Crossover from weakly indirect to direct excitons in atomically thin films of InSe

Document Version
Final published version

Citation for published version (APA):
Ceferino, A., Song, K. W., Magorrian, S. J., Zólyomi, V., & Fal'ko, V. I. (2020). Crossover from weakly indirect to direct excitons in atomically thin films of InSe. *Physical Review B*, *101*(24). https://doi.org/10.1103/PhysRevB.101.245432

Published in:
Physical Review B

OPEN ACCESS

Crossover from weakly indirect to direct excitons in atomically thin films of InSe

Adrián Ceferino [1,2,*], Kok Wee Song [2,†], Samuel J. Magorrian [2], Viktor Zólyomi,[3] and Vladimir I. Fal'ko[1,2,4]

[1]*Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom*
[2]*National Graphene Institute, Booth Street East, Manchester, M13 9PL, United Kingdom*
[3]*STFC Hartree Centre, Daresbury Laboratory, Daresbury, Warrington, WA4 4AD, United Kingdom*
[4]*Henry Royce Institute for Advanced Materials, Manchester, M13 9PL, United Kingdom*

We perform a $\mathbf{k} \cdot \mathbf{p}$ theory analysis of the spectra of the lowest energy and excited states of the excitons in few-layer atomically thin films of InSe taking into account in plane electric polarizability of the film and the influence of the encapsulation environment. For the thinner films, the lowest-energy state of the exciton is weakly indirect in momentum space, with its dispersion showing minima at a layer-number-dependent wave number, due to an inverted edge of a relatively flat topmost valence band branch of the InSe film spectrum, and we compute the activation energy from the momentum dark exciton ground state into the bright state. For the films with more than seven In_2Se_2 layers, the exciton dispersion minimum shifts to Γ point.

Two-dimensional (2D) materials create new opportunities for semiconductor optoelectronics [1–3]. Among those new materials, post-transition metal chalcogenides (InSe and GaSe) occupy a special place, as they offer a flexibility to choose a desirable size of their bandgap (in the range from 3 to 1.3 eV) depending on the number of atomic planes in a thin film [4–8]. While the experimental studies of the band gap and optical properties of few layer films of InSe [7,9–11] and GaSe [12–15] have found a reasonably close quantitative interpretation at the single-particle level, based on density functional theory (DFT) [16–22] and the DFT-parametrized tight-binding model [23], the fine tuning of the theory requires taking into account excitonic effects in the system, which remains an open question for atomically thin InSe films.

Here, we develop a mesoscale theory for the binding energies, dispersions, and excited state spectra of excitons in mono-, bi-, tri-, and few-layer InSe films (γ polytype), taking into account the strongly nonparabolic features of the valence band dispersion in these 2D materials and the influence of various encapsulation environments. In particular, we study the role of a weak inversion of the hole dispersion near the top of the valence band [18,19,24–26], established in the thinnest InSe and GaSe films using angle-resolved photoemission spectroscopy [10,27] and high field magneto-optics studies [23], and analyze the crossover of the excitons from weakly indirect to direct in momentum space, as a function of the InSe film thickness. The crossover of the exciton dispersion from indirect [$\varepsilon(Q) = $ min at $Q \neq 0$] to direct [$\varepsilon(Q) = $ min at $Q = 0$] exciton was found at $L = 7$ layers. For films with $1 \leqslant L \leqslant 10$, we compute the binding energies of the excitons for hBN-encapsulated InSe films and the activation energies from the momentum-dark excitonic bound states, with the results summarized in Fig. 1.

In the analysis presented below, we describe excitons using two-particle wave functions, $\Psi_{\mathbf{Q}}^{\dagger} = \sum_{nm} \int d^2k \psi_{\mathbf{Q},nm}(\mathbf{k}) a_{\mathbf{k}+\mathbf{Q},n}^{\dagger} c_{\mathbf{k},m}$, written in the wave-number representation for the constituent electrons and holes occupying states with wave numbers $\mathbf{k} + \mathbf{Q}$ and $\mathbf{k}$ in sub-bands [11,28] n and m on the conduction ($a_{\mathbf{k}+\mathbf{Q},n}$) and valence ($c_{\mathbf{k},m}$) band side of few-layer InSe film spectrum. Below, we project all electron and hole states onto the lowest sub-bands ($n = 1$) in the film, which is justified by the much larger inter-sub-band energies, as compared to the exciton binding energies in the thin films (with $L \lesssim 10$ see Fig. 2). As a result, the exciton creation operator takes the approximate form $\Psi_{\mathbf{Q}}^{\dagger} = \int d^2k \psi_{\mathbf{Q}}(\mathbf{k}) a_{\mathbf{k}+\mathbf{Q},1}^{\dagger} c_{\mathbf{k},1}$ where $\psi_{\mathbf{Q}} \equiv \psi_{\mathbf{Q},11}$. This gives the Bethe-Salpeter equation [29–33]:

$$\int_{\mathbf{q}} [(\varepsilon_c(\mathbf{k}) - \varepsilon_v(\mathbf{k} - \mathbf{Q}) - \Omega)\delta_{\mathbf{q},0} + V(\mathbf{q})]\psi_{\mathbf{Q}}(\mathbf{k} + \mathbf{q}) = 0, \quad (1)$$

for an effectively 2D exciton with momentum $\mathbf{Q}$ and energy Ω (the latter is a sum $\Omega = E_g + E_b$ of the gap E_g and the binding energy E_b). Here, we use the notation $\int_{\mathbf{q}} \equiv \int \frac{d^2q}{(2\pi)^2}$. The electron-hole (e-h) attraction is accounted for by the Fourier transform of the interaction potential,

$$V(\mathbf{q}) = -\frac{4\pi e^2}{\epsilon_0} \iint |\phi_{e\mathbf{k}_e}(z)|^2 W(\mathbf{q}, z, z') |\phi_{h\mathbf{k}_h}(z')|^2 dz dz',$$

$$W(\mathbf{q}, z, z') = \frac{\cosh\left[\tilde{q}\left(\frac{d}{2} - z\right) + \eta\right]\cosh\left[\tilde{q}\left(\frac{d}{2} + z'\right) + \eta\right]}{\sqrt{\epsilon_{\parallel}\epsilon_z} q \sinh(\tilde{q}d + 2\eta)};$$

$$\tilde{q} = \sqrt{\epsilon_{\parallel}/\epsilon_z} q; \quad \eta = \frac{1}{2}\ln\frac{\sqrt{\epsilon_{\parallel}\epsilon_z} + \sqrt{\kappa_{\parallel}\kappa_z}}{\sqrt{\epsilon_{\parallel}\epsilon_z} - \sqrt{\kappa_{\parallel}\kappa_z}}, \quad (2)$$

designed to take into account both the dielectric polarizability of the 2DM and the dielectric environment [34–37] (e.g., hBN [38,39], with $\kappa_{\parallel} = 6.9$ and $\kappa_z = 3.7$). For L-layer InSe, film thickness is $d = La_z$, where $a_z = 8.32$ Å is the interlayer distance and $\epsilon_{\parallel}$ and ϵ_z are the in- and out-of-plane permittivities

_____
*adrian.ceferino@postgrad.manchester.ac.uk
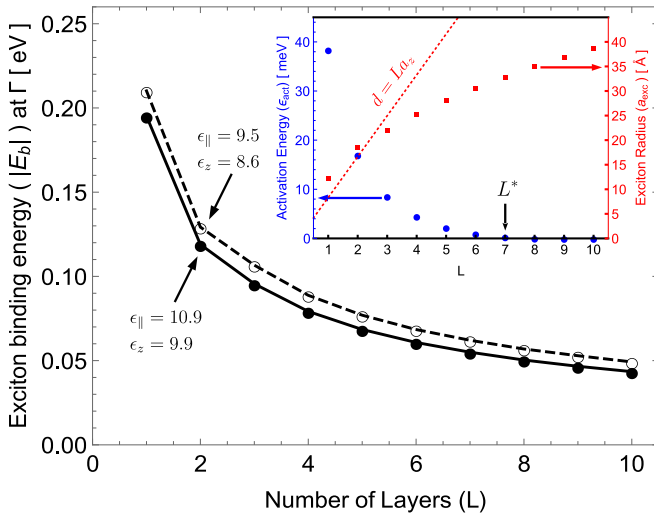†kokwee.song@manchester.ac.uk

FIG. 1. The dependence of the exciton binding energy on the number of layers (L) for hBN encapsulated InSe films. Binding energies at Γ point, $E_b(0)$ are compared for two values of bulk InSe dielectric constants $\epsilon_\parallel$ and ϵ_z. The inset shows the activation energy $\varepsilon_{\rm act} = E_{\rm b}(0) - E_{\rm b}(Q_{\rm min})$ (closed blue circles) where $Q_{\rm min}$ is approximately the wave vector between the Γ point and the edge of the highest valence band. The radius of the exciton, $a_{\rm exc} = \sqrt{\langle |\mathbf{r}_e - \mathbf{r}_h|^2 \rangle}$ is also shown (closed red square). The red dashed line shows the thickness of the InSe film.

of bulk InSe [40]. The above expression takes into account the z dependence of the lowest electron-hole sub-band wave functions $\phi_{e/h,\mathbf{k}}(z)$, and W is quoted for $z \geqslant z'$ (for $z < z'$, z should be interchanged with z'). We note that, for $L = 1$ and 2, the 2D potential $V(\mathbf{q})$ can be simplified to the Keldysh

potential [34,35,37,41],

$$V(\mathbf{q}) \approx -\frac{2\pi e^2}{\sqrt{\kappa_z \kappa_\parallel}} \frac{1}{q(1 + r_* q)}, \qquad r_* = \frac{\sqrt{\epsilon_z \epsilon_\parallel} - 1}{2\sqrt{\kappa_z \kappa_\parallel}} d.$$

In the above equation, r_* is the screening length [42] indicating the region dominated by the logarithmically divergent potential at length scales smaller than r_* and the region dominated by the Coulombic interaction potential at distances greater than r_*. However, for $L \geqslant 3$, the exciton radius ($a_{\rm exc}$) appears to be smaller than the film thickness, so that the electron/hole charge distribution along the z axis in Eq. (2) needs to be taken into account in full details. To do that, we use the quantum-well approximation for the z distribution of the lowest sub-band [28,43], $\phi_{e/h,\mathbf{k}}(z) \approx \sqrt{2/d} \cos(\pi z/d)$. We note that separating wave-function variables and discarding higher energy sub-bands in Eq. (1) is applicable if the quantization energy due to confinement is much larger than the excitonic energy scale, which will be justified later by comparing the inter-sub-band energies to the calculated exciton binding energies.

