Electrical Breakdown of Excitonic Insulators

Yuelin Shao[1,2,*] and Xi Dai[3,†]

[1]*Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*Department of Physics, The Hongkong University of Science and Technology, Clear Water Bay, Kowloon 999077, Hong Kong, China*

(Received 15 April 2023; revised 18 March 2024; accepted 3 May 2024; published 18 June 2024)

We propose a new electrical breakdown mechanism for exciton insulators in the BCS limit, which differs fundamentally from the Zener breakdown mechanism observed in traditional band insulators. Our new mechanism results from the instability of the many-body ground state for exciton condensation, caused by the strong competition between the polarization and condensation energies in the presence of an electric field. We refer to this mechanism as "many-body breakdown." To investigate this new mechanism, we propose a BCS-type trial wave function under finite electric fields and use it to study the many-body breakdown numerically. Our results reveal two different types of electric breakdown behavior. If the system size is larger than a critical value, the Zener tunneling process is first turned on when an electrical field is applied, but the excitonic gap remains until the field strength reaches the critical value of the many-body breakdown, after which the excitonic gap disappears and the system becomes a highly conductive metallic state. However, if the system size is much smaller than the critical value, the intermediate tunneling phase disappears since the many-body breakdown happens before the onset of Zener tunneling. The sudden disappearance of the local gap leads to an "off-on" feature in the current-voltage (I-V) curve, providing a straightforward way to distinguish excitonic insulators from normal insulators.

DOI: 10.1103/PhysRevX.14.021047 Subject Areas: Condensed Matter Physics

I. INTRODUCTION

The excitonic insulator (EI) is an insulating phase where electron-hole pairs condensate [1–3]. Historically, exciton condensation in solid-state systems has been predominantly examined in three distinct types of systems. First, exciton condensation states have been extensively studied in semiconductors with optically pumped electrons and holes [4–6], which is also called exciton-polariton. Although this is essentially a nonequilibrium system, it can be treated as an approximate equilibrium state for a brief period within the lifetime of electrons and holes. The second type of system comprises semimetal materials with equal-sized electron and hole pockets [7–10]. The conservation of electron and hole numbers is ensured by specific symmetries, such as translation symmetry for electron and hole pockets located in different areas of the Brillouin zone or horizontal mirror symmetry for certain two-dimensional materials. The third type of system includes quantum well or double-layer systems separated by an insulating barrier in the middle [11–17]. In these systems, the electrons and holes can be separated on different layers with negligible single-particle tunneling process between them and their densities can be tuned precisely by two independent gates.

We will focus on the third kind of system, where many interesting observations have been reported recently. The real space separation of electrons and holes in these systems provides not only the electrons and holes with a sufficiently long lifetime but also new ways to detect the exciton condensation states, such as perfect Coulomb drag [18–20] and quantum capacitance [17] measurements.

The experimental setup of the double-layer systems, e.g., transition metal dichalcogenides (TMDs) bilayer separated by hexagonal boron nitride (h-BN) or semiconductor quantum well, is illustrated in Fig. 1(a), and the generic model is written as [21–28]

$$H_0 = \sum_{\boldsymbol{k}} \left[\left(\frac{\hbar^2 k^2}{2m_e} - \mu_{\mathrm{ex}} \right) c^\dagger_{e\boldsymbol{k}} c_{e\boldsymbol{k}} - \frac{\hbar^2 k^2}{2m_h} c^\dagger_{h\boldsymbol{k}} c_{h\boldsymbol{k}} \right], \quad (1a)$$

$$H_I = \frac{1}{2\mathcal{V}} \sum_{ss'=eh\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{q}} V_{ss'}(\boldsymbol{q}) c^\dagger_{s\boldsymbol{k}_1} c^\dagger_{s'\boldsymbol{k}_2} c_{s'\boldsymbol{k}_2+\boldsymbol{q}} c_{s\boldsymbol{k}_1-\boldsymbol{q}}, \quad (1b)$$

*ylshao@iphy.ac.cn
†daix@ust.hk

where $c^{\dagger}_{e\boldsymbol{k}}$, $c^{\dagger}_{h\boldsymbol{k}}$ are electron creation operators in the electron and hole layer, respectively, and $\mathcal{V}$ is the area of the 2D system. The single-particle Hamiltonian H_0 describes the electron and hole bands with quadratic dispersion near the valley center with effective mass m_e and m_h. The exciton chemical potential $\mu_{\mathrm{ex}} = eV_b - E_g$ is tuned by the voltage difference V_b between the electron and hole layer. The interlayer and intralayer interaction are taken as the Coulomb ones: $V(r) \equiv V_{s=s'} = e^2/\epsilon r$ and $U(r) \equiv V_{s\neq s'} = e^2/\epsilon\sqrt{r^2+d^2}$ whose Fourier transformations are $V(q) = 2\pi e^2/\epsilon q$, $U(q) = V(q)\mathrm{e}^{-qd}$. ϵ is the dielectric constant.

If the interlayer interaction is absent, the charged bilayer would be expected to exhibit metallic behavior with the coexistence of free electrons and holes. Because of the charge conservations in each layer, the system has a $\mathrm{U}(1) \times \mathrm{U}(1)$ symmetry. However, the attractive interaction $U(r)$ between electrons and holes will drive the system into an exciton condensation state at the charge neutrality point (CNP) which spontaneously breaks the electron-hole U(1) symmetry and leaves only the total charge conservation. In this context, we consider the terms "excitonic insulator" and "exciton condensation" to be synonymous throughout the paper. Furthermore, by tuning the particle-hole density such exciton condensation state will experience a BEC to BCS crossover as illustrated in Fig. 1(b).

Although excitonic insulators have been discussed in the literature for over half a century, very few material systems have been confirmed experimentally to exhibit such exotic states. This is because the exciton condensation only breaks the particle-hole U(1) symmetry, resulting in charge-neutral superfluidity, which is very hard to detect directly through experiments like perfect dragging. In this study, we propose that the excitonic insulator in the BCS limit may possess a unique breakdown mechanism, which can serve as a critical "smoking gun" type of experimental evidence, helping to distinguish an excitonic insulator from ordinary narrow-gap semiconductors.

Recently, there have been experimental evidences showing that the electrical breakdown behavior of an excitonic insulator may largely deviate from the Zener breakdown of normal band insulators [29], e.g., a much smaller critical field strength and an apparent metal-insulator transition [the R-T characteristics show an insulator (metal) feature before (after) the breakdown]. These facts inspire us to investigate the breakdown behaviors of excitonic insulators. The intrinsic breaking-down mechanism for band insulators is attributed to interband Zener tunneling [30–36]. In an infinite system, the total energy becomes unbounded below when a uniform electric field is applied, resulting in the absence of a ground state. However, a finite system can still maintain an insulating stationary state at low electric fields [37,38]. If we take the rigid band assumption and include only the electric field by a positional dependent chemical potential, the single-particle Zener tunneling process can occur when the in-plane bias voltage eFL becomes comparable to the band gap Δ as shown in Fig. 1(c). This means the Zener critical field is inversely proportional to the system size L. To go beyond the rigid band picture, Souza *et al.* [38] considered the polarization of the occupied bands and they found the $1/L$ behavior of the Zener field still stands.

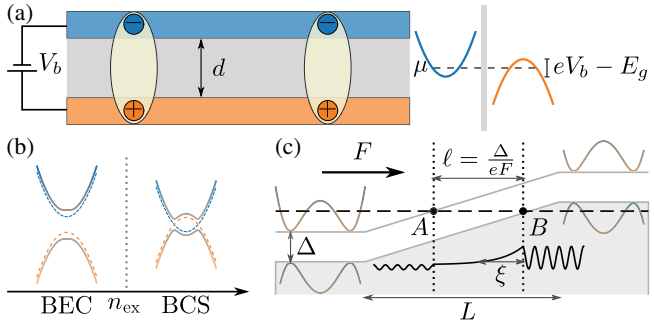

FIG. 1. (a) Setup of the electron-hole bilayer system. The direct gap between electron and hole bands can be tuned by a vertical voltage V_b. (b) At CNP, the interlayer Coulomb interaction between electrons and holes will drive the system into an exciton condensation state and the single-particle gap will be renormalized into an excitonic gap. The electron band is coded by blue and the hole band by orange, where they are mixed when exciton condensation occurs. By increasing the electron and hole densities, the exciton condensation state will experience a BEC to BCS crossover. (c) If an in-plane electrical field F is applied, a Zener tunneling current is expected to appear when the in-plane voltage exceeds the band gap, i.e., $eFL > \Delta$. For any energy-allowed tunneling process, there exists a classically forbidden region (from B to A) with width $\ell = \Delta/eF$ where the wave function decays. The correlation length of the gap ξ characterizes the penetration depth of the wave function into the classically forbidden region.

We would emphasize that this critical field strength denotes the onset of Zener tunneling when a current proportional to the tunneling probability starts to flow. Under WKB approximation, the tunneling probability could be expressed as $\mathrm{e}^{-\ell/\xi}$ [39–41], where ξ is the correlation length determined by the gap Δ and the tunneling length $\ell = \Delta/eF$ is the width of the classically forbidden region for the Zener tunneling process. For an excitonic insulator in the BCS limit, $\xi = 4v_F/\pi\Delta$ is just the coherence length of the exciton condensate (details can be found in Appendix J). When the electric field reaches $\Delta/e\xi$, the tunneling current experiences a sharp increase and the so-called Zener breakdown occurs. Thus, the critical field for Zener breakdown could be roughly estimated as $F^z_c = \Delta/e\xi$.

As pointed out by Zener, the interband tunneling process in normal insulator is just analogous to the autoionization of free atoms by large electric fields [30] and the tunneling probability $\mathrm{e}^{-\ell/\xi} = \mathrm{e}^{-\Delta/eF\xi}$ is similar to the ionization probability of a bound s state with radius ξ and binding energy Δ [42]. In excitonic insulators, the basic ingredients are excitons instead of free atoms, and the Zener breakdown picture will still stand, where the current generation stems

from the field-induced ionization of an exciton to a pair of quasielectron and quasihole. Additionally, there will be a new type of electrical breakdown mechanism which originates from the loss of the stability of the electron-hole pairing ground state due to the competition between polarization and condensation energies; thus we refer to this mechanism as "many-body breakdown."

In this paper, we will show that the many-body breakdown could be interpreted as the collective mode softening in EI whose threshold field strength is much smaller than that of the Zener breakdown in the BCS limit, and the system will encounter the many-body breakdown when the Zener tunneling rate is tiny. Compared to the Zener breakdown, which is due to the ionization probability of each individual exciton, the many-body breakdown is a collective ionization process, in which all the excitons of the system ionize at the same time. Therefore, this unique electric breakdown feature can be considered as an important experimental signal for excitonic insulators, serving as a smoking gun to identify their presence.

II. POLARIZED MEAN-FIELD THEORY

The actual breakdown scenario in excitonic insulators is complex since these two mechanisms could take effect at the same time. To better understand the breakdown of excitonic insulators, we will utilize a self-consistent mean-field theory to analyze the interplay between Zener tunneling and the many-body breakdown.

Although an in-plane field breaks translation symmetry, to describe an insulating ground state, we can always take a trial state that keeps translation symmetry as long as the field is adiabatically added (the proof is in Appendix A). A trial Hartree-Fock (HF) ground state (GS) with translation symmetry at the CNP is $|\mathrm{GS}\rangle = \prod_{\boldsymbol{k}} c^{\dagger}_{v\boldsymbol{k}}|\mathrm{vac}\rangle$, where the valence band $c^{\dagger}_{v\boldsymbol{k}} = \alpha_{\boldsymbol{k}} c^{\dagger}_{e\boldsymbol{k}} + \beta_{\boldsymbol{k}} c^{\dagger}_{h\boldsymbol{k}}$ is a linear combination of the electron and hole band with constraints $|\alpha|^2 + |\beta|^2 = 1$.

Since we choose the exciton chemical potential μ_{ex} as the thermodynamic variable, we are using the grand canonical ensemble for excitons. At zero temperature, the relation between the grand potential and internal energy is $E_G(\mu_{\mathrm{ex}}) = U - \mu_{\mathrm{ex}} N_{\mathrm{ex}}$. By using Dirac notation $|v\boldsymbol{k}\rangle = [\alpha_{\boldsymbol{k}}, \beta_{\boldsymbol{k}}]^T$, the grand potential density becomes a functional of $|v\boldsymbol{k}\rangle$ (see details in Appendix B),

$$\begin{aligned}
&\varepsilon_G[|v\boldsymbol{k}\rangle; F, \mu_{\mathrm{ex}}] \\
&\quad \equiv \frac{1}{\mathcal{V}}\langle \mathrm{GS}|H|\mathrm{GS}\rangle \\
&\quad = \frac{1}{\mathcal{V}}\sum_{s\boldsymbol{k}} h^0_{ss\boldsymbol{k}}\rho_{ss\boldsymbol{k}} + \frac{-eF}{\mathcal{V}\Delta k_{\parallel}}\mathrm{Im}\sum_{\boldsymbol{k}}\log\langle v\boldsymbol{k}|v\boldsymbol{k}+\Delta\boldsymbol{k}_{\parallel}\rangle \\
&\qquad + \frac{2\pi e^2 n_{\mathrm{ex}}^2 d}{\epsilon} - \frac{1}{2\mathcal{V}^2}\sum_{ss'\boldsymbol{k}_1\boldsymbol{k}_2} V_{ss'}(\boldsymbol{k}_1-\boldsymbol{k}_2)\tilde{\rho}_{ss'\boldsymbol{k}_1}\tilde{\rho}_{s's\boldsymbol{k}_2}, \quad (2)
\end{aligned}$$

where $\tilde{\rho} \equiv \rho - \rho^0$ is the density matrix relative to the initial uncharged state $\rho^0_{ss'} = \delta_{ss'}\delta_{sh}$ and ρ is calculated as $\rho_{ss'\boldsymbol{k}} \equiv \langle G|c^{\dagger}_{s'\boldsymbol{k}}c_{s\boldsymbol{k}}|G\rangle = (|v\boldsymbol{k}\rangle\langle v\boldsymbol{k}|)_{ss'}$. The grand potential density depends on μ_{ex} from the single-particle Hamiltonian:

$$h^0_{\boldsymbol{k}} = \begin{bmatrix} \hbar^2k^2/2m_e - \mu_{\mathrm{ex}} & 0 \\ 0 & -\hbar^2k^2/2m_h \end{bmatrix}. \quad (3)$$

So the exciton density n_{ex} is calculated as

$$n_{\mathrm{ex}} = -\partial_{\mu_{\mathrm{ex}}}\varepsilon_G = \frac{1}{\mathcal{V}}\sum_{\boldsymbol{k}}\rho_{ee\boldsymbol{k}}. \quad (4)$$

The four terms in Eq. (2) could be viewed as kinetic, polarization, Hartree, and Fock energies separately. The Hartree energy is just the charging energy of the two-layer capacitor with the charge number density n_{ex}. The relative density matrix $\tilde{\rho}$ is used in the Fock energy expression to avoid the double counting problem [43]. The polarization energy is in principle $-eFP$, where P is electrical polarization which is dependent on the occupied states. For numerical convenience, a periodic boundary condition is assumed, and the polarization is calculated with the help of the expectation value of many-body position operators defined on a ring geometry [44], which is just a discrete form of Berry phase [45–47]:

$$P[|v\boldsymbol{k}\rangle] = \frac{e}{\mathcal{V}\Delta k_{\parallel}}\mathrm{Im}\sum_{\boldsymbol{k}}\log\langle v\boldsymbol{k}|v\boldsymbol{k}+\Delta\boldsymbol{k}_{\parallel}\rangle. \quad (5)$$

This form of polarization energy functional has already been used to calculate the electrical properties of insulators in the literature [38,48,49]. On the other hand, for the open boundary problem, the polarization energy functional should be written in real space by Wannier functions [50,51] which is much more complex technically. However, as long as the system is large enough, the behaviors of the energy functionals for different boundary condition are tested to be identical for topological trivial systems with no edge states [38,50].

The local minimum is found by requiring the first-order derivative of ε_G to be zero, i.e., $\delta\varepsilon_G/\delta\langle v\boldsymbol{k}| = 0$ (details are presented in Appendix C.). This gives the mean-field Hamiltonian $h^{\mathrm{MF}}_{\boldsymbol{k}} \equiv h^0_{\boldsymbol{k}} + h^H + h^F_{\boldsymbol{k}} + h^P_{\boldsymbol{k}}$, where

$$h^H[|v\boldsymbol{k}\rangle] = \frac{4\pi e^2 n_{\mathrm{ex}} d}{\epsilon}(1-\rho^0), \quad (6a)$$

$$h^F_{ss'\boldsymbol{k}}[|v\boldsymbol{k}\rangle] = -\frac{1}{\mathcal{V}}\sum_{\boldsymbol{k}'}V_{s's}(\boldsymbol{k}-\boldsymbol{k}')\tilde{\rho}_{ss'\boldsymbol{k}'}, \quad (6b)$$

$$h^P_{\boldsymbol{k}}[|v\boldsymbol{k}\rangle; F] = \frac{ieF}{2\Delta k_{\parallel}}\sum_{\sigma=\pm}\frac{\sigma|v\boldsymbol{k}+\sigma\Delta\boldsymbol{k}_{\parallel}\rangle\langle v\boldsymbol{k}|}{\langle v\boldsymbol{k}|v\boldsymbol{k}+\sigma\Delta\boldsymbol{k}_{\parallel}\rangle} + \mathrm{H.c.}, \quad (6c)$$

as well as the self-consistent equation,

$$h_{\boldsymbol{k}}^{\mathrm{MF}}[|v\boldsymbol{k}\rangle; F]|v\boldsymbol{k}\rangle = \xi_{v\boldsymbol{k}}|v\boldsymbol{k}\rangle. \tag{7}$$

The definition of the abbreviation σ in Eq. (6c) can be found in Eq. (C10).

In the following, we will take length and energy units as $a_B^* = \epsilon\hbar^2/m^*e^2$ and $\mathrm{Ry}^* = e^2/2\epsilon a_B^*$, where $m^* \equiv m_e m_h/(m_e + m_h)$ is the reduced mass. Then the electrical field strength unit is fixed as $F^* = \mathrm{Ry}^*/ea_B^*$ and the polarization unit is $P^* = e/a_B^*$. At zero temperature, the only independent parameters in the mean-field problem are d/a_B^*, $\mathcal{V}/a_B^2$, F/F^*, and $\mu_{\mathrm{ex}}/\mathrm{Ry}^*$ (or equivalently $n_{\mathrm{ex}}a_B^{*2}$). Typical values of these parameters in a double-layer TMD system, for example, $MoSe_2/WSe_2$ separated by h-BN, are $m_e \approx m_h \approx m = 0.4m_0$, $m^* \approx 0.2m_0$ (m_0 is the bare electron mass) [52], and $\epsilon = 5$ [53]. So the units are calculated as $a_B^* \approx 1.3$ nm, $\mathrm{Ry}^* \approx 108$ meV, and $F^* \approx 8.3 \times 10^5$ V/cm.

III. CRITICAL FIELDS

In the phase diagram depicted in Figs. 2(a) and 2(b), the abscissas represent the system size a_B^*/L_x and exciton density $n_{\mathrm{ex}}a_B^{*2}$ separately, and the vertical axis is the in-plane electric field strength F/F^*. The zero-field band gap Δ^0/Ry (black line, left-hand axis) and the correlation length ξ/a_B^* of the gap (purple line, right-hand axis) estimated by Eq. (J22) are also plotted as functions of system size and exciton density separately in Figs. 2(c) and 2(d).

In the calculation, the interlayer distance is set as $d = 1.875a_B^*$. The momentum space summation in Eq. (2) is restricted in the region $|k_{x,y}| < k_c \approx 2.7a_B^{*-1}$. The numerical results are nearly independent of the cutoff k_c when $k_c \gg k_F$ since the BCS-type condensation occurs only in a small range around $k_F = \sqrt{4\pi n_{\mathrm{ex}}} < 1.3a_B^{*-1}$. The size of the system is defined by the spacing of k mesh as $L = 2\pi/\Delta k$, so the varying of system size is realized by using different sizes of k mesh. The electrical field is applied in the x direction, and the length of the system perpendicular to it is fixed at $L_y = 2\pi N_{k_y}/2k_c \approx 94a_B^*$ ($N_{k_y} = 80$) for numerical convenience. In Figs. 2(a) and 2(c), the exciton density is fixed at $n_{\mathrm{ex}} \approx 0.068a_B^{*-2}$ and the number of k points in the x direction is taken as $N_{k_x} = 40M$ [M is an integer and some used N_{k_x} are marked by red texts above the bottom axis in Fig. 2(c)]. On the contrary, in Figs. 2(b) and 2(d), the system size is fixed (k mesh is fixed at 120×80) and the exciton density varies.

