots n_N\} \rightarrow \begin{matrix} \{n_1 \pm 1, n_2 \mp 1, n_3, n_4, \ldots n_{N-1}, n_N\} \\ \{n_1, n_2 \pm 1, n_3 \mp 1, n_4, \ldots n_{N-1}, n_N\} \\ \{n_1, n_2, n_3 \pm 1, n_4 \mp 1, \ldots n_{N-1}, n_N\} \\ \ldots \\ \{n_1, n_2, n_3, n_4, \ldots n_{N-1} \pm 1, n_N \mp 1\} \end{matrix} \tag{2.6}$$

is connected to 2(*N*-1) neighboring states enumerated here by hopping rates $R_{f,i}$ of the form given in Eq. (2.3). For each step of the random walk, the initial energy $U_i\left(\{n\}\right)$ associated with presently occupied state, and the final energy $U_f\left(\{n\}\right)$ for each of the neighbors, are used to calculate 2(*N*-1) different values of the hopping rate Eq. (2.3) . The final state is then determined from a uniform distribution according to its relative weight, $w_{f,i} = R_{f,i} / \sum_{f'} R_{f,i}$. At the same time, a dwell time associated with the initial state is selected from an exponential distribution of dwell times,

$$\psi_i(t) = \frac{1}{\tau_i} e^{-t/\tau_i} \,, \tag{2.7}$$

where the lifetime $\tau_i = \left(\sum_f R_{f,i}\right)^{-1}$ is the reciprocal of the sum of all 2(*N*-1) rates. In Figure S11, we follow the charging effect on the Stark shift in an *N*-layer system in time for a single random walk, to qualitatively compare results of the simulation to the observed temporal evolution of the spectral jumping. We also make a time-weighted histogram of results, Figure S12, to determine the steady-state probability for an exciton to emit with a particular Stark shift.

Electric Field within a Layer

Explicit calculations of the Stark shift (Figure S11 and Figure S12) require knowledge of average electric field applied to the exciton states in each layer. An exciton embedded in a 2-D charged plane is subjected to the average,

$$E_{j(\text{layer})} = \frac{e}{2\kappa\varepsilon_0 A}\left(\sum_{i=1}^{j-1} n_i + \sum_{i=1}^{j} n_i\right) = \frac{e}{2\kappa\varepsilon_0 A}\left(2\sum_{i=1}^{j} n_i - n_j\right), \qquad (2.8)$$

of the electric field immediately above and below the plane. According to (2.8), if there is no charge in the j^{th} layer, i.e., if $n_j = 0$, then the field in the j^{th} layer is an even multiple of $e/2\kappa\varepsilon_0 A$. If $n_j \neq 0$, then the field in the j-th layer will be an even multiple of $e/2\kappa\varepsilon_0 A$ for even n_j, and an odd multiple of $e/2\kappa\varepsilon_0 A$ for odd n_j. If charge residing within a layer quenches exciton emission such that only uncharged layers participate in the spectrum, the Stokes shift will be quantized in multiples of $e\mu/\kappa\varepsilon_0 A \sim 2\text{meV}$. If charge residing with a layer has no quenching effect, then the Stark shift will be quantized in multiples of $e\mu/2\kappa\varepsilon_0 A \sim 1\text{meV}$. Partial quenching gives a superposition of the two intervals, in agreement with observation. For the plots in Figure S11 and Figure S12, we have introduced partial quenching by allowing layers with no charge and/or with one excess charge to be four times more likely to emit photons than any of the other charge states.