To implement numerical diagonalization of the Bethe-Salpeter equation (1), we use a basis of harmonic oscillator functions for the bound electron-hole states [44], $\psi_{\mathbf{Q}}(\mathbf{k}) = \sum_{0 \leqslant n_x + n_y}^{N_{\rm max}} \mathcal{A}_{n_x,n_y}^{\mathbf{Q}} \varphi_{n_x}(k_x) \varphi_{n_y}(k_y)$ where $\varphi_n(k) = \sqrt{\frac{\lambda}{\pi^{1/2} 2^n n!}} (-i)^n e^{-k^2 \lambda^2/2} H_n(k\lambda)$ and $H_n(x)$ is the nth Hermite polynomial. In the above described basis, the choice of the length λ and the cutoff $N_{\rm max}$ are optimized for speeding up a converging calculation (see Appendix C for details). We also checked the performance of the developed code by comparing its results to the exact solution of the 2D hydrogen problem, aiming at $\lesssim 2\%$ error as compared to the ground-state energy of the Rydberg series. A software package for
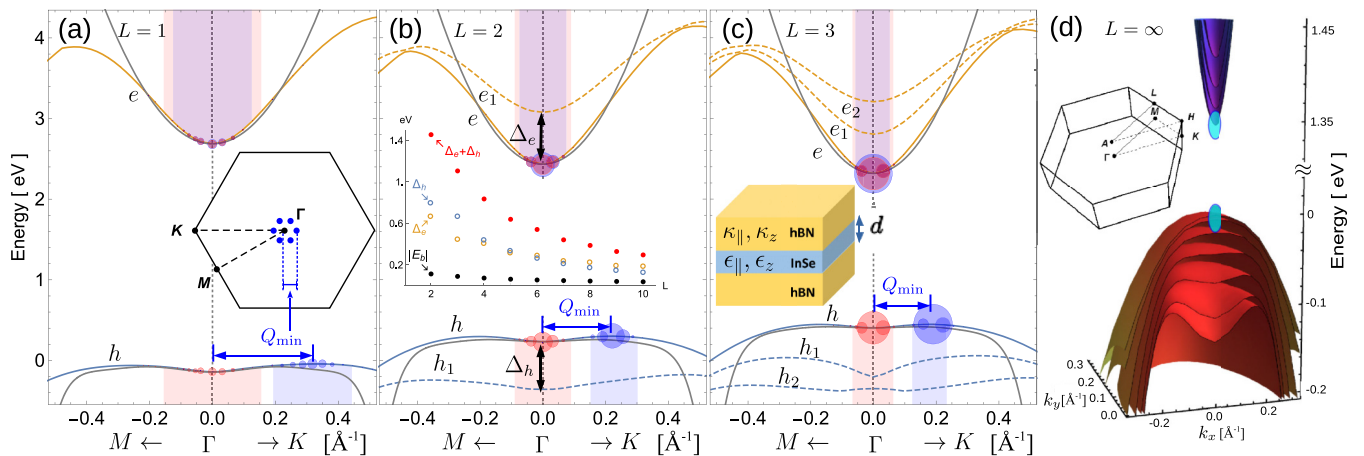

FIG. 2. The bound electron-hole states and the quasiparticle dispersion in momentum space: (a)–(c) The plots for $L = 1$–3 layers film. The solid (dashed) yellow and blue curves are conduction and valence (sub)band tight-binding dispersions [23]. The gray curves are the $\mathbf{k} \cdot \mathbf{p}$ dispersion where ε_c and ε_v are expanded into a polynomial of $\mathbf{k}$. The sizes of the red (blue) circles are proportional to the probability density $|\psi_{\mathbf{Q}}(\mathbf{k})|^2$ with $\mathbf{Q} = 0$ ($\mathbf{Q} = Q_{\rm min}$ being the total momentum of the lowest energy exciton. (d) Plot for bulk InSe dispersion near A point: Each of the bands plotted both in conduction and in valence band correspond to $k_z = 0 - 0.06$ Å^{-1} in steps of 0.01 Å^{-1}. Blue shaded region indicates the region in in-plane momentum and in energy covered by the ground-state exciton in the bulk limit as determined by the size of the Gaussian wave packet and the exciton binding energy. (Insets) (a) Brillouin zone of 2D-InSe, blue circles showing C_6-symmetric localization of holes; (b) a comparison of the $\mathbf{Q} = 0$ exciton binding energy (closed black circles) with the sub-band energy splittings. Δ_e (open yellow circles) and Δ_h (open blue circles) correspond to the e-e_1 and h-h_1 splitting at the Γ point, respectively. The closed black circles are the total splitting, $\Delta_e + \Delta_h$; (c) schematic of hBN-encapsulated InSe, with their different dielectric constants; (d) Brillouin zone of bulk InSe.

the implementation of numerical diagonalization of Eq. (1) with arbitrary parameters for InSe films and encapsulation environment and instructions for interested users are included in the Supplemental Material [45].

With this numerical setting, we solve Eq. (1) using the DFT-parametrized $\mathbf{k} \cdot \mathbf{p}$ theory for InSe films with dispersions illustrated in Fig. 2. In particular, we used a polynomial expansion around Γ point both for the conduction and valence bands [46] computed using the GW-parametrized hybrid $\mathbf{k} \cdot \mathbf{p}$ tight-binding model (see Appendix B), $\varepsilon_{c/v}(\mathbf{k}) = \sum_{i,j=0} A_{ij}^{e/h} k_x^i k_y^j$ also plotted in Figs. 2(a)–2(c). For comparison, in Fig. 2(d), we show the conduction and valence band dispersion of bulk InSe near the band edges (which are at A point in the 3D Brillouin zone), where the inversion of $\varepsilon_v(\mathbf{k}, k_z)$ develops upon the increase of z-axis momentum k_z (counted from A point).

In Fig. 3(a), we show the first eight bound states energies of the Γ-point exciton with $Q = 0$ (solid line) [47] and the lowest-energy momentum-dark state of the exciton at $Q = Q_{\min}$ (dashed line) for $1 \leqslant L \leqslant 3$. A minimum at $Q = Q_{\min}$ in the exciton dispersion for each state [Fig. 3(b)] is due to the sombrero of the h band (see Fig. 2). The non-hydrogen-like energy sequence [48,49] is due to the 2D screening of the e-h interaction in the film leading to a Keldysh-like potential for $L = 1$ and 2.

To illustrate the layer-number dependence of the exciton dispersion, we compare the exciton binding energy at the Γ point, $E_b(0)$, and at its dispersion minimum, $E_b(Q_{\min})$, for L up to 10 layers. In the inset in Fig. 1, we plot the activation energy ε_{act} from the dark exciton state (at $Q = Q_{\min}$) to the optically active state (at $Q = 0$). We find that $\varepsilon_{act} \to 0$ at $L^* = 7$, which indicates that an indirect to direct crossover for the exciton occurs before the expectation based on a single-particle valence band dispersion (at $L^* = 10$, see Appendix B). For completeness, we also analyzed excitons in bulk 3D InSe using bulk band dispersions shown in Fig. 2(d) and $V(\mathbf{q}, q_z) = -\frac{4\pi e^2/\epsilon_0}{\epsilon_\parallel q^2 + \epsilon_z q_z^2}$. We solve the Bethe-Salpeter equation for bulk InSe using the 3D harmonic oscillator basis, and use dielectric constants [40] $\epsilon_\parallel = 10.9$, $\epsilon_z = 9.9$ for InSe, together with the GW-computed valence band masses ($m_{v\parallel} = -5.35 m_0$, $m_{vz} = -0.078 m_0$) and the conduction band masses ($m_{c\parallel}/m_0 = 0.16$ and $m_{cz}/m_0 = 0.086$ where m_0 is the free electron mass) which are close to those measured in cyclotron resonance experiments [50]. The examples of computed bulk exciton dispersions, $E_{3D}(\mathbf{Q}, Q_z)$ are shown in Fig. 3(c). Using ($Q_z \approx \frac{\pi}{L a_z}$) for the quantization of the transverse exciton motion, we find that the crossover into indirect spectrum should be expected at $L \approx 6$–7 layers, in agreement with the transition number of layers L^* found in the layer dependence of the activation energy ε_{act} (inset in Fig. 1). We note that the computed bulk (3D) exciton binding energy is about 30% lower than the experimentally claimed [51,52] values of 13–15 meV. Binding energy can be increased to 14.6 meV by choosing $\epsilon_\parallel = 9.5$, $\epsilon_z = 8.6$ (with $\sqrt{\epsilon_z/\epsilon_\parallel} = 0.95$ as in Ref. [40]). For this reason we computed and compared the exciton spectra in the films using two choices of dielectric parameters $\epsilon_\parallel = 10.9$, $\epsilon_z = 9.9$ and $\epsilon_\parallel = 9.5$, $\epsilon_z = 8.6$. We find that in thin films $L \leqslant 10$ such a variation of InSe dielectric parameters has a much weaker influence on the exciton bindings than in the
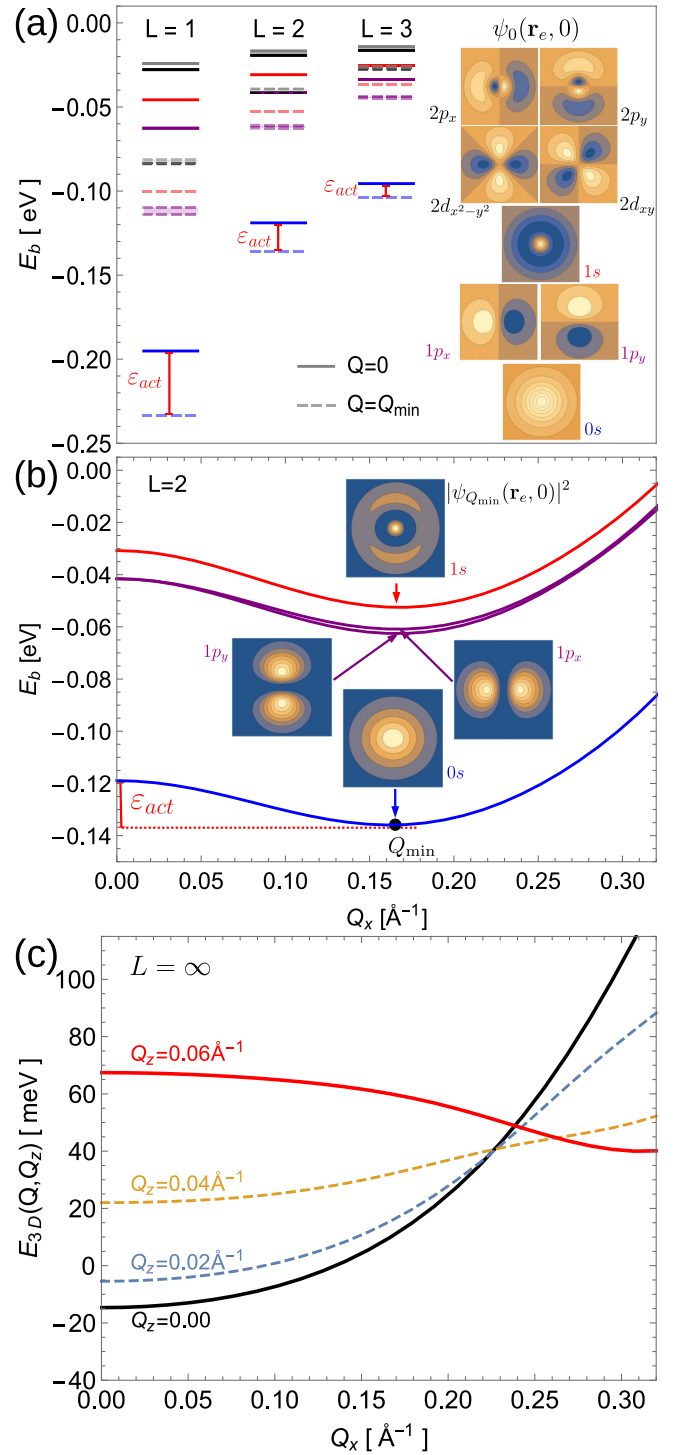

