Supplementary Information

for

Flat-band-induced many-body interactions and exciton complexes in a layered semiconductor

Gabriele Pasquale[1,2*], Zhe Sun[1,2*†], Kristiāns Čerņevičs[3], Raul Perea-Causin[5], Fedele Tagarelli[1,2], Kenji Watanabe[6], Takashi Taniguchi[7], Ermin Malic[4,5], Oleg V. Yazyev[3], Andras Kis[1,2†]

[1]*Institute of Electrical and Microengineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[2]*Institute of Materials Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[3]*Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[4]*Philipps-Universität Marburg, Department of Physics, Renthof 7, D-35032 Marburg, Germany*
[5]*Chalmers University of Technology, Department of Physics, 412 96 Gothenburg, Sweden*
[6]*Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan*
[7]*International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan*

***** *These authors contribute equally to this work.*
[†] *Correspondence should be addressed to: Zhe Sun (sunacv@bc.edu) and Andras Kis (andras.kis@epfl.ch)*

1. First-principles calculations of band diagram for a pristine InSe slab and InSe with a selenium vacancy

One of the goals in this section is to calculate the position of defect energy levels in a multilayer InSe. First-principles calculations of a 6L-InSe slab with a selenium (Se) vacancy utilizing either hybrid or meta-GGA functionals to obtain correct band gap are computationally very expensive due the large size of the supercell. Therefore, as a representative example, we consider the Se vacancy in the thinnest multi-layer system, 2L-InSe.

Figure S1a & b depicts a calculated band diagram and density of states (DOS) for a pristine 2L-InSe slab. The valence band maximum (VBM) is about 30 meV higher than the Γ point, in good agreement with literature[1]. The Mexican-hat dispersion induces a peak in DOS, referred

to as the van Hove singularity[2], as shown in Figure S1b. The energy offset between the VBM and the valence band energy at the Γ point can be regarded as the bandwidth of the flat band dispersion. As the thickness increases, the bandwidth decreases. When the thickness of InSe is higher than seven layers, the Mexican-hat dispersion almost vanishes[1]. In experiments, we use a 6L-InSe whose VBM is only about 5 meV higher than the Γ point in the valence band (Figure 1b).

To calculate the position of defect energy levels, we applied the modified Becke-Johnson (MBJ) functional on a 4×4 supercell with a Se vacancy located on the top atomic surface.

Figure S1c shows the bands induced by localized states are about 100-150 meV lower than the conduction band minimum. This is consistent with the energy difference between the D_0 and X_0' emissions. We argue that our calculation results for a 2L-InSe can be extended to the case of 6 layers, since we assume the Se vacancy is on the top atomic surface.

We extract the effective mass of electrons and holes from the calculated band diagram by fitting it to a parabola. Due to the band inversion near the VBM, the effective mass of holes in thin InSe is heavier and more complicated. For example, an electronlike effective mass at the Γ point with $|m_{h,\Gamma}| \approx 0.9\, m_0$ and a holelike effective mass at the VBM with $|m_{h,VBM}| \approx 1.9\, m_0$ is predicted for InSe monolayer[3]. Since the radiative recombination of excitons is more related to the holes near the band edge, we use the value of the holelike effective mass at the VBM. Figure S1d presents the effective mass as a function of the layer number. The results are in good agreement with literature[3,4].

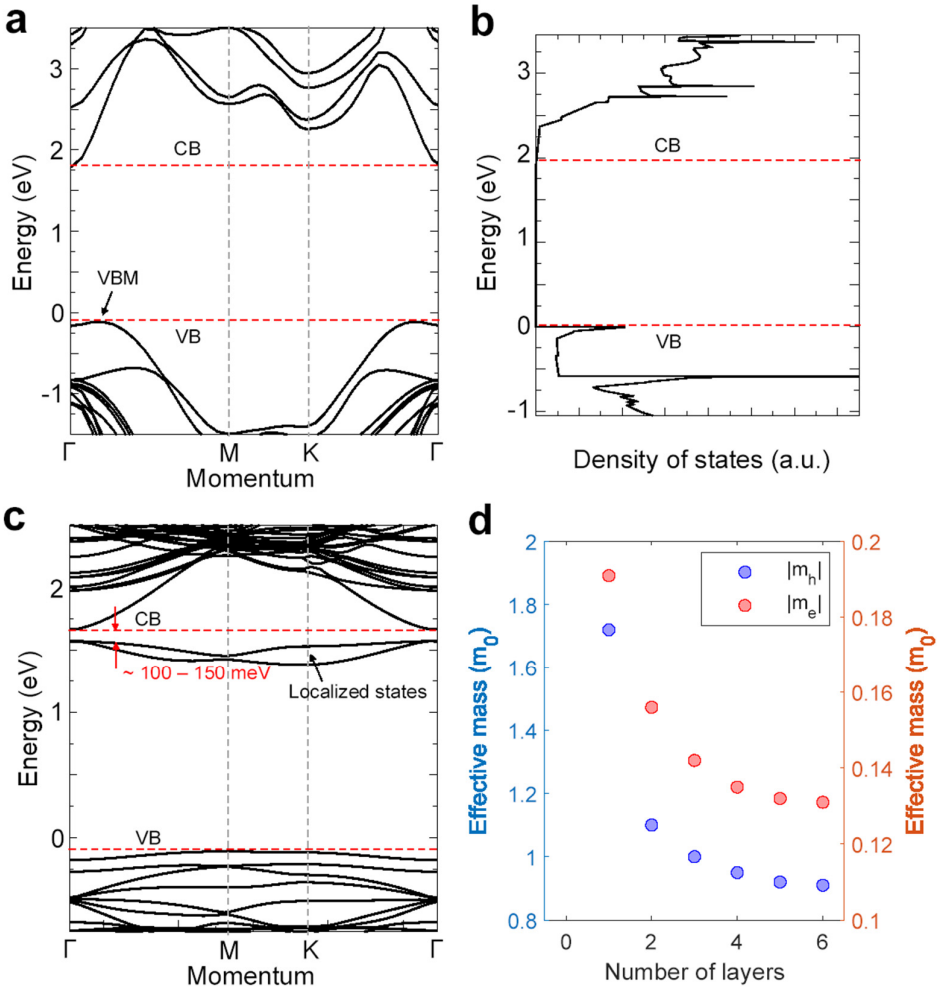

Figure S1. (a) Calculated band diagram for a pristine 2L-InSe slab. CB: conduction band; VB: valence band; VBM: valence band maximum. The Mexican-hat dispersion appears near the Γ point. (b) Calculated density of states (DOS) for a pristine 2L-InSe slab. The sharp feature near the VBM in the DOS spectra is a van Hove singularity. (c) Calculated band diagram for 2L-InSe slab with a Selenium (Se) vacancy located on the top atomic surface. (d) Effective mass of electrons and holes as a function of the layer number, extracted from the band diagram calculations. Red: electron mass; blue: hole mass.

2. Experimental setup

All optical measurements are performed in a helium flow cryostat at 4.5 K in a setup schematically presented in Figure S2. A confocal microscope is used to excite excitons optically and collect the emitted photons through the same objective with a working distance (WD) of ~ 4.5 mm and NA = 0.65. The excitation light sources are coupled into a single-mode fiber linked to the excitation arm of the confocal microscope setup. Long-pass (LP) filters are located at the collection arm in front of a spectrometer or avalanche photodiode (APD) to block the excitation laser. A charge-coupled device (CCD) camera is used to image the surface of the heterostructure.

In the time-resolved PL (TRPL) measurements, an APD (Excelitas Technologies, SPCM-AQRH-16) shielded from stray light is used to collect the PL photons. The output of the APD is connected to a time-correlated photon counting module (TCPCM) with a resolution of 12 ps r.m.s. (PicoQuant, PicoHarp 300), which measures the arrival time of each photon. The single-photon timing resolution of the APD is about ~ 300 ps, which is the time limitation for the whole setup. In TRPL measurements, a tunable Ti:Sapphire laser (Coherent Chameleon), which generates sub-picosecond (ps) pulses with a repetition rate of 80 MHz, is used to excite the sample. The wavelength is set to 720 nm.