As is shown in Figs. 2(c) and 2(d), the correlation length of the gap ξ [evaluated by Eq. (J22)] is about $4a_B^*$ within the range of the parameters we consider. The correlation length ξ is much smaller than the system size L_x along the direction of the electrical field, which means tunneling current at the onset of Zener tunneling $I \propto \mathrm{e}^{-L_x/\xi}$ is negligible. Additionally, the fact that $L_x \gg \xi$ also indicates that the assumption of the translation symmetry and periodic boundary condition are reasonable.

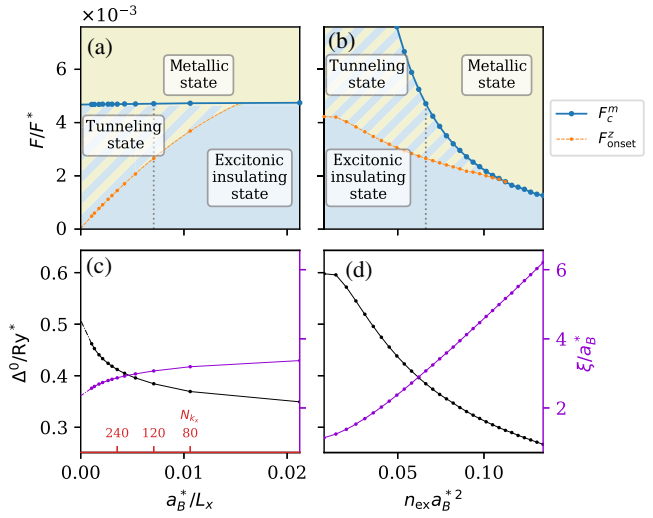

FIG. 2. (a) Phase diagram as a function of the in-plane electrical field F and system size $1/L_x$. (b) Phase diagram as a function of the in-plane electrical field F and exciton density n_{ex}. The critical field F_c^m for many-body breakdown (solid blue lines) firstly divides the entire region into a locally gapped phase and a metallic phase. The onset filed for Zener tunneling F_{onset}^z (dashed orange lines) further separates the locally gapped phase into an excitonic insulating phase and tunneling phase. (c),(d) Zero-field band gap Δ^0 (black lines, left-hand axis) and the correlation length ξ (purple lines, right-hand axis) as functions of system size and exciton density. The red labels above the bottom axis of (c) mark the number of k points used for the corresponding system size.

To overcome the Zener instability of the energy functional for the electrical field in the range $\Delta/eL_x \sim \Delta/e\xi$, the polarization Hamiltonian $h_{\boldsymbol{k}}^P$ and the polarization energy are always evaluated on the coarse 40×80 mesh. For an original $40M \times 80$ k mesh, this is equivalent to dividing the system into M copies with size $L_x = L_0 \approx 47a_B^*$. Thus the Zener tunneling process whose tunneling length ℓ satisfies $ML_0 > \ell > L_0 > \xi$ is ignored. This approximation is reasonable since the tunneling probability $\mathrm{e}^{-\ell/\xi}$ for such process is smaller than $\mathrm{e}^{-L_0/\xi} \approx 10^{-3}$. Although this approximation method was first developed to calculate the higher-order susceptibilities in the zero-field limit [38], this does not mean that the finite field solution has no physical meanings. Souza *et al.* [54] rederived the effective Hamiltonian for polarization Eq. (6c) from the time-dependent dynamics of density matrix, and the solution from the minimization of the energy functional was found to be a resonance state with very long lifetime in the thermodynamic limit.

The blue lines in Figs. 2(a) and 2(b) represent the critical field F_c^m accounting for the many-body breakdown of the excitonic gap, which divides the entire region into a

metallic phase and a locally gapped phase. In the metallic phase, the self-consistent equations have no solutions with excitonic order parameter, while in the locally gapped phase, such solutions always exist. By solving $eFL_x = \Delta(F) \approx \Delta_0$ ($L_x = 2\pi/\Delta k_x$ is the system size assumed in the calculation which is inverse proportional to the k-mesh spacing), the onset field for Zener tunneling F^z_{onset} is obtained and plotted by the orange lines and further separates the locally gapped phase into an excitonic insulating phase and a tunneling phase. In the excitonic insulating phase, the system is fully gapped, and no current flows. In the tunneling phase, an exponentially small Zener tunneling current appears while the system is still locally gapped. In the metallic phase, the excitonic gap is destroyed, the system becomes highly conductive, and the resistivity-temperature (R-T) curve becomes typical metallic.

To illustrate how the critical field F_c^m for the many-body breakdown is extracted, let us investigate the effect of electrical field on some physical quantities. Assume we are in the region of insulating state, so the local minimum $|v\boldsymbol{k}; F\rangle$ of the energy functional Eq. (2) could be found by our self-consistent procedure. The self-consistent equation at the mean-field solution reads $h^{\text{MF}}_{\boldsymbol{k}}[|v\boldsymbol{k}; F\rangle; F]|i\boldsymbol{k}; F\rangle = \xi_{i\boldsymbol{k},F}|i\boldsymbol{k}; F\rangle$, where $|c\boldsymbol{k}; F\rangle, |v\boldsymbol{k}; F\rangle$ are conduction and valence bands with band energies $\xi_{c\boldsymbol{k},F} > \xi_{v\boldsymbol{k},F}$. Then the mean-field gap is just defined as $\Delta(F) = \min(\xi_{c\boldsymbol{k},F} - \xi_{v\boldsymbol{k},F})$. Additionally, the polarization density is obtained from Eq. (5) as $P(F) \equiv P[|v\boldsymbol{k}; F\rangle]$ and the electrical susceptibility could also be defined as $\chi(F) \equiv \partial P(F)/\partial F$.

On a 120×80 k mesh with exciton density $n_{\text{ex}} \approx 0.068 a_B^{*-2}$ [dashed gray line in Figs. 2(a) and 2(b)]. Some physical quantities are plotted in Fig. 3 as functions of field strength. Figure 3(a) shows the mean-field band structure at zero field and the critical field strength. The results indicate that the electrical field has little effect on the band structure and as a result the mean-field gap barely changes with the increase of the field strength, as is shown in Fig. 3(b). To determine the boundary of the locally gapped phase, the polarization P and susceptibility χ are plotted in Figs. 3(c) and 3(d). When approaching the critical field strength F_c^m, χ^{-1} continuously goes to 0, which means the electrical susceptibility χ diverges and the system will transition into a metallic phase. Additionally, the momentum space distributions of the interband coherence h^{MF}_{cv} at $F = 0$ and $F = F_c$ are also shown in Fig. 4. With the increase of electric field strength, the amplitude slightly shrinks while the phase varies dramatically.

IV. FLUCTUATIONS AND COLLECTIVE MODES

In addition to the nonanalytic behaviors of macroscopic physical quantities, the breakdown phase transition could also be understood by examining the stability of the local

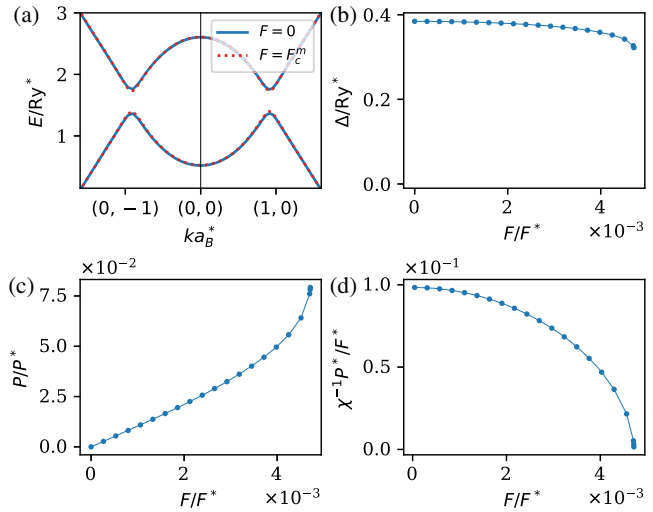

FIG. 3. (a) Mean-field band structure at zero field (blue solid line) and critical field strength (red dotted line). We can see that the electrical field has little effect on the mean-field band structure. (b) Mean-field gap as a function of electrical field strength. When reaching the critical field strength F_c^m, the mean-field gap goes to zero discontinuously. (c) Electrical field induced polarization P as a function of field strength F. Nonanalytic behavior appears when reaching the critical field strength. (d) To see the nonanalytic behavior clearly, inverse of the susceptibility χ^{-1} is plotted. χ^{-1} goes to zero means χ diverges and the system turns into a metallic state. These data are generated on a 120×80 k mesh with exciton density $n_{\text{ex}} \approx 0.068 a_B^{*-2}$ [along the dashed gray line in Figs. 2(a) and 2(b)].

minimum to fluctuations. At the local minimum, the trial HF state with fluctuations could be written as [55]

$$|v'\boldsymbol{k}; F\rangle = (|v\boldsymbol{k}; F\rangle + z_{\boldsymbol{k}}|c\boldsymbol{k}; F\rangle)\Big/\sqrt{1 + |z_{\boldsymbol{k}}|^2}, \quad (8)$$

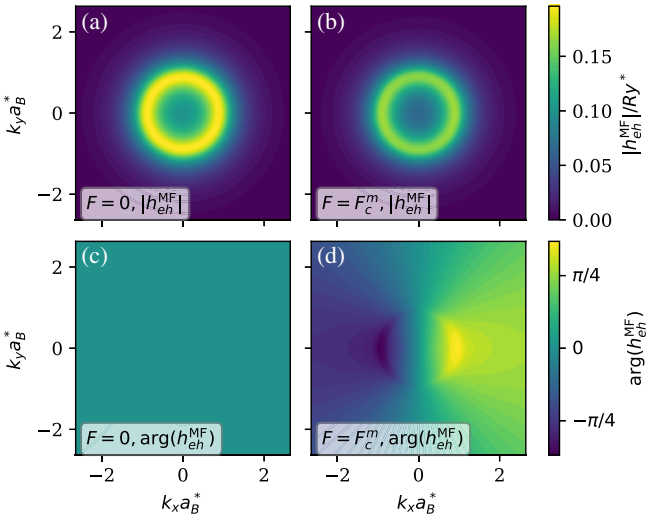

FIG. 4. (a),(b) Amplitude and (c),(d) phase distributions of the exciton order parameter (interband coherence h^{MF}_{eh}) in momentum space at $F = 0$ and $F = F_c^m$. These data are also generated on a 120×80 k mesh with exciton density $n_{\text{ex}} \approx 0.068 a_B^{*-2}$.

where the fluctuation variables $z_{\boldsymbol{k}} = x_{\boldsymbol{k}} + iy_{\boldsymbol{k}}$ are arbitrary complex-valued functions. As derived in Appendix E, when the gauge of the mean-field conduction and valence band wave functions are taken as Eq. (E1), the real and imaginary parts of the fluctuation variables $z_{\boldsymbol{k}}$ are directly related to the density and phase fluctuations of the EI order parameter $\rho_{eh\boldsymbol{k}}$. Then the grand potential becomes a functional of $x_{\boldsymbol{k}}, y_{\boldsymbol{k}}$, i.e., $E_G[x_{\boldsymbol{k}}, y_{\boldsymbol{k}}; F] \equiv E_G[|v'\boldsymbol{k}; F\rangle; F]$, and up to the second order of the fluctuation variables $x_{\boldsymbol{k}}$ and $y_{\boldsymbol{k}}$, the grand potential could be approximated as

$$E_G \approx E_G[|v\boldsymbol{k}; F\rangle] + \sum_{\boldsymbol{k}\boldsymbol{k}'} \left[x_{\boldsymbol{k}} \mathcal{K}^{(+)}_{\boldsymbol{k}\boldsymbol{k}'} x_{\boldsymbol{k}'} + y_{\boldsymbol{k}} \mathcal{K}^{(-)}_{\boldsymbol{k}\boldsymbol{k}'} y_{\boldsymbol{k}'} + 2x_{\boldsymbol{k}} \mathcal{K}^{(X)}_{\boldsymbol{k}\boldsymbol{k}'} y_{\boldsymbol{k}'} \right], \quad (9)$$

where the specific expression of the kernel matrix $\mathcal{K}_{\boldsymbol{k}\boldsymbol{k}'}$ can be found in Appendix D. In the absence of electrical field, the cross term $\mathcal{K}^{(X)}$ is exactly zero, which recovers the expression in Wu *et al.* [24]. However, when the electrical field is added, the density and phase fluctuations will be coupled together and $\mathcal{K}_{\boldsymbol{k}\boldsymbol{k}'}$ is not zero.

Stability of the mean-field ground state against fluctuations requires the eigenvalues of the Hessian matrix,

$$\mathcal{H} \equiv \begin{bmatrix} \mathcal{K}^{(+)} & \mathcal{K}^{(X)} \\ (\mathcal{K}^{(X)})^T & \mathcal{K}^{(-)} \end{bmatrix}, \quad (10)$$

to be non-negative, where the eigenvalues and fluctuation eigenmodes are defined by the eigenvalue equation,

$$\sum_{\boldsymbol{k}'} \mathcal{H}_{\boldsymbol{k}\boldsymbol{k}'} \begin{bmatrix} x^{\lambda}_{\boldsymbol{k}'} \\ y^{\lambda}_{\boldsymbol{k}'} \end{bmatrix} = \lambda \begin{bmatrix} x^{\lambda}_{\boldsymbol{k}} \\ y^{\lambda}_{\boldsymbol{k}} \end{bmatrix}, \quad (11)$$

and the superscript λ in $z^{\lambda}_{\boldsymbol{k}} = x^{\lambda}_{\boldsymbol{k}} + iy^{\lambda}_{\boldsymbol{k}}$ means it is the fluctuation eigenmode with respect to the eigenvalue λ. For convenience, the eigenmodes in the following text are normalized by $z^{\lambda}_{\boldsymbol{k}} \to z^{\lambda}_{\boldsymbol{k}} / \sqrt{\sum_{\boldsymbol{k}} |z^{\lambda}_{\boldsymbol{k}}|^2}$. Still along the dashed gray line in Figs. 2(a) and 2(b), the stability of the ground state is analyzed, and the results are shown in Fig. 5. In Fig. 5(a), we plot the smallest few eigenvalues λ_{0-4} of the Hessian matrix Eq. (10) as functions of field strength. By taking the trial HF state as $|v'\boldsymbol{k}; F; \theta_{\lambda_i}\rangle \propto |v\boldsymbol{k}; F\rangle + \theta_{\lambda_i} z^{\lambda_i}_{\boldsymbol{k}} |c\boldsymbol{k}; F\rangle$, the grand potential difference between the trial state and the HF ground state along the directions $z^{\lambda_i}_{\boldsymbol{k}}$ in the variational parameter space is evaluated as

$$\Delta E_G(F, \theta_{\lambda_i}) \equiv E_G[|v'\boldsymbol{k}; F; \theta_{\lambda_i}\rangle] - E_G[|v\boldsymbol{k}; F\rangle]. \quad (12)$$

Using the lowest three eigenmodes $z^{\lambda_{0,1,2}}_{\boldsymbol{k}}$, for example, the grand potential difference as a function of the electric field F and excitation amplitudes θ_{λ_i} is plotted in Figs. 5(b)–5(d). In these plots, the horizontal axes are the amplitudes of those eigenmodes, while different electric field strengths are represented by different color lines.

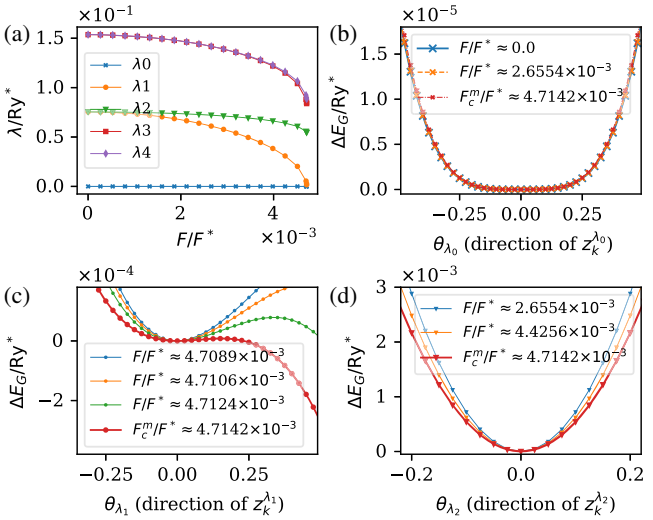

FIG. 5. (a) The smallest five eigenvalues of the Hessian matrix Eq. (10) as a function of the electric field. (b)–(d) Grand potential difference Eq. (12) as a function of the electric field and excitation amplitudes along the directions $z^{\lambda_{0,1,2}}_{\boldsymbol{k}}$ in the variational parameter space. The excitation amplitudes are used as the horizontal axes while different field strengths are represented by different color lines. These data are also generated along the dashed gray line in Figs. 2(a) and 2(b).

There is a consistent zero mode λ_0 for any electric field strength, as shown in Fig. 5(a). However, the behavior of the energy functional along the direction $z^{\lambda_0}_{\boldsymbol{k}}$ in Fig. 5(b) indicates that it is not a "breaking-down mode" because the high-order derivatives of the energy functional along this direction are always positive. Such a zero mode is exactly the Goldstone mode related to phase fluctuation of the exciton condensate and accounts for the exciton superfluidity (see details in Appendix F). The real breaking-down direction in parameter space is $z^{\lambda_1}_{\boldsymbol{k}}$ as shown in Fig. 5(c). When the electric field is small, all eigenvalues of the Hessian matrix (except the Goldstone mode λ_0) satisfy $\lambda > \lambda_1 > 0$, which means the solution is indeed a local minimum. As the electric field approaches the critical field strength F^m_c, the eigenvalue of the breakdown mode λ_1 approaches 0 and the excitonic insulator ground state becomes unstable as the local minimum turns into a saddle point. Further investigations on the breakdown mode $z^{\lambda_1}_{\boldsymbol{k}}$ in Appendix G reveal that it accounts for the polarization fluctuation δP_x which couples with the electrical field in the x direction.

To find the collective modes, we also need to include the fluctuation dynamics. In Appendix H, the dynamics

equation of fluctuation variable $z_{\boldsymbol{k}}$ is derived from the time-dependent HF equation as

$$-\partial_t y_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'}\left[\mathcal{K}^{(+)}_{\boldsymbol{k}\boldsymbol{k}'}x_{\boldsymbol{k}'} + \mathcal{K}^{(X)}_{\boldsymbol{k}\boldsymbol{k}'}y_{\boldsymbol{k}'}\right], \tag{13a}$$

$$\partial_t x_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'}\left[\mathcal{K}^{(X)}_{\boldsymbol{k}'\boldsymbol{k}}x_{\boldsymbol{k}'} + \mathcal{K}^{(-)}_{\boldsymbol{k}\boldsymbol{k}'}y_{\boldsymbol{k}'}\right], \tag{13b}$$

which is consistent with the previous study by Wu *et al.* [24]. After the Fourier transformation to the frequency domain, the collective modes should be obtained by solving the generalized eigenvalue problem:

$$\sum_{\boldsymbol{k}'}\mathcal{H}_{\boldsymbol{k}\boldsymbol{k}'}\begin{bmatrix}x^{\omega}_{\boldsymbol{k}'}\\ y^{\omega}_{\boldsymbol{k}'}\end{bmatrix} = i\omega\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\begin{bmatrix}x^{\omega}_{\boldsymbol{k}}\\ y^{\omega}_{\boldsymbol{k}}\end{bmatrix}. \tag{14}$$

As proved in Appendix H, the eigenvalues ω are either zero or appear in pairs as $\pm\omega$, which are the excitation energies of the collective modes.

In general, the fluctuation eigenmodes solved by Eq. (11) are not necessarily identical with the collective modes solved by Eq. (14). But the fluctuation eigenmode with eigenvalue $\lambda = 0$ is always a collective mode with zero excitation energy $\omega = 0$. This means that when the eigenvalue λ_1 in Fig. 5(a) becomes zero when approaching the critical field F_c^m, there must exist another collective mode with zero excitation energy in addition to the Goldstone mode. In Fig. 6(a), the collective modes spectra in the long wavelength limit (zero momentum excitations) are plotted as functions of exciton density at zero electrical field. Because of the rotational symmetry, the collective modes could be labeled by their angular momentums. In Fig. 6(a), the s-wave collective mode with zero angular momentum is indicated by the blue line with cross markers, which is exactly the zero-energy Goldstone mode. Additionally, the two degenerated p-wave collective modes with angular momentum $l_z = \pm 1$ are indicated by the orange line with dot markers. In Fig. 6(b), the same quantities are plotted as functions of electrical field strength at a fixed exciton density $n_{\mathrm{ex}} \approx 0.068 a_B^{*-2}$. Since the electrical field breaks the rotational symmetry, the degeneracy of the two p-wave collective modes is lifted. And the p_x mode which couples directly with the electrical field in x direction gradually softens when approaching the critical field strength.