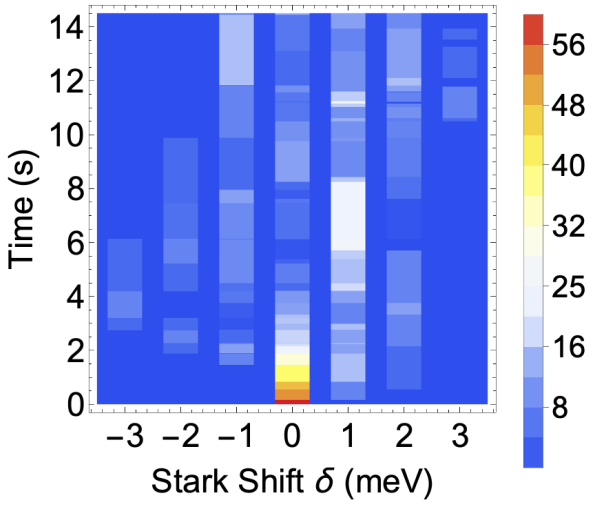

Figure S11. **Spectral diffusion.** Shown is an illustration of spectral diffusion versus time, following the evolution of the continuous time random walk for 50 hops for a nanoflake with 15 layers**.** The horizontal axis displays the Stark shift δ, and time is displayed on the vertical axis. The intensity of emission is indicated by the color temperature on a scale from 0 (blue) to a maximum of 60 (red). Abrupt changes in color indicate hops. Initially δ =0, none of the layers has yet to acquire excess charge, and the intensity maximum. After 50 hops, δ ranges from -3 meV to 3 meV. The dispersion in δ is controlled by a single dimensionless parameter $(e^2/2C)/kT = 0.02$. For $e^2/2C = 2\,\text{meV}$ this corresponds to $kT \simeq 0.1\text{eV}$. The Stark shift is quantized in multiples of $\mu e/2\kappa\varepsilon_0 A = 1\,\text{meV}$, while the length of the time-axis determined by the hopping rate. Here $R = 0.1\text{s}^{-1}$.

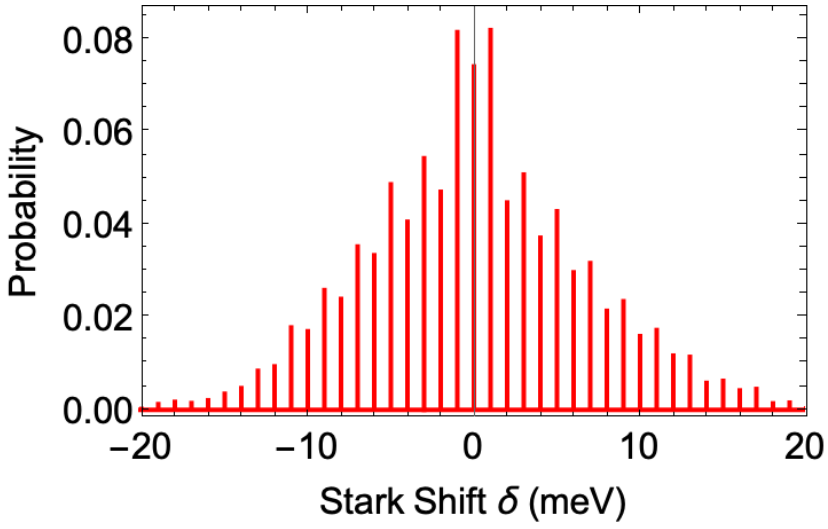

Figure S12. Stark shift comb. The figure shows the steady-state probability to observe different values of δ for a 15-layer nanoflake. The same parameter values were used for Figure S11. The histogram displays the time-weighted spectrum for 20,000 hops. The spectrum shifts in multiples of 1.0 meV. We have taken the probability for radiative decay to be 4 times more likely in layers with no charge, $n = 0$, and/or one charge, $n = \pm 1$, and this weighting gives rise to the alternating pattern shown here. Correlation between layers having few charges and layers having low electric fields causes an enhanced probability in the neighborhood of $\delta \simeq 0$.