In the PL excitation (PLE) measurements, a narrow-linewidth tunable continuous-wave laser (MSquared) is used. A long pass 900 nm (LP900) filter is inserted in the collection arm to monitor the intensity from the D_0 peak.

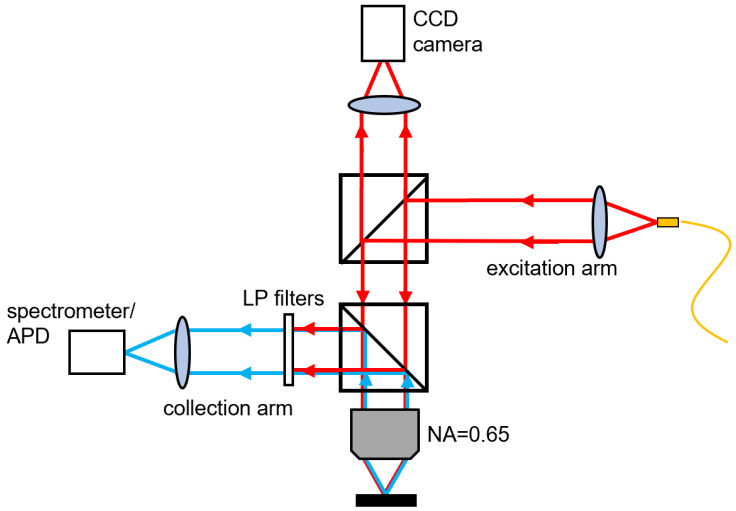

Figure S2. Sketch of the confocal microscope setup. Red: excitation laser and laser reflection; blue: PL photons.

3. Photoluminescence linewidth

Figure S3 summarizes the PL peak energy and linewidth of InSe excitons in literature. In our work, the exciton peak at 1.48 eV has a linewidth of 15 meV, which so far is among the narrowest for excitons in a thin InSe. The narrow linewidth allows us to resolve previously unresolved spectral features.

Reference	Thickness/Number of layers	Linewidth (meV)	Peak Energy (eV)	Temperature (K)
1) Balakrishnan *et al.*, 2D Mater. 4 025043. (2017)	24 nm 9 nm 5 nm	~ 50 ~ 50 ~ 100	1.25 1.32 1.47	300 300 300
2) Song *et al.*, ACS Appl. Mater. Interfaces 10, 3994 (2018)	bulk 3L to 8L	~ 50 ~ 60 to > 100	1.25 1.3 to ~ 1.7	300
3) Bandurin *et al.*, Nature Nano. 12, 223-227 (2017)	bulk 1L to 5L	100 ~ 80 to > 300	1.25 1.48 to ~ 3	300
4) Sánchez-Royo *et al.*, Nano Research 7, 1556-1568 (2014)	2 nm to 12 nm	~ 50 to ~ 100	1.27 to 1.44	4.2
5) Mudd *et al.*, Adv. Mater., 25, 5714-5718. (2013)	6.5 nm to 19.5 nm	~ 60 to > 100	1.27 to ~ 1.48	300
6) Ubrig *et al.*, Nature Materials 19, 299–304 (2020)	2L	~ 100	1.7 to 1.9	5
7) Shubina *et al.*, Nature Comm 10, 3479 (2019)	bulk	~ 10	1.3	10
8) Zultak *et al.*, Nature Comm. 11, 125 (2020)	5L to 7L	~ 30 to ~ 50	1.38 to ~ 1.5	4.2
9) Venanzi *et al.*, Phys. Rev. Mat. 4 044001 (2020)	9L	~ 20	1.42	4
10) Henck *et al.*, arxiv: 2201.01264 (2022)	5L	7.5	1.47	5
Our work	6L	15	1.48	4.5

Figure S3. Summary of PL peak energy and linewidth of InSe excitons in literature.

4. Device A: reproducibility of spectral features

Figure S4b presents a spatial map of PL intensity for device A. In the middle of device A, the PL intensity is quenched due to the graphene contact on the flake. In Figure S4c-e, we plot the PL intensity, acquired on position A, B & C, as a function of the emission energy and gate voltage using $P = 50$ μW. All the features presented in Figure 2a are reproducible. In addition, the redshift energy in the *p*-doped regime as a function of the Fermi energy shown in Figure 5a is extracted from Figure S4c-e.

Figure S4f shows PL emission spectra under gate voltages V_g from -5 V to 5 V using a laser power $P = 50$ μW, measured on position A. In the *n*-doped regime ($V_g > 0$ V), the peak energy of X_- almost remains constant before it vanishes, contrary to the strong redshift observed in the *p*-doped regime ($V_g < -5.8$ V).

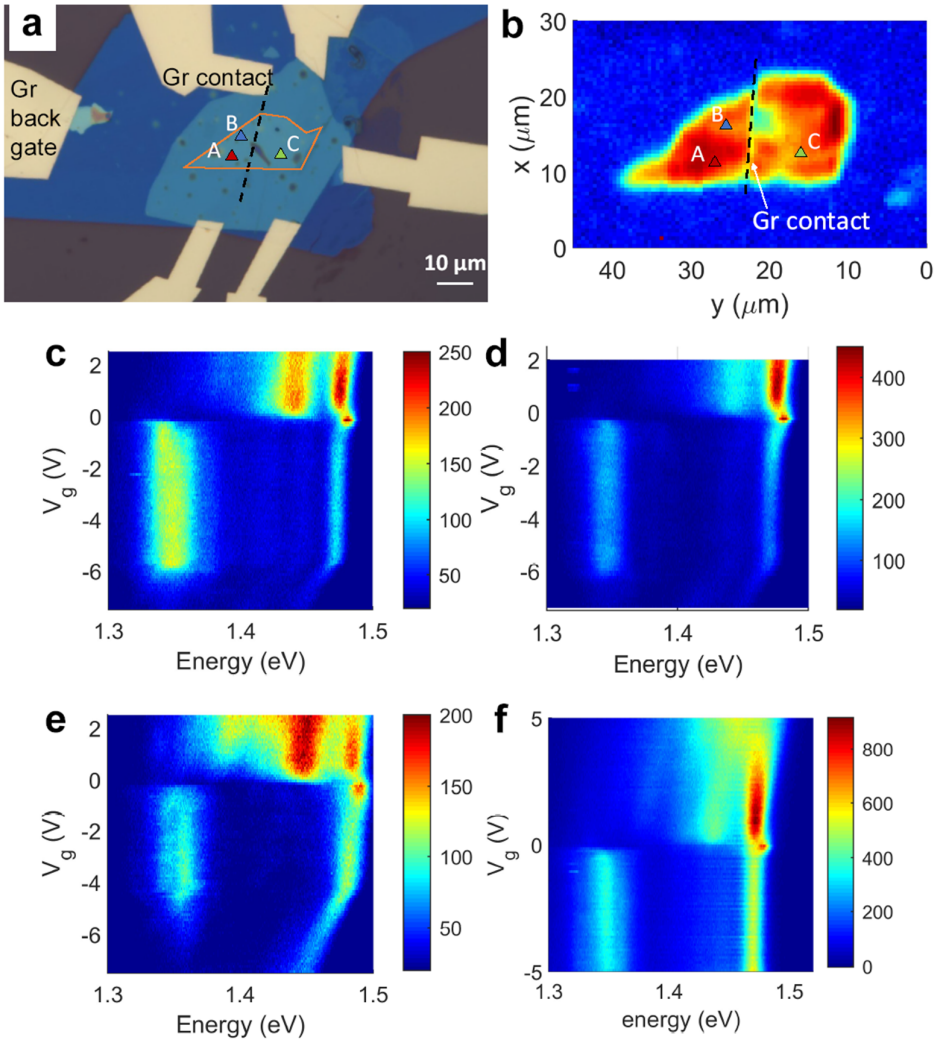

Figure S4. (a) Optical microscope image of device A. The orange contour encloses the region of InSe flake. The dashed line indicates the position of the few-layer graphene contact on the flake. Scale bar, 10 µm. (b) Spatial map of PL intensity for device A at $V_g = 0$ V. (c), (d) & (e) PL count rate as a function of the emission energy and gate voltage using $P = 50$ µW, measured on position A, B & C. (f) PL emission spectra under gate voltages V_g from -5 V to 5 V using a laser power $P = 50$ µW, measured on position A.