FIG. 3. (a) The first eight low-energy states of the exciton at $Q = 0$, $Q_{\min}$ in hbN-encapsulated InSe. Insets on the right are the schematic exciton wave functions $\psi_0(\mathbf{r}_e, 0)$ in the real space ($\psi_\mathbf{Q}(\mathbf{r}_e, \mathbf{r}_h) \equiv \sum_\mathbf{k} \psi_\mathbf{Q}(\mathbf{k}) e^{i\mathbf{k} \cdot (\mathbf{r}_e - \mathbf{r}_h)}$) which are sorted from higher to lower binding energies (bottom to top). The dark and bright region correspond to a negative and positive value for the wave-function amplitude. (b) The exciton dispersion for the first four low-energy states. The schematic real-space probability distributions of a bound electron, $|\psi_{Q_{\min}}(\mathbf{r}_e, 0)|^2$, are illustrated by the insets. (c) The energy dispersion of the exciton ground state in bulk InSe with $Q_y = 0$. The different Q_z values are indicated in the plot. We use the adjusted dielectric parameters in the plot which are $\epsilon_\parallel = 9.5$ and $\epsilon_\parallel = 8.6$.

TABLE I. Polynomial fit of the sombrero dispersion as for the topmost valence band and parabolic dispersion of the lowest conduction band. m_0 is the free electron mass.

L	A_8^h (eVÅ^8)	A_6^h (eVÅ^6)	A_4^h (eVÅ^4)	A_2^h (eVÅ^2)	m_c/m_0
1	−1188.591	471.809	−68.601	3.674	0.266
2	−1210.270	388.158	−49.004	1.989	0.223
3	−1308.626	371.401	−43.048	1.372	0.207
4	−1411.696	364.846	−39.437	0.985	0.198
5	−1565.869	366.036	−36.797	0.703	0.193
6	−1745.505	368.254	−34.556	0.487	0.189
7	−1938.337	369.112	−32.543	0.316	0.187
8	−2130.725	367.119	−30.684	0.179	0.184
9	−2302.573	361.073	−28.941	0.068	0.183
10	−2085.138	331.905	−27.004	−0.026	0.181

bulk material. These calculated binding energies compare well with the values observed in the recent experiments [11] on hBN-encapsulated thin InSe films. In summary, we present a mesoscale theory which is particularly useful for investigating the energy spectrum of a Wannier-Mott exciton in large gap semiconductors ($E_g \gg |E_b|$). Most interestingly, this theoretical framework can also be applied to study direct and indirect excitons in complex van der Waals heterostructures [53–61].

ACKNOWLEDGMENTS

APPENDIX A: PARAMETRIZATION OF ELECTRON-HOLE DISPERSION IN *L*-LAYER INSE

Here, we give the details on the parametrized electron and hole dispersion by using polynomial fit. The conduction and valence band dispersions near the Γ point are approximated by

$$\varepsilon_c(\mathbf{k}) = \frac{1}{2m_c}\mathbf{k}^2, \qquad (A1)$$

$$\varepsilon_v(\mathbf{k}) = A_2^h\mathbf{k}^2 + A_4^h\mathbf{k}^4 + A_6^h\mathbf{k}^6 + A_8^h\mathbf{k}^8, \qquad (A2)$$

where the hexagonal wrapping terms are ignored, because the exciton wave function is strongly localized in k-space. These polynomials are obtained by fitting to bands from the GW-parametrized hybrid $\mathbf{k}\cdot\mathbf{p}$ tight-binding (HkpTB) model, Appendix B, and the fitted values for 1- to 10-layer InSe film are listed in Table I. We note for $L \leqslant 9$ that the quadratic term in the valence band dispersion corresponds to negative effective hole masses. This yields a sombrero-shaped dispersion in the valence band and requires one to retain higher-order terms

in the expansion for fitting. The hole mass becomes positive at $L = 10$. For the 3D bulk dispersion near the conduction and valence band edges of γ-InSe, we employ the following polynomial of the form,

$$\varepsilon_c(\mathbf{k}, k_z) = \frac{k^2}{2m_{c\parallel}} + \frac{k_z^2}{2m_{cz}}, \qquad (A3)$$

$$\varepsilon_v(\mathbf{k}, k_z) = \frac{\mathbf{k}^2}{2m_{v\parallel}} + \frac{k_z^2}{2m_{vz}} + \gamma\mathbf{k}^4 + \alpha\mathbf{k}^2k_z^2 + \gamma_zk_z^4. \qquad (A4)$$

Here, k_z is measured from the A point. In the fit, obtained using GW-DFT computed bands, the effective in-plane and out-of-plane masses for the electron are $m_{c\parallel} = 0.16m_0$ and $m_{cz} = 0.086m_0$, close to the experimentally measured [50] values of $m_{c\parallel} \approx 0.14m_0$ and $m_{cz} \approx 0.08m_0$, respectively. For the valence band, the fitted parameters are as follows: $m_{v\parallel} = -5.35m_0$, $m_{vz} = -0.078m_0$, $\gamma = -10.84$ eVÅ^4, $\alpha = 1074$ eVÅ^4, and $\gamma_z = 1688$ eVÅ^4.

APPENDIX B: HYBRID MULTIBAND k · p TIGHT-BINDING THEORY WITH PARAMETERS FROM QUASIPARTICLE SELF-CONSISTENT GW CALCULATIONS

1. Hybrid multiband k · p tight-binding model

The model used in this study is built using two main components: a multiband $\mathbf{k}\cdot\mathbf{p}$ model describing the monolayer bands (following Refs. [20,62]), and interlayer coupling in few-layer and bulk systems, described using a tight-binding approach based on the monolayer $\mathbf{k}\cdot\mathbf{p}$ bands (similar to the hybrid $\mathbf{k}\cdot\mathbf{p}$ tight-binding approach taken in Refs. [28,63]).

In this description we model the bands of few-layer and bulk InSe near the Γ point using a Hamiltonian with the form,

$$H = \sum_{\mathbf{k},\sigma}\left[\sum_{n=1}^{N}H_{ML,\mathbf{k},\sigma}^n + \sum_{n=1}^{N-1}H_{IL,\mathbf{k},\sigma}^{n,n+1} + \text{H.c.}\right], \qquad (B1)$$

where $H_{ML,\mathbf{k},\sigma}^n$ is the monolayer $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian on layer n of the N-layer crystal, at $\mathbf{k}$ with z projection of spin $\sigma = \pm\frac{1}{2}$. $H_{I-L}^{n,n+1}$ includes the interlayer tight-binding hops between the monolayer bands.

a. Monolayer k · p Hamiltonian

The monolayer Hamiltonian follows the multiband $\mathbf{k}\cdot\mathbf{p}$ approaches of Refs. [20,62]. While in our previous works [23,28] the basis of monolayer bands was a basis of single-band $\mathbf{k}\cdot\mathbf{p}$ expansions, so that matrix elements such as couplings to electromagnetic fields and the interlayer hops mentioned above had to depend on k, here we follow the multiband approach and take as our basis the bands at Γ, and introduce k-dependent off-diagonal terms to account for the variation of the bands with k. At the expense of an increase in the dimensionality of the parameter space, this allows us to make the approximation that the interlayer hops are independent of k, and assists in the capture of higher-order effects, such as the offset valence band maximum, while keeping the $\mathbf{k}\cdot\mathbf{p}$ expansions to order k^2. The monolayer Hamiltonian for layer

n of an N-layer crystal takes the form,

$$
\begin{aligned}
H_{ML,\mathbf{k},\sigma}^n = {} & (\varepsilon_{c_1} + \alpha_{c_1}k^2)a_{n,c_1,\mathbf{k}}^{\sigma\dagger}a_{n,c_1,\mathbf{k}}^{\sigma} + (\varepsilon_c + \alpha_c k^2)a_{n,c,\mathbf{k}}^{\sigma\dagger}a_{n,c,\mathbf{k}}^{\sigma} + (\varepsilon_v + \alpha_v k^2)a_{n,v,\mathbf{k}}^{\sigma\dagger}a_{n,v,\mathbf{k}}^{\sigma} \\
& + \left(\varepsilon_{v_1} + \left(\alpha_{v_1}k^2 + \alpha'_{v_1}(k_x^2 - k_y^2)\right)\right)a_{n,v_{1x},\mathbf{k}}^{\sigma\dagger}a_{n,v_{1x},\mathbf{k}}^{\sigma} + \left(\varepsilon_{v_1} + \alpha_{v_1}k^2 + \alpha'_{v_1}(k_y^2 - k_x^2)\right)a_{n,v_{1y},\mathbf{k}}^{\sigma\dagger}a_{n,v_{1y},\mathbf{k}}^{\sigma} \\
& + \left(\varepsilon_{v_2} + (\alpha_{v_2}k^2 + \alpha'_{v_2}(k_x^2 - k_y^2))\right)a_{n,v_{2x},\mathbf{k}}^{\sigma\dagger}a_{n,v_{2x},\mathbf{k}}^{\sigma} + \left(\varepsilon_{v_2} + \alpha_{v_2}k^2 + \alpha'_{v_2}(k_y^2 - k_x^2)\right)a_{n,v_{2y},\mathbf{k}}^{\sigma\dagger}a_{n,v_{2y},\mathbf{k}}^{\sigma} + 2\alpha'_{v_1}k_x k_y a_{n,v_{1x},\mathbf{k}}^{\sigma\dagger}a_{n,v_{1y},\mathbf{k}}^{\sigma} \\
& + 2\alpha'_{v_2}k_x k_y a_{n,v_{2x},\mathbf{k}}^{\sigma\dagger}a_{n,v_{2y},\mathbf{k}}^{\sigma} + \beta_{c_1,v}k^2 a_{n,c_1,\mathbf{k}}^{\sigma\dagger}a_{n,v,\mathbf{k}}^{\sigma} + i\beta_{c_1,v_2}\left(k_x a_{n,c_1,\mathbf{k}}^{\sigma\dagger}a_{n,v_{2x},\mathbf{k}}^{\sigma} + k_y a_{n,c_1,\mathbf{k}}^{\sigma\dagger}a_{n,v_{2y},\mathbf{k}}^{\sigma}\right) \\
& + i\beta_{v,v_2}\left(k_x a_{n,v,\mathbf{k}}^{\sigma\dagger}a_{n,v_{2x},\mathbf{k}}^{\sigma} + k_y a_{n,v,\mathbf{k}}^{\sigma\dagger}a_{n,v_{2y},\mathbf{k}}^{\sigma}\right) + i\beta_{c,v_1}\left(k_x a_{n,c,\mathbf{k}}^{\sigma\dagger}a_{n,v_{1x},\mathbf{k}}^{\sigma} + k_y a_{n,c,\mathbf{k}}^{\sigma\dagger}a_{n,v_{1y},\mathbf{k}}^{\sigma}\right) \\
& - 2\lambda_{v_{1,2}}i\sigma\left(a_{n,v_{1x},\mathbf{k}}^{\sigma\dagger}a_{n,v_{1y},\mathbf{k}}^{\sigma} + a_{n,v_{2x},\mathbf{k}}^{\sigma\dagger}a_{n,v_{2y},\mathbf{k}}^{\sigma}\right) + \lambda_{v,v_1}\left(-2\sigma a_{n,v,\mathbf{k}}^{\sigma\dagger}a_{n,v_{1x},\mathbf{k}}^{-\sigma} + ia_{n,v,\mathbf{k}}^{\sigma\dagger}a_{n,v_{1y},\mathbf{k}}^{-\sigma}\right).
\end{aligned}
\tag{B2}
$$