In the zero-field limit, due to the angular momentum conservation, only the p-wave modes with angular momentum $l_z = \pm 1$ indicated by the orange lines in Fig. 6 can couple with the electrical field directly. Additionally, the softened p_x mode is highly related to the breakdown mode $z^{\lambda_1}_{\boldsymbol{k}}$ in Fig. 5, which is proven in Appendix G to be the polarization fluctuation δP_x arisen from the relative motion of electrons and holes. As the Goldstone mode can be viewed as the analogy to the acoustical phonon mode of ionic crystals, the breakdown mode is then similar to the optical modes.

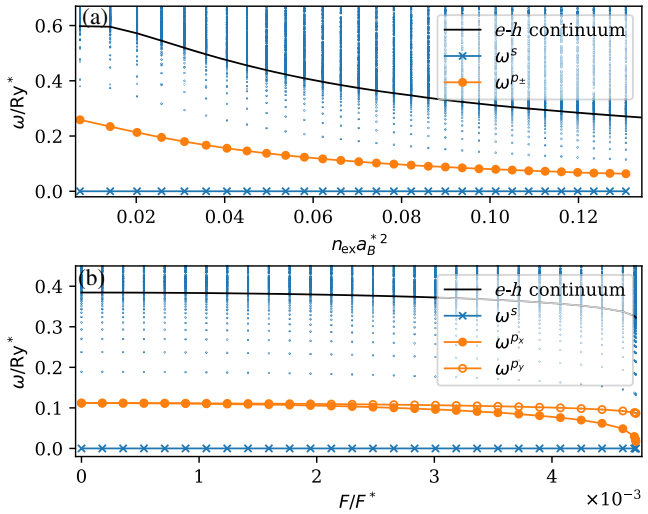

FIG. 6. (a) Collective modes spectra in the long wavelength limit (zero momentum excitations) as functions of the excitation density at zero electrical field. The black solid line represents the mean-field gap, which marks the boundary between collective modes and quasiparticle electron-hole continuum. Because of the rotational symmetry, the collective modes could be labeled by their angular momentum. The s-wave collective mode with zero angular momentum is indicated by the blue line with cross markers. The two degenerated p-wave collective modes with angular momentum $l_z = \pm 1$ are indicated by the orange line with dot markers. Collective modes with higher angular momentums are not explicitly marked. (b) At $n_{\mathrm{ex}} \approx 0.068 a_B^{*-2}$, the collective modes spectra are also plotted as functions of the electrical field. Since the electrical field in x direction breaks the rotational symmetry, the two degenerated p-wave collective modes split into the p_x and p_y modes.

Because of inversion symmetry, the excitation energy of the breakdown mode ω^{p_x} should be an even function of the electrical field strength F. Near zero-field strength, the excitation energy could be approximated by

$$\omega^{p_x}(F) \approx \omega_0^{p_x} - \frac{\eta_0}{2}F^2, \tag{15}$$

where $\omega_0^{p_x}$ is the excitation energy at zero field and $\eta_0 \equiv -\partial_F^2\omega^{p_x}(F)|_{F=0}$ is the polarizability. Then the condition of the many-body breakdown is just $\omega^{p_x} \sim 0$, which means the critical field is approximately $\sqrt{2\omega_0^{p_x}/\eta_0}$. Detailed analyses in Appendix I give a more accurate estimation of the critical field as

$$F_c^m \approx \sqrt{\omega_0^{p_x}/2\eta_0}, \tag{16}$$

which is only related to the zero-field excitation energy $\omega_0^{p_x}$ and its polarizability η_0. Near zero field, the polarizability of the breakdown mode is calculated and plotted in Fig. 7(a) as a function of exciton density. In Fig. 7(b), the critical

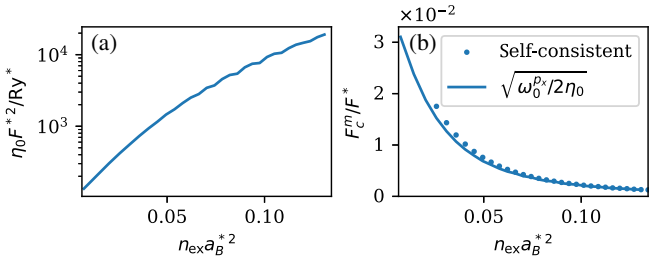

FIG. 7. (a) Zero-field polarizability of the breakdown mode. (b) The blue dots are the critical field calculated from the self-consistent procedure, while the blue line is estimated from the zero-field quantities by Eq. (16), which show good agreement.

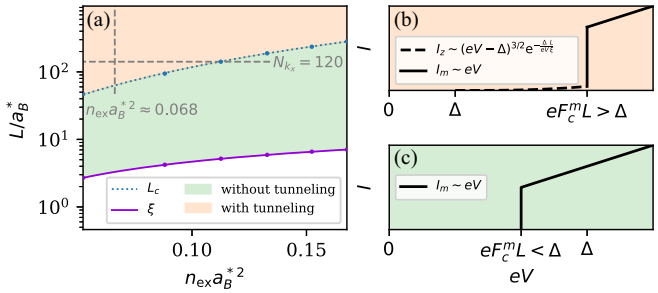

FIG. 8. (a) The critical length L_c where the onset field for Zener tunneling F^z_{onset} equals the critical field F_c^m for the many-body breakdown is plotted as a function of exciton density by the dotted blue line, which separates the $n_{\mathrm{ex}} - L$ plane into two regions, i.e., the green region where the many-body breakdown occurs without Zener tunneling and the orange region where Zener tunneling current appears before the many-body breakdown. The correlation length of the excitonic gap ξ given by Eq. (J22) is also plotted by the purple line for reference. And we only focus on the case with $L > \xi$. The two dashed gray lines mark the paths along which Fig. 2 is generated. (b),(c) I-V characteristics for the excitonic insulator in the two regions in (a).

fields for the many-body breakdown calculated from the self-consistent procedure are also compared with the estimated values by Eq. (16), which show good agreement.

Such a many-body breakdown mechanism is completely different from traditional Zener tunneling and the corresponding critical field strength can be much weaker than the one for Zener tunneling, as discussed in the following section.

V. DISCUSSION

With the increase in exciton density, the zero-field excitation energy of the breakdown mode decreases as shown in Fig. 6(a), while the polarizability grows exponentially as shown in Fig. 6(b). Thus the critical field for the many-body breakdown also decreases dramatically according to the estimation formula Eq. (16). This is reasonable since with the increase in exciton density, the binding between electron and hole becomes weaker and the excitonic insulator will turn into a quantum electron-hole plasma state [56–61].

From the physical picture of Zener tunneling, the tunneling current exists only when the gate voltage is larger than the single-particle gap. Assume the distance between electrodes is L, then at the critical field of many-body breakdown F_c^m, the gate voltage is $eF_c^m L$. Comparing the critical voltage $eF_c^m L$ with the single-particle gap Δ gives a critical value for the electrodes distance:

$$L_c \sim \Delta / eF_c^m. \tag{17}$$

Below the critical distance L_c, there will be no Zener tunneling even when the many-body breakdown occurs, which also indicates that the many-body breakdown mechanism is distinct from the Zener tunneling and breakdown physics. The ratio $\xi/L_c \sim F_c^m/F_c^z$ roughly measures the relative magnitudes between the critical fields of many-body breakdown and Zener breakdown. In Fig. 8(a), L_c and ξ are plotted as functions of exciton density. The ratio ξ/L_c decreases with the increase of exciton density and in the region included in Fig. 8(a) $\xi/L_c \sim 10^{-1}$–10^{-2}, which means the critical field strength for the many-body breakdown is about 10 to 100 times smaller than Zener breakdown. Such a small field is expected to serve as a smoking gun to identify the excitonic gap in the BCS limit.

To study the many-body breakdown, the most ideal case is to avoid the Zener tunneling effect by reducing the electrode distance. In the green colored region in Fig. 8(a), where $L_c > L > \xi$, the excitonic gap is disrupted before the onset of interband Zener tunneling. For small field strength, the system is purely insulating at zero temperature and no current flows. As the electrical field increases, the BCS-type exciton condensation wave function will lose stability and exhibit a typical first-order transition feature. After this transition, the system becomes gapless and highly conductive, and a quasilinear metallic current $I_m \propto eV$ will flow in the system. Thus a discontinuous switching phenomenon is expected in the I-V characteristic as shown in Fig. 8(c).

However, most experimental setups fall into the orange colored region where $L \gg L_c \gg \xi$, and a tunneling current will first appear when the in-plane bias voltage exceeds the band gap. For voltage in the range $\Delta \ll eV < eF_c^m L$, this current is in the form of

$$I_z(eV \equiv eFL) \sim (eV - \Delta)^{3/2} \mathrm{e}^{-(\Delta/\mathrm{eV})(L/\xi)}. \tag{18}$$

The exponential factor $\mathrm{e}^{-\Delta L/eV\xi}$ is the WKB tunneling probability and the power term $(eV-\Delta)^{3/2}$ arises from the density of states of the tunneling channels in 2D systems (details can be found in Appendix J). The tunneling current persists until the field strength reaches the critical field of many-body breakdown, after which the excitonic gap disappears and a metallic current $I_m \propto eV$ appears replacing the Zener tunneling current I_z. However, even at the critical field F_c^m, the tunneling current $I_z(F = F_c^m) \propto \mathrm{e}^{-L_c/\xi}$ in the BCS

limit is still exponentially small as the critical length L_c is nearly 2 orders larger than correlation length ξ, as is shown in Fig. 8(a). This means the discontinuity of the I-V curve like Fig. 8(b) is still observable.

In addition to the smaller value of critical field strength in the BCS limit, from the discussions above we conclude that the discontinuity of the I-V characteristic at nearly zero temperature is also an important feature of the many-body breakdown since the tunneling current increase smoothly in the Zener breakdown picture. This discontinuity arises from the gap closing, and the induced metal-insulator transition could be identified by investigating the R-T characteristic; i.e., before and after the many-body breakdown, the R-T characteristics should behave like a semiconductor and a metal, respectively, while for Zener breakdown, the local gap always exists and the R-T curve always shows a semiconductor feature. Additionally, the gap closing after the many-body breakdown may also be identified by charge compressibility measurements. In the excitonic insulator phase, the system is charge incompressible when chemical potential lies between the gap [17], while in the metallic phase, absence of local gap makes the system charge compressible.

We note that Sugimoto *et al.* [41] also proposed a breaking-down mechanism in correlated insulators which has a threshold field much smaller than that for Zener breakdown. However, the mechanism in their work is distinct from the many-body breakdown mechanism proposed in our work. The many-body breakdown is intrinsic for an excitonic insulator while the critical field in their work is related to the extrinsic relaxation time. Additionally, the typical I-V curve for an excitonic insulator as illustrated in Figs. 8(b) and 8(c) has size dependence which is already observed by the experiments of Yang *et al.* [29].

Finally, the many-body breakdown mechanism is a breakdown of the electronic band structure and has nearly no influence on the lattice, which means the breaking-down process is reversible and the switching phenomenon of the I-V characteristic is promising for practical usage.

ACKNOWLEDGMENTS

We thank Professor Zheng Vitto Han, Naoto Nagaosa, and Wan Yao for their helpful discussions. X. D. acknowledges financial support from the Hong Kong Research Grants Council (Project No. 16309020). The work described in this paper was supported by a fellowship award and a CRF award from the Research Grants Council of the Hong Kong Special Administrative Region, China (Projects No. HKUST SRFS2324-6S01 and No. C7037-22GF).

APPENDIX A: TRIAL STATE

We first prove that under a uniform electric field, the many-body state will keep its lattice translation symmetry at all times.

A many-body state $|\Psi;t\rangle$ is said to have lattice translation symmetry if and only if the wave function satisfies

$$\Psi(\boldsymbol{r}_1+\boldsymbol{R}_0,\ldots,\boldsymbol{r}_{N_e}+\boldsymbol{R}_0;t)=\mathrm{e}^{i\phi}\Psi(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t),$$

where N_e is the total number of electrons and $\boldsymbol{R}_0$ is arbitrary lattice vector.

The many-body Schrödinger equation in length gauge (using a scalar field $\varphi=e\boldsymbol{F}\cdot\boldsymbol{r}$ to include electric field) is written as

$$\begin{aligned}&i\partial_t\Psi^E(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t)\\&\quad=\Bigg\{\sum_{i=1}^{N_e}[h^0(-i\nabla_{\boldsymbol{r}_i},\boldsymbol{r}_i)+e\boldsymbol{F}\cdot\boldsymbol{r}_i]\\&\qquad+\sum_{1\le i<j\le N_e}V(\boldsymbol{r}_i-\boldsymbol{r}_j)\Bigg\}\Psi^E(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t),\end{aligned}\tag{A1}$$

which seems to break lattice translation symmetry. However, by taking gauge transformation of the electric field $\partial_t\boldsymbol{A}(t)=-\boldsymbol{F}$ and defining

$$\Psi^A(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t)=\mathrm{e}^{-i\sum_{i=1}^{N_e}e\boldsymbol{A}(t)\cdot\boldsymbol{r}_i}\Psi^E(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t),\tag{A2}$$

we find that the Schrödinger equation for $|\Psi^A\rangle$ becomes

$$\begin{aligned}&i\partial_t\Psi^A(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t)\\&\quad=\Bigg[\sum_{i=1}^{N_e}h^0[-i\nabla_{\boldsymbol{r}_i}+e\boldsymbol{A}(t),\boldsymbol{r}_i]\\&\qquad+\sum_{1\le i<j\le N_e}V(\boldsymbol{r}_i-\boldsymbol{r}_j)\Bigg]\Psi^A(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t),\end{aligned}\tag{A3}$$

which keeps the lattice translation symmetry. So starting from a many-body state $|\Psi^0\rangle$ with lattice translation symmetry, the many-body state $|\Psi^A;t\rangle$ as well as $|\Psi^E;t\rangle$ will have lattice translation symmetry at any time:

$$\begin{aligned}&\Psi^E(\boldsymbol{r}_1+\boldsymbol{R}_0,\ldots,\boldsymbol{r}_{N_e}+\boldsymbol{R}_0;t)\\&\quad=\mathrm{e}^{i\sum_{i=1}^{N_e}e\boldsymbol{A}(t)\cdot(\boldsymbol{r}_i+\boldsymbol{R}_0)}\Psi^A(\boldsymbol{r}_1+\boldsymbol{R}_0,\ldots,\boldsymbol{r}_{N_e}+\boldsymbol{R}_0;t)\\&\quad=\mathrm{e}^{iN_e e\boldsymbol{A}(t)\cdot\boldsymbol{R}_0+i\phi_A}\mathrm{e}^{i\sum_{i=1}^{N_e}\boldsymbol{A}(t)\cdot\boldsymbol{r}_i}\Psi^A(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t)\\&\quad=\mathrm{e}^{iN_e e\boldsymbol{A}(t)\cdot\boldsymbol{R}_0+i\phi_A}\Psi^E(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{N_e};t).\end{aligned}\tag{A4}$$

When treating a static uniform electric field, as long as the field is adiabatically turned on, a trial HF state with lattice translation symmetry could be safely assumed. For insulators, this state is written as

$$|\mathrm{GS}\rangle=\prod_{n=1}^{n_e}\prod_{\boldsymbol{k}\in\mathrm{BZ}}c_{n\boldsymbol{k}}^\dagger|\mathrm{vac}\rangle,\tag{A5}$$

where n_e is electron per cell and $|\mathrm{vac}\rangle$ is vacuum state. $c_{n\boldsymbol{k}}^\dagger$ is creation operators of Bloch electron with wave function

$$\psi_{nk}(\boldsymbol{r}) = \{\Psi(\boldsymbol{r}), c^\dagger_{nk}\} = \frac{1}{\sqrt{\mathcal{V}}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nk}(\boldsymbol{r}), \tag{A6}$$

where $\mathcal{V} = \mathcal{N} v_c$ is the volume of the system, $\mathcal{N}$ is the number of unit cells, and v_c is cell volume.

As electron creation operators, $c^\dagger_{nk}$ should satisfy

$$\{c^\dagger_{mk}, c_{nk'}\} = \delta_{mn}\delta_{kk'}, \tag{A7}$$

which forces the corresponding Bloch functions to be orthonormal; i.e.,

$$\langle \psi_{mk} | \psi_{nk'} \rangle = \int d\boldsymbol{r}\, \psi^*_{mk}(\boldsymbol{r}) \psi_{nk'}(\boldsymbol{r}) = \delta_{mn}\delta_{kk'}, \tag{A8a}$$

$$\langle u_{mk} | u_{nk} \rangle = \frac{1}{v_c} \int_{\text{cell}} d\boldsymbol{r}\, u^*_{mk}(\boldsymbol{r}) u_{nk}(\boldsymbol{r}) = \delta_{mn}. \tag{A8b}$$

APPENDIX B: POLARIZED HF ENERGY FUNCTIONAL

In this appendix, a general form of the polarized HF energy as a functional of occupied Bloch states will be derived.

Using field operator $\Psi(\boldsymbol{r})$, the second quantization form of the single-particle (kinetic and potential energy), polarization, and interaction Hamiltonians are written as

$$H_0 = \int d\boldsymbol{r}\, \Psi^\dagger(\boldsymbol{r}) h^0(-i\nabla_{\boldsymbol{r}}, \boldsymbol{r}) \Psi(\boldsymbol{r}), \tag{B1}$$

$$H_P = e\boldsymbol{F} \cdot \int d\boldsymbol{r}\, \Psi^\dagger(\boldsymbol{r}) \boldsymbol{r} \Psi(\boldsymbol{r}), \tag{B2}$$

$$H_I = \frac{1}{2} \int d\boldsymbol{r}_1 d\boldsymbol{r}_2 \Psi^\dagger(\boldsymbol{r}_1) \Psi^\dagger(\boldsymbol{r}_2) V(\boldsymbol{r}_1 - \boldsymbol{r}_2) \Psi(\boldsymbol{r}_2) \Psi(\boldsymbol{r}_1). \tag{B3}$$

Matrix elements of the single-particle density operator $\hat{\rho}$ under position basis are calculated as

$$\begin{aligned}\rho(\boldsymbol{r}, \boldsymbol{r}') &= \langle \text{GS} | \Psi^\dagger(\boldsymbol{r}') \Psi(\boldsymbol{r}) | \text{GS} \rangle \\ &= \sum_{n=1}^{n_e} \sum_{\boldsymbol{k} \in \text{BZ}} \{c_{nk}, \Psi^\dagger(\boldsymbol{r}')\} \{\Psi(\boldsymbol{r}), c^\dagger_{nk}\} \\ &= \sum_{n=1}^{n_e} \sum_{\boldsymbol{k} \in \text{BZ}} \psi_{nk}(\boldsymbol{r}) \psi^*_{nk}(\boldsymbol{r}').\end{aligned} \tag{B4}$$

Then its k-dependent counterpart is defined by

$$\hat{\rho}_{\boldsymbol{k}} = \mathcal{N} e^{-i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}} \hat{\rho} e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}} = \sum_{i=1}^{n_e} |u_{nk}\rangle\langle u_{nk}|. \tag{B5}$$

It is important to note that the single-particle Hilbert space $\mathcal{H}$ of $\hat{\rho}$ is all kinds of functions while the Hilbert space $\mathcal{H}_{\boldsymbol{k}}$ of $\hat{\rho}_{\boldsymbol{k}}$ is only the cell-periodic functions. That is why the prefactor $\mathcal{N}$, number of cells, appears in the definition of $\hat{\rho}_{\boldsymbol{k}}$ in Eq. (B5). And we will see the single-particle and interaction energies could be expressed as functionals of $\hat{\rho}_{\boldsymbol{k}}$ and therefore functionals of occupied states $|u_{nk}\rangle$.