5. Device B: electrical transport and photoluminescence

To correlate the spectral features with the charge configurations, we fabricated another device (device B) with two graphene contacts on the flake to measure its d.c. electrical responses. Our device B consists of a 7L.-InSe flake separated from the graphene bottom gate by a hBN spacer of 20 nm. We observed similar spectral features by performing PL spectroscopy on device B, as shown in Figure S5b. Figure S5a exhibits the source-drain current (I_{sd}) as a function of the source-drain bias (V_{SD}) and gate voltage (V_g). We identify that the InSe flake is in the n-doped regime when $V_g > -0.4\,V$ and the p-doped regime when $V_g < -3.5\,V$. By comparing PL

spectrums from devices A and B, we conclude that the D_0 emission only appear in the undoped regime, and the exciton peak tends to redshift as the sample enters the p-doped regime. We notice that the p-type current is about five orders of magnitude lower than the n-type current, mainly due to the large effective mass of holes in the valence band.

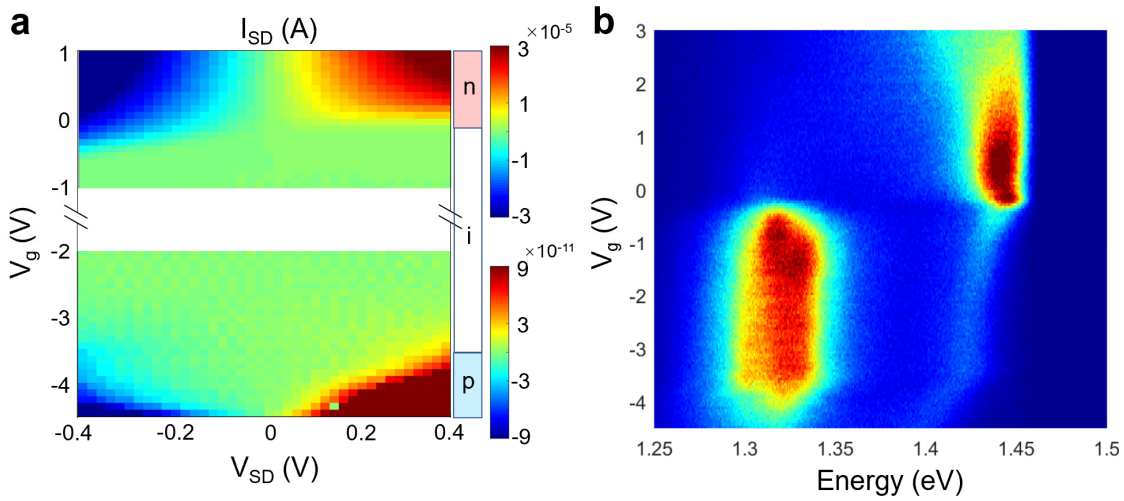

Figure S5. (a) Source-drain current I_{SD} as a function of the source-drain bias V_{SD} and gate voltage V_g. To show the n- and p-doped regime clearly, the upper and lower part of the 2D scan is plotted using two different color scales. (b) PL intensity as a function of the emission energy and gate voltage V_g using P = 10 μW.

6. Out-of-plane photocurrent - an alternative approach for determining charge configurations

We also acquired the out-of-plane photocurrent (I_g in Figure 1c) as a function of V_g, at the same time when we performed the V_g-dependent PL measurements. The red curve in Figure S6a, for instance, shows that the $I_g - V_g$ relation has three regimes. For $V_g > 0\ V$, I_g is almost the same as the value in the absence of laser (black curve in Figure S6a). I_g is enhanced when $-5.8\ \mathrm{V} < V_g < 0\ V$, which corresponds to the undoped regime. As the gate voltage is tuned to $V_g < -5.8\ \mathrm{V}$, I_g starts to increase with a higher slope. The three stages of I_g match perfectly with the three charging regimes determined by in-plane electrical transport and PL measurements. This is also true for the data acquired using both devices A and B. We conclude that the out-of-plane photocurrent is an alternative to determine charge configurations. The out-of-plane photocurrent in the undoped and p-doped regime is characterized by different power dependences, as shown in Figure S6b. In the p-doped regime ($V_g = -7\ V$), I_g grows linearly with power, while I_g in the undoped regime ($V_g = -3\ V$) increases sublinearly.

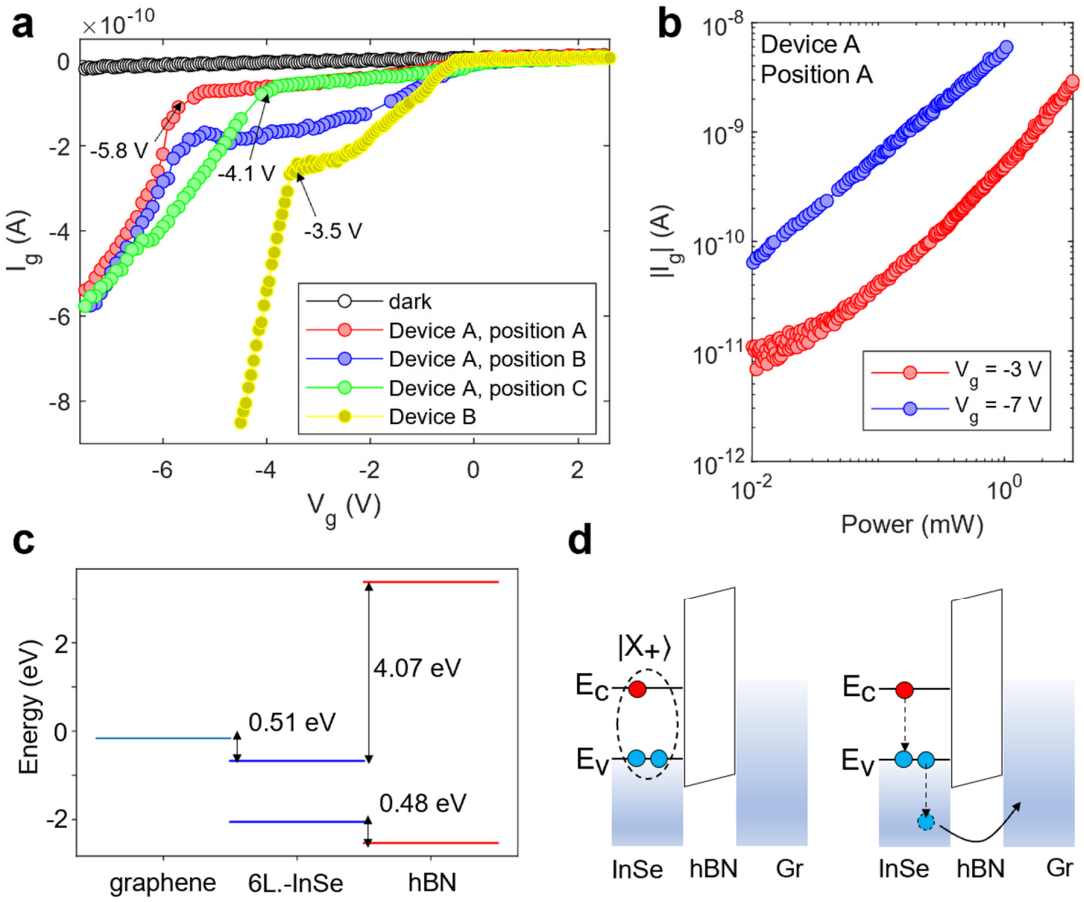

Figure S6. (a) Out-of-plane photocurrent I_g measured together with the gate voltage dependent PL measurements in Figure S4 and Figure S5. Red: Figure S4c; blue: Figure S4d; green: Figure S4e; yellow: Figure S5b; black: I_g in the absence of laser. (b) I_g as a function of excitation power for $V_g = -3$ V (red) and $V_g = -7$ V (blue). (c) Calculated band alignment between a 6L-InSe and hBN. (d) Schematic of the out-of-plane photocurrent generation. Red ball: electron; blue ball: hole.