The bands which form the basis of the model are the monolayer Γ-point bands in the absence of spin-orbit coupling (SOC) [23]. The operator $a_{n,j,\mathbf{k}}^{\sigma(\dagger)}$ annihilates (creates) an electron in layer n, band j, with spin $\sigma = \pm\frac{1}{2}$ and in-plane momentum $\mathbf{k}$. As singly degenerate bands which are totally in-plane symmetric at Γ, bands c_1, c, v are assigned Γ-point energies $\varepsilon_{c_1,c,v}$ with quadratic "onsite" dispersions with respective coefficients $\alpha_{c_1,c,v}$. In contrast, in the absence of SOC bands v_1 and v_2, being dominated by p_x and p_y orbitals, are twice degenerate at Γ with energies ε_{v_1,v_2}. The dispersions of their two light- and heavy-hole branches are handled using two components corresponding to a basis of their p_x and p_y components, with quadratic intra- and intercomponent contributions with coefficients $\alpha_{v_{1,2}}^{(\prime)}$. In the multiband $\mathbf{k} \cdot \mathbf{p}$ picture away from Γ the bands are modified by off-diagonal terms between them. These terms must preserve the σ_h symmetry of the monolayer, so only involve the pairs c_1, v, c, v_1 and v, v_2. Of these, c_1, v is between bands which are totally in-plane symmetric at Γ, so the off-diagonal term is quadratic, while terms involving the x and y components of $v_{1,2}$ are linear in k_x and k_y, respectively. The coefficients of these terms are denoted as $\beta_{c_1,v}$, β_{c,v_1}, and β_{v,v_2}, respectively. Finally, spin-orbit coupling (SOC) is included within the components of v_1 and v_2 ($l_z s_z$ with coupling strength $\lambda_{v_{1,2}}$) and between v_1 and v (the "spin-flip" $l_x s_x + l_y s_y$ with coupling strength λ_{v,v_1}). Cross-gap "spin-flip" terms are neglected.

b. Interlayer tight-binding hops

The nonzero interlayer tight-binding hops between the monolayer bands, and their form, can be inferred from the symmetries of the bands involved in the hop. The resulting interlayer contribution to the Hamiltonian takes the form,

$$
\begin{aligned}
H_{IL,\mathbf{k},\sigma}^{n,n+1} = {} & \sum_{j=c_1,c,v} t_j a_{n,j,\mathbf{k}}^{\sigma\dagger}a_{n+1,j,\mathbf{k}}^{\sigma} \\
& + t_{c_1,c}\left(a_{n,c_1,\mathbf{k}}^{\sigma\dagger}a_{n+1,c,\mathbf{k}}^{\sigma} - a_{n,c,\mathbf{k}}^{\sigma\dagger}a_{n+1,c_1,\mathbf{k}}^{\sigma}\right) \\
& + t_{c,v}\left(a_{n,c,\mathbf{k}}^{\sigma\dagger}a_{n+1,v,\mathbf{k}}^{\sigma} - a_{n,v,\mathbf{k}}^{\sigma\dagger}a_{n+1,c,\mathbf{k}}^{\sigma}\right) \\
& + t_{v_{1,2}}\sum_{i=x,y}\left(a_{n,v_{1i},\mathbf{k}}^{\sigma\dagger}a_{n+1,v_{1i},\mathbf{k}}^{\sigma} - a_{n,v_{1i},\mathbf{k}}^{\sigma\dagger}a_{n+1,v_{2i},\mathbf{k}}^{\sigma}\right. \\
& \left. - a_{n,v_{2i},\mathbf{k}}^{\sigma\dagger}a_{n+1,v_{2i},\mathbf{k}}^{\sigma} + a_{n,v_{2i},\mathbf{k}}^{\sigma\dagger}a_{n+1,v_{1i},\mathbf{k}}^{\sigma}\right).
\end{aligned}
\tag{B3}
$$

Since the γ stacking preserves the C_3 rotational symmetry of the monolayer, the bands may be divided into two groups, with no hopping between the singly and doubly degenerate basis bands, with the x and y components also not mixed by the interlayer hops. We have made the approximation that, since interlayer hops are dominated by interlayer Se-Se pairs [23], they may be taken as $z/-z$ symmetric. As a result, hops between c_1 and c, and between c and v, which are pairs of bands with opposing symmetry under $z/-z$ reflection in the monolayer, are antisymmetric under exchange of layers. We neglect the hop $t_{c_1,v}$ as the bands are well separated in energy, and interlayer hops involving c_1 are expected to be weak owing to the dominance of the c_1 wave function by orbitals on the indium atoms in the center of each layer. Finally, using the domination of v_1 and v_2 by Se p_x and p_y orbitals, we assume that all hops within and between v_1 and v_2 are of the same magnitude, $t_{v_{1,2}}$.

2. Parametrization—bulk γ-InSe

Since DFT can often underestimate band gaps, significantly so in the case of thicker 2D and bulk InSe, a means by which one may obtain spectra of more use in comparison with experiments is the use of a "scissor operator"—a rigid shift upwards in energy of the unoccupied bands with respect to the occupied bands. In other words, one assumes that features of the DFT bands, such as effective masses, band widths, matrix elements, and so on, are all correct, other than the size of the gap itself. This has been shown to be a useful procedure in theoretical studies of semiconductors [64–66], and in 2D InSe [7,23]. However, the magnitude of the underestimation of the gap (approaching a factor of $\sim$4 in the bulk limit) for InSe can make the procedure more complex. For example, a straight scissor correction without taking into account other effects of the underestimation of the gap can lead to an overestimation of the interband out-of-plane electric dipole matrix element [23], or an underestimation of the band-edge effective masses in the bulk case and hence an overestimation of the splitting of sub-bands in the few-layer case [28]. While there are means by which some of these problems may be overcome (for example, the out-of-plane effective mass was corrected in Ref. [28] by applying a scissor correction to the monolayer bands *after* parametrization of the interlayer hops), the presence of cross-gap off-diagonal matrix elements in even the monolayer Hamiltonian presents challenges in the determination of the appropriate means of compensating for an underestimation of the band gap in a DFT reference.

In this case, therefore, we take as our first-principles reference a quasiparticle self-consistent GW (QSGW) calculation for the bulk crystal. For this we use the QUESTAAL package

[67,68], using the Bethe-Salpeter equation (BSE) to determine the polarization in the calculation of W. Since the bands and gaps of InSe have been shown to be sensitive to strain [69,70], we use an experimental lattice with crystal structure parameters found using x-ray diffraction [71]. The DFT part of the calculation sampled the Brillouin zone with a $24 \times 24 \times 24$ **k**-point grid, while for the QSGW part a $6 \times 6 \times 6$ grid was used. In the calculation of W, nine occupied bands and fifteen unoccupied bands were handled using the BSE, while the rest of the bands were handled at the random-phase approximation level. In the DFT part, the G-vector cutoff for the interstitial density mesh was 9.1 $Ry^{1/2}$, while in the QSGW part the cutoffs for the plane-wave expansions of the eigenfunctions and the Coulomb integrals were 3.4 a.u. and 2.9 a.u., respectively. The QSGW calculation of the self-energy is carried out without taking SOC into account, with the effects of SOC included at the DFT level afterwards. We choose a calculation of the bulk crystal as a reference for finding model parameters as a QSGW calculation for few-layer InSe would be prohibitively expensive given the number of atoms in a unit cell. The calculation gives a quasiparticle band gap of 1.367 eV for the bulk, close to the experimentally obtained 1.351 eV [72].

In the case of the model, Eq. (B1) is amended to describe the bulk with a unit cell corresponding to a single layer as

$$H = H_{1L}^1 + H_{IL}^{1,1} e^{ik_z a_z} + \text{H.c.}, \tag{B4}$$

where k_z is the out-of-plane momentum and $a_z = 8.315$ Å is the distance between successive layers [71]. The parametrization is carried out in two steps. First, we fit bands for 50 k_z points between $k_z = 0$ and $k_z = \pi/a_z$ for **k** = 0, as we show in Fig. 4, then holding the 2D Γ-point parameters fixed, we fit the in-plane dispersions for small **k** near Γ up to $k = K/5$ for each k_z used in the first stage of the fitting. In Fig. 4 we show the in-plane QSGW and model dispersions for $k_z a_z = 0, \pi/2, \pi$. The model parameters are given in Table II.

3. Few-layer bands

Having found a parameter set for the model, we now explore its behavior in the few-layer case, with an overview of some of the key features of the bands of few-layer InSe shown in Fig. 5. The dispersive nature of the bulk conduction and valence bands, arising from the strong interlayer hops t_c, t_v, t_{cv} between bands with strong wave-function contributions from selenium p_z orbitals, translate to large splittings between sub-bands in the few-layer case. It is this strong interlayer hybridization which is responsible for the large variation of band gap with crystal thickness [7,23], reaching >2.8 eV for monolayer films. In contrast, v_1 and v_2, being dominated by $p_{x,y}$ orbitals which lie mostly in the 2D crystal plane, have weak interband hops and exhibit much weaker splitting. As a consequence when the conduction and valence bands acquire contributions from v_1 and v_2 (due to, in the model, interband **k** · **p** mixing) away from Γ their splitting becomes weaker. In the conduction band this manifests itself as a difference between the effective masses of successive sub-bands, which in Ref. [28] was handled by a k-dependent t_c.