Fine Structure of PL Spectrum

The simplest model for the emission spectrum is to assume linear exciton-phonon coupling to a single dispersionless optical phonon band with frequency ω_0.[3] Neglecting gauge terms generated in the Lang-Firsov transformation, the low-temperature spectrum (follows from the Holstein Model in S1 with infinitesimal line widths),

$$S(\omega) = e^{-g^2} \sum_{m=0}^{\infty} \frac{g^{2m}}{m!} \delta\left(\omega - \Delta + m\omega_0\right), \tag{2.9}$$

is a set of equally spaced lines, beginning with the 0-0 line at $\omega = \Delta$, with successively lower energy phonons in multiples of ω_0 weighted by a Poissonian distribution determined by the square of the dimensionless coupling constant $g = \delta x / \sqrt{2} x_0$ measuring the oscillator displacement δx in multiples of the zero point amplitude x_0.[3] Let us assume that the exciton is linearly coupled to three optical phonons, with energies $\hbar\omega_1 = 31\,\text{meV}$, $\hbar\omega_2 = 68\,\text{meV}$, and $\hbar\omega_3 = 168\,\text{meV}$. (The phonon energies are slightly tweaked from the values obtained from the low-resolution PL spectra fit in Table S1.) A line spectrum can be constructed that reproduces most of the experimental features in Figure 3a (except for the P4 line) by choosing coupling constants $g_1 = 0.49$, $g_2 = 0.38$

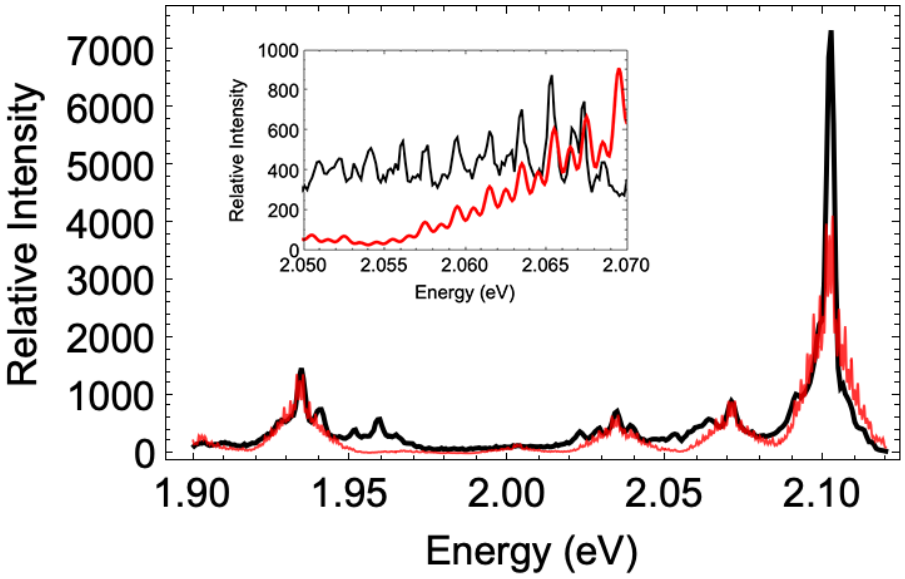

Figure S13. Spectrum with three commensurate phonons. Shown is a calculated spectrum in which the exciton is weakly coupled to three different phonons, with frequencies that are approximately multiples of 33 meV, as described in the text. This has been superimposed on the measured spectrum shown in Figure 2a, with alignment of the 0-0 lines at 2.102 eV. The only source of broadening in the calculated spectrum is the inhomogeneous broadening provided by the different electrical environments in the different layers, and this has been incorporated by convoluting the line spectra with the Stark shift comb shown in Figure S12. A magnification (inset) of the calculated spectrum (red) reveals the periodicity of the comb. The energy spacing between high peaks is 2 meV, while the energy spacing from low peak to high peak 1 meV. The alternating structure is similar to that seen in the measured spectrum at high resolution (black).

, and $g_3 = 0.58$, respectively. The Stark shift comb described above and shown in Figure S11 has been convoluted to provide inhomogeneous broadening. The smoothed histogram was made with Gaussian bins having standard deviation 0.35 meV. The result of the calculation is shown in Figure S13 (red line) demonstrating reasonable agreement with the experimental data (black line).