The enhanced out-of-plane photocurrent in the *p*-doped regime has been observed in a monolayer WSe_2 embedded in a field-effect structure[5], and can be interpreted as the consequence of the Auger recombination of a single exciton. An exciton has a probability to recombine non-radiatively, leading to an energy transfer to a free carrier. Suppose the band offset between the semiconductor and hBN is much smaller than the exciton energy (~1.48 eV). In this case, it is energetically possible for the free carriers to go through the hBN spacer forming an out-of-plane current flow. In Figure S6c, our first-principles calculation of band alignment between a 6L-InSe and hBN indeed shows a slight offset in the valence band (~0.48 eV). In the *p*-doped regime, positively charged trions are formed. Auger recombination of one electron and hole creates energy that can excite the additional hole to become a hot carrier (Figure S6d). However, this process does not happen in the *n*-doped regime because of the

large band offset in the conduction band (~ 4.07 eV). One evidence of this mechanism is the linear power dependence of I_g in the p-doped regime, as the process involves only one exciton. The out-of-plane photocurrent in the undoped regime, which does not strictly satisfy the linear power dependence, might be related to the exciton-exciton annihilation associated with ionized donors[6]. The mechanism of the out-of-plane photocurrent is out of the scope of this work and requires further discussions in the future.

7. Time-resolved photoluminescence

To determine the lifetime of the D_0 (~ 920 nm) and X_0' (~ 845 nm) emissions at $V_g = -3$ V, besides inserting a LP800 filter in the collection arm, we use either a LP900 or a SP900 filter to select the PL photons further. Figure S7a presents normalized PL intensity of the D_0 (red circles) and X_0' (blue circles) emissions as a function of time. The PL intensity has been normalized using the highest intensity of the time trace. By fitting the data using a biexponential curve, we extract that for the D_0 emission, the decay is characterized by two time scales: 1 ns and 7.4 ns. For the X_0' emission, the longer time scale is 1.6 ns. The time scale of the fast-decay component is shorter than our time resolution, which is about 300 ps. Therefore, we can clearly see that the lifetime of the D_0 photons is much longer than that of the X_0' photons.

At positive gate voltages, since the emissions of X_- (~ 860 nm) and D_- (~ 845 nm) are very close in energy, it is difficult to isolate the two parts of photons completely. We send all photons to the APD and measure the time trace as a function of the gate voltage, as shown in Figure S7c. When the X_0 peak is pronounced in the range $V_g \in [-0.5\ \mathrm{V}, 0\ \mathrm{V}]$, both the D_0 and D_- emissions become much weaker. I_{X_0} clearly shows a faster decay compared with the decay of $I_{D_0} + I_{X_0'}$ and $I_{X_-} + I_{D_-}$. This reveals that both the D_0 and D_- emissions have a longer lifetime, indicating that both originate from bound exciton complexes.

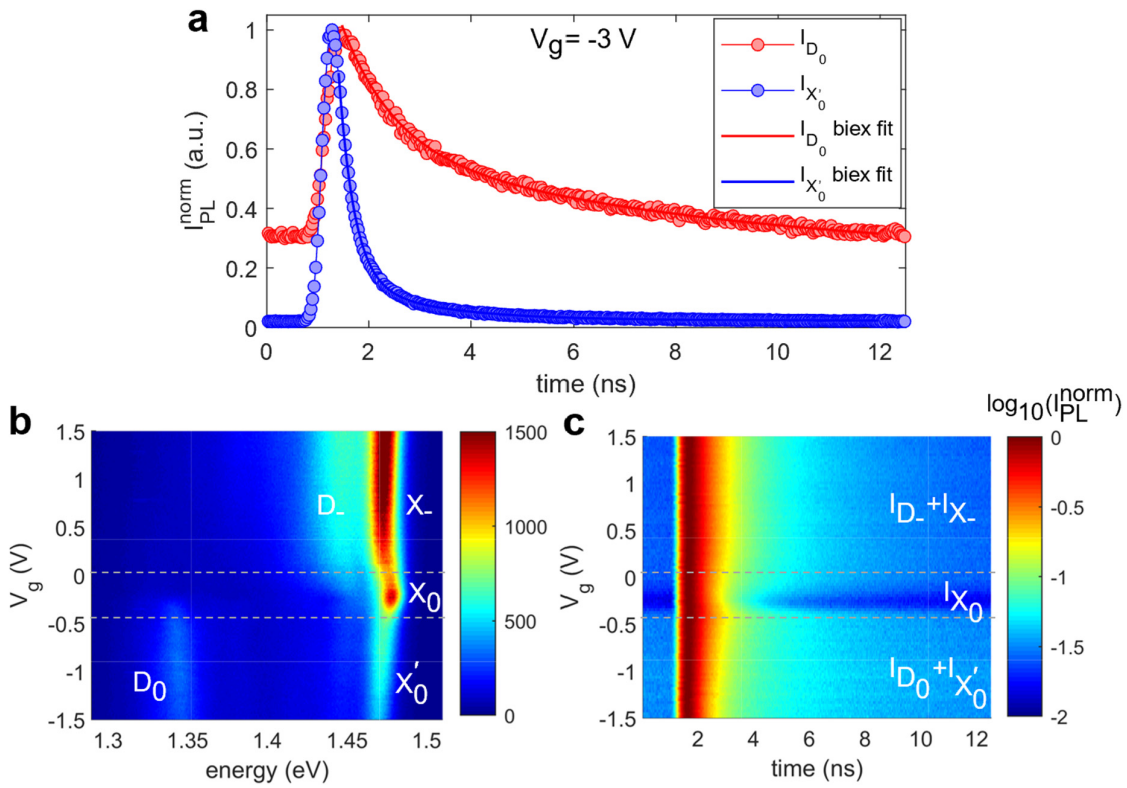

Figure S7. (a) Normalized PL intensity of the D_0 (red circles) and X_0' (blue circles) emissions as a function of time at $V_g = -3$ V. The red and blue solid curves are biexponential fittings to the data. (b) PL intensity as a function of the emission energy and gate voltage V_g using $P = 100$ μW. (c) Normalized intensity of all PL photons at each gate voltage V_g as a function of time using $P = 100$ μW. Logarithmic scale.

8. White light reflectance measurement

The experimental setup for the reflectance measurement is almost the same as the setup for PL measurements. Instead of using a laser, we couple a broadband white light into a fiber and connect it to the excitation arm of the confocal microscope setup. In the collection setup, we remove all spectral filters. The differential reflectance spectrum (dR) is acquired using $dR = (R_{HT} - R_{hBN})/R_{hBN}$, where R_{HT} is the reflectance spectrum on the heterostructure consisting of a 6L. InSe with the top and bottom hBN and R_{hBN} is the reflectance spectrum taken from a position where there is only the top and bottom hBN.

Figure S8 depicts the dR signal as a function of the gate voltage and energy. The PL emission energy of excitons is around 1.48 eV, whereas the dR spectrum shows a broad peak at about 1.7 eV. Moreover, the peak in the dR spectrum does not change with the gate voltage. We performed the same measurement on another sample with an InSe flake of different thicknesses to understand this. Unlike the emission energy, which is sensitive to flake thickness, we found the peak in the reflectance spectrum always located at about 1.7 eV. The dR contrast is smaller

for a thinner InSe. Therefore, we conclude that the peak shown in the dR spectrum might come from the surface reflection, which is irrelevant to the exciton resonance.

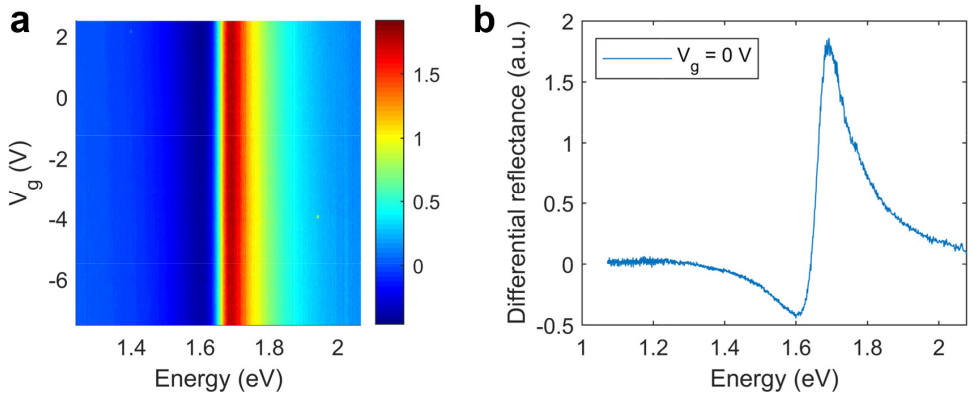

Figure S8. (a) Differential reflectance as a function of the gate voltage and energy. (b) differential reflectance as a function of the gate voltage at $V_g = 0$ V.