For the valence band the situation is more complex. As has been theoretically predicted [17,19] and shown in ARPES
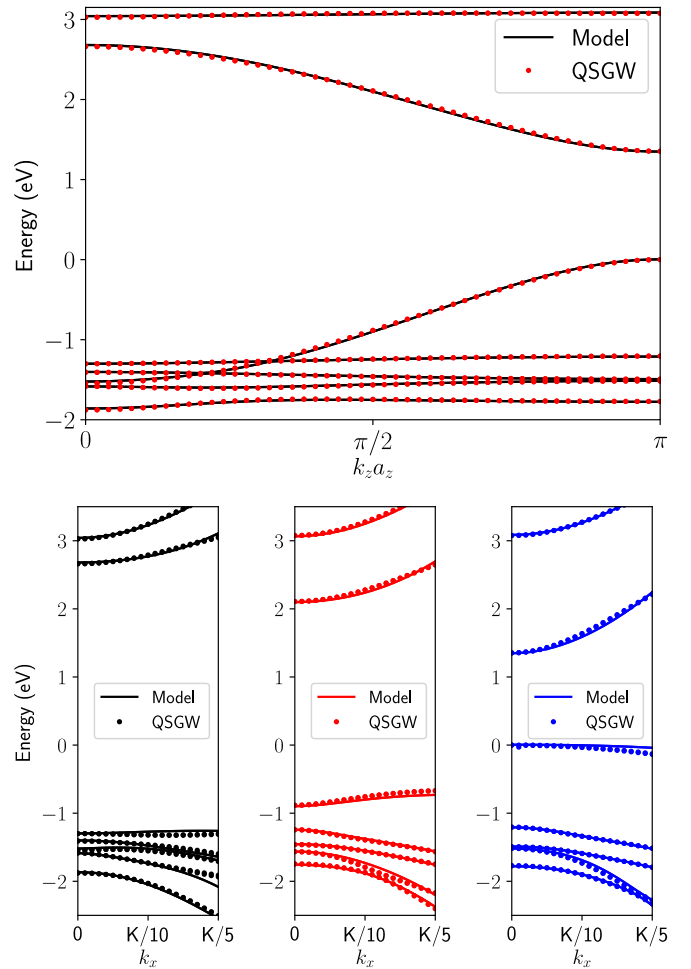

FIG. 4. Upper panel: QSGW (dots) and fitted model (lines) out-of-plane dispersions for bulk γ-InSe, for in plane momentum **k** = 0. Lower panels: in-plane dispersions (along k_x) for (from left to right) $k_z a_z = 0, \pi/2, \pi$. 0 eV set to valence band edge in all cases.

experiments [10], for the thinnest films an offset in the valence band maximum develops, leading to a slightly indirect band gap, in contrast to the direct gap found in thicker films and in the bulk crystal. In the multiband **k** · **p** picture a key contribution to this phenomenon can be understood [62] as repulsion away from Γ between bands v and v_2. When in the few-layer case v splits much more than v_2 this repulsion becomes much weaker. Coupled with a weaker splitting of v itself at larger k in a similar manner to that of the conduction band, this causes the depth and radius of the "Mexican hat" offset to decrease rapidly with increasing crystal thickness, ultimately leading to a direct gap in the model for $N \geqslant 10$ layers.
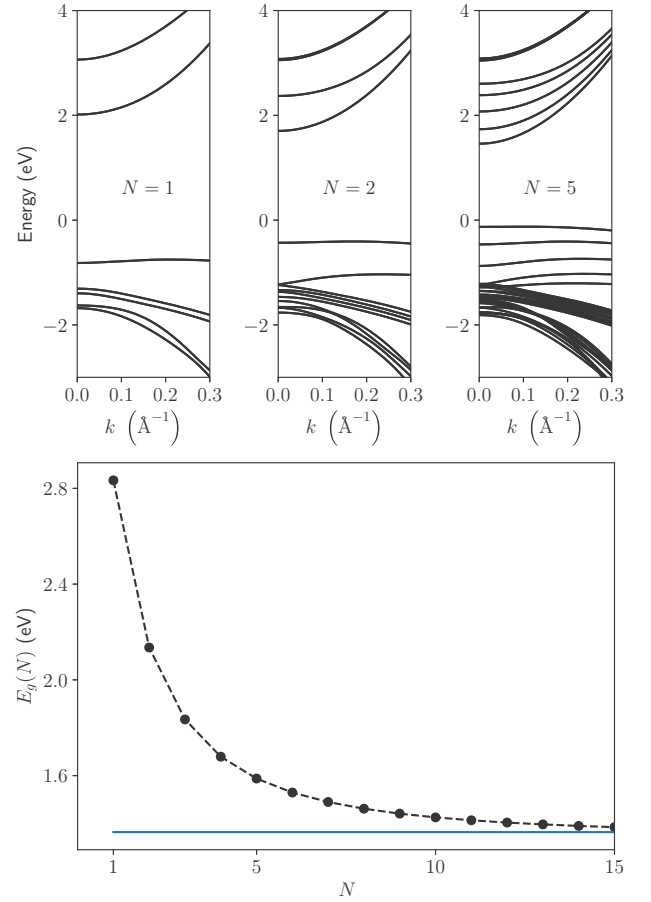
APPENDIX C: NUMERICAL IMPLEMENTATION OF HARMONIC OSCILLATOR BASIS

In the harmonic oscillator basis described in the text, the BSE (1) takes the form,

$$\sum_{n'_x n'_y} \left[\mathcal{H}^0_{n_x n_y; n'_x n'_y} - \mathcal{V}_{n_x n_y; n'_x n'_y} \right] \mathcal{A}^{\mathbf{Q}}_{n'_x n'_y} = \Omega \mathcal{A}^{\mathbf{Q}}_{n_x n_y}, \tag{C1}$$

TABLE II. Model parameters for Eqs. (B2) and (B3) fitted to QSGW bands for bulk InSe. 0 eV is set to the valence band edge in the bulk.

ε_{c_1}	3.064 eV
ε_c	2.015 eV
ε_v	−0.855 eV
ε_{v_1}	−1.449 eV
ε_{v_2}	−1.538 eV
$\lambda_{v_{1,2}}$	0.142 eV
λ_{v,v_1}	0.119 eV
t_{c_1}	−0.011 eV
t_c	0.333 eV
t_v	−0.420 eV
$t_{v_{1,2}}$	−0.048 eV
$t_{c_1,c}$	0.019 eV
$t_{c,v}$	0.251 eV
α_{c_1}	1.54 eVÅ^2
α_c	−18.7 eV Å^2
α_v	−4.95 eVÅ^2
α_{v_1}	6.48 eVÅ^2
α'_{v_1}	−10.51 eVÅ^2
α_{v_2}	−0.28 eVÅ^2
α'_{v_2}	−4.20 eVÅ^2
$\beta_{c_1,v}$	3.77 eVÅ^2
β_{c_1,v_2}	8.51 eVÅ
β_{c,v_1}	10.54 eVÅ
β_{v,v_2}	−2.78 eVÅ

with the kinetic energy matrix,

$$\mathcal{H}^0_{n_x n_y; n'_x n'_y}(\mathbf{Q}) = \int d^2k [\varepsilon_c(\mathbf{k}) - \varepsilon_v(\mathbf{k}-\mathbf{Q})]$$
$$\times \varphi^*_{n_x}(k_x)\varphi^*_{n_y}(k_y)\varphi_{n'_x}(k_x)\varphi_{n'_y}(k_y), \quad \text{(C2)}$$

and the interaction matrix,

$$\mathcal{V}_{n_x n_y; n'_x n'_y} = \int \frac{d^2k d^2q}{(2\pi)^2} V(\mathbf{q})$$
$$\times \varphi^*_{n_x}(k_x)\varphi^*_{n_y}(k_y)\varphi_{n'_x}(k_x + q_x)\varphi_{n'_y}(k_y + q_y). \quad \text{(C3)}$$

In the following, we explain how to choose an optimal harmonic oscillator basis set to speed up the convergence in a calculation. We also give the details for how to construct the matrix equation in Eq. (C1).

Choice of basis set. To diagonalize the BSE in Eq. (C1), we first need to specify the harmonic oscillator basis set which is determined by the parameters: λ, the length scale of the oscillator, and $N_{\max}$, the cutoff of the oscillator modes with $n_x + n_y \leqslant N_{\max}$. In principle, λ can be arbitrary since a unique result can be obtained provided that $N_{\max}$ is large enough. In practice, working with a large basis set is undesirable because large matrix diagonalization is a very demanding computational task. In the following, we show that a good convergent result can be obtained with a relatively small basis set if a proper choice of λ is used.

The procedure for obtaining the optimal λ is to maximize the exciton binding energy against λ (inset of Fig. 6). This λ corresponds to the optimal coverage of the exciton by the basis set in the momentum/real space. In Fig. 6, we demonstrate
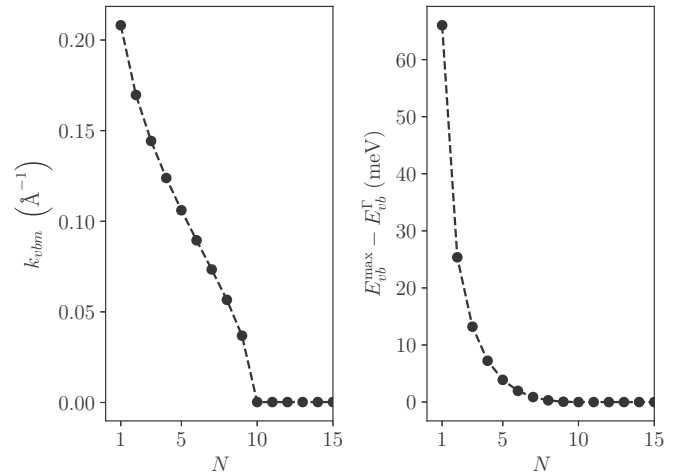

FIG. 5. (Upper panels) 2D model dispersions for monolayer, bilayer, and five-layer InSe. (Middle panel) Vertical band gaps at Γ for $N = 1$- to 15-layer InSe. Solid line is the bulk band gap. (Lower panels) Position (left) and magnitude (right) of offset of valence band maximum from Γ point for $N = 1$- to 15-layer InSe, showing indirect-direct gap transition at 10 layers.

how the binding energy depends on λ of a finite basis set with $n_x + n_y \leqslant N_{\max} = 12$. We note that the optimal λ for each different states need to be determined separately since each states have a very different characteristic localized length scale. In Fig. 6, one can see that once the optimal λ is determined, we obtain a good convergent result for the binding energy at
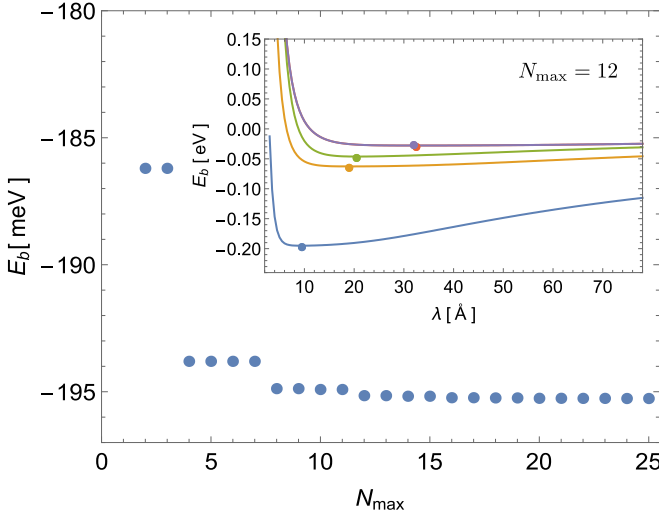
FIG. 6. The convergence in the calculation of exciton ground-state binding energy by using the optimal λ for a monolayer InSe with Keldysh potential. The optimal λ for each state is determined by maximizing their corresponding binding energy which are marked by the closed circles in the inset. The optimization for the first four Γ-point exciton binding energy in the inset is performed with fixed $N_{\max} = 12$.