The single-particle part is

$$\begin{aligned}E_0 &\equiv \langle \text{GS} | H_0 | \text{GS} \rangle \\ &= \int d\boldsymbol{r} d\boldsymbol{r}' \delta(\boldsymbol{r} - \boldsymbol{r}') h^0(-i\nabla_{\boldsymbol{r}}, \boldsymbol{r}) \rho(\boldsymbol{r}, \boldsymbol{r}') \\ &= \sum_{n=1}^{n_e} \sum_{\boldsymbol{k} \in \text{BZ}} \int d\boldsymbol{r}\, \psi^*_{nk}(\boldsymbol{r}) h^0(-i\nabla_{\boldsymbol{r}}, \boldsymbol{r}) \psi_{nk}(\boldsymbol{r}) \\ &= \sum_{\boldsymbol{k} \in \text{BZ}} \text{Tr}\left[\hat{h}^0_{\boldsymbol{k}} \hat{\rho}_{\boldsymbol{k}}\right],\end{aligned} \tag{B6}$$

where $\hat{h}^0_{\boldsymbol{k}} = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \hat{h}^0(\hat{\boldsymbol{p}}, \hat{\boldsymbol{r}}) e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}} = \hat{h}^0(\hat{\boldsymbol{p}} + \boldsymbol{k}, \hat{\boldsymbol{r}})$ is the k-dependent single-particle Hamiltonian acting on cell-periodic functions with matrix elements:

$$h^0_{mnk} \equiv \frac{1}{v_c} \int_{\text{cell}} d\boldsymbol{r}\, u^*_{mk}(\boldsymbol{k}) h^0(-i\nabla_{\boldsymbol{r}} + \boldsymbol{k}, \boldsymbol{r}) u_{nk}(\boldsymbol{r}). \tag{B7}$$

Similarly, the interaction part is evaluated with the help of Wick's theorem,

$$\begin{aligned}&\langle \text{GS} | H_I | \text{GS} \rangle \\ &= \frac{1}{2} \int d\boldsymbol{r} d\boldsymbol{r}' V(\boldsymbol{r} - \boldsymbol{r}') \langle \Psi^\dagger(\boldsymbol{r}) \Psi^\dagger(\boldsymbol{r}') \Psi(\boldsymbol{r}') \Psi(\boldsymbol{r}) \rangle \\ &= \frac{1}{2} \int d\boldsymbol{r} d\boldsymbol{r}' V(\boldsymbol{r} - \boldsymbol{r}') [\rho(\boldsymbol{r}, \boldsymbol{r}) \rho(\boldsymbol{r}', \boldsymbol{r}') - \rho(\boldsymbol{r}', \boldsymbol{r}) \rho(\boldsymbol{r}, \boldsymbol{r}')] \\ &= \frac{1}{2\mathcal{V}} \sum_{\boldsymbol{q}} V(\boldsymbol{q}) \int d\boldsymbol{r} d\boldsymbol{r}' e^{i\boldsymbol{q}\cdot(\boldsymbol{r} - \boldsymbol{r}')} \rho(\boldsymbol{r}, \boldsymbol{r}) \rho(\boldsymbol{r}', \boldsymbol{r}') \\ &\quad - \frac{1}{2\mathcal{V}} \sum_{\boldsymbol{q}} V(\boldsymbol{q}) \int d\boldsymbol{r} d\boldsymbol{r}' e^{i\boldsymbol{q}\cdot(\boldsymbol{r} - \boldsymbol{r}')} \rho(\boldsymbol{r}', \boldsymbol{r}) \rho(\boldsymbol{r}, \boldsymbol{r}'),\end{aligned} \tag{B8}$$

where $V(\boldsymbol{q}) \equiv \int d\boldsymbol{r} V(\boldsymbol{r}) e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}$ is the Fourier transformation of $V(\boldsymbol{r})$. The first part in Eq. (B8) is the Hartree energy and is simplified as

$$\begin{aligned}E_H &= \frac{1}{2\mathcal{V}} \sum_{\boldsymbol{q}} V(\boldsymbol{q}) \int d\boldsymbol{r} d\boldsymbol{r}' e^{i\boldsymbol{q}\cdot(\boldsymbol{r} - \boldsymbol{r}')} \rho(\boldsymbol{r}, \boldsymbol{r}) \rho(\boldsymbol{r}' \boldsymbol{r}') \\ &= \frac{1}{2\mathcal{V}} \sum_{\boldsymbol{q}} V(\boldsymbol{q}) \int d\boldsymbol{r} e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \sum_{n=1}^{n_e} \sum_{\boldsymbol{k}_1 \in BZ} \psi^*_{nk_1}(\boldsymbol{r}) \psi_{nk_1}(\boldsymbol{r}) \\ &\quad \times \int d\boldsymbol{r}' e^{-i\boldsymbol{q}\cdot\boldsymbol{r}'} \sum_{m=1}^{n_e} \sum_{\boldsymbol{k}_2 \in BZ} \psi^*_{mk_2}(\boldsymbol{r}') \psi_{mk_2}(\boldsymbol{r}') \\ &= \frac{1}{2\mathcal{V}} \sum_{\boldsymbol{k}_i \in \text{BZ}} \sum_{\boldsymbol{q}} V(\boldsymbol{q}) \delta_{\boldsymbol{q}\boldsymbol{G}} \text{Tr}\left[e^{i\boldsymbol{q}\cdot\hat{\boldsymbol{r}}} \hat{\rho}_{\boldsymbol{k}_1}\right] \text{Tr}\left[e^{-i\boldsymbol{q}\cdot\hat{\boldsymbol{r}}} \hat{\rho}_{\boldsymbol{k}_2}\right] \\ &= \frac{1}{2\mathcal{V}} \sum_{\boldsymbol{k}_i \in \text{BZ}, \boldsymbol{G}} V(\boldsymbol{G}) \text{Tr}\left[e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}} \hat{\rho}_{\boldsymbol{k}_1}\right] \text{Tr}\left[e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}} \hat{\rho}_{\boldsymbol{k}_2}\right],\end{aligned} \tag{B9}$$

where $\boldsymbol{G}$ is reciprocal vector. The $e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}$ term in Eq. (B9) should be understood as a single-particle operator that acts on $|u_{n\boldsymbol{k}}\rangle$ as

$$\langle \boldsymbol{r}|e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}|u_{n\boldsymbol{k}}\rangle = e^{i\boldsymbol{G}\cdot\boldsymbol{r}}u_{n\boldsymbol{k}}(\boldsymbol{r}) = u_{n\boldsymbol{k}-\boldsymbol{G}}(\boldsymbol{r}) = \langle \boldsymbol{r}|u_{n\boldsymbol{k}-\boldsymbol{G}}\rangle. \tag{B10}$$

The second part in Eq. (B8) is the Fock energy:

$$\begin{aligned} E_F &= -\frac{1}{2\mathcal{V}}\sum_{\boldsymbol{q}} V(\boldsymbol{q}) \int d\boldsymbol{r}d\boldsymbol{r}' e^{i\boldsymbol{q}\cdot(\boldsymbol{r}-\boldsymbol{r}')} \\ &\quad \times \sum_{m,n=1}^{n_e}\sum_{\boldsymbol{k}_i\in \mathrm{BZ}} \psi_{n\boldsymbol{k}_1}(\boldsymbol{r}')\psi^*_{n\boldsymbol{k}_1}(\boldsymbol{r})\psi_{m\boldsymbol{k}_2}(\boldsymbol{r})\psi^*_{m\boldsymbol{k}_2}(\boldsymbol{r}') \\ &= \frac{-1}{2\mathcal{V}}\sum_{\boldsymbol{k}_i\in \mathrm{BZ},\boldsymbol{q}} V(\boldsymbol{q})\delta_{\boldsymbol{q}\boldsymbol{k}_1-\boldsymbol{k}_2+\boldsymbol{G}}\mathrm{Tr}\left[e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}\hat{\rho}_{\boldsymbol{k}_1}e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}\hat{\rho}_{\boldsymbol{k}_2}\right] \\ &= \frac{-1}{2\mathcal{V}}\sum_{\boldsymbol{k}_i\in \mathrm{BZ},\boldsymbol{G}} V(\boldsymbol{k}_1-\boldsymbol{k}_2+\boldsymbol{G})\mathrm{Tr}\left[e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}\hat{\rho}_{\boldsymbol{k}_1}e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}\hat{\rho}_{\boldsymbol{k}_2}\right]. \end{aligned} \tag{B11}$$

The polarization energy cannot be expressed by density operator $\hat{\rho}_{\boldsymbol{k}}$ but is still a functional of occupied states:

$$\begin{aligned} E_P &\equiv \langle \mathrm{GS}|H_P|\mathrm{GS}\rangle \\ &= e\boldsymbol{F}\cdot\int d\boldsymbol{r}\boldsymbol{r}\rho(\boldsymbol{r},\boldsymbol{r}) \\ &= e\boldsymbol{F}\cdot\sum_{n=1}^{N_e}\sum_{\boldsymbol{k},\boldsymbol{k}'}\delta_{\boldsymbol{k},\boldsymbol{k}'}\int d\boldsymbol{r}\psi^*_{n\boldsymbol{k}'}(\boldsymbol{r})\boldsymbol{r}\psi_{n\boldsymbol{k}}(\boldsymbol{r}) \\ &= e\boldsymbol{F}\cdot\sum_{n=1}^{N_e}\sum_{\boldsymbol{k},\boldsymbol{k}'}\delta_{\boldsymbol{k}\boldsymbol{k}'}\times\left[-i\nabla_{\boldsymbol{k}}\int d\boldsymbol{r}\psi^*_{n\boldsymbol{k}'}(\boldsymbol{r})\psi_{n\boldsymbol{k}}(\boldsymbol{r})\right. \\ &\quad \left.+\frac{1}{\mathcal{V}}\int d\boldsymbol{r}e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}}u^*_{n\boldsymbol{k}'}(\boldsymbol{r})i\nabla_{\boldsymbol{k}}u_{n\boldsymbol{k}}(\boldsymbol{r})\right] \\ &= e\boldsymbol{F}\cdot\sum_{n=1}^{N_e}\sum_{\boldsymbol{k},\boldsymbol{k}'}\delta_{\boldsymbol{k}\boldsymbol{k}'}[-i\nabla_{\boldsymbol{k}}\delta_{\boldsymbol{k}\boldsymbol{k}'}+\langle u_{n\boldsymbol{k}}|i\nabla_{\boldsymbol{k}}u_{n\boldsymbol{k}}\rangle] \\ &= \sum_{n=1}^{n_e}\sum_{\boldsymbol{k}}\langle u_{n\boldsymbol{k}}|ie\boldsymbol{F}\cdot\nabla_{\boldsymbol{k}}|u_{n\boldsymbol{k}}\rangle. \end{aligned} \tag{B12}$$

This result is consistent with the Berry phase definition of polarization. For a finite-size system with periodic boundary conditions, the polarization and the polarization energy should be written with the discrete form of Berry phase as [44]

$$E_P = \frac{-eF}{\Delta k_\parallel}\mathrm{Im}\sum_{\boldsymbol{k}}\log\det S(\boldsymbol{k},\boldsymbol{k}+\Delta\boldsymbol{k}_\parallel), \tag{B13}$$

where $|\Delta\boldsymbol{k}_\parallel| = 2\pi/L$ and is along the direction of electric field. The overlap matrix S is defined as

$$S_{mn}(\boldsymbol{k},\boldsymbol{k}') = \langle u_{m\boldsymbol{k}}|u_{n\boldsymbol{k}'}\rangle, \qquad m,n = 1,2,\ldots,n_e. \tag{B14}$$

APPENDIX C: MEAN-FIELD HAMILTONIAN AND SELF-CONSISTENT EQUATION

The total energy as a functional of occupied bands $\{|u_{n\boldsymbol{k}}\rangle\}_{n=1}^{n_e}$ is written as

$$E_{\mathrm{tot}}[|u_{n\boldsymbol{k}}\rangle;F] = E_0[\hat{\rho}_{\boldsymbol{k}}] + E_{\mathrm{HF}}[\hat{\rho}_{\boldsymbol{k}}] + E_P[|u_{n\boldsymbol{k}}\rangle;F], \tag{C1}$$

and the stationary state is found by minimizing E_{tot} with constraints

$$\langle u_{m\boldsymbol{k}}|u_{n\boldsymbol{k}}\rangle = \delta_{mn}. \tag{C2}$$

By introducing Lagrange multipliers $\xi_{n\boldsymbol{k}}$, the constrained minimization of E_{tot} is transformed into an unconstrained minimization of

$$F[|u_{n\boldsymbol{k}}\rangle;F] \equiv E_{\mathrm{tot}}[|u_{n\boldsymbol{k}}\rangle;F] + \sum_{n\boldsymbol{k}}\xi_{n\boldsymbol{k}}(1-\langle u_{n\boldsymbol{k}}|u_{n\boldsymbol{k}}\rangle). \tag{C3}$$

Let us calculate the unconstrained derivatives of F with respect to $\langle u_{n\boldsymbol{k}}|$. We first show that

$$\begin{aligned} \frac{\delta\mathrm{Tr}[\hat{\rho}_{\boldsymbol{k}_2}\hat{o}_{\boldsymbol{k}_2}]}{\delta\langle u_{n\boldsymbol{k}_1}|} &= \frac{\delta}{\delta\langle u_{n\boldsymbol{k}_1}|}\sum_m\langle u_{n\boldsymbol{k}_2}|\hat{o}_{\boldsymbol{k}_2}|u_{n\boldsymbol{k}_2}\rangle \\ &= \delta_{\boldsymbol{k}_1\boldsymbol{k}_2}\hat{o}_{\boldsymbol{k}_2}|u_{n\boldsymbol{k}_2}\rangle. \end{aligned} \tag{C4}$$

The single-particle, Hartree, and Fock energy functionals all take this form and thus are easily evaluated:

$$\frac{\delta E_0}{\delta\langle u_{n\boldsymbol{k}}|} = \hat{h}^0_{\boldsymbol{k}}|u_{n\boldsymbol{k}}\rangle, \tag{C5}$$

$$\frac{\delta E_H}{\delta\langle u_{n\boldsymbol{k}}|} = \frac{1}{\mathcal{V}}\sum_{\boldsymbol{k}_2\in\mathrm{BZ},\boldsymbol{G}} V(\boldsymbol{G})\mathrm{Tr}[\hat{\rho}_{\boldsymbol{k}_2}e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}]e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}|u_{n\boldsymbol{k}}\rangle, \tag{C6}$$

$$\frac{\delta E_F}{\delta\langle u_{n\boldsymbol{k}}|} = -\frac{1}{\mathcal{V}}\sum_{\boldsymbol{k}_2\in\mathrm{BZ},\boldsymbol{G}} V(\boldsymbol{k}-\boldsymbol{k}_2+\boldsymbol{G})e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}\hat{\rho}_{\boldsymbol{k}_2}e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}|u_{n\boldsymbol{k}}\rangle. \tag{C7}$$

From the expression above, we could define the Hartree and Fock Hamiltonian as

$$\hat{h}^H_{\boldsymbol{k}}[\hat{\rho}_{\boldsymbol{k}}] = \frac{1}{\mathcal{V}}\sum_{\boldsymbol{k}_2\in\mathrm{BZ},\boldsymbol{G}} V(\boldsymbol{G})\mathrm{Tr}[\hat{\rho}_{\boldsymbol{k}_2}e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}]e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}, \tag{C8}$$

$$\hat{h}^F_{\boldsymbol{k}}[\hat{\rho}_{\boldsymbol{k}}] = -\frac{1}{\mathcal{V}}\sum_{\boldsymbol{k}_2\in\mathrm{BZ},\boldsymbol{G}} V(\boldsymbol{k}-\boldsymbol{k}_2+\boldsymbol{G})e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}\hat{\rho}_{\boldsymbol{k}_2}e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}. \tag{C9}$$

As functionals of gauge invariant single-particle density operator $\hat{\rho}_{\boldsymbol{k}}$, the Hartree and Fock Hamiltonian defined in Eqs. (C8) and (C9) are also invariant under k-space gauge transform of the occupied bands.

As for the polarization term, we start from the discrete form Eq. (B13) and take the thermodynamic limit later. Let us first rewritte Eq. (B13) as

$$\begin{aligned}E_P &= \frac{ieF}{2\Delta k_\parallel}\sum_{\boldsymbol{k}}[\log\det S(\boldsymbol{k},\boldsymbol{k}+\Delta\boldsymbol{k}_\parallel)-\log\det S^\dagger(\boldsymbol{k},\boldsymbol{k}+\Delta\boldsymbol{k}_\parallel)]\\ &= \frac{ieF}{2\Delta k_\parallel}\sum_{\boldsymbol{k}}[\log\det S(\boldsymbol{k},\boldsymbol{k}+\Delta\boldsymbol{k}_\parallel)-\log\det S(\boldsymbol{k}+\Delta\boldsymbol{k}_\parallel,\boldsymbol{k})]\\ &= \frac{ieF}{2\Delta k_\parallel}\sum_{\boldsymbol{k}}[\log\det S(\boldsymbol{k},\boldsymbol{k}+\Delta\boldsymbol{k}_\parallel)-\log\det S(\boldsymbol{k},\boldsymbol{k}-\Delta\boldsymbol{k}_\parallel)]\\ &= \frac{ieF}{2\Delta k_\parallel}\sum_{\boldsymbol{k},\sigma=\pm}\sigma\log\det S(\boldsymbol{k},\boldsymbol{k}+\sigma\Delta\boldsymbol{k}_\parallel).\end{aligned} \tag{C10}$$

Then the unconstrained derivatives of E_P is

$$\begin{aligned}\frac{\delta E_P}{\delta\langle u_{n\boldsymbol{k}}|} &= \frac{ieF}{2\Delta k_\parallel}\frac{\delta}{\delta\langle u_{n\boldsymbol{k}}|}\left[\sum_{\sigma=\pm}\sigma\sum_{\boldsymbol{k}}\log\det S(\boldsymbol{k},\boldsymbol{k}_\sigma)\right]\\ &= \frac{ieF}{2\Delta k_\parallel}\frac{\delta}{\delta\langle u_{n\boldsymbol{k}}|}\left[\sum_{\sigma=\pm}\sigma\sum_{\boldsymbol{k}}\mathrm{Tr}\log S(\boldsymbol{k},\boldsymbol{k}_\sigma)\right]\\ &= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm}\sigma\mathrm{Tr}\left[\frac{\delta S(\boldsymbol{k},\boldsymbol{k}_\sigma)}{\delta\langle u_{n\boldsymbol{k}}|}S^{-1}(\boldsymbol{k},\boldsymbol{k}_\sigma)\right]\\ &= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm}\sigma\sum_{m=1}^{n_e}|u_{m\boldsymbol{k}_\sigma}\rangle S^{-1}_{mn}(\boldsymbol{k},\boldsymbol{k}_\sigma),\end{aligned} \tag{C11}$$

where abbreviation $\boldsymbol{k}_\sigma = \boldsymbol{k}+\sigma\Delta\boldsymbol{k}_\parallel$ is used for simplicity. Denote $|D_{n\boldsymbol{k}}\rangle = \delta E_P/\delta\langle u_{n\boldsymbol{k}}|$. It is easy to see that

$$\begin{aligned}\langle u_{l\boldsymbol{k}}|D_{n\boldsymbol{k}}\rangle &= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma}\sigma\sum_{m=1}^{n_e}S_{lm}(\boldsymbol{k},\boldsymbol{k}_\sigma)S^{-1}_{mn}(\boldsymbol{k},\boldsymbol{k}_\sigma)\\ &= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma}\sigma\delta_{ln}\\ &= 0.\end{aligned} \tag{C12}$$

So the polarization Hamiltonian could be defined as

$$\hat{h}^P_{\boldsymbol{k}}[|u_{n\boldsymbol{k}}\rangle;F] = \sum_{n=1}^{n_e}|D_{n\boldsymbol{k}}\rangle\langle u_{n\boldsymbol{k}}| + \mathrm{H.c.} \tag{C13}$$

and satisfies

$$\hat{h}^P_{\boldsymbol{k}}|u_{n\boldsymbol{k}}\rangle = \sum_{m=1}^{n_e}|D_{m\boldsymbol{k}}\rangle\delta_{mn} = |D_{n\boldsymbol{k}}\rangle = \frac{\delta E_P}{\delta\langle u_{n\boldsymbol{k}}|}. \tag{C14}$$

Before processing, one should verify that this definition of polarization Hamiltonian is a gauge invariant. By denoting $\Phi^\dagger_{\boldsymbol{k}} = [|u_{1\boldsymbol{k}}\rangle,\ldots,|u_{n_e\boldsymbol{k}}\rangle]$, the polarization Hamiltonian is written in a neater form:

$$\hat{h}^P_{\boldsymbol{k}} = \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm}\sigma\Phi_{\boldsymbol{k}_\sigma}(\Phi^\dagger_{\boldsymbol{k}}\Phi_{\boldsymbol{k}_\sigma})^{-1}\Phi^\dagger_{\boldsymbol{k}} + \mathrm{H.c.} \tag{C15}$$

A k-space gauge transformation $(U_{\boldsymbol{k}})_{n_e\times n_e}$ on occupied bands will transform $\Phi_{\boldsymbol{k}}$ into $\Phi_{\boldsymbol{k}}U_{\boldsymbol{k}}$, and the polarization Hamiltonian becomes

$$\begin{aligned}(\hat{h}^P_{\boldsymbol{k}})' &= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm}\sigma\Phi_{\boldsymbol{k}_\sigma}U_{\boldsymbol{k}_\sigma}(U^\dagger_{\boldsymbol{k}}\Phi^\dagger_{\boldsymbol{k}}\Phi_{\boldsymbol{k}_\sigma}U_{\boldsymbol{k}_\sigma})^{-1}U^\dagger_{\boldsymbol{k}}\Phi^\dagger_{\boldsymbol{k}}+\mathrm{H.c.}\\ &= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm}\sigma\Phi_{\boldsymbol{k}_\sigma}U_{\boldsymbol{k}_\sigma}U^{-1}_{\boldsymbol{k}_\sigma}(\Phi^\dagger_{\boldsymbol{k}}\Phi_{\boldsymbol{k}_\sigma})^{-1}(U^\dagger_{\boldsymbol{k}})^{-1}U^\dagger_{\boldsymbol{k}}\Phi^\dagger_{\boldsymbol{k}}+\mathrm{H.c.}\\ &= \hat{h}^P_{\boldsymbol{k}},\end{aligned} \tag{C16}$$

which is invariant.