9. Extraction of the binding energy using Elliott's theory

To extract the exciton binding energy from the PLE spectrums (absorption spectrums), we use the 2D Elliott model, which takes the form[7,8]:

$$\alpha(\hbar\omega) \propto \sum_{m=1}^{\infty} \frac{4E_R}{\left(m-\frac{1}{2}\right)^3}\delta\left(\hbar\omega - E_g + \frac{E_R}{\left(m-\frac{1}{2}\right)^2}\right) + \frac{2}{1+exp\left(-2\pi\sqrt{\frac{E_R}{\hbar\omega - E_g}}\right)}\Theta(\hbar\omega - E_g)$$

where $\alpha(\hbar\omega)$ is the absorption coefficient as a function of the photon energy, E_g is the single particle band gap, $\delta(x)$ is a Dirac delta function, and $\Theta(x)$ is a Heaviside function. The first term describes the absorption caused by the discrete excitonic transitions, and the second term represents the unbound continuum absorption. The binding energy of m-*th* exciton resonance is $E_b^m = \frac{E_R}{\left(m-\frac{1}{2}\right)^2}(m = 1, 2\ldots)$. Here, E_R is the Rydberg energy which has the form of $E_R = \frac{m_r e^4}{2\epsilon_{eff}^2\hbar^2}$. Considering the finite linewidth of the exciton and the continuum part, we replace $\delta\left(\hbar\omega - E_g + \frac{E_R}{\left(m-\frac{1}{2}\right)^2}\right)$ with a hyperbolic secant function $sech\left(\frac{\hbar\omega - E_g + E_R/\left(m-\frac{1}{2}\right)^2}{\Gamma_{ex}}\right)$, where Γ_{ex} is the linewidth of the exciton resonance[9]. Similarly, $\Theta\left(\hbar\omega - E_g\right)$ is replaced by

$\int_{E_g}^{\infty} sech\left(\frac{\hbar\omega-\varepsilon}{\Gamma_c}\right)\, d\varepsilon$, where Γ_c is the linewidth of the electron-hole continuum. We use the following expression to fit:

$\alpha(\hbar\omega) \propto$

$$\sum_{m=1}^{\infty} \frac{4E_R}{\left(m-\frac{1}{2}\right)^3} sech\left(\frac{\hbar\omega-E_g+E_R/\left(m-\frac{1}{2}\right)^2}{\Gamma_{ex}}\right) + \frac{2}{1+exp\left(-2\pi\sqrt{\frac{E_R}{\hbar\omega-E_g}}\right)} \int_{E_g}^{\infty} sech\left(\frac{\hbar\omega-\varepsilon}{\Gamma_c}\right) \times \frac{1}{1-\alpha(\varepsilon-E_g)}\, d\varepsilon.$$

$\frac{1}{1-\alpha(\varepsilon-E_g)}$ is a factor to account for the non-parabolicity of the bands, as it becomes important for $\varepsilon \gg E_g$.

We calculate the absorption spectrum as a function of E_b using the above expression, as shown in Figure S9a. The relative amplitude between the first term and the second term directly reflects the value of E_b. Therefore, the higher the contrast of the exciton resonance in the absorption spectrum, the higher the binding energy.

In the fittings in Figure 3c and Figure S9b, we use $\Gamma_{ex} = 4.5$ meV and $\Gamma_c = 10$ meV. For X_0 at $V_g = -0.2$ V, the fitting parameters are $E_g = 1.493$ eV and $E_b = 12$ meV; For X_0' at $V_g = -3$ V, the fitting parameters are $E_g = 1.478$ eV and $E_b = 6$ meV; For X_+ at $V_g = -6$ V, the fitting parameters are $E_g = 1.472$ eV and $E_b = 4.5$ meV. The fitting for X_+ agrees with the band gap renormalization picture which we use to explain its redshift. In the presence of a Fermi reservoir, dynamical screening of the electron-hole interactions leads to a lower binding energy $E_b' = E_b - \delta E_b$ ($\delta E_b > 0$). Meanwhile, the many-body screening effect also renormalizes the particle self-energies and results in a reduced band gap $E_g' = E_g - \delta E_g$ ($\delta E_g > 0$). δE_b is smaller than δE_g since its value is limited by E_b. The energy of PL emission E_{PL} can be described as $E_{PL} = E_g - E_b$. Therefore, as we increase the carrier density, the X_+ emission exhibits an overall redshift. For excitons in TMDCs, due to the large E_b, δE_b and δE_g can compensate each other, therefore making E_{PL} unchanged.

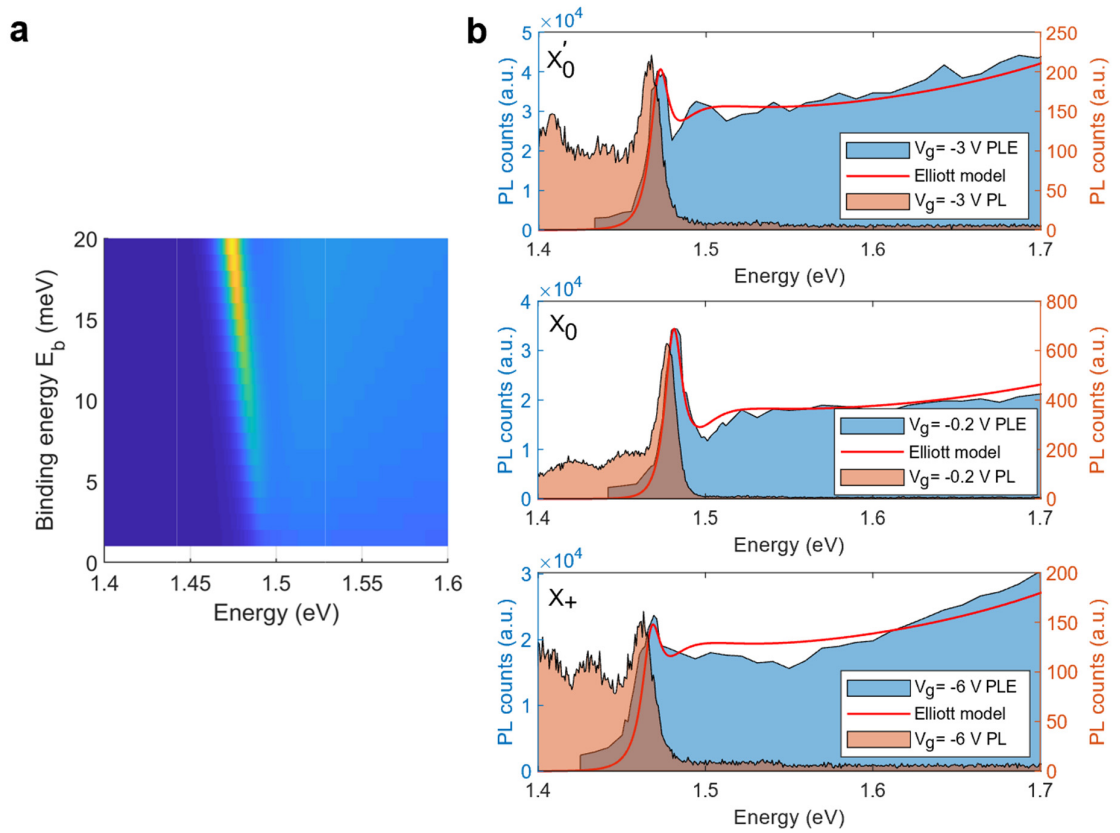

Figure S9. (a) Calculation of the absorption spectrum as a function of the exciton binding energy using the 2D Elliot model. (b) Integrated PL intensity of D_0 as a function of excitation energy (PLE spectrum, blue shaded area) and PL intensity as a function of emission energy (orange shaded area) at $V_g = -3$ V and -0.2 V. The red solid lines are fittings using the 2D Elliott model. For X_0 ($V_g = -0.2$ V), the fitting parameters are $E_g = 1.493$ eV and $E_b = 12$ meV; For X_0' ($V_g = -3$ V), the fitting parameters are $E_g = 1.478$ eV and $E_b = 6$ meV; For X_+ ($V_g = -6$ V), the fitting parameters are $E_g = 1.472$ eV and $E_b = 4.5$ meV.

