Dynamical screening and excitonic bound states in biased bilayer graphene

Harley D. Scammell and Oleg P. Sushkov
School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia

(Received 7 July 2022; revised 19 January 2023; accepted 23 January 2023; published 6 February 2023)

Excitonic bound states are characterized by a binding energy ϵ_b and a single-particle band gap Δ_b. This work provides a theoretical description for both strong ($\epsilon_b \sim \Delta_b$) and weak ($\epsilon_b \ll \Delta_b$) excitonic bound states, with particular application to biased bilayer graphene. Standard description of excitons is based on a wave function that is determined by a Schrödinger-like equation with screened attractive potential. The wave function approach is valid only in the weak-binding regime $\epsilon_b \ll \Delta_b$. The screening depends on frequency, i.e., dynamical screening, and this implies retardation. In the case of strong binding, $\epsilon_b \sim \Delta_b$, a wave function description is not possible due to the retardation. Instead we appeal to the Bethe-Salpeter equation, written in terms of the electron-hole Green's function, to solve the problem. So far only the weak-binding regime has been achieved experimentally. Our analysis demonstrates that the strong-binding regime is also possible and we specify conditions in which it can be achieved for the prototypical example of biased bilayer graphene. The conditions concern the bias, the configuration of gates, and the substrate material. To verify the accuracy of our analysis we compare with available data for the weak-binding regime. We anticipate applying the developed dynamical screening Bethe-Salpeter techniques to various 2D materials with strong binding.

DOI: 10.1103/PhysRevB.107.085104

I. INTRODUCTION

Graphene layers, and the manipulation thereof, are the model hunting ground for peculiar single-particle quantum phases of matter, such as topological insulators, as well as many-body phases, including superconductivity. An important many-body phase actively pursued in graphene layers is the excitonic insulator [1–3]—a many-body ground state comprising condensed particle-hole pairs. This phase holds promise for novel superfluidity that could be harnessed for low-energy technology [4–8]. Understanding of an isolated exciton is a necessary step for understanding the exciton condensation.

An exciton is a particle-hole bound state in a band insulator. Excitons in biased bilayer graphene (BBG) have been observed several years ago [9]. Theoretically the exciton problem in BBG has been considered in Refs. [10–13]. These works ultimately employ the instantaneous screened Coulomb approximation to find the binding energy and the wave function of the exciton. Often this approach is referred to as the Bethe-Salpeter equation (BSE); however, it is necessary to clarify the terminology: For an instantaneous interaction, a Hamiltonian approach is valid, known generically as the Lippmann-Schwinger equation (LSE). In the case of retardation the approach is the BSE. The distinction is important; the LSE provides a relatively simple wave function description, whereas for the BSE a wave function is not possible, and instead the correct object is the two-particle Green's function [14].

In this work we address the issue of retardation in electron-hole binding in BBG. There are two main parameters in the problem, exciton binding energy ϵ_b and single-particle band gap Δ_b induced by bias. In the weak-binding limit, $\epsilon_b \ll \Delta_b$, retardation is negligible. This is the limit addressed in the existing experiment [9] as well in previous theoretical works [10–13]. Contrary to this, we find that in the case of strong binding, $\epsilon_b \sim \Delta_b$, the retardation is non-negligible, and acts to significantly enhance the binding energy. Notably, the strong-binding regime is essential to understand the possibility of exciton condensation. However, we leave the pursuit of condensation for future work. The importance of retardation in some two-dimensional semiconductor exciton problems has been previously pointed out in Ref. [15]. The authors of Ref. [15] replace the zero frequency in the screened potential by some effective frequency, and ultimately solve the LSE.

Full solution of the BSE is numerically challenging, and presents a bottleneck. In this work we develop a systematic method to account for retardation, and at a low numerical cost. This is achieved through a perturbative expansion of the BSE. Employing such techniques in the case of BBG, our analysis demonstrates that the strong-binding regime in BBG is possible and we specify conditions in which it can be achieved. The conditions concern the bias, the configuration of gates, and the substrate material. To confirm the validity of our perturbative treatment, we also implement the numerically exact solution to the BSE; comparison shows that the perturbation theory (to first order) accounts for the majority retardation correction in the weak-to-intermediate binding regime, and still captures approximately 50%, or more, of the correction deep in the strong-binding regime $\epsilon_b \sim \Delta_b$. The first-order perturbation theory is therefore an efficient means to establish the importance of retardation.