It is easier to see this gauge invariance in the thermodynamic limit $L\to\infty$ and $dk = \Delta k_\parallel \to 0$. In this limit,

$$S_{mn}(\boldsymbol{k},\boldsymbol{k}_\sigma) = \delta_{mn} + \sigma\langle u_{m\boldsymbol{k}}|\partial_{k_\parallel}u_{n\boldsymbol{k}}\rangle dk, \tag{C17a}$$

$$S^{-1}_{mn}(\boldsymbol{k},\boldsymbol{k}_\sigma) = \delta_{mn} - \sigma\langle u_{m\boldsymbol{k}}|\partial_{k_\parallel}u_{n\boldsymbol{k}}\rangle dk, \tag{C17b}$$

so

$$\begin{aligned}|D_{n\boldsymbol{k}}\rangle &= \frac{ieF}{2dk}\sum_{\sigma=\pm}\sigma\sum_{m=1}^{n_e}(|u_{m\boldsymbol{k}}\rangle + \sigma|\partial_{k_\parallel}u_{m\boldsymbol{k}}\rangle dk)\\ &\quad\times(\delta_{mn}-\sigma\langle u_{m\boldsymbol{k}}|\partial_{k_\parallel}u_{n\boldsymbol{k}}\rangle dk)\\ &= ieF\sum_{m=1}^{n_e}[|\partial_{k_\parallel}u_{m\boldsymbol{k}}\rangle\delta_{mn} - |u_{m\boldsymbol{k}}\rangle\langle u_{m\boldsymbol{k}}|\partial_{k_\parallel}u_{n\boldsymbol{k}}\rangle]\\ &= ieF(1-\hat{\rho}_{\boldsymbol{k}})|\partial_{k_\parallel}u_{n\boldsymbol{k}}\rangle,\end{aligned} \tag{C18}$$

and the polarization Hamiltonian in the thermodynamic limit is written as

$$\begin{aligned}\lim_{dk\to0}\hat{h}^P_{\boldsymbol{k}} &= ieF\sum_{n}^{N_e}(1-\hat{\rho}_{\boldsymbol{k}})|\partial_{k_\parallel}u_{n\boldsymbol{k}}\rangle\langle u_{n\boldsymbol{k}}| + \mathrm{H.c.}\\ &= ieF(1-\hat{\rho}_{\boldsymbol{k}})\partial_{k_\parallel}\hat{\rho}_{\boldsymbol{k}} + \mathrm{H.c.}\\ &= ie\boldsymbol{F}\cdot[\nabla_{\boldsymbol{k}}\hat{\rho}_{\boldsymbol{k}},\hat{\rho}_{\boldsymbol{k}}].\end{aligned} \tag{C19}$$

The thermodynamic limit expression Eq. (C19) is only a functional of the gauge invariant $\hat{\rho}_{\boldsymbol{k}}$ and thus is also a gauge invariant.

Finally, minimization of $F[|u_{n\boldsymbol{k}}\rangle;F]$ gives us the self-consistent equation,

$$\frac{\delta F}{\delta\langle u_{n\boldsymbol{k}}|} = 0 \Rightarrow \hat{h}^{\mathrm{MF}}_{\boldsymbol{k}}[|u_{n\boldsymbol{k}}\rangle;F]|u_{n\boldsymbol{k}}\rangle = \xi_{n\boldsymbol{k}}|u_{n\boldsymbol{k}}\rangle, \tag{C20}$$

where the mean-field Hamiltonian is

$$\hat{h}_{\boldsymbol{k}}^{\mathrm{MF}} = \hat{h}_{\boldsymbol{k}}^{0} + \hat{h}_{\boldsymbol{k}}^{H}[\hat{\rho}_{\boldsymbol{k}}] + \hat{h}_{\boldsymbol{k}}^{F}[\hat{\rho}_{\boldsymbol{k}}] + \hat{h}_{\boldsymbol{k}}^{P}[|u_{n\boldsymbol{k}}\rangle; F]. \tag{C21}$$

APPENDIX D: HESSIAN MATRIX

Assume $F < F_c^m$, and the self-consistent equation has solutions

$$h_{\boldsymbol{k}}^{\mathrm{MF}}[|v\boldsymbol{k}; F\rangle]|i\boldsymbol{k}; F\rangle = \xi_{i\boldsymbol{k};F}|i\boldsymbol{k}; F\rangle, \qquad i = c, v. \tag{D1}$$

The valence band $|v\boldsymbol{k}; F\rangle$ is chosen as the one with lower band energy, i.e., $\xi_{v\boldsymbol{k};F} < \xi_{c\boldsymbol{k};F}$, The F label in wave functions and band energies means they are converged solutions.

At the converged point (local minimum of the energy functional), the trial HF state could be reparametrized as

$$|v'\boldsymbol{k}; F\rangle = \frac{|v\boldsymbol{k}; F\rangle + z_{\boldsymbol{k}}|c\boldsymbol{k}; F\rangle}{\sqrt{1 + |z_{\boldsymbol{k}}|^2}}, \tag{D2}$$

where $z_{\boldsymbol{k}}$ is an arbitrary complex-valued function defined on Brillouin zone. This parametrization is unconstrained and complete, and the grand potential then becomes functional of $z_{\boldsymbol{k}}$ as

$$E_G[z_{\boldsymbol{k}}^*, z_{\boldsymbol{k}}; F] \equiv E_G[|v'\boldsymbol{k}; F\rangle; F]. \tag{D3}$$

By writing $z_{\boldsymbol{k}} = x_{\boldsymbol{k}} + iy_{\boldsymbol{k}}$, the Hessian matrix is defined as

$$\mathcal{H}_{\boldsymbol{k}\boldsymbol{k}'} = \frac{1}{2}\begin{bmatrix} \frac{\delta^2 E_G}{\delta x_{\boldsymbol{k}}\delta x_{\boldsymbol{k}'}} & \frac{\delta^2 E_G}{\delta x_{\boldsymbol{k}}\delta y_{\boldsymbol{k}'}} \\ \frac{\delta^2 E_G}{\delta y_{\boldsymbol{k}}\delta x_{\boldsymbol{k}'}} & \frac{\delta^2 E_G}{\delta y_{\boldsymbol{k}}\delta y_{\boldsymbol{k}'}} \end{bmatrix}. \tag{D4}$$

To be consistent with the notation in Wu *et al.* [24], we will also denote the diagonal part of the Hessian matrix as $\mathcal{K}^{(+)}$ and $\mathcal{K}^{(-)}$. Additionally, the upper off-diagonal part is denoted as $\mathcal{K}^{(X)}$. In summary, the Hessian matrix is written in the form of

$$\mathcal{H} = \begin{bmatrix} \mathcal{K}^{(+)} & \mathcal{K}^{(X)} \\ (\mathcal{K}^{(X)})^T & \mathcal{K}^{(-)} \end{bmatrix}. \tag{D5}$$

For simplicity, the F label will be omitted in the following derivations.

We first calculate the derivatives of $|v'\boldsymbol{k}\rangle$ with respect to $x_{\boldsymbol{k}}$ and $y_{\boldsymbol{k}}$ for further usage.

$$\frac{\delta|v'\boldsymbol{k}\rangle}{\delta x_{\boldsymbol{k}}} = \frac{-x_{\boldsymbol{k}}|v\boldsymbol{k}\rangle + (1 - iy_{\boldsymbol{k}}z_{\boldsymbol{k}})|c\boldsymbol{k}\rangle}{(1 + |z_{\boldsymbol{k}}|^2)^{3/2}}, \tag{D6}$$

$$\frac{\delta|v'\boldsymbol{k}\rangle}{\delta y_{\boldsymbol{k}}} = \frac{-y_{\boldsymbol{k}}|v\boldsymbol{k}\rangle + i(1 + x_{\boldsymbol{k}}z_{\boldsymbol{k}})|c\boldsymbol{k}\rangle}{(1 + |z_{\boldsymbol{k}}|^2)^{3/2}}. \tag{D7}$$

At $z_{\boldsymbol{k}} = 0$, they are simplified as

$$\left.\frac{\delta|v'\boldsymbol{k}\rangle}{\delta x_{\boldsymbol{k}}}\right|_{z_{\boldsymbol{k}}=0} = |c\boldsymbol{k}\rangle, \qquad \left.\frac{\delta|v'\boldsymbol{k}\rangle}{\delta y_{\boldsymbol{k}}}\right|_{z_{\boldsymbol{k}}=0} = i|c\boldsymbol{k}\rangle. \tag{D8}$$

The second-order derivatives of $|v'\boldsymbol{k}\rangle$ at $z_{\boldsymbol{k}} = 0$ are

$$\left.\frac{\delta^2|v'\boldsymbol{k}\rangle}{\delta x_{\boldsymbol{k}}\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} = -\delta_{\boldsymbol{k}\boldsymbol{k}'}|v\boldsymbol{k}\rangle, \tag{D9}$$

$$\left.\frac{\delta^2|v'\boldsymbol{k}\rangle}{\delta y_{\boldsymbol{k}}\delta y_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} = -\delta_{\boldsymbol{k}\boldsymbol{k}'}|v\boldsymbol{k}\rangle, \tag{D10}$$

$$\left.\frac{\delta^2|v'\boldsymbol{k}\rangle}{\delta x_{\boldsymbol{k}}\delta y_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} = 0. \tag{D11}$$

The first-order derivative of E_G defined by Eq. (2) is

$$\begin{aligned}\frac{\delta E_G}{\delta x_{\boldsymbol{k}}} &= \frac{\delta\langle v'\boldsymbol{k}|}{\delta x_{\boldsymbol{k}}}\frac{\delta E_G}{\delta\langle v'\boldsymbol{k}|} + \text{c.c.} \\ &= \frac{\delta\langle v'\boldsymbol{k}|}{\delta x_{\boldsymbol{k}}} h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]|v'\boldsymbol{k}\rangle + \text{c.c.},\end{aligned} \tag{D12a}$$

$$\frac{\delta E_G}{\delta y_{\boldsymbol{k}}} = \frac{\delta\langle v'\boldsymbol{k}|}{\delta y_{\boldsymbol{k}}} h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]|v'\boldsymbol{k}\rangle + \text{c.c.} \tag{D12b}$$

We use the definition of mean-field Hamiltonian $h_{\boldsymbol{k}}^{\mathrm{MF}}[|v\boldsymbol{k}\rangle]|v\boldsymbol{k}\rangle \equiv \delta E_G/\delta\langle v\boldsymbol{k}|$ in Eq. (D12). At $z_{\boldsymbol{k}} = 0$, the first-order derivative is just

$$\begin{aligned}\left.\frac{\delta E_G}{\delta x_{\boldsymbol{k}}}\right|_{z_{\boldsymbol{k}}=0} &= \langle c\boldsymbol{k}|h_{\boldsymbol{k}}^{\mathrm{MF}}[|v\boldsymbol{k}\rangle]|v\boldsymbol{k}\rangle + \text{c.c.} = 0, \\ \left.\frac{\delta E_G}{\delta y_{\boldsymbol{k}}}\right|_{z_{\boldsymbol{k}}=0} &= -i\langle c\boldsymbol{k}|h_{\boldsymbol{k}}^{\mathrm{MF}}[|v\boldsymbol{k}\rangle]|v\boldsymbol{k}\rangle + \text{c.c.} = 0,\end{aligned}$$

which is consistent with the fact that $|v\boldsymbol{k}\rangle$ is a local minimum.

Then let us evaluate second-order derivatives of E_G:

$$\begin{aligned}&\left.\frac{\delta^2 E_G}{\delta x_{\boldsymbol{k}}\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} \\ &= \left.\frac{\delta\langle v'\boldsymbol{k}|}{\delta x_{\boldsymbol{k}}}\right|_{z_{\boldsymbol{k}}=0}\left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]}{\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0}|v'\boldsymbol{k}\rangle \\ &\quad + \left.\frac{\delta\langle v'\boldsymbol{k}|}{\delta x_{\boldsymbol{k}}}\right|_{z_{\boldsymbol{k}}=0} h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]\left.\frac{\delta|v'\boldsymbol{k}\rangle}{\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} \\ &\quad + \left.\frac{\delta^2\langle v'\boldsymbol{k}|}{\delta x_{\boldsymbol{k}}\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]|v'\boldsymbol{k}\rangle + \text{c.c.} \\ &= \delta_{\boldsymbol{k}\boldsymbol{k}'}(\xi_{c\boldsymbol{k}} - \xi_{v\boldsymbol{k}}) + \langle c\boldsymbol{k}|\left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]}{\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0}|v\boldsymbol{k}\rangle + \text{c.c.}\end{aligned} \tag{D13}$$

Similarly,

$$\left.\frac{\delta^2 E_G}{\delta y_{\boldsymbol{k}} \delta y_{\boldsymbol{k}'}}\right|_{z_k=0} = \delta_{\boldsymbol{k}\boldsymbol{k}'}(\xi_{c\boldsymbol{k}} - \xi_{v\boldsymbol{k}}) - i\langle c\boldsymbol{k}| \left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]}{\delta y_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle + \text{c.c.} \tag{D14}$$

and

$$\left.\frac{\delta^2 E_G}{\delta x_{\boldsymbol{k}} \delta y_{\boldsymbol{k}'}}\right|_{z_k=0} = \langle c\boldsymbol{k}| \left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k}\rangle]}{\delta y_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle + \text{c.c.} \tag{D15}$$

So the final task is to evaluate the derivatives of $h_{\boldsymbol{k}}^{\mathrm{MF}}$ with respect to $x_{\boldsymbol{k}}$ and $y_{\boldsymbol{k}}$.

The derivatives of Hartree Hamiltonian h^H are

$$\begin{aligned}
&\langle c\boldsymbol{k}| \left.\frac{\delta h^H}{\delta x_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle \\
&= \frac{4\pi e^2 d \langle c\boldsymbol{k}|e\rangle\langle e|v\boldsymbol{k}\rangle}{\epsilon} \left.\frac{\delta n_{\mathrm{ex}}}{\delta x_{\boldsymbol{k}'}}\right|_{z_k=0} \\
&= \frac{4\pi e^2 d \langle c\boldsymbol{k}|e\rangle\langle e|v\boldsymbol{k}\rangle}{\epsilon} \left[\frac{\delta\langle v'\boldsymbol{k}|}{\delta x_{\boldsymbol{k}'}} \left.\frac{\delta n_{\mathrm{ex}}}{\delta\langle v'\boldsymbol{k}|}\right|_{z_k=0} + \text{c.c.}\right] \\
&= \frac{4\pi e^2 d \langle c\boldsymbol{k}|e\rangle\langle e|v\boldsymbol{k}\rangle}{\epsilon} \left[\frac{1}{\mathcal{V}}\langle c\boldsymbol{k}'|e\rangle\langle e|v\boldsymbol{k}'\rangle + \text{c.c.}\right] \\
&= \frac{2}{\mathcal{V}} \frac{4\pi e^2 d \langle c\boldsymbol{k}|e\rangle\langle e|v\boldsymbol{k}\rangle}{\epsilon} \mathrm{Re}[\langle c\boldsymbol{k}'|e\rangle\langle e|v\boldsymbol{k}'\rangle]
\end{aligned}$$

and

$$\begin{aligned}
&\langle c\boldsymbol{k}| \left.\frac{\delta h^H}{\delta y_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle \\
&= \frac{4\pi e^2 d \langle c\boldsymbol{k}|e\rangle\langle e|v\boldsymbol{k}\rangle}{\epsilon} \left[-i\frac{1}{\mathcal{V}}\langle c\boldsymbol{k}'|e\rangle\langle e|v\boldsymbol{k}'\rangle + \text{c.c.}\right] \\
&= \frac{2}{\mathcal{V}} \frac{4\pi e^2 d \langle c\boldsymbol{k}|e\rangle\langle e|v\boldsymbol{k}\rangle}{\epsilon} \mathrm{Im}[\langle c\boldsymbol{k}'|e\rangle\langle e|v\boldsymbol{k}'\rangle].
\end{aligned}$$

The derivatives of Fock Hamiltonian $h_{\boldsymbol{k}}^F$ are

$$\begin{aligned}
&\langle c\boldsymbol{k}| \left.\frac{\delta h_{\boldsymbol{k}}^F}{\delta x_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle \\
&= -\frac{1}{\mathcal{V}}\sum_{ss'} V_{s's}(\boldsymbol{k}-\boldsymbol{k}')\langle c\boldsymbol{k}|s\rangle\langle s'|v\boldsymbol{k}\rangle \left.\frac{\delta\rho_{ss'\boldsymbol{k}'}}{\delta x_{\boldsymbol{k}'}}\right|_{z_k=0} \\
&= -\frac{1}{\mathcal{V}}\sum_{ss'} V_{s's}(\boldsymbol{k}-\boldsymbol{k}')\langle c\boldsymbol{k}|s\rangle\langle s'|v\boldsymbol{k}\rangle \\
&\quad \times (\langle v\boldsymbol{k}'|s'\rangle\langle s|c\boldsymbol{k}'\rangle + \langle c\boldsymbol{k}'|s'\rangle\langle s|v\boldsymbol{k}'\rangle)
\end{aligned}$$

and

$$\begin{aligned}
&\langle c\boldsymbol{k}| \left.\frac{\delta h_{ss'\boldsymbol{k}}^F}{\delta y_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle \\
&= -\frac{i}{\mathcal{V}}\sum_{ss'} V_{s's}(\boldsymbol{k}-\boldsymbol{k}')\langle c\boldsymbol{k}|s\rangle\langle s'|v\boldsymbol{k}\rangle \\
&\quad \times (\langle v\boldsymbol{k}'|s'\rangle\langle s|c\boldsymbol{k}'\rangle - \langle c\boldsymbol{k}'|s'\rangle\langle s|v\boldsymbol{k}'\rangle).
\end{aligned}$$

As for the polarization term, we use the fact that

$$\langle c\boldsymbol{k}|h_{\boldsymbol{k}}^P \left.\frac{\delta|v'\boldsymbol{k}\rangle}{\delta z_{\boldsymbol{k}'}}\right|_{z_k=0} \propto \delta_{\boldsymbol{k}\boldsymbol{k}'}\langle c\boldsymbol{k}|h_{\boldsymbol{k}}^P|c\boldsymbol{k}\rangle = 0,$$

where we have

$$\begin{aligned}
&\langle c\boldsymbol{k}| \left.\frac{\delta h_{\boldsymbol{k}}^P}{\delta x_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle \\
&= \langle c\boldsymbol{k}| \left.\frac{\delta(h_{\boldsymbol{k}}^P|v'\boldsymbol{k}\rangle)}{\delta x_{\boldsymbol{k}'}}\right|_{z_k=0} \\
&= \langle c\boldsymbol{k}| \frac{\delta}{\delta x_{\boldsymbol{k}'}} \left[\frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm1}\frac{\sigma|v'\boldsymbol{k}_\sigma\rangle}{\langle v'\boldsymbol{k}|v'\boldsymbol{k}_\sigma\rangle}\right] \Bigg|_{z_k=0} \\
&= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm}\sigma\left[\delta_{\boldsymbol{k}'\boldsymbol{k}_\sigma}\frac{\langle c\boldsymbol{k}|c\boldsymbol{k}_\sigma\rangle}{\langle v\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle} - \delta_{\boldsymbol{k}'\boldsymbol{k}}\frac{(\langle c\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle)^2}{(\langle v\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle)^2}\right. \\
&\quad \left. - \delta_{\boldsymbol{k}'\boldsymbol{k}_\sigma}\frac{\langle c\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle\langle v\boldsymbol{k}|c\boldsymbol{k}_\sigma\rangle}{(\langle v\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle)^2}\right]
\end{aligned}$$

and

$$\begin{aligned}
&\langle c\boldsymbol{k}| \left.\frac{\delta h_{\boldsymbol{k}}^P}{\delta y_{\boldsymbol{k}'}}\right|_{z_k=0} |v\boldsymbol{k}\rangle \\
&= \frac{ieF}{2\Delta k_\parallel}\sum_{\sigma=\pm}\sigma\left[i\delta_{\boldsymbol{k}'\boldsymbol{k}_\sigma}\frac{\langle c\boldsymbol{k}|c\boldsymbol{k}_\sigma\rangle}{\langle v\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle} + i\delta_{\boldsymbol{k}'\boldsymbol{k}}\frac{(\langle c\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle)^2}{(\langle v\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle)^2}\right. \\
&\quad \left. - i\delta_{\boldsymbol{k}'\boldsymbol{k}_\sigma}\frac{\langle c\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle\langle v\boldsymbol{k}|c\boldsymbol{k}_\sigma\rangle}{(\langle v\boldsymbol{k}|v\boldsymbol{k}_\sigma\rangle)^2}\right].
\end{aligned}$$

APPENDIX E: PHASE AND DENSITY FLUCTUATIONS

In this appendix, we will prove that under a proper gauge for the mean-field conduction and valence band wave functions,

$$|v\boldsymbol{k}\rangle = \begin{bmatrix} e^{i\phi_k/2}\alpha_{\boldsymbol{k}} \\ e^{-i\phi_k/2}\beta_{\boldsymbol{k}} \end{bmatrix}, \qquad |c\boldsymbol{k}\rangle = \begin{bmatrix} e^{i\phi_k/2}\beta_{\boldsymbol{k}} \\ -e^{-i\phi_k/2}\alpha_{\boldsymbol{k}} \end{bmatrix}, \tag{E1}$$

where α, $\beta > 0$ and $\alpha^2 + \beta^2 = 1$, the real and imaginary part of fluctuation variables $z_{\boldsymbol{k}}$ introduced in the main text

by Eq. (8) are exactly the density and phase fluctuations of the EI state.