10. Analysis on bound exciton complexes

In this section, we present more detailed considerations to support the picture of donor ionization in the main text.

A. Energy shift between X_0 and X_0' as a function of excitation power

We extract the peak energy of X_0 and X_0' at $V_g = -0.2$ V and $V_g = -3$ V. The red curve in Figure 4c depicts the energy shift $\Delta E = E_{X_0} - E_{X_0'}$ as a function of the excitation power. The energy shift vanishes when $P \gtrsim 800$ μW. As a comparison, we also plot the PL intensity of D_0 at $V_g = -3$ V as a function of the laser power (blue curve in Figure 4c), after removing the linear background. The PL intensity of D_0 saturates at a similar laser power ($P_{sat} = 400$ μW). The energy shift that vanishes at large powers constitutes solid evidence for our interpretation.

B. Gate voltage range and energy level diagram

The gate voltage range for the sharp peak X_0 is $V_g \in [-0.5\ \mathrm{V}, 0\ \mathrm{V}]$. This range should correspond to the situation of $E_d < E_F < E_c$, where E_c (E_v) denotes the energy of the conduction (valence) band edge. According to the energy splitting between X_0' and D_0, we estimate $E_c - E_d \approx 0.12$ eV. Meanwhile, the gate voltage range for the undoped regime is $V_g \in [-5.8\ \mathrm{V}, 0\ \mathrm{V}]$, which directly relates to the band gap energy $E_g = E_c - E_v \approx 1.49$ eV. The ratio of the V_g range matches well the ratio of the energy range: $0.5\ \mathrm{V}/5.8\ \mathrm{V} = 0.086$ and $0.12\ \mathrm{eV}/1.49\ \mathrm{eV} = 0.081$. Therefore, the proportion of the gate voltage range matches well the energy level diagram.

C. Estimation of donor density

We estimate the donor density n_D from the power-dependent data presented in Figure 4c. Based on the analysis in the main text, the power at which I_{D_0} saturates and the energy shift between X_0 and X_0' becomes vanished should roughly correspond to $n_D \approx n_X$, where n_X is the free exciton density. Therefore, by estimating n_X at $P \sim 500$ μW, we can also estimate n_D. Consider a f_{rep} = 80 MHz pulse laser with an average power of P = 500 μW. The energy per pulse is P/f_{rep} = 6.25 × 10^{-12} J / pulse. The wavelength of the excitation laser is 720 nm, and thus the number of photons per pulse is 2.3 × 10^7. The laser is focused on a spot with an area of πw^2 (w = 1 μm). The photon density is calculated to be around 7.2 × 10^{14} photon / cm^2. The exciton density n_X should be on the order of $10^{12} - 10^{13}$ cm^{-2}, if we assume the absorptance of InSe is about 0.01−0.1 (according to the white light reflection data, the surface of InSe is highly reflective at 1.72 eV. Hence, we expect the absorptance is low). As a result, we expect that the donor density is on the order of $n_D \sim 10^{12} - 10^{13}$ cm^{-2}.

D. Energy difference between spectral lines

The energy difference between D_0 and X_0 (or X_0') should correspond to the binding energy of electrons to donors E_b^e. E_b^e is larger than the binding energy of quasi-particles to donors E_b^Q, in compliance with Hayne's rule[10]. Typically, the ratio E_b^Q/E_b^e is a constant coefficient on the order of 0.1, depending only on material parameters such as the effective masses[11,12]. For the $|De; X\rangle$ complex, the emission energy is 30 meV lower than X_-, thus revealing that the Hayne's

coefficient is about 0.25. Further theoretical efforts are required to provide a microscopic picture for this energy difference[13].

E. Exclusion of the acceptor

We exclude that the acceptors are responsible for the spectral features based on the following reasons:

I). In principle, the D_0 emission could also result from excitons bound to ionized acceptors $|Ah;e\rangle$. Here, $|A\rangle$ is used to express the ionized acceptor. However, in theory the stability of $|De;h\rangle$ and $|Ah;e\rangle$ relies on the mass ratio m_e/m_h. When m_e/m_h is small, excitons are more likely to be bound with an ionized donor[11,14].

II). In the *p*-doped regime, we did not observe broad PL emissions. The formation of a positively charged defect-bound trion $|Ah;X\rangle$ requires acceptor energy levels near the valence band[7,12,15].

III). In the metal chalcogenide semiconductors like MoS_2 and WSe_2, chalcogen vacancies are the most commonly observed defects due to their lower formation energy[16,17]. Chalcogen vacancies behave as electron donors due to unsaturated bonds, whereas metal vacancies usually lead to electron acceptors resulting in the *p*-type doping.

11. Theoretical model of band gap renormalization

In order to directly compare our experimental results with the theoretical model, we first quantify the redshift in the *p*-doped regime as a function of the Fermi energy E_F. Based on the thickness (t) and dielectric constant (ε_{hBN}) of hBN, we could estimate the geometrical capacitance per unit area: $C_g = \epsilon_0\epsilon_{hBN}/t = e\, dn_h/dV_g = 0.0012\ F/m^2$, where n_h is the carrier density of holes. The density of states (DOS) in the valence band dn_h/dE_F can be expressed using $dn_h/dE_F = m_h/\pi\hbar^2 = 4.2\times10^{14}\ \mathrm{cm^{-2}eV^{-1}}$. Combining the two equations, we obtain $dE_F/dV_g = 0.0018\, e$. We extract the relation between the redshift energy (E_{red}) and gate voltage to be $dE_{red}/dV_g = -0.017\ e$. Here, the minus sign denotes the redshift of the PL energy. Therefore, we find E_{red} is about one order of magnitude larger than the Fermi energy $dE_{red}/dE_F \approx -10$, as shown in Figure 5a in the main text.

The model presented in this section is derived based on Ref. [7,18,19].

A. 2D model

We begin with a Coulomb potential in real space $V_r = \frac{e^2}{4\pi\epsilon_r\epsilon_0 r} = \frac{e^2}{\epsilon r}$, where ϵ_r is the dielectric constant of the semiconducting material and ϵ_0 is the vacuum permittivity.

By applying a 2D Fourier transform $V_q = \int d^2r\, V_r\, e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}$, we obtain the Coulomb potential in $\boldsymbol{k}$-space

$$V_q = \frac{e^2}{2\epsilon_r\epsilon_0 q} = \frac{2\pi e^2}{\epsilon q}. \tag{X.1}$$

We introduce a W_q as the dynamically screened Coulomb potential, which is related to V_q via the dynamical dielectric function

$$\epsilon(q,\ \omega) = \frac{V_q}{W_q}. \tag{X.2}$$

Using random phase approximation, $\epsilon(q,\ \omega)$ can be expressed as

$$\epsilon(q,\ \omega) = 1 - V_q \sum_{\boldsymbol{k}} \frac{f_{k-q} - f_k}{\hbar\omega + E_{k-q} - E_k}, \tag{X.3}$$

where $E_k = \frac{\hbar^2 k^2}{2m}$ and f_k is the Fermi-Dirac distribution ($f_k = \frac{1}{e^{\beta(E_k-\mu)}+1}$, where μ is the chemical potential and $\beta = k_B T$ is the product of the Boltzmann constant and temperature).

In the static limit $\omega = 0,\ \epsilon(q,\ \omega)$ yields

$$\epsilon(q,\ 0) = 1 + V_q \frac{\partial n}{\partial \mu}, \tag{X.4}$$

where n is the carrier density, and thus $\frac{\partial n}{\partial \mu}$ is the density of states (DOS).