In BBG, the single-particle band gap Δ_b is proportional to a perpendicular (to the BBG plane) displacement field, which

is generated via metallic gates above and below the plane; e.g., [16]. We will assume that the gates are symmetrically placed. Screening, in general, has a significant affect on the excitonic binding energy, ϵ_b. There are three sources of screening: (i) dielectric due to a material between BBG and the gates, (ii) metallic gates, and (iii) in-plane, captured using the random phase approximation (RPA). We will see that to get to the strong-coupling regime, $\epsilon_b \sim \Delta_b$, it is necessary to eliminate the dielectric material and use suspended BBG; suspended BBG has been experimentally achieved [17,18]. Typical energies that we consider are below 100 meV. Within this range the gate metallic screening is practically frequency independent, but (iii) is frequency dependent. This is the origin of the effect that we address.

The rest of the paper is structured as follows: In Sec. II we establish the single-particle Hamiltonian and the behavior of the screened Coulomb interaction. Section III introduces the LSE approach, which allows for a particularly straightforward treatment of the two-body exciton problem, without account of retardation. Section IV moves onto the BSE, which accounts for retardation (due to dynamical screening). The approach is more demanding numerically, so we first develop a perturbative expansion which allows for a relatively simple numerical implementation. We subsequently provide a numerically exact treatment of the BSE. Together, we find that with account of retardation, binding energies are significantly enhanced and we predict that the strong binding/exciton condensation is possible to implement experimentally. As an additional consistency check we analyze existing experimental data for weakly bound excitons in BG. Without introducing fitting parameters, we show excellent quantitative agreement. Finally, we discuss our findings and their relation to future experiments in Sec. V.

II. SINGLE-PARTICLE HAMILTONIAN AND SCREENED COULOMB INTERACTION

A. Single-particle Hamiltonian

We will see that the spatial size of the exciton is about $r \sim 1/\sqrt{m\Delta_b}$, where $m \sim 3 \times 10^{-2} m_e$ is the effective mass; hereafter we set $\hbar = 1$. Even at a large band gap $\Delta_b = 100\,\text{meV}$ the size is about 5 nm; the continuum approximation is therefore sufficient for analysis of the problem. The low-energy single-particle Hamiltonian of BBG is [19]

$$H_0 = \begin{pmatrix} \Delta & -\frac{p_-^2}{2m} \\ -\frac{p_+^2}{2m} & -\Delta \end{pmatrix}, \tag{1}$$

which is written in terms of $\{A_1, B_2\}$ orbitals, with A, B referring to graphene sublattice and subscripts 1,2 referring to layers. Here $p_\pm = \tau p_x \pm i p_y$, $\boldsymbol{p}$ is the in-plane momentum, $\tau = \pm 1$ the valley quantum number, m the effective mass, and $\Delta = \Delta_b/2$, which is proportional to the bias electric field [16].

There are known corrections to this Hamiltonian [20], which we gather as a perturbation,

$$\delta H = \begin{pmatrix} \frac{p^2}{2M}\left(1 - \frac{\Delta}{\Delta_0}\right) & \frac{P_0 p_+}{2m} \\ \frac{P_0 p_-}{2m} & \frac{p^2}{2M}\left(1 + \frac{\Delta}{\Delta_0}\right) \end{pmatrix}. \tag{2}$$

This captures three types of perturbations, $\frac{P_0 p_\pm}{2m}$, $\frac{p^2}{2M}$, and $\frac{p^2}{2M}\frac{\Delta}{\Delta_0}$; the first perturbation controls trigonal warping, the second controls particle-hole asymmetry, and the third does not break symmetries, but is nonetheless treated as a small correction.