When the gauge is fixed, the relation between the single-particle density matrix and valence band wave function is a one-to-one correspondence. For the valence band wave function $|v\boldsymbol{k}\rangle$ in Eq. (E1), the density matrix is just

$$\rho_{\boldsymbol{k}} = |v\boldsymbol{k}\rangle\langle v\boldsymbol{k}| = \begin{bmatrix} \alpha_{\boldsymbol{k}}^2 & e^{i\phi_{\boldsymbol{k}}}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}} \\ e^{-i\phi_{\boldsymbol{k}}}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}}^2 \end{bmatrix}, \tag{E2}$$

where the off-diagonal part $\rho_{eh\boldsymbol{k}} = e^{i\phi_{\boldsymbol{k}}}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}}$ is the EI order parameter.

Let us first consider a phase fluctuation to the EI order parameter $\rho_{eh\boldsymbol{k}} \to \rho'_{eh\boldsymbol{k}} = e^{i(\phi_{\boldsymbol{k}}+\delta\phi_{\boldsymbol{k}})}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}}$. Then the valence band wave function becomes

$$|v\boldsymbol{k}\rangle \to |v'\boldsymbol{k}\rangle = \begin{bmatrix} e^{i(\phi_{\boldsymbol{k}}+\delta\phi_{\boldsymbol{k}})/2}\alpha_{\boldsymbol{k}} \\ e^{-i(\phi_{\boldsymbol{k}}+\delta\phi_{\boldsymbol{k}})/2}\beta_{\boldsymbol{k}} \end{bmatrix}, \tag{E3}$$

which could be written as linear combinations of $|v\boldsymbol{k}\rangle$ and $|c\boldsymbol{k}\rangle$ as

$$\begin{aligned} |v'\boldsymbol{k}\rangle &= \langle v\boldsymbol{k}|v'\boldsymbol{k}\rangle|v\boldsymbol{k}\rangle + \langle c\boldsymbol{k}|v'\boldsymbol{k}\rangle|c\boldsymbol{k}\rangle \\ &= (\alpha_{\boldsymbol{k}}^2 e^{i\delta\phi_{\boldsymbol{k}}/2} + \beta_{\boldsymbol{k}}^2 e^{-i\delta\phi_{\boldsymbol{k}}/2})|v\boldsymbol{k}\rangle \\ &\quad + \alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}}(e^{i\delta\phi_{\boldsymbol{k}}/2} - e^{-i\delta\phi_{\boldsymbol{k}}/2})|c\boldsymbol{k}\rangle \\ &\approx |v\boldsymbol{k}\rangle + i\delta\phi_{\boldsymbol{k}}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}}|v\boldsymbol{k}\rangle. \end{aligned} \tag{E4}$$

Comparing Eqs. (E4) and (D2) we find that the phase fluctuation $\delta\phi_{\boldsymbol{k}}$ of the EI order parameter is directly related to the fluctuation variable $z_{\boldsymbol{k}} = i\delta\phi_{\boldsymbol{k}}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}}$, which is pure imaginary.

Then let us consider the density fluctuation $\rho_{eh\boldsymbol{k}} \to \rho'_{eh\boldsymbol{k}} = e^{i\phi_{\boldsymbol{k}}}(\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}} + \delta n_{\boldsymbol{k}})$. Assume $\alpha_{\boldsymbol{k}}$ and $\beta_{\boldsymbol{k}}$ transform to $\alpha'_{\boldsymbol{k}} = \alpha_{\boldsymbol{k}} + \delta\alpha_{\boldsymbol{k}}$ and $\beta'_{\boldsymbol{k}} = \beta_{\boldsymbol{k}} + \delta\beta_{\boldsymbol{k}}$, then up to linear order of $\delta n_{\boldsymbol{k}}$, $\delta\alpha_{\boldsymbol{k}}$ and $\delta\beta_{\boldsymbol{k}}$ should satisfy

$$\alpha'_{\boldsymbol{k}}\beta'_{\boldsymbol{k}} = \alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}} + \delta n_{\boldsymbol{k}} \Rightarrow \alpha_{\boldsymbol{k}}\delta\beta_{\boldsymbol{k}} + \beta_{\boldsymbol{k}}\delta\alpha_{\boldsymbol{k}} = \delta n_{\boldsymbol{k}}, \tag{E5a}$$

$$(\alpha'_{\boldsymbol{k}})^2 + (\beta'_{\boldsymbol{k}})^2 = 1 \Rightarrow \alpha_{\boldsymbol{k}}\delta\alpha_{\boldsymbol{k}} + \beta_{\boldsymbol{k}}\delta\beta_{\boldsymbol{k}} = 0. \tag{E5b}$$

And $\delta\alpha_{\boldsymbol{k}}$, $\delta\beta_{\boldsymbol{k}}$ are solved as

$$\delta\alpha_{\boldsymbol{k}} = -\frac{\beta_{\boldsymbol{k}}\delta n_{\boldsymbol{k}}}{\alpha_{\boldsymbol{k}}^2 - \beta_{\boldsymbol{k}}^2}, \qquad \delta\beta_{\boldsymbol{k}} = \frac{\alpha_{\boldsymbol{k}}\delta n_{\boldsymbol{k}}}{\alpha_{\boldsymbol{k}}^2 - \beta_{\boldsymbol{k}}^2}. \tag{E6}$$

The valence band wave function just transforms to

$$|v\boldsymbol{k}\rangle \to |v'\boldsymbol{k}\rangle = \begin{bmatrix} e^{i\phi_{\boldsymbol{k}}/2}(\alpha_{\boldsymbol{k}} + \delta\alpha_{\boldsymbol{k}}) \\ e^{-i\phi_{\boldsymbol{k}}/2}(\beta_{\boldsymbol{k}} + \delta\beta_{\boldsymbol{k}}) \end{bmatrix}, \tag{E7}$$

which could be written as linear combinations of $|v\boldsymbol{k}\rangle$ and $|c\boldsymbol{k}\rangle$ as

$$\begin{aligned} |v'\boldsymbol{k}\rangle &= \langle v\boldsymbol{k}|v'\boldsymbol{k}\rangle|v\boldsymbol{k}\rangle + \langle c\boldsymbol{k}|v'\boldsymbol{k}\rangle|c\boldsymbol{k}\rangle \\ &= (\alpha_{\boldsymbol{k}}^2 + \alpha_{\boldsymbol{k}}\delta\alpha_{\boldsymbol{k}} + \beta_{\boldsymbol{k}}^2 + \beta_{\boldsymbol{k}}\delta\beta_{\boldsymbol{k}})|v\boldsymbol{k}\rangle \\ &\quad + (\beta_{\boldsymbol{k}}\delta\alpha_{\boldsymbol{k}} - \alpha_{\boldsymbol{k}}\delta\beta_{\boldsymbol{k}})|c\boldsymbol{k}\rangle \\ &= |v\boldsymbol{k}\rangle - \delta n_{\boldsymbol{k}}/(\alpha_{\boldsymbol{k}}^2 - \beta_{\boldsymbol{k}}^2)|c\boldsymbol{k}\rangle. \end{aligned} \tag{E8}$$

Comparing Eqs. (E8) and (D2) we find that the density fluctuation $\delta n_{\boldsymbol{k}}$ of the EI order parameter is directly related to the fluctuation variable $z_{\boldsymbol{k}} = -\delta n_{\boldsymbol{k}}/(\alpha_{\boldsymbol{k}}^2 - \beta_{\boldsymbol{k}}^2)$, which is pure real.

APPENDIX F: GOLDSTONE MODE

The many-body Hamiltonian Eq. (1) is invariant under gauge transformations of the electron creation operators: $c_{e\boldsymbol{k}}^\dagger \to e^{i\phi_e}c_{e\boldsymbol{k}}^\dagger$, $c_{h\boldsymbol{k}}^\dagger \to e^{i\phi_h}c_{h\boldsymbol{k}}^\dagger$. This $U(1)\times U(1)$ symmetry corresponds to the charge conservation in each layer.

After this global gauge transformation, the valence band electron creation operator becomes

$$(c_{v\boldsymbol{k}}^\dagger)' = \alpha_{\boldsymbol{k}}e^{i(\phi_{\boldsymbol{k}}/2+\delta\phi_e)}c_{e\boldsymbol{k}}^\dagger + \beta_{\boldsymbol{k}}e^{i(-\phi_{\boldsymbol{k}}/2+\delta\phi_h)}c_{h\boldsymbol{k}}^\dagger, \tag{F1}$$

which gives a new trial wave function of the valence band as

$$|v'\boldsymbol{k}\rangle = \begin{bmatrix} e^{i(\phi_{\boldsymbol{k}}/2+\delta\phi_e)}\alpha_{\boldsymbol{k}} \\ e^{i(-\phi_{\boldsymbol{k}}/2+\delta\phi_h)}\beta_{\boldsymbol{k}} \end{bmatrix} = e^{i\delta\phi}\begin{bmatrix} e^{i(\phi_{\boldsymbol{k}}+\delta\phi_{\rm ex})/2}\alpha_{\boldsymbol{k}} \\ e^{-i(\phi_{\boldsymbol{k}}+\delta\phi_{\rm ex})/2}\beta_{\boldsymbol{k}} \end{bmatrix}, \tag{F2}$$

where $\delta\phi = (\phi_e + \phi_h)/2$, $\delta\phi_{\rm ex} = (\phi_e - \phi_h)/2$ are related to the conservation of total charge and exciton number, respectively. The relative density matrix $\tilde{\rho} = \rho - \rho^0$ transforms into

$$\tilde{\rho}'_{\boldsymbol{k}} = \begin{bmatrix} \alpha_{\boldsymbol{k}}^2 & e^{i(\phi_{\boldsymbol{k}}+\delta\phi_{\rm ex})}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}} \\ e^{-i(\phi_{\boldsymbol{k}}+\delta\phi_{\rm ex})}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}} & \beta_{\boldsymbol{k}}^2 - 1 \end{bmatrix}, \tag{F3}$$

Additionally, the overlap matrix $S(\boldsymbol{k},\boldsymbol{k}) = \langle v\boldsymbol{k}|v\boldsymbol{k}'\rangle$ becomes

$$\begin{aligned} S'(\boldsymbol{k},\boldsymbol{k}') &\equiv \langle v'\boldsymbol{k}|v'\boldsymbol{k}'\rangle \\ &= e^{i(\phi_{\boldsymbol{k}'}-\phi_{\boldsymbol{k}})/2}\alpha_{\boldsymbol{k}}\alpha_{\boldsymbol{k}'} + e^{-i(\phi_{\boldsymbol{k}'}-\phi_{\boldsymbol{k}})/2}\beta_{\boldsymbol{k}}\beta_{\boldsymbol{k}'} \\ &= S(\boldsymbol{k},\boldsymbol{k}'). \end{aligned} \tag{F4}$$

Substituting Eqs. (F3) and (F4) into the grand potential expression Eq. (2) we find that $\varepsilon_G[|v'\boldsymbol{k}\rangle; F] = \varepsilon_G[|v\boldsymbol{k}\rangle; F]$; i.e., the grand potential is invariant under the transformation $|v\boldsymbol{k}\rangle \to |v'\boldsymbol{k}\rangle$.

The U(1) symmetry related to exciton conservation (phase fluctuation $\delta\phi_{\rm ex} = \phi_e - \phi_h$ of electron-hole pairing condensate $\rho_{eh\boldsymbol{k}}$) gives a zero-energy Goldstone mode to

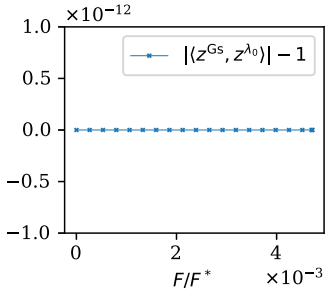

FIG. 9. Overlap between the Goldstone mode and the zero mode of the Hessian matrix.

the valence band fluctuation. Compare Eqs. (E3) and (F2), and the Goldstone mode is directly obtained from Eq. (E4) as $z_{\boldsymbol{k}}^{\mathrm{Gs}} \propto i\delta\phi_{\mathrm{ex}}\alpha_{\boldsymbol{k}}\beta_{\boldsymbol{k}}$.

The overlap between the Goldstone mode $z_{\boldsymbol{k}}^{\mathrm{Gs}}$ and the zero mode $z_{\boldsymbol{k}}^{\lambda_0}$ of the Hessian matrix Eq. (10) is calculated as

$$I = |\langle z_{\boldsymbol{k}}^{\mathrm{Gs}}, z_{\boldsymbol{k}}^{\lambda_0}\rangle| = \left|\sum_{\boldsymbol{k}} (z_{\boldsymbol{k}}^{\mathrm{Gs}})^* z_{\boldsymbol{k}}^{\lambda_0}\right| \tag{F5}$$

and plotted in Fig. 9. The results show that I is equal to 1 in numerical precision, which means the zero mode of the Hessian matrix is indeed the Goldstone mode $z_{\boldsymbol{k}}^{\mathrm{Gs}}$ discussed in this appendix.

APPENDIX G: BREAKDOWN MODE

In this appendix, we will prove that, in the zero-field limit $F = 0$, the breakdown mode $z_{\boldsymbol{k}}^{\lambda_1}$ shown in Fig. 5 is the only fluctuation eigenmode which accounts for the polarization fluctuation in x direction (the direction of electrical field).

At zero electrical field strength, the bilayer model has a continuous rotation symmetry. In addition, the phase of the EI order parameter $\rho_{eh\boldsymbol{k}}$ is a constant, as shown by Fig. 4, which could be chosen as zero due to the electron-hole U(1) symmetry. At this time, the valence band wave function could be written in the form

$$|v\boldsymbol{k}\rangle = \begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix}, \tag{G1}$$

where $\alpha_k, \beta_k > 0$ are only functions of the norm of $\boldsymbol{k}$. When phase and density fluctuations $\delta\phi_{\boldsymbol{k}}$ and $\delta n_{\boldsymbol{k}}$ are considered, the valence band wave functions becomes

$$|v'\boldsymbol{k}\rangle = \begin{bmatrix} e^{i\delta\phi_k/2}(\alpha_k + \delta\alpha_{\boldsymbol{k}}) \\ e^{-i\delta\phi_k/2}(\beta_k + \delta\beta_{\boldsymbol{k}}) \end{bmatrix}, \tag{G2}$$

where the relation between $\delta\alpha_{\boldsymbol{k}}, \delta\beta_{\boldsymbol{k}}$ and $\delta n_{\boldsymbol{k}}$ is given by Eq. (E6). Then the ground state polarization density in x direction becomes

$$\begin{aligned} P_x &= \int \frac{d^2k}{(2\pi)^2} \langle v'\boldsymbol{k}|i\partial_{k_x}|v'\boldsymbol{k}\rangle \\ &= \int \frac{d^2k}{(2\pi)^2}\frac{1}{2}[(\alpha_k + \delta\alpha_{\boldsymbol{k}})^2 - (\beta_k + \delta\beta_{\boldsymbol{k}})^2]\partial_{k_x}\delta\phi_{\boldsymbol{k}}. \end{aligned} \tag{G3}$$

To first order of $\delta\phi_{\boldsymbol{k}}$ and $\delta n_{\boldsymbol{k}}$, the polarization fluctuation is written as

$$\delta P_x = \int \frac{d^2k}{(2\pi)^2}\frac{\alpha_k^2 - \beta_k^2}{2}\partial_{k_x}\delta\phi_{\boldsymbol{k}}, \tag{G4}$$

where only the phase fluctuation leads to the fluctuation of polarization δP_x. Because of the rotational symmetry at zero field, the phase fluctuation $\delta\phi_{\boldsymbol{k}}$ could be expanded into channels with different angular momentum as

$$\delta\phi_{\boldsymbol{k}} = \sum_n \delta\phi_k^l e^{il\theta}, \tag{G5}$$

where θ is the angle of $\boldsymbol{k}$. Since $\delta\phi_{\boldsymbol{k}}$ is real, the expansion coefficient satisfies $\delta\phi_k^l = (\delta\phi_k^{-l})^*$. Then $\partial_{k_x}\delta\phi_{\boldsymbol{k}}$ becomes

$$\begin{aligned} \partial_{k_x}\delta\phi_{\boldsymbol{k}} &= \sum_l \left(\frac{\partial k}{\partial k_x}\partial_k\delta\phi_k^l + il\delta\phi_k^l\frac{\partial\theta}{\partial k_x}\right)e^{il\theta} \\ &= \sum_l \left(\cos\theta\partial_k\delta\phi_k^l - \frac{in\delta\phi_k^l\sin\theta}{k}\right)e^{in\theta}, \end{aligned} \tag{G6}$$

and the polarization fluctuation is rewritten as

$$\begin{aligned} \delta P_x &= \sum_l \frac{1}{8\pi^2}\left[\int kdk(\alpha_k^2 - \beta_k^2)\partial_k\delta\phi_k^l\int_0^{2\pi} d\theta\cos\theta e^{il\theta}\right. \\ &\quad \left. - \mathrm{in}\int dk(\alpha_k^2 - \beta_k^2)\delta\phi_k^l\int_0^{2\pi} d\theta\sin\theta e^{il\theta}\right] \\ &= \frac{1}{4\pi}\int kdk(\alpha_k^2 - \beta_k^2)\partial_k\mathrm{Re}\delta\phi_k^{l=1}, \end{aligned} \tag{G7}$$

which means only the real part of $\delta\phi_k^{l=1}$ can contribute to the polarization fluctuation in x direction. Keeping only the real part of $\delta\phi_k^{l=1}$ in Eq. (G5), the phase fluctuation related to the polarization fluctuation is in the form of

$$\delta\phi_{\boldsymbol{k}} \sim 2\mathrm{Re}\delta\phi_k^{l=1}\cos\theta. \tag{G8}$$

In Fig. 10, the real (density fluctuation) and imaginary (phase fluctuation) part of the breakdown mode $z_{\boldsymbol{k}}^{\lambda_1} = x_{\boldsymbol{k}}^{\lambda_1} + iy_{\boldsymbol{k}}^{\lambda_1}$ are plotted at zero field and the critical field strength. At zero field $F = 0$, the off-diagonal part $\mathcal{K}^{(X)}$ of the Hessian matrix Eq. (10) is zero. So the density and phase fluctuations are decoupled. Figures 10(a) and 10(c) show that the breakdown mode at zero field is a pure phase

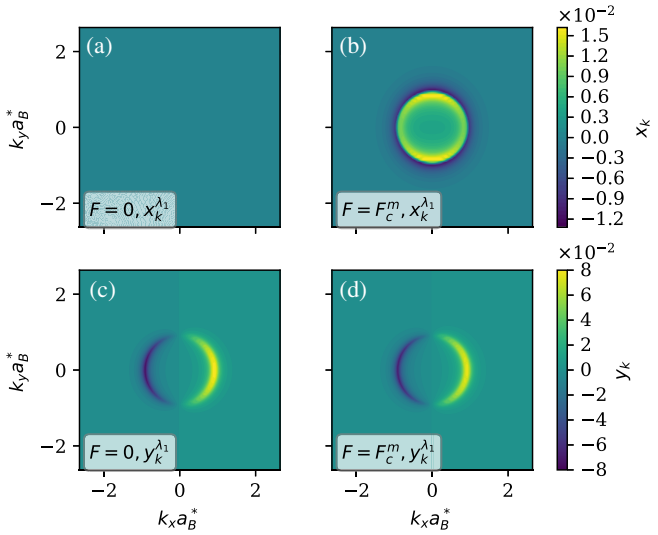

FIG. 10. (a),(b) Real part (density fluctuation component) of the breakdown mode $z_{\boldsymbol{k}}^{\lambda_1}$ at zero electrical field $F = 0$ and the critical field strength $F = F_c^m$. (c),(d) Imaginary part (phase fluctuation component) of the breakdown mode $z_{\boldsymbol{k}}^{\lambda_1}$. These plots are generated on the 120×80 k mesh with exciton density $n_{\mathrm{ex}} \approx 0.068 a_B^{*-2}$.

fluctuation of the type of Eq. (G8), which exactly corresponds to the polarization fluctuation δP_x. When electrical field is turned on, $\mathcal{K}^{(X)}$ becomes nonzero, which will mix the density and phase fluctuations. As a consequence, the breakdown mode will gain some density fluctuation component, while the main component is still the phase fluctuation of the type of Eq. (G8) as illustrated by Figs. 10(b) and 10(d).