By introducing a screening wave number $\kappa = \frac{2\pi e^2}{\epsilon}\frac{\partial n}{\partial \mu}$, the dynamical dielectric function takes the form of

$$\epsilon(q, 0) = 1 + \frac{\kappa}{q}. \tag{X.5}$$

For the chemical potential of a 2DEG, we have $\frac{\mu}{k_B T} = \ln\left(e^{\frac{\beta\hbar^2\pi n}{m}} - 1\right)$. Therefore, the 2D screening wave number has a simple form of

$$\kappa = \frac{2me^2}{\epsilon\hbar^2} f_{k=0}. \tag{X.6}$$

For a degenerate 2DEG at low temperatures, $n = \frac{m}{\pi\hbar^2} E_F$.

The carrier-induced BGR is dominated by the Coulomb-hole term (Σ_{CH}). The Coulomb-hole term can be calculated using

$$\Sigma_{CH} = \frac{1}{2}\sum_q (W_q - V_q). \tag{X.7}$$

In 2D, the summation in $\boldsymbol{k}$-space can be expressed using an integral and therefore we get

$$\Sigma_{CH} = -\frac{e^2}{2\epsilon}\int_0^{q_c} \left(1 + \frac{q}{\kappa} + c(q)\right)^{-1} dq. \tag{X.8}$$

The integral is divergent, if $q_c \to \infty$. However, the plasma of the Fermi gas whose energy is much larger than the Fermi energy experience Landau damping. This sets an upper bound for $\frac{\hbar^2 q_c^2}{2m}$. Based on Ref. [18], we simply set this value to $\frac{\hbar^2 q_c^2}{2m}$ = 0.05 eV. $c(q)$ is a term that compensates the static approximation which leads to an overestimation of Σ_{CH}. In the calculations presented in Figure 5c, we set $c = 0$, $\epsilon_r = 10$ [1] and T = 5 K as parameters.

Figure S10 depicts the screening wave number κ as a function of the carrier density for $m = 0.1\, m_0$ and $m = m_0$, where m_0 is the bare electron mass. For $m = m_0$, at high carrier densities, the inverse of κ is about 0.3 nm. The inverse of the screening wave number can be regarded as the effective screening length. If the flake thickness is much higher than the 2D effective screening length, the third dimension needs to be considered. Therefore, a 3D model is derived to describe the dynamical screening in 3D.

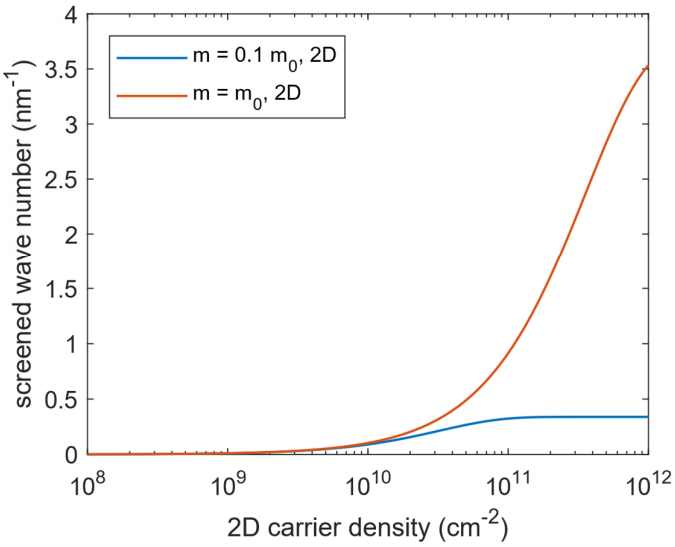

Figure S10. Calculated screening wave number κ as a function of the carrier density for $m = 0.1\, m_0$ (blue curve) and $m = m_0$ (red curve).

B. 3D model

Similar to the 2D model, we could rewrite the Coulomb potential in $\boldsymbol{k}$-space using a 3D Fourier transform

$$V_q = \int d^3r\, V_r\, e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} = \frac{e^2}{\epsilon_r \epsilon_0 q} = \frac{4\pi e^2}{\epsilon q^2}. \tag{X.9}$$

Therefore,

$$\epsilon(q,\ 0) = 1 + V_q \frac{\partial n}{\partial \mu} = 1 + \frac{4\pi e^2}{\epsilon q^2}\frac{\partial n}{\partial \mu}. \tag{X.10}$$

We again introduce a 3D screening wave number $\kappa = \sqrt{\frac{4\pi e^2}{\epsilon}\frac{\partial n}{\partial \mu}}$, such that $\epsilon(q,\ 0)$ takes the form of

$$\epsilon(q,0) = 1 + \frac{\kappa^2}{q^2}. \tag{X.11}$$

In 3D, the relation between n and μ has a form of

$$n = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \int_0^\infty \sqrt{E}\ \frac{1}{e^{\beta(E-\mu)}+1}\ dE. \tag{X.12}$$

$\frac{\partial n}{\partial \mu}$ cannot be expressed analytically. Here, for simplicity, we use the Boltzmann distribution to replace the Fermi-Dirac distribution. We obtain an analytical expression of κ at a finite temperature

$$\kappa = \sqrt{\frac{4\pi e^2 n\beta}{\epsilon}}. \tag{X.13}$$

For a degenerate 3DEG at low temperatures,

$$n = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \frac{2}{3}E_F^{3/2}. \tag{X.14}$$

We again using (X.7) and derive the Coulomb-hole energy shift Σ_{CH} in 3D

$$\Sigma_{CH} = -\frac{e^2}{\pi\epsilon}\int_0^{q_c}\left(1 + \frac{q^2}{\kappa^2} + c(q)\right)^{-1}\ dq. \tag{X.15}$$

If we assume $q_c \to \infty$, (X.15) simply becomes $\Sigma_{CH} = -\frac{e^2}{2\epsilon}\kappa$. Here, we still set $\frac{\hbar^2 q_c^2}{2m}$ = 0.05 eV to describe a cut-off energy for the Landau damping of plasma.

The discrepancy between our data and theoretical model comes from two aspects: (i) using a parallel capacitor model to estimate the carrier density neglects the electrostatic screening leading to an overestimation, especially if the semiconducting flake is thicker than a monolayer. As a result, the Fermi energy (*x*-axis of Figure 5a) is overcalculated; (ii) as mentioned above, in the static approximation, the calculated Coulomb-hole term (*y*-axis of Figure 5b) is overestimated.

12. Discussion on the Fermi-polaron picture

In our work, we do not use Fermi-polarons to explain the redshift in the *p*-doped regime. The neutral exciton and trion in monolayer TMDCs are sometimes considered as repulsive and attractive Fermi-polarons which are quasi-particles formed by excitons coupling to a Fermi reservoir[20,21]. The polaron effects are usually observed in the reflectance spectrum for monolayer TMDCs. The spectral peak in an emission spectrum usually locates at the lowest

energy where the population is largest, while the contrast in a reflectance spectrum quantifies the oscillator strength of a specific many-body state. The state with high oscillator strength does not necessarily show intense PL emissions because the energy is not the lowest. However, the state with high oscillator strength can couple with a cavity mode to form polaritons that exhibits anti-crossing features in the spectrum. This has been reported by Sidler *et al*.[20] The underlying reasons are as follows. The coupling strength of a trion to the radiative field is proportional to $\left(k_{ph}a_T\right)^2$, where k_{ph} is the momentum of the radiative field that is on the order of the light-cone size $k_{ph}{\sim}E_g/\hbar c \approx 7.6\times10^{-4}$ Å^{-1}, and a_T is the trion Bohr radius. Due to the coupling to the Fermi reservoir, the polaron has a much larger oscillator strength that is proportional to $(k_F a_T)^2$, where $k_F=\sqrt{2m_hE_F}/\hbar$ is the momentum at the Fermi surface[22]. In our experiment, we tune the chemical potential by about 5 meV. We estimate $k_F{\sim}3.6\times10^{-2}$ Å^{-1}. As all spectral features reported in our work are shown in the PL emission spectra, we tend to use the BGR picture to explain the redshift. Different from the polaron picture where the exciton is directly coupled with the Fermi reservoir forming new quasi-particles, the BGR picture can be considered as using exciton (or trion) to probe the change of band gap due to the many-body screening effect in the Fermi reservoir.