S11. First-principles calculations of excitons in bulk $NiPS_3$

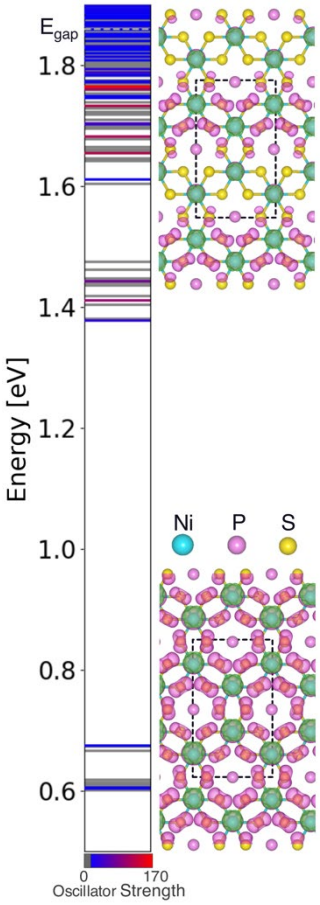

Figure S14. Exciton binding characters in Bulk $NiPS_3$. The exciton state energy levels for $\boldsymbol{Q} = 0$, where the color goes as the oscillator strength associated with a given exciton; see colorbar. The electron (light-green shading) and hole (pink shading) wave functions are drawn within the $NiPS_3$ unit cell (black dashed line) for both high and low energy electron-hole pairs. For low energy excitons, the hole density (pink shading) is spread uniformly over nickel, sulfur, and phosphorous atomic sites. In contrast, the electron wave function (green shading) is tightly localized to the nickel atomic centers. As the exciton energy increases, the hole density becomes more anisotropic, with the majority of its weight on the opposite magnetic sublattice. Concomitantly, the electron density gradually shifts weight between sublattices. We note that parts of the hole wave function are not visible for the low energy excitons since the electron and hole wavefunctions are co-localized (overlap) on the same atomic sites.

S12. No discernible spectral jumps in CVT-Bulk SO-X exciton

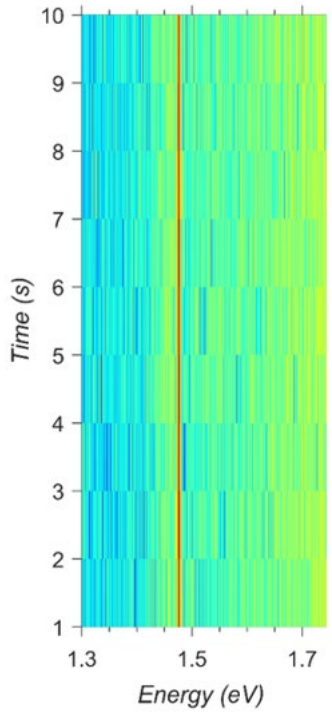

Figure S15. PL intensity-time trace in a CVT-Bulk. PL from a CVT-Bulk collected every 1s at 8 K. No discernible spectral jumps noticed for SO-X. Our charging model for nanoflakes predicts the area dependence of field dependence and applying that to the micrometer sized CVT-Bulk would only result in a non-resolvable minute jumps.