$N_{\max} \sim 12$. Increasing the number of basis beyond $N_{\max} = 12$ only leads to no more than 2 meV correction.

Kinetic energy matrix. For a general band dispersion such as those in tight-binding model, analytical expression may not be available and the query of the band energy may be computationally expansive. Therefore, a straightforward numerical integration in Eq. (C2) is not a practical approach. A feasible numerical method is to expand the band dispersion (periodic function) into a fast convergent Fourier series. Namely,

$$\varepsilon_c(\mathbf{k}) - \varepsilon_v(\mathbf{k} - \mathbf{Q}) = \sum_{\mathbf{s}=-\infty}^{\infty} [C_{\mathbf{s}} - V_{\mathbf{s}} e^{i2\pi \mathbf{s}\cdot\bar{\mathbf{Q}}}] e^{-i2\pi \mathbf{s}\cdot\bar{\mathbf{k}}}, \quad \text{(C4)}$$

where $\mathbf{s} = (s_x, s_y)$, $\bar{\mathbf{k}} = (k_x/T_x, k_y/T_y)$, and $\bar{\mathbf{Q}} = (Q_x/T_x, Q_y/T_y)$ with the (T_x, T_y) are the periodicity of the dispersion in each dimensions. The Fourier coefficients are therefore defined as

$$\begin{bmatrix} C_{\mathbf{s}} \\ V_{\mathbf{s}} \end{bmatrix} = \int_{-\frac{T_x}{2}}^{\frac{T_x}{2}} \frac{dk_x}{T_x} \int_{-\frac{T_y}{2}}^{\frac{T_y}{2}} \frac{dk_y}{T_y} e^{i2\pi \mathbf{s}\cdot\bar{\mathbf{k}}} \begin{bmatrix} \varepsilon_c(\mathbf{k}) \\ \varepsilon_v(\mathbf{k}) \end{bmatrix}. \quad \text{(C5)}$$

With this expansion, we can integrate out the momentum explicitly. Hence, the band energy matrix in Eq. (C2) become

$$\mathcal{H}^0_{n_x n_y; n'_x n'_y}(\mathbf{Q})$$

$$= \sum_{s_x, s_y = -\infty}^{\infty} [C_{\mathbf{s}} - V_{\mathbf{s}} e^{i2\pi \mathbf{s}\cdot\bar{\mathbf{Q}}}] \prod_j^{x,y} \frac{2^{\zeta_j - \frac{1}{2}(n_j + n'_j)} \zeta_j! (\bar{a}_j s_j)^{\Delta_j}}{i^{n'_j - n_j + \Delta_j} \sqrt{n_j! n'_j!}}$$

$$\times e^{-\frac{1}{4}\bar{a}_j^2 s_j^2} L_{\zeta_j}^{\Delta_j}\left(\frac{1}{2}\bar{a}_j^2 s_j^2\right), \quad \text{(C6)}$$

with $\zeta_j = \min[n_j, n'_j]$, $\Delta_j = |n'_j - n_j|$, $\bar{a}_j = 2\pi/(T_j\lambda)$, and $L_n^\alpha(x)$ is the associated Laguerre polynomial. Since the band dispersion is periodic, only a few of the Fourier modes are
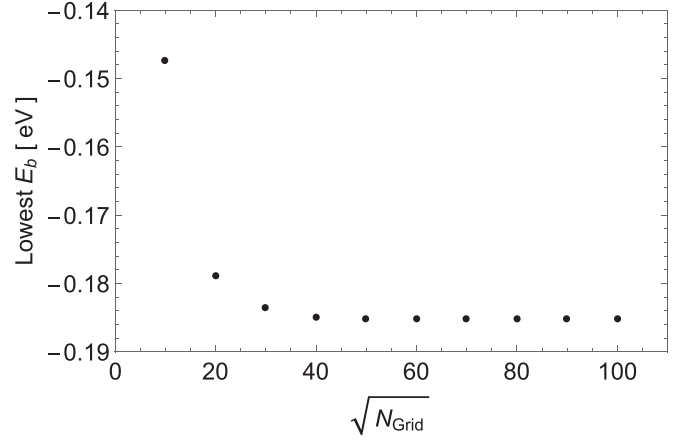

FIG. 7. The convergence of the $\mathbf{Q} = 0$ exciton ground-state energy with different grid size for constructing the Fourier series of ε_c and ε_v.

relevant to the series. Moreover, we note that the higher order term in the sum are exponentially suppressed. This implies that we have transformed the numerical integration problem into a fast convergent summation.

To calculate the Fourier coefficients, we can approximate the integral in Eq. (C5) as a Riemann sum by discretizing the momentum space into a uniform grid. The calculation of Riemann sum is the same as calculating the discrete Fourier transformation which can be very efficiently evaluated by the fast-Fourier transformation. In this numerical approach, the tight-binding Hamiltonian only needs to be diagonalized once in constructing the uniform grid. Depending on the smoothness of the band structure, typically, the grid size greater than 50×50 points is good enough for a desirable convergent result (see Fig. 7). In this paper, we use 100×100 grid points for the calculation.

Although we have used a straightforward method with fast-Fourier transformation. The idea of our method is essentially the same as K-point sampling in Refs. [73–75]. The K-point sampling method is much more efficient since it utilizes all the symmetry in the function and regrouping the Fourier series into a faster convergent series. The Fourier coefficient in the series can be very efficiently calculated by the Monkhorst-Pack grid in the reduced Brillouin zone. This method was originally discussed in Ref. [75] as a "hybrid method."

We can further simplify the calculation in Eq. (C2) if only the low-energy exciton is in our interest. As indicated in Fig. 2, only the low-energy electronic modes (red/blue shaded region) which are well described by the $\mathbf{k} \cdot \mathbf{p}$ model are relevant for exciton binding. In this low-energy regime, one may approximate $\varepsilon_{c/v}$ by expanding it into polynomial. Thus, in this approach, we can use the following identity to calculate Eq. (C2) analytically,

$$\int dk k^l \frac{e^{-k^2} H_m(k) H_n(k)}{\sqrt{\pi m! n! 2^{m+n}}}$$

$$= \sqrt{\frac{n!}{m!}} \sum_{r=0}^{\lfloor l/2 \rfloor} \sum_{s=0}^{\min[m, l-2r]} \binom{m}{s} \frac{2^{s-l-\frac{1}{2}(m-n)} l!}{r!(l-2r-s)!} \delta_{l+m-2r-2s, n}, \quad \text{(C7)}$$

TABLE III. Comparison of binding energies in meV as obtained from the harmonic oscillator basis against analytical and calculated results [49] for suspended MoS_2. Basis size used in the comparison with MoS_2 monolayer corresponded to $N_{max} = 12$ (basis size=91 states) and λ was optimized. For the 2D hydrogen atom with a reduced effective mass of $\mu = 0.14$ and $\epsilon = 9$ the basis size used for the comparison was $N_{max} = 20$ for every states in the table except the $1s$. As for the $1s$ state a greater basis size of $N_{max} = 24$ as used.

	HO basis	2D hydrogen	HO basis	Suspended MoS_2 [49]
$0s$	92.5	94.2	554	555
$1p_{x,y}$	10.38	10.47	315	316
$1s$	9.6	10.47	257	258
$2d_{xy}$	3.76	3.77	209	209
$2p_{x,y}$	3.69	3.77	184	185

where $\lfloor l/2 \rfloor$ is the largest integer that is equal or smaller than $l/2$.

Interaction matrix. In this paper, we assume in-plane rotational symmetry in the e-h interaction. Hence, the k integration in Eq. (C3) can be carried out explicitly by using $H_n(x + y) = \sum_{s=0}^{n} \binom{n}{s} H_s(x)(2y)^{n-s}$ and this yields

$$\mathcal{V}_{n_x n_y; n'_x n'_y} = \int \frac{q dq}{(2\pi)^2} V(q) e^{-\frac{1}{4} q^2 \lambda^2} \prod_{j}^{x,y} \sum_{s_j=0}^{\min[m_j, n_j]} (\lambda q)^{\sigma_j},$$

$$\times \binom{n_j}{s_j} \binom{m_j}{s_j} \frac{(-\frac{1}{2})^{\frac{1}{2}\sigma_j} s_j!}{\sqrt{m_j! n_j!}} 2B\left(\frac{\sigma_x + 1}{2}, \frac{\sigma_y + 1}{2}\right),$$
(C8)

where $\sigma_j = n_j + m_j - 2s_j$ and $B(x, y)$ is the beta function. For the Keldysh potential, $V(q) = -\frac{2\pi e^2}{\sqrt{\kappa_\parallel \kappa_z} q (1 + r_* q)}$, we have the following analytical expression for

$$\int \frac{q dq}{(2\pi)^2} V(q) (\lambda q)^{\sigma_x + \sigma_y} e^{-\frac{1}{4} q^2 \lambda^2}$$

$$= \frac{-e^2}{2\pi \sqrt{\kappa_\parallel \kappa_z}} \left(-\frac{\lambda}{r_*}\right)^{\sigma_x + \sigma_y} \left\{ \frac{e^{-\frac{\lambda^2}{4r_*^2}}}{2r_*/\lambda} \left[\pi \,\mathrm{erf}\left(\frac{\lambda}{2r_*}\right) - \mathrm{Ei}\left(\frac{\lambda^2}{4r_*^2}\right)\right] \right.$$

$$\left. - \sum_{j=0}^{\sigma_x + \sigma_y - 1} \Gamma\left(\frac{j+1}{2}\right) \left(-\frac{2r_*}{\lambda}\right)^j \right\},$$

where $\Gamma(x)$ is the gamma function, $\mathrm{Ei}(x)$ is the exponential integral, and $\mathrm{erf}(x)$ is the error function.