APPENDIX H: FLUCTUATION DYNAMICS AND COLLECTIVE MODES

To find the collective modes, we also need to study the dynamics of the fluctuation variables $z_{\boldsymbol{k}}$.

The time-dependent trial HF occupied states is defined as

$$|v'\boldsymbol{k};t\rangle = \frac{|v\boldsymbol{k};t\rangle + z_{\boldsymbol{k}}(t)\mathrm{e}^{i(\xi_{c\boldsymbol{k}}-\xi_{v\boldsymbol{k}})t}|c\boldsymbol{k};t\rangle}{\sqrt{1+|z_{\boldsymbol{k}}(t)|^2}}, \tag{H1}$$

where $|c/v\boldsymbol{k};t\rangle = \mathrm{e}^{-i\xi_{c/v\boldsymbol{k}}t}|c/v\boldsymbol{k}\rangle$. In the definition of $z_{\boldsymbol{k}}(t)$ in Eq. (H1), the dynamical phases $\mathrm{e}^{i(\xi_{c\boldsymbol{k}}-\xi_{v\boldsymbol{k}})t}$ from the time evolutions of $|v\boldsymbol{k};t\rangle$ and $|c\boldsymbol{k};t\rangle$ are subtracted. The time evolution of Eq. (H1) should satisfy the time-dependent HF equation:

$$i\partial_t|v'\boldsymbol{k};t\rangle = h_{\boldsymbol{k}}^{\mathrm{MF}}[|v'\boldsymbol{k};t\rangle]|v'\boldsymbol{k};t\rangle. \tag{H2}$$

To zeroth order of $z_{\boldsymbol{k}}$, Eq. (H2) becomes

$$h_{\boldsymbol{k}}^{\mathrm{MF}}[|v\boldsymbol{k};t\rangle]|v\boldsymbol{k};t\rangle = i\partial_t|v\boldsymbol{k};t\rangle = \xi_{v\boldsymbol{k}}|v\boldsymbol{k};t\rangle, \tag{H3}$$

which is exactly the self-consistent equation Eq. (7). To first order of $z_{\boldsymbol{k}}$, Eq. (H2) becomes

$$[i\partial_t z_{\boldsymbol{k}}(t) - (\xi_{c\boldsymbol{k}} - \xi_{v\boldsymbol{k}})]\mathrm{e}^{i(\xi_{c\boldsymbol{k}}-\xi_{v\boldsymbol{k}})t}|c\boldsymbol{k};t\rangle = \sum_{\boldsymbol{k}'}\left[\left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}}{\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} x_{\boldsymbol{k}'}(t) + \left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}}{\delta y_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} y_{\boldsymbol{k}'}(t)\right]|v\boldsymbol{k};t\rangle. \tag{H4}$$

Or, equivalently,

$$i\partial_t z_{\boldsymbol{k}}(t) = \xi_{c\boldsymbol{k}} - \xi_{v\boldsymbol{k}} + \sum_{\boldsymbol{k}'}\langle c\boldsymbol{k}|\left[\left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}}{\delta x_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} x_{\boldsymbol{k}'}(t) + \left.\frac{\delta h_{\boldsymbol{k}}^{\mathrm{MF}}}{\delta y_{\boldsymbol{k}'}}\right|_{z_{\boldsymbol{k}}=0} y_{\boldsymbol{k}'}(t)\right]|v\boldsymbol{k}\rangle. \tag{H5}$$

By taking real and imaginary parts of the previous equation and using the definition of the Hessian matrix Eqs. (D4), (D5), (D13)–(D15), we finally get the dynamics equation of the fluctuation variables as

$$-\partial_t y_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'}\left[\mathcal{K}_{\boldsymbol{k}\boldsymbol{k}'}^{(+)} x_{\boldsymbol{k}'} + \mathcal{K}_{\boldsymbol{k}\boldsymbol{k}'}^{(X)} y_{\boldsymbol{k}'}\right], \tag{H6a}$$

$$\partial_t x_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'}\left[\mathcal{K}_{\boldsymbol{k}'\boldsymbol{k}}^{(X)} x_{\boldsymbol{k}'} + \mathcal{K}_{\boldsymbol{k}\boldsymbol{k}'}^{(-)} y_{\boldsymbol{k}'}\right], \tag{H6b}$$

which recovers the dynamics equation in Wu *et al.* [24] (the dynamics equation in their paper is derived from an effective field theory and there is a sign error when they apply the Euler-Lagrange equation).

To solve the dynamics equation, let us omit the subscript $\boldsymbol{k}$ and write the dynamics equation in a neater form as

$$\partial_t\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & \mathcal{I} \\ -\mathcal{I} & 0 \end{bmatrix}\mathcal{H}\begin{bmatrix} x \\ y \end{bmatrix}, \tag{H7}$$

where $\mathcal{I}$ is the identity matrix. Since the Hessian matrix $\mathcal{H}$ is real symmetric and non-negative, the square root of $\mathcal{H}$ is well defined and is also real symmetric. Define $u = \sqrt{\mathcal{H}}(x,y)^T$, then Eq. (H7) could be written as

$$\partial_t u = \mathcal{D}u, \tag{H8}$$

where the coefficient matrix $\mathcal{D}$ is defined as

$$\mathcal{D} = \sqrt{\mathcal{H}}\begin{bmatrix} 0 & \mathcal{I} \\ -\mathcal{I} & 0 \end{bmatrix}\sqrt{\mathcal{H}}. \tag{H9}$$

It is easy to verify that $\mathcal{D}^T = -\mathcal{D}$, which means $\mathcal{D}$ is a real and antisymmetric matrix. As an antisymmetric matrix, the eigenvalues can only be zero or pure imaginary numbers. As a real matrix, the imaginary eigenvalues must appear in pairs as $\pm i\omega$, where ω could be viewed as the excitation energies of collective modes.

The fluctuation eigenmodes (x^λ, y^λ) of the Hessian matrix $\mathcal{H}$ are not necessary to be the collective modes since the collective modes are eigenvectors of the $\mathcal{D}$ matrix defined by Eq. (H9). However, the zero mode of Hessian matrix is always a collective mode with zero excitation energy. Assume (x^λ, y^λ) is a zero mode of the Hessian matrix such that

$$\mathcal{H}\begin{bmatrix} x^\lambda \\ y^\lambda \end{bmatrix} = 0, \tag{H10}$$

then we can verify that

$$\mathcal{D}u^\lambda = \mathcal{D}\sqrt{\mathcal{H}}\begin{bmatrix} x^\lambda \\ y^\lambda \end{bmatrix} = \sqrt{\mathcal{H}}\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}\mathcal{H}\begin{bmatrix} x^\lambda \\ y^\lambda \end{bmatrix} = 0. \tag{H11}$$

APPENDIX I: ESTIMATION OF THE BREAKDOWN FIELD FROM ZERO-FIELD QUANTITIES

Because of inversion symmetry, the excitation energy of the breakdown mode ω^{p_x} should be an even function of the electrical field F. Additionally, near the critical field strength $F \sim F_c^m$, the critical behavior of the excitation energy should be

$$\omega^{p_x}(F \to F_c^m + 0^-) \sim (1 - F/F_c^m)^\nu, \tag{I1}$$

where ν is the critical exponent of the many-body breakdown phase transition. Define zero-field excitation energy as $\omega_0^{p_x} \equiv \omega^{p_x}(F = 0)$, then $(\omega^{p_x}/\omega_0^{p_x})^4$ is replotted as a function of $(F/F_c^m)^2$ in Fig. 11(a), which shows a good linearity. This indicates that the critical exponent is $\nu = 1/4$, and the excitation energy as a function of electrical field strength could be fitted by

$$\omega^{p_x}(F) = \omega_0^{p_x}[1 - (F/F_c^m)^2]^{1/4}, \tag{I2}$$

which is also shown in Fig. 11(b).

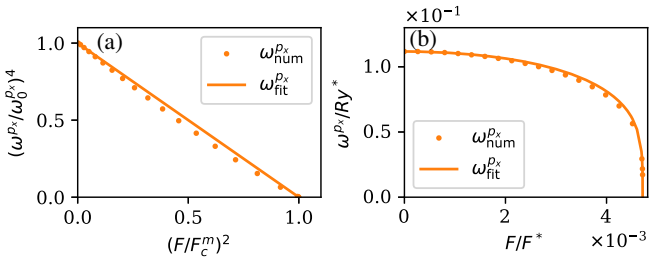

FIG. 11. (a) $(\omega^{p_x}/\omega_0^{p_x})^4$ as a function of $(F/F_c^m)^2$ which shows a good linearity. The orange dots are numerically solved data while the solid line is the linear fit. (b) Fit of excitation energy of the breakdown mode ω^{p_x} by the function form $\omega^{p_x}(F) = \omega_0^{p_x}[1 - (F/F_c^m)^2]^{1/4}$.

Near zero-field strength, Eq. (I2) is approximated as

$$\omega^{p_x}(F \to 0) \approx \omega_0^{p_x} - \frac{\omega_0^{p_x}}{4(F_c^m)^2}F^2, \tag{I3}$$

and the polarizability is just $\eta_0 \equiv -\partial_F^2\omega^{p_x}(F)|_{F=0} = \omega_0^{p_x}/2(F_c^m)^2$. This means the critical field for the many-body breakdown could be estimated from the zero-field excitation energy $\omega_0^{p_x}$ and the polarizability η_0 as

$$F_c^m = \sqrt{\omega_0^{p_x}/2\eta_0}. \tag{I4}$$

APPENDIX J: INTERBAND ZENER TUNNELING

Consider the interband tunneling problem of the 2D continuous model,

$$\hat{h} = \begin{bmatrix} -\frac{\partial_x^2}{2m} - \frac{\partial_y^2}{2m} - \frac{\mu_{ex}^0}{2} & \frac{\Delta}{2} \\ \frac{\Delta}{2} & \frac{\partial_x^2}{2m} + \frac{\partial_y^2}{2m} + \frac{\mu_{ex}^0}{2} \end{bmatrix} + V(x), \tag{J1}$$

where the barrier potential $V(x)$ is defined as

$$V(x) = \begin{cases} eFL/2, & x \le -L/2 \\ -eFx, & -L/2 \le x \le L/2 \\ -eFL/2, & x \ge L/2. \end{cases} \tag{J2}$$

For a given tunneling energy E, the Schrödinger equation is

$$\hat{h}|\Psi; E\rangle = E|\Psi; E\rangle. \tag{J3}$$

Since the electrical field is applied only along the x direction, translation symmetry in the y direction still holds and k_y is a good quantum number. Following Zener and Fowler [30], we could write the approximated WKB wave function as

$$|\Psi_{k_y}; E\rangle \propto \exp\left[ik_y y + i\int_{-\infty}^{x} k(x')dx'\right]|\tilde{u}_{k(x)k_y}\rangle. \tag{J4}$$

If $k(x)$ is slow varying so that $\partial_x k(x)$ could be neglected, substituting Eq. (J4) into the Schrödinger equation we find that

$$h_{k(x)k_y}|\tilde{u}_{k(x)k_y}\rangle = (E - V(x))|\tilde{u}_{k(x)k_y}\rangle, \tag{J5}$$

where

$$h_{k(x)k_y} = \begin{bmatrix} \frac{k^2(x)}{2m} - \frac{\mu_{ex}(k_y)}{2} & \frac{\Delta}{2} \\ \frac{\Delta}{2} & -\frac{k^2(x)}{2m} + \frac{\mu_{ex}(k_y)}{2} \end{bmatrix}, \tag{J6}$$

and $\mu_{\rm ex}(k_y)=\mu_{\rm ex}^0-k_y^2/m$. Solving the secular equation (J5) gives the relation between the complex wave vector $k(x)$ and position x:

$$[k^2(x)-m\mu_{\rm ex}]^2+(m\Delta)^2=\{2m[E-V(x)]\}^2. \quad \text{(J7)}$$

Things are different for $\mu_{\rm ex}>0$ and $\mu_{\rm ex}<0$ and should be discussed separately. The condition $\mu_{\rm ex}(k_y)=0$ gives a critical k_y as

$$\mu_{\rm ex}(k_y)=\mu_{\rm ex}^0-k_y^2/m=0\Rightarrow k_{y,c}=\sqrt{m\mu_{\rm ex}^0}. \quad \text{(J8)}$$

The tunneling scenario for $\mu_{\rm ex}(k_y)>0$ (or equivalently $k_y^2\le k_{y,c}^2$) is illustrated in Fig. 12(a). As a tunneling state propagating to the right, $|\Psi_{k_y};E\rangle$ should behave like a valence band electron in the region $x\ll -L/2$ [$k_L^{\pm}$ states in Fig. 12(a)] and like a conduction band electron in the region $x\gg L/2$ [$k_R^{\pm}$ states in Fig. 12(a)]. This places a restriction on the tunneling energy $(\Delta-eFL)/2\le E\le -(\Delta-eFL)/2$, which further demands that $eFL\ge\Delta$. In other words, the interband Zener tunneling occurs only when the in-plane bias voltage exceeds the band gap.

As the electron propagates to the right in the region $|x|\le L/2$, the complex wave vector $k(x)$ will travel from k_L^{σ} to k_R^{σ} in the complex plane along the line [34]

$$\mathrm{Im}[k^2(x)-m\mu_{\rm ex}]^2=0. \quad \text{(J9)}$$

Equation (J9) is just the imaginary part of Eq. (J7) and is solved as

$$\mathrm{Im}k\times\mathrm{Re}k\times[(\mathrm{Re}k)^2-(\mathrm{Im}k)^2-m\mu_{\rm ex}]=0. \quad \text{(J10)}$$

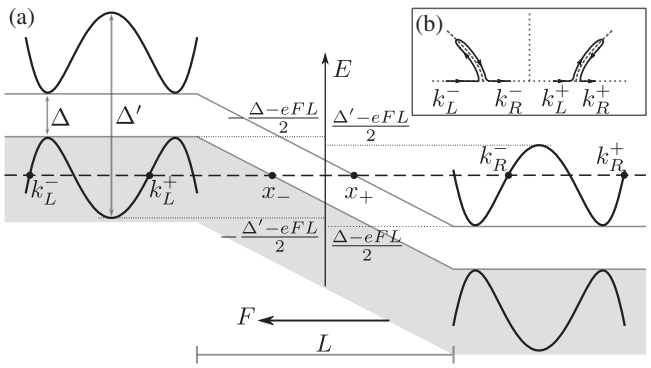

FIG. 12. (a) Tunneling scenario for $\mu_{\rm ex}(k_y)>0$. The tunneling channels $k_L^{\pm}\to k_R^{\pm}$ exist only when in-plane bias voltage overcomes the band gap, i.e., $eFL>\Delta$. Under WKB approximation, the valence band k_L^{σ} states in the region $x\le -L/2$ will continuously turn into the conduction band k_R^{σ} states in the region $x\ge L/2$ as propagating to the right. $x_{\pm}=(\pm\Delta/2-E)/eF$ marks the classical turning points. (b) The paths of the complex wave vectors $k^{\sigma}(x)$ in the complex plane are indicated by the black arrow lines.

The solutions of Eq. (J10) in the complex plane are represented by dashed gray lines in Fig. 12(b). The paths of $k^{\sigma}(x)$ in the complex plane are also illustrated by solid black arrow lines in Fig. 12(b). This analysis means that $k^+(x)$ and $k^-(x)$ are two independent tunneling channels.

It is important to note that the tunneling channel $k^+(x)$ exists only for tunneling energy $E\ge -(\Delta'-eFL)/2$, where $\Delta'=\sqrt{\mu_{\rm ex}^2+\Delta^2}$. This is because there is no k_L^+ state in the region $x\ll -L/2$ when $E<-(\Delta'-eFL)/2$, as is shown in Fig. 12(a). So the allowed tunneling energy range for $k^+(x)$ channel is $E_{\rm max}^+=(eFL-\Delta)/2$ and $E_{\rm min}^+=\max[-(\Delta'-eFL)/2,(\Delta-eFL)/2]$. Similarly, the tunneling channel $k^-(x)$ exists only when tunneling energy is in the range $E_{\rm max}^-=\min[(\Delta'-eFL)/2,(eFL-\Delta)/2]$ and $E_{\rm min}^-=(\Delta-eFL)/2$.

Once these energy conditions are satisfied, one can calculate the tunneling probability under WKB approximation directly by

$$P^{\rm WKB}_{k_L^{\sigma}k_R^{\sigma},k_y}(E)=\frac{|\Psi_{k_y}(x=L/2;E)|^2}{|\Psi_{k_y}(x=-L/2;E)|^2}=\mathrm{e}^{-2\zeta^{\sigma}_{k_y}(E)},$$

where $\zeta^{\sigma}_{k_y}(E)$ is the Zener parameter defined by

$$\zeta^{\sigma}_{k_y}(E)\equiv\int_{x_-}^{x_+}dx|\mathrm{Im}k^{\sigma}(x)|. \quad \text{(J11)}$$

The lower and upper limits $x_{\pm}=(\pm\Delta/2-E)/eF$ of the integration are the classical turning points. Only in the range $x_-\le x\le x_+$, $k^{\sigma}(x)$ has an imaginary part:

$$|\mathrm{Im}k^{\sigma}(x)|=\sqrt{\frac{m}{2}}\sqrt{\sqrt{\mu_{\rm ex}^2+\Delta^2-4(E+eFx)^2}-\mu_{\rm ex}}.$$

So the Zener parameter is calculated as

$$\begin{aligned}\zeta^{\sigma}_{k_y}(E)&=\frac{\sqrt{m}}{2\sqrt{2}eF}\int_{-\Delta}^{\Delta}dE\sqrt{\sqrt{\mu_{\rm ex}^2+\Delta^2-E^2}-\mu_{\rm ex}}\\&=\frac{\sqrt{m}\Delta^{3/2}}{\sqrt{2}eF}\int_0^1 d\varepsilon\sqrt{\sqrt{\tilde{\mu}_{\rm ex}^2+1-\varepsilon^2}-\tilde{\mu}_{\rm ex}},\end{aligned} \quad \text{(J12)}$$

where $\tilde{\mu}_{\rm ex}=\mu_{\rm ex}(k_y)/\Delta=(\mu_{\rm ex}^0-k_y^2/m)/\Delta>0$. One can see that $\zeta^{\sigma}_{k_y}(E)=\zeta(k_y)$ is only a function of k_y. So the transition probability is also only a function of k_y; i.e., $P^{\rm WKB}_{k_L^{\sigma}k_R^{\sigma},k_y}(E)=P(k_y)=\mathrm{e}^{-2\zeta(k_y)}$.

The current contributed by state $|\Psi_{k_y};k_L^{\sigma}\to k_R^{\sigma}\rangle$ is calculated by multiplying the tunneling probability with the velocity $v_{c,k_R^{\sigma}k_y}=\partial_{k_R^{\sigma}}\varepsilon_{c,k_R^{\sigma}k_y}$ of the final state. Sum all possible final states k_R^{σ} together and we get

$$j(k_y) = -e\sum_\sigma \int \frac{dk_R^\sigma}{2\pi} P^{\mathrm{WKB}}_{k_L^\sigma k_R^\sigma, k_y}(E)\partial_{k_R^\sigma}\varepsilon_{c,k_R^\sigma k_y}$$
$$= -\frac{eP(k_y)}{2\pi}\sum_\sigma \int_{E^\sigma_{\min}}^{E^\sigma_{\max}} dE$$
$$= -\frac{eP(k_y)}{2\pi}\delta E(k_y), \tag{J13}$$

where $\delta E(k_y) = \min[2(eFL-\Delta), \Delta'-\Delta]$.