S13. Excitation energy dependence of PL peaks in a nanoflake

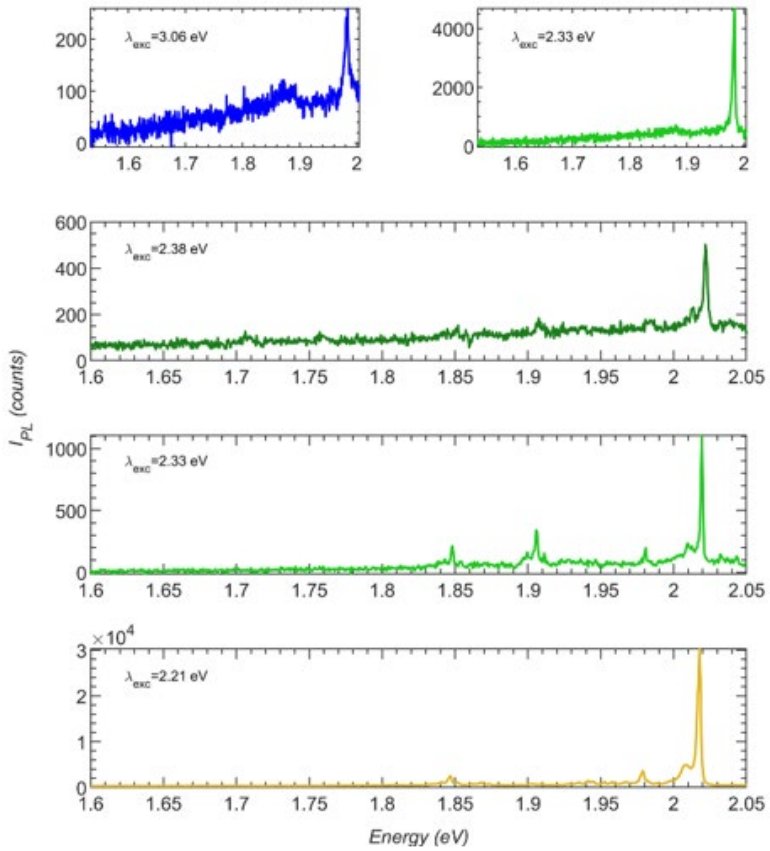

Figure S16. Excitation laser energy dependence of PL in a nanoflake. PL collected with different excitation laser energies under confocal excitation with a diffraction limited spot size for a nanoflake $NiPS_3$ with the same power density of about 70 W/cm^2. Two datasets from two different nanoflakes are shown. The topmost dataset shows a nanoflake with PL collected with excitation energies of 3.06 eV and 2.33 eV. The bottom three spectra are collected from an another nanoflake at specified excitation energies. The PL peak energies are not shifting with the excitation energies ruling out the Raman lines. All data are collected at 10 K.