Comparison with hydrogenlike exciton levels for $V \propto -1/r$ and Keldysh interaction. As shown in Table III, in comparing the binding energy as obtained from the harmonic oscillator basis with the analytically obtained 2D hydrogen atom energy levels, the discrepancy between the two was found smaller than 2% as for the ground-state energy and even lower for the states with $l \neq 0$. The higher excited states with $l = 0$ required a very large basis size in order to accurately calculate the binding energy due to the very sharp singularity of the wave function appearing at $r = 0$ (Kato cusp). This situation is similar to the well-known problem in the Slater-type versus Gaussian-type orbitals in quantum chemistry [76], since the

harmonic oscillator is essentially a Gaussian basis. Such a sharp feature in the excitonic wave function is mitigated in the Keldysh potential as the $1/r$ divergence becomes logarithmic. In this case, the harmonic oscillator basis yields better accuracy for each binding state in the spectrum. In comparing our binding energy calculation with the calculated bindings for MoS_2, the error was significantly reduced for the same basis size with <0.3% as for the ground state and lower for the $l \neq 0$ states.

Connection to the real-space formalism. It is also instructive to describe the excitonic problem in terms of real space. To do this, we can Fourier transform the Bethe-Salpeter equation in (1) by using

$$\psi(\mathbf{r}_e, \mathbf{r}_h) = \sum_{\mathbf{k}_e, \mathbf{k}_h} \psi_\mathbf{Q}(\mathbf{k}) e^{i(\mathbf{k}_e \cdot \mathbf{r}_e - \mathbf{k}_h \cdot \mathbf{r}_h)}.$$
(C9)

We remind that $\mathbf{k}_e = \mathbf{k}$ and $\mathbf{k}_h = \mathbf{k} - \mathbf{Q}$. This transformation turns all momentum in the dispersion in Eq. (1) into derivative operators and yields

$$[\varepsilon_c(-i\nabla_{\mathbf{r}_e}) - \varepsilon_v(i\nabla_{\mathbf{r}_h}) - \Omega + V(\mathbf{r}_e - \mathbf{r}_h)]\psi(\mathbf{r}_e, \mathbf{r}_h) = 0,$$
(C10)

where $V(\mathbf{r}) = \int_\mathbf{q} e^{i\mathbf{r}\cdot\mathbf{q}} V(\mathbf{q})$ is the Fourier transformation of the potential. The above equation yields the Mott-Wannier model if only the quadratic mass term in $\varepsilon_{c/v}$ is kept. However, in our model, we need to retain higher-order terms in the hole dispersion.

Similar to the Mott-Wannier model, Eq. (C10) can be reduced to a one-body problem by using the canonical transformation. Since the hole effective mass is not well defined due to the sombrero-shaped dispersion, instead of using the center-of-mass frame coordinate system, we choose

$$\begin{bmatrix} X \\ x \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{r}_e + \mathbf{r}_h \\ \mathbf{r}_e - \mathbf{r}_h \end{bmatrix}, \quad \begin{bmatrix} \hat{\mathbf{P}} \\ \hat{\mathbf{p}} \end{bmatrix} = -\frac{i}{\sqrt{2}} \begin{bmatrix} \nabla_{\mathbf{r}_e} + \nabla_{\mathbf{r}_h} \\ \nabla_{\mathbf{r}_e} - \nabla_{\mathbf{r}_h} \end{bmatrix}.$$

The crucial requirement for this transformation is that the new coordinate system satisfies $[x_j, \hat{p}_j] = [X_j, \hat{P}_j] = i$ such that the physical phase space volume is preserved. Using the (X, x) coordinate, Eq. (C10) in the real space reads

$$\left[\varepsilon_c\left(\frac{\hat{\mathbf{p}} + \hat{\mathbf{P}}}{\sqrt{2}}\right) - \varepsilon_v\left(\frac{\hat{\mathbf{p}} - \hat{\mathbf{P}}}{\sqrt{2}}\right) - \Omega - V(\sqrt{2}x)\right]\psi(X, x) = 0.$$
(C11)

We note that, in this coordinate system, the correspondence between momentum and real space representation of the exciton momentum is $\mathbf{Q} = \mathbf{k}_e - \mathbf{k}_h \leftrightarrow -i\nabla_{\mathbf{r}_e} - i\nabla_{\mathbf{r}_h} = \sqrt{2}\hat{\mathbf{P}}$.

First, it is noted that $[\hat{\mathbf{P}}, H] = 0$ where H (independent of X) is the electron-hole two-particle Hamiltonian in (C10) indicating that $\mathbf{P}$ is a well-defined quantum number which gives the exciton momentum $\mathbf{Q} = \sqrt{2}\mathbf{P}$. Therefore, the wave function is uniquely dependent on x:

$$\psi(X, x) = e^{i\mathbf{P}\cdot X} \chi(x),$$
(C12)

which is the eigenfunction of H. Substituting the above ansatz wave function into Eq. (C10), we reduce the equation into a one-body Shrödinger equation as

$$\left[\varepsilon_c \left(\frac{\sqrt{2}\hat{\mathbf{p}} + \mathbf{Q}}{2} \right) \right.$$
$$\left. - \varepsilon_v \left(\frac{\sqrt{2}\hat{\mathbf{p}} - \mathbf{Q}}{2} \right) - \Omega + V(\sqrt{2}\mathbf{x}) \right] \chi(\mathbf{x}) = 0. \quad \text{(C13)}$$

Expanding $\chi(\mathbf{x})$ into the harmonic oscillator basis as $\chi(\mathbf{x}) = \sum_{\mathbf{n}} \mathcal{C}^{\mathbf{Q}}_{n_x n_y} \varphi_{n_x}(\rho_x) \varphi_{n_y}(\rho_y)$ where $\boldsymbol{\rho} = \mathbf{r}_e - \mathbf{r}_h = \sqrt{2}\mathbf{x}$ is the relative coordinate of electron and hole. The real-space basis function, $\varphi_n(\rho)$, is the Fourier transformation of $\varphi_n(k)$ which is also a harmonic oscillator. Therefore, the matrix representation for Eq. (C13) is

$$\sum_{n'_x n'_y} \left[\mathcal{H}^0_{n_x n_y; n'_x n'_y} + \mathcal{V}_{n_x n_y; n'_x n'_y} \right] \mathcal{C}^{\mathbf{Q}}_{n'_x n'_y} = \Omega \mathcal{C}^{\mathbf{Q}}_{n_x n_y}, \quad \text{(C14)}$$

with the kinetic Hamiltonian,

$$\mathcal{H}^0_{n_x n_y; n'_x n'_y}(\mathbf{Q}) = \int d^2\rho \, \varphi_{n_x}(\rho_x) \varphi_{n_y}(\rho_y) \left[\varepsilon_c \left(\frac{\sqrt{2}\hat{\mathbf{p}} + \mathbf{Q}}{2} \right) \right.$$
$$\left. - \varepsilon_v \left(\frac{\sqrt{2}\hat{\mathbf{p}} - \mathbf{Q}}{2} \right) \right] \varphi_{n'_x}(\rho_x) \varphi_{n'_y}(\rho_y), \quad \text{(C15)}$$

and the Coulomb interaction matrix,

$$\mathcal{V}_{n_x n_y; n'_x n'_y} = \int d^2\rho \, V(\boldsymbol{\rho}) \varphi_{n_x}(\rho_x) \varphi_{n_y}(\rho_y) \varphi_{n'_x}(\rho_x) \varphi_{n'_y}(\rho_y).$$

The integration in Eq. (C15) can be carried out exactly by using the chain rule to rewrite $\hat{\mathbf{p}} = -i\nabla_x = \sqrt{2}(-i\nabla_\rho)$ and then using the recursive relation of the Hermite functions. Alternatively, one may also calculate it by turning $-i\nabla_\rho$ into the simple harmonic ladder operators and carrying out the commutation algebra. Nevertheless, the calculated result from both methods is identical to the momentum space calculation in Eq. (C7).

[1] G. R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M. S. Strano, V. R. Cooper, L. Liang, S. G. Louie, E. Ringe, W. Zhou, S. S. Kim, R. R. Naik, B. G. Sumpter, H. Terrones, F. Xia, Y. Wang, J. Zhu, D. Akinwande, N. Alem, J. A. Schuller, R. E. Schaak, M. Terrones, and J. A. Robinson, ACS Nano **9**, 11509 (2015).

[2] X. Qian, Y. Wang, W. Li, J. Lu, and J. Li, 2D Mater. **2**, 032003 (2015).

[3] K. S. Novoselov, A. Mishchenko, A. Carvalho, and A. H. C. Neto, Science **353**, aac9439 (2016).

[4] G. W. Mudd, S. A. Svatek, T. Ren, A. Patanè, O. Makarovsky, L. Eaves, P. H. Beton, Z. D. Kovalyuk, G. V. Lashkarev, Z. R. Kudrynskyi *et al.*, Adv. Mater. **25**, 5714 (2013).

[5] S. Lei, L. Ge, S. Najmaei, A. George, R. Kappera, J. Lou, M. Chhowalla, H. Yamaguchi, G. Gupta, R. Vajtai, A. D. Mohite, and P. M. Ajayan, ACS Nano **8**, 1263 (2014).

[6] M. Brotons-Gisbert, D. Andres-Penares, J. Suh, F. Hidalgo, R. Abargues, P. J. Rodríguez-Cantó, A. Segura, A. Cros, G. Tobias, E. Canadell, P. Ordejón, J. Wu, J. P. Martínez-Pastor, and J. F. Sánchez-Royo, Nano Lett. **16**, 3221 (2016).

[7] D. A. Bandurin, A. V. Tyurnina, G. L. Yu, A. Mishchenko, V. Zólyomi, S. V. Morozov, R. K. Kumar, R. V. Gorbachev, Z. R. Kudrynskyi, S. Pezzini, Z. D. Kovalyuk, U. Zeitler, K. S. Novoselov, A. Patanè, L. Eaves, I. V. Grigorieva, V. I. Fal'ko, A. K. Geim, and Y. Cao, Nat. Nanotechnol. **12**, 223 (2016).

[8] D. J. Terry, V. Zólyomi, M. Hamer, A. V. Tyurnina, D. G. Hopkinson, A. M. Rakowski, S. J. Magorrian, N. Clark, Y. M. Andreev, O. Kazakova, K. Novoselov, S. J. Haigh, V. I. Fal'ko, and R. Gorbachev, 2D Mater. **5**, 041009 (2018).

[9] J. F. Sánchez-Royo, G. Muñoz-Matutano, M. Brotons-Gisbert, J. P. Martínez-Pastor, A. Segura, A. Cantarero, R. Mata, J. Canet-Ferrer, G. Tobias, E. Canadell, J. Marqués-Hueso, and B. D. Gerardot, Nano Research **7**, 1556 (2014).

[10] M. J. Hamer, J. Zultak, A. V. Tyurnina, V. Zólyomi, D. Terry, A. Barinov, A. Garner, J. Donoghue, A. P. Rooney, V. Kandyba, A. Giampietri, A. Graham, N. Teutsch, X. Xia, M. Koperski, S. J. Haigh, V. I. Fal'ko, R. V. Gorbachev, and N. R. Wilson, ACS Nano **13**, 2136 (2019).