On the other hand, the tunneling scenario for the case $\mu_{\mathrm{ex}} < 0$ (or equivalently $k_y^2 > k_{y,c}^2$) is shown in Fig. 13(a). Different from the case $\mu_{\mathrm{ex}} > 0$, there exists one and only one tunneling channel $|\Psi_{k_y}; k_L \to k_R\rangle$ for tunneling energy in the range $E_{\min} = (\Delta' - eFL)/2 \le 0$ and $E_{\max} = (eFL-\Delta')/2 \ge 0$. And the path of the wave vector $k(x)$ in the complex plane is indicated by the black solid arrow line in Fig. 13(b). The existence of tunneling channels requires $eFL \ge \Delta'(k_y) = \sqrt{\mu_{\mathrm{ex}}^2(k_y) + \Delta^2}$, which gives an upper bound for k_y^2:

$$k_y^2 \le k_{y,\max}^2 = m\left[\mu_{\mathrm{ex}}^0 + \sqrt{(eFL)^2 - \Delta^2}\right]. \tag{J14}$$

In this case, the classical turning points are $x'_\pm = (\pm\Delta'/2 - E)/eF$. In addition to the region $x_- \le x \le x_+$, the complex wave vector $k(x)$ also has an imaginary part in the region $x'_- \le x \le x_-$ and $x_+ \le x \le x'_+$, which is

$$|\mathrm{Im}k(x)| = \sqrt{m}\sqrt{|\mu_{\mathrm{ex}}| - \sqrt{4(E+eFx)^2 - \Delta^2}}.$$

The Zener parameter in this case is

$$\zeta(k_y) = \frac{\sqrt{m}\Delta^{3/2}}{\sqrt{2}eF}\left[\int_0^1 d\varepsilon\sqrt{\sqrt{\tilde{\mu}_{\mathrm{ex}}^2 + 1 - \varepsilon^2} + |\tilde{\mu}_{\mathrm{ex}}|}\right.$$
$$\left. + \sqrt{2}\int_1^{\sqrt{1+\tilde{\mu}_{\mathrm{ex}}^2}} d\varepsilon\sqrt{|\tilde{\mu}_{\mathrm{ex}}| - \sqrt{\varepsilon^2 - 1}}\right], \tag{J15}$$

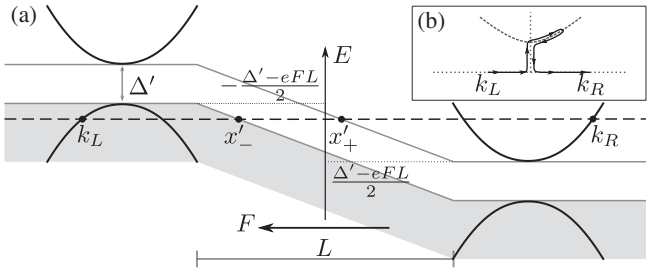

FIG. 13. (a) Tunneling scenario for $\mu_{\mathrm{ex}}(k_y) < 0$. In this case, there is no band inversion and the band gap becomes $\Delta' = \sqrt{\Delta^2 + \mu_{\mathrm{ex}}^2}$. There exists one and only one tunneling channel $k_L \to k_R$ when tunneling energy satisfies $|E| \le (eFL - \Delta')/2$. (b) The path of the complex wave vector $k(x)$ in the complex plane is indicated by the black arrow line.

where $\tilde{\mu}_{\mathrm{ex}} = \mu_{\mathrm{ex}}(k_y)/\Delta = (\mu_{\mathrm{ex}}^0 - k_y^2/m)/\Delta < 0$. Then the WKB tunneling probability is $P(k_y) = \mathrm{e}^{-2\zeta(k_y)}$ and the current density is

$$j(k_y) = -\frac{eP(k_y)}{2\pi}\delta E(k_y), \tag{J16}$$

where $\delta E(k_y) = eFL - \Delta'$.

Combining Eqs. (J13) and (J16) and integrating over k_y gives the final expression for the tunneling current density,

$$j = -\frac{e}{(2\pi)^2}\int_{-k_{y,\max}}^{k_{y,\max}} dk_y \mathrm{e}^{-2\zeta(k_y)}\delta E(k_y), \tag{J17}$$

where $\delta E(k_y) = \min[2(eFL-\Delta), \Delta'-\Delta]$ for $k_y^2 \le m\mu_{\mathrm{ex}}^0$ and is $eFL - \Delta'$ for $k_y^2 > m\mu_{\mathrm{ex}}^0$. In addition, the Zener parameter $\zeta(k_y)$ is given by Eq. (J12) for $k_y^2 \le m\mu_{\mathrm{ex}}^0$ and is Eq. (J15) for $k_y^2 > m\mu_{\mathrm{ex}}^0$.

The integration in Eq. (J17) could not be solved analytically, but we can give an upper estimation for the tunneling current. The Zener parameter $\zeta(k_y)$ is a monotonically increasing function of k_y^2; thus,

$$\zeta \ge \frac{\sqrt{m}\Delta^{3/2}}{\sqrt{2}eF}\int_0^1 d\varepsilon\sqrt{\sqrt{(\tilde{\mu}_{\mathrm{ex}}^0)^2 + 1 - \varepsilon^2} - \tilde{\mu}_{\mathrm{ex}}^0}, \tag{J18}$$

where $\tilde{\mu}_{\mathrm{ex}}^0 = \mu_{\mathrm{ex}}^0/\Delta$. It is convenient to define the correlation length of the gap (penetration depth of the electron wave function into the classically forbidden region),

$$\xi^{-1} = \sqrt{2m\Delta}\int_0^1 d\varepsilon\sqrt{\sqrt{(\tilde{\mu}_{\mathrm{ex}}^0)^2 + 1 - \varepsilon^2} - \tilde{\mu}_{\mathrm{ex}}^0}, \tag{J19}$$

and the tunneling length $\ell \equiv \Delta/eF$. Then the tunneling probability is approximated as $P = \mathrm{e}^{-2\zeta} \le \mathrm{e}^{-\ell/\xi}$. Additionally, one could verify that $\delta E(k_y) \le 2(eFL - \Delta)$, so an upper bound for the current density is estimated as

$$|j| < \frac{e(eFL-\Delta)\mathrm{e}^{-\ell/\xi}}{\pi^2}\sqrt{m\left[\mu_{\mathrm{ex}}^0 + \sqrt{(eFL)^2 - \Delta^2}\right]}, \tag{J20}$$

which generates a Zener tunneling current in the form of

$$I_z \sim (eFL - \Delta)^{3/2}\mathrm{e}^{-\ell/\xi} \tag{J21}$$

in the thermodynamic limit $eFL \gg \Delta$.

In excitonic insulators, μ_{ex}^0 appearing in this appendix should be understood as the exciton chemical potential normalized by original band gap E_g and the HF self-energy Σ^{HF}, i.e., $\mu'_{\mathrm{ex}} = \mu_{\mathrm{ex}}^0 - E_g - \mathrm{Tr}(\Sigma^{\mathrm{HF}}\sigma_z)$, which is roughly

related with exciton density as $\mu'_{\rm ex} = k_F^2/m = 4\pi n_{\rm ex}/m$. Then the correlation length as a function of exciton density is

$$\xi^{-1} = \sqrt{2m\Delta}\int_0^1 d\varepsilon\sqrt{\sqrt{(\mu'_{\rm ex})^2/\Delta^2 + 1 - \varepsilon^2} - \mu'_{\rm ex}/\Delta} = \frac{m\Delta}{\sqrt{2\pi n_{\rm ex}}}\int_0^1 d\varepsilon\frac{\sqrt{1-\varepsilon^2}}{\sqrt{\sqrt{1+(1-\varepsilon^2)(m\Delta/4\pi n_{\rm ex})^2}+1}}. \tag{J22}$$

In the high exciton density limit,

$$\xi^{-1} \approx \frac{m\Delta}{\sqrt{2\pi n_{\rm ex}}}\frac{\int_0^1 d\varepsilon\sqrt{1-\varepsilon^2}}{\sqrt{2}} = \frac{\pi m\Delta}{8\sqrt{\pi n_{\rm ex}}} = \frac{\pi\Delta}{4v_F}. \tag{J23}$$

[1] A. N. Kozlov and L. A. Maksimov, *The metal-dielectric divalent crystal phase transition*, Sov. J. Exp. Theor. Phys. **21**, 790 (1965).

[2] L. V. Keldysh and Y. V. Kopaev, *Possible instability of the semimetallic state toward Coulomb interaction*, Sov. Phys. Solid State **6**, 2219 (1965).

[3] D. Jérome, T. M. Rice, and W. Kohn, *Excitonic insulator*, Phys. Rev. **158**, 462 (1967).

[4] A. Imamoğlu, R. J. Ram, S. Pau, and Y. Yamamoto, *Nonequilibrium condensates and lasers without inversion: Exciton-polariton lasers*, Phys. Rev. A **53**, 4250 (1996).

[5] D. Snoke, *Spontaneous Bose coherence of excitons and polaritons*, Science **298**, 1368 (2002).

[6] M. Richard, J. Kasprzak, A. Baas, K. Lagoudakis, M. Wouters, I. Carusotto, R. André, B. Deveaud, S. Le, and D. Le Si, *Exciton-polariton Bose-Einstein condensation: Advances and issues*, Int. J. Nanotechnol. **7**, 668 (2010).

[7] H. Cercellier, C. Monney, F. Clerc, C. Battaglia, L. Despont, M. G. Garnier, H. Beck, P. Aebi, L. Patthey, H. Berger, and L. Forró, *Evidence for an excitonic insulator phase in* $1 \times T - \mathrm{TiSe_2}$, Phys. Rev. Lett. **99**, 146403 (2007).

[8] Y. F. Lu, H. Kono, T. I. Larkin, A. W. Rost, T. Takayama, A. V. Boris, B. Keimer, and H. Takagi, *Zero-gap semiconductor to excitonic insulator transition in* $\mathrm{Ta_2NiSe_5}$, Nat. Commun. **8**, 14408 (2017).

[9] P. A. Volkov, M. Ye, H. Lohani, I. Feldman, A. Kanigel, and G. Blumberg, *The bulk-corner correspondence of time-reversal symmetric insulators*, npj Quantum Mater. **6**, 1 (2021).

[10] Y. Jia, P. Wang, C.-L. Chiu, Z. Song, G. Yu, B. Jäck, S. Lei, S. Klemenz, F. A. Cevallos, M. Onyszczak *et al.*, *Evidence for a monolayer excitonic insulator*, Nat. Phys. **18**, 87 (2022).

[11] L. V. Butov, A. Zrenner, G. Abstreiter, G. Böhm, and G. Weimann, *Condensation of indirect excitons in coupled* AlAs/GaAs *quantum wells*, Phys. Rev. Lett. **73**, 304 (1994).

[12] A. A. High, J. R. Leonard, A. T. Hammack, M. M. Fogler, L. V. Butov, A. V. Kavokin, K. L. Campman, and A. C. Gossard, *Spontaneous coherence in a cold exciton gas*, Nature (London) **483**, 584 (2012).

[13] M. M. Fogler, L. V. Butov, and K. S. Novoselov, *High-temperature superfluidity with indirect excitons in van der Waals heterostructures*, Nat. Commun. **5**, 4555 (2014).

[14] L. Du, X. Li, W. Lou, G. Sullivan, K. Chang, J. Kono, and R.-R. Du, *Evidence for a topological excitonic insulator in InAs/GaSb bilayers*, Nat. Commun. **8**, 1971 (2017).

[15] J. I. A. Li, T. Taniguchi, K. Watanabe, J. Hone, and C. R. Dean, *Excitonic superfluid phase in double bilayer graphene*, Nat. Phys. **13**, 751 (2017).

[16] Z. Wang, D. A. Rhodes, K. Watanabe, T. Taniguchi, J. C. Hone, J. Shan, and K. F. Mak, *Evidence of high-temperature exciton condensation in two-dimensional atomic double layers*, Nature (London) **574**, 76 (2019).

[17] L. Ma, P. X. Nguyen, Z. Wang, Y. Zeng, K. Watanabe, T. Taniguchi, A. H. MacDonald, K. F. Mak, and J. Shan, *Strongly correlated excitonic insulator in atomic double layers*, Nature (London) **598**, 585 (2021).

[18] D. Nandi, A. D. K. Finck, J. P. Eisenstein, L. N. Pfeiffer, and K. W. West, *Exciton condensation and perfect Coulomb drag*, Nature (London) **488**, 481 (2012).

[19] X. Liu, K. Watanabe, T. Taniguchi, B. I. Halperin, and P. Kim, *Quantum Hall drag of exciton condensate in graphene*, Nat. Phys. **13**, 746 (2017).

[20] X. Liu, J. I. A. Li, K. Watanabe, T. Taniguchi, J. Hone, B. I. Halperin, P. Kim, and C. R. Dean, *Crossover between strongly coupled and weakly coupled exciton superfluids*, Science **375**, 205 (2022).

[21] X. Zhu, P. B. Littlewood, M. S. Hybertsen, and T. M. Rice, *Exciton condensate in semiconductor quantum well structures*, Phys. Rev. Lett. **74**, 1633 (1995).

[22] P. B. Littlewood and X. Zhu, *Possibilities for exciton condensation in semiconductor quantum-well structures*, Phys. Scr. **1996**, 56 (1996).

[23] P. B. Littlewood, P. R. Eastham, J. M. J. Keeling, F. M. Marchetti, B. D. Simons, and M. H. Szymanska, *Models of coherent exciton condensation*, J. Phys. Condens. Matter **16**, S3597 (2004).

[24] F.-C. Wu, F. Xue, and A. H. MacDonald, *Theory of two-dimensional spatially indirect equilibrium exciton condensates*, Phys. Rev. B **92**, 165121 (2015).

[25] D. I. Pikulin and T. Hyart, *Interplay of exciton condensation and the quantum spin Hall effect in InAs/GaSb bilayers*, Phys. Rev. Lett. **112**, 176403 (2014).

[26] M. Xie and A. H. MacDonald, *Electrical reservoirs for bilayer excitons*, Phys. Rev. Lett. **121**, 067702 (2018).

[27] Q. Zhu, M. W.-Y. Tu, Q. Tong, and W. Yao, *Gate tuning from exciton superfluid to quantum anomalous Hall in van der Waals heterobilayer*, Sci. Adv. **5**, eaau6120 (2019).

[28] Y. Zeng and A. H. MacDonald, *Electrically controlled two-dimensional electron-hole fluids*, Phys. Rev. B **102**, 085154 (2020).

[29] K. Yang, X. Gao, Y. Wang, T. Zhang, Y. Gao, X. Lu, S. Zhang, J. Liu, P. Gu, Z. Luo *et al.*, *Unconventional correlated insulator in CrOCl-interfaced Bernal bilayer graphene*, Nat. Commun. **14**, 2136 (2023).

[30] C. Zener and R. H. Fowler, *A theory of the electrical breakdown of solid dielectrics*, Proc. R. Soc. A **145**, 523 (1934).

[31] L. Esaki, *New phenomenon in narrow germanium $p-n$ junctions*, Phys. Rev. **109**, 603 (1958).

[32] G. H. Wannier, *Wave functions and effective Hamiltonian for Bloch electrons in an electric field*, Phys. Rev. **117**, 432 (1960).

[33] E. O. Kane, *Zener tunneling in semiconductors*, J. Phys. Chem. Solids **12**, 181 (1960).

[34] E. O. Kane and E. I. Blount, in *Tunneling Phenomena in Solids: Lectures Presented at the 1967/NATO Advanced Study Institute, Risö, Denmark*, edited by E. Burstein and S. Lundqvist (Springer US, Boston, 1969), pp. 79–91.

[35] A. C. Seabaugh and Q. Zhang, *Low-voltage tunnel transistors for beyond CMOS logic*, Proc. IEEE **98**, 2095 (2010).

[36] N. Ma and D. Jena, *Interband tunneling in two-dimensional crystal semiconductors*, Appl. Phys. Lett. **102**, 132102 (2013).

[37] G. Nenciu, *Dynamics of band electrons in electric and magnetic fields: Rigorous justification of the effective Hamiltonians*, Rev. Mod. Phys. **63**, 91 (1991).

[38] I. Souza, J. Íñiguez, and D. Vanderbilt, *First-principles approach to insulators in finite electric fields*, Phys. Rev. Lett. **89**, 117602 (2002).

[39] V. Y. Irkhin and M. I. Katsnelson, *Theory of intermediate-valence semiconductors*, Sov. Phys. JETP **63**, 631 (1986).

[40] M. Holthaus, *Bloch oscillations and Zener breakdown in an optical lattice*, J. Opt. B **2**, 589 (2000).

[41] N. Sugimoto, S. Onoda, and N. Nagaosa, *Field-induced metal-insulator transition and switching phenomenon in correlated insulators*, Phys. Rev. B **78**, 155104 (2008).

[42] L. D. Landau and E. M. Lifshits, *Quantum Mechanics: Non-Relativistic Theory* (Elsevier, Amsterdam, 2103), pp. 294–295.

[43] Y.-P. Shim and A. H. MacDonald, *Spin-orbit interactions in bilayer exciton-condensate ferromagnets*, Phys. Rev. B **79**, 235329 (2009).

[44] R. Resta, *Quantum-mechanical position operator in extended systems*, Phys. Rev. Lett. **80**, 1800 (1998).

[45] J. Zak, *Berry's phase for energy bands in solids*, Phys. Rev. Lett. **62**, 2747 (1989).

[46] D. Vanderbilt and R. D. King-Smith, *Electric polarization as a bulk quantity and its relation to surface charge*, Phys. Rev. B **48**, 4442 (1993).

[47] R. D. King-Smith and D. Vanderbilt, *Theory of polarization of crystalline solids*, Phys. Rev. B **47**, 1651(R) (1993).

[48] R. W. Nunes and X. Gonze, *Berry-phase treatment of the homogeneous electric field perturbation in insulators*, Phys. Rev. B **63**, 155107 (2001).

[49] J. Íñiguez, D. Vanderbilt, and L. Bellaiche, *First-principles study of $(BiScO_3)_{1-x}-(PbTiO_3)_x$ piezoelectric alloys*, Phys. Rev. B **67**, 224107 (2003).

[50] R. W. Nunes and D. Vanderbilt, *Real-space approach to calculation of electric polarization and dielectric constants*, Phys. Rev. Lett. **73**, 712 (1994).

[51] P. Fernández, A. Dal Corso, and A. Baldereschi, *Ab initio study of the dielectric properties of silicon and gallium arsenide using polarized Wannier functions*, Phys. Rev. B **58**, R7480 (1998).

[52] A. Kormányos, G. Burkard, M. Gmitra, J. Fabian, V. Zólyomi, N. D. Drummond, and V. Fal'ko, $k\cdot p$ *theory for two-dimensional transition metal dichalcogenide semiconductors*, 2D Mater. **2**, 022001 (2015).

[53] A. Laturia, M. L. Van de Put, and W. G. Vandenberghe, *Dielectric properties of hexagonal boron nitride and transition metal dichalcogenides: From monolayer to bulk*, npj 2D Mater. Appl. **2**, 1 (2018).

[54] I. Souza, J. Íñiguez, and D. Vanderbilt, *Dynamics of Berry-phase polarization in time-dependent electric fields*, Phys. Rev. B **69**, 085106 (2004).

[55] G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, England, 2005).

[56] L. Liu, L. Świerkowski, and D. Neilson, *Exciton and charge density wave formation in spatially separated electron–hole liquids*, Physica (Amsterdam) **249–251B**, 594 (1998).

[57] S. De Palo, F. Rapisarda, and G. Senatore, *Excitonic condensation in a symmetric electron-hole bilayer*, Phys. Rev. Lett. **88**, 206401 (2002).

[58] V. V. Nikolaev and M. E. Portnoi, *Theory of the excitonic Mott transition in quasi-two-dimensional systems*, Superlattices Microstruct. **43**, 460 (2008).

[59] R. Maezono, P. López Ríos, T. Ogawa, and R. J. Needs, *Excitons and biexcitons in symmetric electron-hole bilayers*, Phys. Rev. Lett. **110**, 216407 (2013).

[60] D. Neilson, A. Perali, and A. R. Hamilton, *Excitonic superfluidity and screening in electron-hole bilayer systems*, Phys. Rev. B **89**, 060502(R) (2014).

[61] K. Asano and T. Yoshioka, *Exciton–Mott physics in two-dimensional electron–hole systems: Phase diagram and single-particle spectra*, J. Phys. Soc. Jpn. **83**, 084702 (2014).