The single-particle band gap $\Delta_b = 2\Delta$ encodes the applied displacement field, D. The conversion is observed to be approximately linear [16], with

$$eD/(2\Delta) \approx 10.4\ \text{nm}^{-1} \tag{3}$$

taken as a linear fit to experimental data [16] in the range $D \in (0, 1.9)\,\text{V nm}^{-1}$. Meanwhile, all other parameters of the single-particle Hamiltonian can be related to the standard Slonczewski-Weiss-McClure (SWM) parameters of BG [20]

$$\begin{aligned} m &= \frac{2\gamma_1}{3a^2\gamma_0^2} \approx 0.032 m_e, \\ M &= \frac{m}{2\gamma_4/\gamma_0 + \Delta'/\gamma_1} \approx \frac{m}{0.146} \approx 0.22 m_e, \\ \Delta_0 &= \frac{\gamma_1}{2}\frac{m}{M} \approx 28\ \text{meV}, \\ P_0 &= \frac{2\gamma_1\gamma_3}{\sqrt{3}\gamma_0^2 a} \approx 0.068\ \text{nm}^{-1}. \end{aligned} \tag{4}$$

To obtain the numerical values, we have taken the SWM parameters established by the experiment and analysis of Ref. [21],

$$\begin{aligned} a = 2.46\ \text{Å}, \quad \gamma_0 = 3.16\ \text{eV}, \quad \gamma_1 = 0.381\ \text{eV}, \\ \gamma_3 = 0.38\ \text{eV}, \quad \gamma_4 = 0.14\ \text{eV}, \quad \Delta' = 0.022\ \text{eV}. \end{aligned} \tag{5}$$

Taken together, this uniquely fixes the single-particle Hamiltonian; we therefore have not introduced fitting parameters.

The Hamiltonian (1) [with/without the corrections (2)] determines the dispersion of the valence, $\epsilon_{\boldsymbol{p}}^{(-)}$, and conduction, $\epsilon_{\boldsymbol{p}}^{(+)}$, bands, as well as the corresponding wave functions $\psi_{\boldsymbol{p},\tau}^{(\pm)}$ (spin index is idle, and so is not included). We will perform calculations for the Hamiltonian (1) with and without the corrections (2); we will see that perturbations influence results rather weakly for the parameter range of interest. Therefore, for presentation we omit the correction (2) everywhere except of the comparison with experiment (in Fig. 6). Excluding the small corrections (2), it is convenient to perform the analysis in rescaled units, whereby energy is measured in units of Δ and momentum in units of $p_0 = \sqrt{2m\Delta}$,

$$\begin{aligned} \overline{\epsilon} &= \epsilon/\Delta, \\ \overline{p} &= p/p_0 = p/\sqrt{2m\Delta}. \end{aligned} \tag{6}$$

In these units the Hamiltonian (1) reads

$$H_0 = \begin{pmatrix} 1 & -\overline{p}_-^2 \\ -\overline{p}_+^2 & -1 \end{pmatrix}, \tag{7}$$

with eigenenergies $\epsilon_q^{(\pm)} = \pm\sqrt{1 + \overline{p}^4}$.

B. Screened Coulomb interaction

The many-body RPA screening of Coulomb interaction in BG is significant. We recall the essential details. The

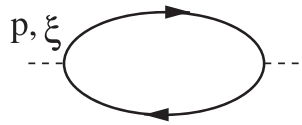

FIG. 1. Electron polarization operator.

polarization operator at an imaginary frequency ξ, as given by the diagram in Fig. 1, reads (for generic RPA details see, e.g., [22])

$$\Pi(\boldsymbol{p}, i\xi) = 2 \times 2 \times \int D_{\boldsymbol{q}} \frac{2(\epsilon^{(-)}_{\boldsymbol{q}} - \epsilon^{(+)}_{\boldsymbol{q}+\boldsymbol{p}}) F^{+-}_{\boldsymbol{q},\boldsymbol{q}+\boldsymbol{p}}}{(\epsilon^{(-)}_{\boldsymbol{q}} - \epsilon^{(+)}_{\boldsymbol{q}+\boldsymbol{p}})^2 + \xi^2}. \tag{8}$$