[11] J. Zultak, S. J. Magorrian, M. Koperski, A. Garner, M. J. Hamer, E. Tóvári, K. S. Novoselov, A. A. Zhukov, Y. Zou, N. R. Wilson, S. J. Haigh, A. V. Kretinin, V. I. Fal'ko, and R. Gorbachev, Nat. Commun. **11**, 125 (2020).

[12] X. Li, M.-W. Lin, A. A. Puretzky, J. C. Idrobo, C. Ma, M. Chi, M. Yoon, C. M. Rouleau, I. I. Kravchenko, D. B. Geohegan, and K. Xiao, Sci. Rep. **4**, 5497 (2014).

[13] Z. B. Aziza, D. Pierucci, H. Henck, M. G. Silly, C. David, M. Yoon, F. Sirotti, K. Xiao, M. Eddrief, J.-C. Girard, and A. Ouerghi, Phys. Rev. B **96**, 035407 (2017).

[14] Z. B. Aziza, V. Zólyomi, H. Henck, D. Pierucci, M. G. Silly, J. Avila, S. J. Magorrian, J. Chaste, C. Chen, M. Yoon, K. Xiao, F. Sirotti, M. C. Asensio, E. Lhuillier, M. Eddrief, V. I. Fal'ko, and A. Ouerghi, Phys. Rev. B **98**, 115405 (2018).

[15] A. Budweg, D. Yadav, A. Grupp, A. Leitenstorfer, M. Trushin, F. Pauly, and D. Brida, Phys. Rev. B **100**, 045404 (2019).

[16] D. V. Rybkovskiy, N. R. Arutyunyan, A. S. Orekhov, I. A. Gromchenko, I. V. Vorobiev, A. V. Osadchy, E. Y. Salaev, T. K. Baykara, K. R. Allakhverdiev, and E. D. Obraztsova, Phys. Rev. B **84**, 085314 (2011).

[17] D. V. Rybkovskiy, A. V. Osadchy, and E. D. Obraztsova, Phys. Rev. B **90**, 235302 (2014).

[18] V. Zólyomi, N. D. Drummond, and V. I. Fal'ko, Phys. Rev. B **87**, 195403 (2013).

[19] V. Zólyomi, N. D. Drummond, and V. I. Fal'ko, Phys. Rev. B **89**, 205416 (2014).

[20] M. Zhou, R. Zhang, J. Sun, W.-K. Lou, D. Zhang, W. Yang, and K. Chang, Phys. Rev. B **96**, 155430 (2017).

[21] M. Zhou, D. Zhang, S. Yu, Z. Huang, Y. Chen, W. Yang, and K. Chang, Phys. Rev. B **99**, 155402 (2019).

[22] Y. Sun, S. Luo, X.-G. Zhao, K. Biswas, S.-L. Li, and L. Zhang, Nanoscale **10**, 7991 (2018).

[23] S. J. Magorrian, V. Zólyomi, and V. I. Fal'ko, Phys. Rev. B **94**, 245431 (2016).

[24] T. Cao, Z. Li, and S. G. Louie, Phys. Rev. Lett. **114**, 236602 (2015).

[25] M. Trushin, Phys. Rev. B **99**, 205307 (2019).

[26] B. Skinner, Phys. Rev. B **93**, 235110 (2016).

[27] G. W. Mudd, M. R. Molas, X. Chen, V. Zólyomi, K. Nogajewski, Z. R. Kudrynskyi, Z. D. Kovalyuk, G. Yusa, O. Makarovsky, L. Eaves, M. Potemski, V. I. Fal'ko, and A. Patanè, Sci. Rep. **6**, 39619 (2016).

[28] S. J. Magorrian, A. Ceferino, V. Zólyomi, and V. I. Fal'ko, Phys. Rev. B **97**, 165304 (2018).

[29] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[30] J. Jiang, R. Saito, G. G. Samsonidze, A. Jorio, S. G. Chou, G. Dresselhaus, and M. S. Dresselhaus, Phys. Rev. B **75**, 035407 (2007).

[31] M. L. Trolle, G. Seifert, and T. G. Pedersen, Phys. Rev. B **89**, 235410 (2014).

[32] F. Wu, F. Qu, and A. H. MacDonald, Phys. Rev. B **91**, 075310 (2015).

[33] T. Deilmann and K. S. Thygesen, 2D Mater. **6**, 035003 (2019).

[34] L. V. Keldysh, Sov. J. Expt. Theor. Phys. Lett. **29**, 658 (1979).

[35] N. S. Rytova, Moscow Univ. Phys. Bull. **3**, 18 (1967).

[36] S. Latini, T. Olsen, and K. S. Thygesen, Phys. Rev. B **92**, 245123 (2015).

[37] M. L. Trolle, T. G. Pedersen, and V. Véniard, Sci. Rep. **7**, 39844 (2017).

[38] R. Geick, C. H. Perry, and G. Rupprecht, Phys. Rev. **146**, 543 (1966).

[39] A. Laturia, M. L. Van de Put, and W. G. Vandenberghe, npj 2D Mater. Appl. **2**, 6 (2018).

[40] N. Kuroda and Y. Nishina, Solid State Commun. **34**, 481 (1980).

[41] P. Cudazzo, I. V. Tokatly, and A. Rubio, Phys. Rev. B **84**, 085406 (2011).

[42] $r_* \approx L \times 7.7$ Å.

[43] F. García Flórez, L. D. A. Siebbeles, and H. T. C. Stoof, arXiv:2002.5921.

[44] The bound electron-hole states in an exciton resemble the two-body system in a hydrogen atom. Therefore, these bound states may be well described by an atomic orbital basis set. For the sake of numerical simplicity, we use harmonic oscillator basis function which is an orthogonal Gaussian-type orbital basis.

[45] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.101.245432 for the software package. The instruction is included in the file.

[46] In the expansion, one has to ensure that the truncated polynomials must not allow $\varepsilon_c(\mathbf{k}_e) - \varepsilon_v(\mathbf{k}_h)$ smaller than the bandgap for all $\mathbf{k}_{e/h}$.

[47] The degeneracies for the non-s-wave states are lifted by less than 1 meV in the tight-binding model which only has the sixfold rotational symmetry.

[48] Z. Ye, T. Cao, K. O'Brien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, Nature **513**, 214 (2014).

[49] J.-Z. Zhang and J.-Z. Ma, J. Phys. Condens. Matter **31**, 105702 (2019).

[50] E. Kress-Rogers, R. Nicholas, J. Portal, and A. Chevy, Solid State Commun. **44**, 379 (1982).

[51] N. Kuroda, Y. Nishina, H. Hori, and M. Date, J. Phys. Soc. Jpn. **47**, 1373 (1979).

[52] A. Schindlmayr, Eur. J. Phys. **18**, 374 (1997).

[53] H. Fang, C. Battaglia, C. Carraro, S. Nemsak, B. Ozdol, J. S. Kang, H. A. Bechtel, S. B. Desai, F. Kronast, A. A. Unal, G. Conti, C. Conlon, G. K. Palsson, M. C. Martin, A. M. Minor, C. S. Fadley, E. Yablonovitch, R. Maboudian, and A. Javey, Proc. Natl. Acad. Sci. **111**, 6198 (2014).

[54] P. Rivera, J. R. Schaibley, A. M. Jones, J. S. Ross, S. Wu, G. Aivazian, P. Klement, K. Seyler, G. Clark, N. J. Ghimire, J. Yan, D. G. Mandrus, W. Yao, and X. Xu, Nat. Commun. **6**, 6242 (2015).

[55] P. Rivera, K. L. Seyler, H. Yu, J. R. Schaibley, J. Yan, D. G. Mandrus, W. Yao, and X. Xu, Science **351**, 688 (2016).

[56] Z. Wang, Y.-H. Chiu, K. Honz, K. F. Mak, and J. Shan, Nano Lett. **18**, 137 (2017).

[57] A. Ciarrocchi, D. Unuchek, A. Avsar, K. Watanabe, T. Taniguchi, and A. Kis, Nat. Photonics **13**, 131 (2018).

[58] E. V. Calman, M. M. Fogler, L. V. Butov, S. Hu, A. Mishchenko, and A. K. Geim, Nat. Commun. **9**, 1895 (2018).

[59] N. Ubrig, E. Ponomarev, J. Zultak, D. Domaretskiy, V. Zólyomi, D. Terry, J. Howarth, I. Gutiérrez-Lezama, A. Zhukov, Z. R. Kudrynskyi, Z. D. Kovalyuk, A. Patané, T. Taniguchi, K. Watanabe, R. V. Gorbachev, V. I. Fal'ko, and A. F. Morpurgo, Nat. Mater. **19**, 299 (2020).

[60] K. F. Mak and J. Shan, Nat. Nanotechnol. **13**, 974 (2018).

[61] M. M. Fogler, L. V. Butov, and K. S. Novoselov, Nat. Commun. **5**, 4555 (2014).

[62] P. Li and I. Appelbaum, Phys. Rev. B **92**, 195129 (2015).

[63] D. A. Ruiz-Tijerina, M. Danovich, C. Yelgel, V. Zólyomi, and V. I. Fal'ko, Phys. Rev. B **98**, 035411 (2018).

[64] V. Fiorentini and A. Baldereschi, Phys. Rev. B **51**, 17196 (1995).

[65] K. A. Johnson and N. W. Ashcroft, Phys. Rev. B **58**, 15548 (1998).

[66] N. Bernstein, M. J. Mehl, and D. A. Papaconstantopoulos, Phys. Rev. B **66**, 075212 (2002).

[67] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B **76**, 165106 (2007).

[68] D. Pashov, S. Acharya, W. R. Lambrecht, J. Jackson, K. D. Belashchenko, A. Chantis, F. Jamet, and M. van Schilfgaarde, Comput. Phys. Commun. **249**, 107065 (2020).

[69] Y. Li, T. Wang, M. Wu, T. Cao, Y. Chen, R. Sankar, R. K. Ulaganathan, F. Chou, C. Wetzel, C.-Y. Xu, S. G. Louie, and S.-F. Shi, 2D Materials **5**, 021002 (2018).

[70] C. Song, F. Fan, N. Xuan, S. Huang, G. Zhang, C. Wang, Z. Sun, H. Wu, and H. Yan, ACS Appl. Mater. Interfaces **10**, 3994 (2018).

[71] J. Rigoult, A. Rimsky, and A. Kuhn, Acta Crystallogr. B **36**, 916 (1980).

[72] J. Camassel, P. Merle, H. Mathieu, and A. Chevy, Phys. Rev. B **17**, 4718 (1978).

[73] D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973).

[74] S. L. Cunningham, Phys. Rev. B **10**, 4988 (1974).

[75] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[76] A. L. Magalhães, J. Chem. Educ. **91**, 2124 (2014).