REFERENCES

(1) Ceferino, A.; Song, K. W.; Magorrian, S. J.; Zólyomi, V.; Fal'ko, V. I. Crossover from Weakly Indirect to Direct Excitons in Atomically Thin Films of InSe. *Phys. Rev. B* **2020**, *101* (24), 245432. https://doi.org/10.1103/PhysRevB.101.245432.

(2) Li, G.; Luican, A.; Lopes dos Santos, J. M. B.; Castro Neto, A. H.; Reina, A.; Kong, J.; Andrei, E. Y. Observation of Van Hove Singularities in Twisted Graphene Layers. *Nat. Phys.* **2010**, *6* (2), 109–113. https://doi.org/10.1038/nphys1463.

(3) Ben Aziza, Z.; Zólyomi, V.; Henck, H.; Pierucci, D.; Silly, M. G.; Avila, J.; Magorrian, S. J.; Chaste, J.; Chen, C.; Yoon, M.; Xiao, K.; Sirotti, F.; Asensio, M. C.; Lhuillier, E.; Eddrief, M.; Fal'ko, V. I.; Ouerghi, A. Valence Band Inversion and Spin-Orbit Effects in the Electronic Structure of Monolayer GaSe. *Phys. Rev. B* **2018**, *98* (11), 115405. https://doi.org/10.1103/PhysRevB.98.115405.

(4) Ceferino, A.; Magorrian, S. J.; Zólyomi, V.; Bandurin, D. A.; Geim, A. K.; Patanè, A.; Kovalyuk, Z. D.; Kudrynskyi, Z. R.; Grigorieva, I. V.; Fal'ko, V. I. Tunable Spin-Orbit Coupling in Two-Dimensional InSe. *Phys. Rev. B* **2021**, *104* (12), 125432. https://doi.org/10.1103/PhysRevB.104.125432.

(5) Chow, C. M. E.; Yu, H.; Schaibley, J. R.; Rivera, P.; Finney, J.; Yan, J.; Mandrus, D.; Taniguchi, T.; Watanabe, K.; Yao, W.; Cobden, D. H.; Xu, X. Monolayer Semiconductor Auger Detector. *Nano Lett.* **2020**, *20* (7), 5538–5543. https://doi.org/10.1021/acs.nanolett.0c02190.

(6) Linardy, E.; Yadav, D.; Vella, D.; Verzhbitskiy, I. A.; Watanabe, K.; Taniguchi, T.; Pauly, F.; Trushin, M.; Eda, G. Harnessing Exciton–Exciton Annihilation in Two-Dimensional Semiconductors. *Nano Lett.* **2020**, *20* (3), 1647–1653. https://doi.org/10.1021/acs.nanolett.9b04756.

(7) Peyghambarian, N.; Koch, S. W.; Mysyrowicz, A. *Introduction to Semiconductor Optics*; Prentice Hall, 1993.

(8) Passarelli, J. V.; Mauck, C. M.; Winslow, S. W.; Perkinson, C. F.; Bard, J. C.; Sai, H.; Williams, K. W.; Narayanan, A.; Fairfield, D. J.; Hendricks, M. P.; Tisdale, W. A.; Stupp, S. I. Tunable Exciton Binding Energy in 2D Hybrid Layered Perovskites through Donor–Acceptor Interactions within the Organic Layer. *Nat. Chem.* **2020**, *12* (8), 672–682. https://doi.org/10.1038/s41557-020-0488-2.

(9) Neutzner, S.; Thouin, F.; Cortecchia, D.; Petrozza, A.; Silva, C.; Srimath Kandada, A. R. Exciton-Polaron Spectral Structures in Two-Dimensional Hybrid Lead-Halide Perovskites. *Phys. Rev. Mater.* **2018**, *2* (6), 064605. https://doi.org/10.1103/PhysRevMaterials.2.064605.
(10) Haynes, J. R. Experimental Proof of the Existence of a New Electronic Complex in Silicon. *Phys. Rev. Lett.* **1960**, *4* (7), 361–363. https://doi.org/10.1103/PhysRevLett.4.361.
(11) Meyer, B. K.; Sann, J.; Lautenschläger, S.; Wagner, M. R.; Hoffmann, A. Ionized and Neutral Donor-Bound Excitons in ZnO. *Phys. Rev. B* **2007**, *76* (18), 184120. https://doi.org/10.1103/PhysRevB.76.184120.
(12) Klingshirn, C. F. *Semiconductor Optics*; Springer Science & Business Media, 2006.
(13) Mostaani, E.; Szyniszewski, M.; Price, C. H.; Maezono, R.; Danovich, M.; Hunt, R. J.; Drummond, N. D.; Fal'ko, V. I. Diffusion Quantum Monte Carlo Study of Excitonic Complexes in Two-Dimensional Transition-Metal Dichalcogenides. *Phys. Rev. B* **2017**, *96* (7), 075431. https://doi.org/10.1103/PhysRevB.96.075431.
(14) Sharma, R. R.; Rodriguez, S. Theory of Excitons Bound to Ionized Impurities in Semiconductors. *Phys. Rev.* **1967**, *153* (3), 823–827. https://doi.org/10.1103/PhysRev.153.823.
(15) Rivera, P.; He, M.; Kim, B.; Liu, S.; Rubio-Verdú, C.; Moon, H.; Mennel, L.; Rhodes, D. A.; Yu, H.; Taniguchi, T.; Watanabe, K.; Yan, J.; Mandrus, D. G.; Dery, H.; Pasupathy, A.; Englund, D.; Hone, J.; Yao, W.; Xu, X. Intrinsic Donor-Bound Excitons in Ultraclean Monolayer Semiconductors. *Nat. Commun.* **2021**, *12* (1), 871. https://doi.org/10.1038/s41467-021-21158-8.
(16) Zhou, W.; Zou, X.; Najmaei, S.; Liu, Z.; Shi, Y.; Kong, J.; Lou, J.; Ajayan, P. M.; Yakobson, B. I.; Idrobo, J.-C. Intrinsic Structural Defects in Monolayer Molybdenum Disulfide. *Nano Lett.* **2013**, *13* (6), 2615–2622. https://doi.org/10.1021/nl4007479.
(17) Zhang, S.; Wang, C.-G.; Li, M.-Y.; Huang, D.; Li, L.-J.; Ji, W.; Wu, S. Defect Structure of Localized Excitons in a WSe_2 Monolayer. *Phys. Rev. Lett.* **2017**, *119* (4), 046101. https://doi.org/10.1103/PhysRevLett.119.046101.
(18) Scharf, B.; Tuan, D. V.; Žutić, I.; Dery, H. Dynamical Screening in Monolayer Transition-Metal Dichalcogenides and Its Manifestations in the Exciton Spectrum. *J. Phys.: Condens. Matter* **2019**, *31* (20), 203001. https://doi.org/10.1088/1361-648X/ab071f.
(19) Haug, H.; Koch, S. W. *Quantum Theory of the Optical and Electronic Properties of Semiconductors*; World Scientific, 2009.
(20) Sidler, M.; Back, P.; Cotlet, O.; Srivastava, A.; Fink, T.; Kroner, M.; Demler, E.; Imamoglu, A. Fermi Polaron-Polaritons in Charge-Tunable Atomically Thin Semiconductors. *Nat. Phys.* **2017**, *13* (3), 255–261. https://doi.org/10.1038/nphys3949.
(21) Efimkin, D. K.; MacDonald, A. H. Many-Body Theory of Trion Absorption Features in Two-Dimensional Semiconductors. *Phys. Rev. B* **2017**, *95* (3), 035417. https://doi.org/10.1103/PhysRevB.95.035417.
(22) Imamoglu, A.; Cotlet, O.; Schmidt, R. Exciton–Polarons in Two-Dimensional Semiconductors and the Tavis–Cummings Model. *Comptes Rendus Phys.* **2021**, *22* (S4), 1–8. https://doi.org/10.5802/crphys.47.