Throughout the paper we use the notation $D_{\boldsymbol{q}} = d^2q/(2\pi)^2$. The prefactor 2×2 in (8) comes from spin and valley degeneracy; the vertex factor is the overlap of the conduction and valence single-particle wave functions, $F^{+-}_{\boldsymbol{q},\boldsymbol{q}+\boldsymbol{p}} = |\langle \psi^{(+)}_{\boldsymbol{q}+\boldsymbol{p},\tau} | \psi^{(-)}_{\boldsymbol{q},\tau} \rangle|^2$. We do not account for vertices that change the valley; the resulting Coulomb interaction would be significantly suppressed due to the momentum ratio $q/K \ll 1$, where K is the valley momentum. The vertex factor is zero at $\boldsymbol{q} = \boldsymbol{0}$. The polarization operator has dimension of mass; rewriting Eq. (8) in dimensionless units (6) gives

$$\Pi(\boldsymbol{p}, i\xi) = 2m\overline{\Pi}(\overline{\boldsymbol{p}}, i\overline{\xi}), \tag{9}$$

where $\overline{\Pi}$ is given by the same Eq. (8), but all the variables replaced by those with bars. At $\overline{q} \gg 1$ the polarization operator is $\overline{\Pi} = \frac{\ln 4}{\pi} \approx 0.441$, which reduces to the $\Delta = 0$ case considered in Ref. [23]. In the general case of arbitrary Δ, the operator is straightforwardly computed numerically. Plots of $-\overline{\Pi}$ versus $\overline{p}$ for different values of $\overline{\xi}$ are presented in Fig. 2(a). The polarization operator is zero at $\overline{p} = 0$ and it is approaching $(\ln 4)/\pi$ at large $\overline{p}$. The frequency dependence of the polarization operator becomes significant when the frequency is comparable and larger than the band gap, $\xi \gtrsim \Delta_b = 2\Delta$. In the Appendix we discuss the validity of the polarization operator (8) at large frequencies.

There are two points to note on the behavior of the polarization operator: (i) The polarization screening becomes significant at $\overline{p} > 1$. The scale $\overline{p} \sim 1$ will determine the size of the exciton, $1/r \sim p_0 = \sqrt{2m\Delta}$; and (ii) the (imaginary) frequency monotonically reduces screening, i.e., $\overline{\Pi}(\overline{\boldsymbol{p}}, i\overline{\xi}) < \overline{\Pi}(\overline{\boldsymbol{p}}, 0)$ for all $\xi > 0$. We would like to stress that this is true only for imaginary frequency; for real frequencies the screening properties are complicated and obscured. It is convenient to work with imaginary frequency.

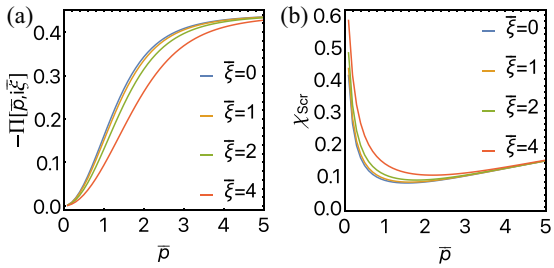

FIG. 2. Polarization operator $-\overline{\Pi}$ (a) and the Coulomb interaction screening factor $\chi_{\rm scr}$ (b) versus $\overline{p}$ for different values of $\overline{\xi}$. The screening factor is presented for $\Delta = 15$ meV and $\epsilon_r = 1$.

Using notation $e^2 = e_0^2/\epsilon_r$, where e_0 is the bare charge and ϵ_r is the dielectric constant, the screened Coulomb interaction is

$$\begin{aligned} V_p(\xi) &= -\frac{2\pi e^2}{p - 2\pi e^2 \Pi(\boldsymbol{p}, i\xi)} = \frac{1}{2m}\overline{V}_{\overline{p}}(\overline{\xi}), \\ \overline{V}_{\overline{p}}(\overline{\xi}) &= -\frac{4\pi}{s\overline{p} - 4\pi\overline{\Pi}(\overline{\boldsymbol{p}}, i\overline{\xi})}, \\ s &= \sqrt{\frac{\epsilon_r^2 \Delta}{m e_0^4/2}}. \end{aligned} \tag{10}$$