S14. Multiple PL emissions from same nanoflake

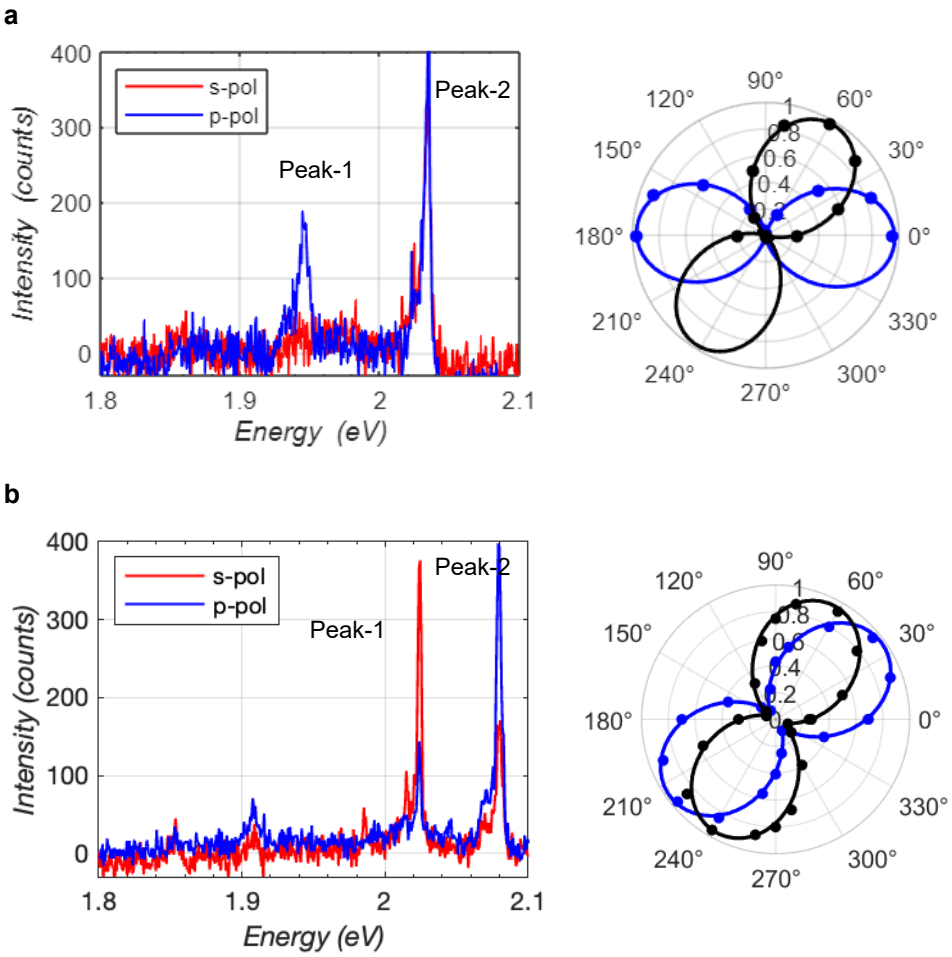

Figure S17. Multiple emitting centers in same nanoflake. (a,b) Two examples for PL from same nanoflake showing 2 PL peaks. The right-side polar plots show the 2 PL peaks are oriented different from each other. We attribute some nanoflakes could have multiple grains or that different edges could have resulted in this behavior.

S15. PL experimental methods

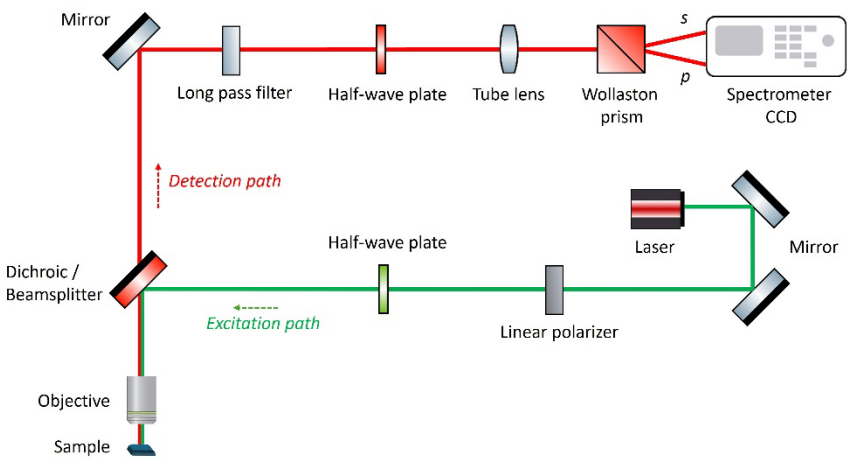

Figure S18. PL experimental setup. PL excitation and detection scheme is shown.

A continuous wave solid state laser with an energy of 2.33 eV is used for sample excitation. The laser is reflected through a dichroic beamsplitter and then focused onto the sample (placed in Oxford MicrostatHires liquid Helium flow cryostat) to a diffraction limited spot size using a 50×, 0.7 NA Olympus objective microscope, which is used to both excite the sample and collect the PL. Collected photons after passing through the same dichroic beamsplitter and a 545 nm long-pass filter go to a spectrometer + LN-cooled CCD combo (Acton SP2300i/Pylon 100BR or Acton SP2500i/Pylon 2KB) with 300 g/mm or 1200 g/mm or 2400 g/mm gratings.

For polarization-resolved spectroscopy, a non-polarizing beamsplitter is used in place of dichroic beamsplitter and a Wollaston prism is placed just before the spectrometer + CCD combo to split the signal into *s* and *p* polarized components. To investigate the PL polarization, a half-wave plate is used as an analyzer either in the detection path after the beamsplitter or the excitation path before the beamsplitter. For every rotation angle of the half-wave plate analyzer either in the detection path or the excitation path, the PL signal containing *s* and *p* polarized components are recorded in the spectrometer + CCD. The data are post-processed and displayed as polar plots in the figures.

References:

1 Lane, C.; Zhu, J.-X. Thickness dependence of electronic structure and optical properties of a correlated van der Waals antiferromagnetic $NiPS_3$ thin film. *Phys. Rev. B* **2020**, *102* (7), 075124.

2 Young, W. M.; Elcock, E. W. Monte Carlo studies of vacancy migration in binary ordered alloys: I. *Proc. Phys. Soc.* **1966**, *89* (3), 735-746.

3 Huang, K.; Rhys, A.; Mott, N. F. Theory of light absorption and non-radiative transitions in F-centres. *Proc. R. Soc. Lond. A.* **1950**, *204* (1078), 406-423.