In these dimensionless units the interaction depends only on the ratio of Δ over "Rydberg," $\mathrm{Ry} = \frac{me^4}{2}$. At $\epsilon_r = 1$ the Ry value is $\mathrm{Ry} \approx 435$ meV. Hence, for $\Delta = 15$ meV the parameter s is $s = 0.186$. In Fig. 2(b) we plot the screening factor of the bare Coulomb interaction,

$$\chi_{\rm scr} = \frac{-\overline{V}_{\overline{p}}}{(4\pi/s\overline{p})} = \frac{s\overline{p}}{s\overline{p} - 4\pi\overline{\Pi}(\overline{\boldsymbol{p}}, i\overline{\xi})}, \tag{11}$$

for these parameters. Figure 2(b) further emphasizes the behavior seen in Fig. 2(a), i.e., that (i) screening is the most significant at $\overline{p} \sim 1$; and (ii) screening is significantly reduced at high imaginary frequency $\overline{\xi} \gg 1$.

III. LIPPMANN-SCHWINGER EQUATION

We define quasimomentum, $\boldsymbol{p}$, with respect to the valley minimum (K point) and only consider bound states with zero total quasimomentum, i.e., electron in valley τ, with momentum $\boldsymbol{p} + \tau\boldsymbol{K}$ pairing with a hole in valley τ', with total momentum $-\boldsymbol{p} - \tau'\boldsymbol{K}$. This means that the total momentum, i.e., as defined with respect to the Γ point, is zero if the electron and hole pair in the same valley (intravalley exciton) and it is equal to $\pm 2\boldsymbol{K}$ if they pair in different valleys (intervalley exciton). Optically, only the intravalley exciton can be excited.

LSE is a result of summation of ladder diagrams for an instantaneous interaction between the particles. To apply the LSE, we consider the interaction (10), and set $\xi = 0$. For the purposes of presentation, we do not include the small corrections (2), and therefore $\epsilon^{(+)}_{\boldsymbol{p}} = -\epsilon^{(-)}_{-\boldsymbol{p}} \equiv \epsilon_{\boldsymbol{p}} = \sqrt{\Delta^2 + \frac{p^4}{4m^2}}$. For some of our numerics, we account for the corrections (2). The bound state equation reads

$$\begin{aligned} (E_0 - 2\epsilon_{\boldsymbol{p}})\Psi_{\boldsymbol{p}} &= \int V_{\boldsymbol{p}-\boldsymbol{k}}(0) Z_{\boldsymbol{p},\boldsymbol{k}} \Psi_{\boldsymbol{k}} D_{\boldsymbol{k}}, \\ E_0 &= 2\Delta - \epsilon_b^{(0)}. \end{aligned} \tag{12}$$

Here $\Psi_{\boldsymbol{p}}$ is the exciton wave function, and $\epsilon_b^{(0)}$ is the exciton binding energy (the subscript/superscript "0" is used to distinguish from the case with account of retardation, to be discussed in Sec. IV). The vertex form factors are given by $Z^{\tau',\tau}_{\boldsymbol{p},\boldsymbol{k}} = \langle \psi^{(-)}_{\boldsymbol{p},\tau'} | \psi^{(-)}_{\boldsymbol{k},\tau} \rangle \langle \psi^{(+)}_{\boldsymbol{k},\tau} | \psi^{(+)}_{\boldsymbol{p},\tau} \rangle$. The form factors do not distinguish spin, yet they weakly distinguish between intra- and intervalley excitons. Note that τ corresponds to the valley where the electron is located and τ' corresponds to the valley

FIG. 3. Bilayer graphene with two metallic gates, placed at a distance d above and below.

where the hole is located. Explicitly, the expression is

$$Z^{\tau',\tau}_{\boldsymbol{p},\boldsymbol{k}} = \frac{(1+\epsilon_{\boldsymbol{p}}\epsilon_{\boldsymbol{k}}e^{2i\tau'\theta})}{1+\epsilon^2_{\boldsymbol{p}}}\frac{(1+\epsilon_{\boldsymbol{p}}\epsilon_{\boldsymbol{k}}e^{2i\tau\theta})}{1+\epsilon^2_{\boldsymbol{k}}}. \tag{13}$$

Here $\theta = \theta_{\boldsymbol{k}} - \theta_{\boldsymbol{p}}$ and $\epsilon_{\boldsymbol{p}} = 2m(\varepsilon_{\boldsymbol{p}} - \Delta)/p^2$. Hence, for the intervalley exciton, $\tau' = -\tau$, the Z factor is real,

$$Z^{-\tau,\tau}_{\boldsymbol{p},\boldsymbol{k}} = \frac{1+\epsilon^2_{\boldsymbol{p}}\epsilon^2_{\boldsymbol{k}} + 2\epsilon_{\boldsymbol{p}}\epsilon_{\boldsymbol{k}}\cos(2\theta)}{\big(1+\epsilon^2_{\boldsymbol{p}}\big)\big(1+\epsilon^2_{\boldsymbol{k}}\big)}. \tag{14}$$

At the same time for the intravalley exciton, $\tau' = \tau$, the Z factor is complex,

$$Z^{\tau,\tau}_{\boldsymbol{p},\boldsymbol{k}} = \frac{1+\epsilon^2_{\boldsymbol{p}}\epsilon^2_{\boldsymbol{k}}e^{4i\tau\theta} + 2\epsilon_{\boldsymbol{p}}\epsilon_{\boldsymbol{k}}e^{2i\tau\theta}}{\big(1+\epsilon^2_{\boldsymbol{p}}\big)\big(1+\epsilon^2_{\boldsymbol{k}}\big)}. \tag{15}$$

In dimensionless units (6), Eq. (12) reads

$$(\overline{E}_0 - 2\overline{\epsilon}_{\overline{p}})\Psi_{\overline{p}} = \int \overline{V}_{\overline{p}-\overline{k}}(0)Z_{\overline{p},\overline{k}}\Psi_{\overline{k}}D_{\overline{k}}. \tag{16}$$

The interaction $\overline{V}$ is defined in (10). For brevity, we suppress valley indices on the Z factors. As already pointed out, the solution of (16) depends only on the dimensionless parameter s, defined in (10). The interaction in (10) has not accounted for an important effect—screening due to the metallic gates. We consider a setup shown in Fig. 3, whereby the top and bottom gates are a distance d from the BG plane. Accounting for gate screening, via the method of images, the interaction in Eq. (10) is replaced by

$$\overline{V}_{\overline{p}}(\overline{\xi}) = -\frac{4\pi}{s\overline{p}/\Upsilon_{\overline{p}} - 4\pi\,\overline{\Pi}(\overline{\boldsymbol{p}}, i\overline{\xi})},$$

$$\Upsilon_{\overline{p}} = \tanh(pd) = \tanh(p_0\overline{p}d). \tag{17}$$

Without account of the small trigonal warping $\propto P_0$ in (2), we may classify excitonic states $\Psi_{\boldsymbol{p}}$ via 2D angular harmonics $e^{i\ell\theta_p}$, using

$$\Psi_{\overline{\boldsymbol{p}}} = \sum_{\ell}\frac{1}{\sqrt{\overline{p}}}\,\psi^{\ell}_{\overline{p}}\,e^{i\ell\theta_{\overline{p}}}, \tag{18}$$

where $\psi^{\ell}_{\overline{p}}$ depends only on the absolute value of momentum. In a channel with a given orbital momentum l Eq. (16) is reduced to

$$(\overline{E}_0 - 2\overline{\epsilon}_{\overline{p}})\psi^{\ell}_{\overline{p}} = \int_0^{\infty} V^{\ell}_{\overline{p},\overline{k}}\,\psi^{\ell}_{\overline{k}}\,d\overline{k},$$

$$V^{\ell}_{\overline{p},\overline{k}} = \frac{\sqrt{p\overline{k}}}{(2\pi)^2}\int_0^{2\pi} e^{-i\ell\theta_{\overline{p}}}\,\overline{V}_{\overline{p}-\overline{k}}(0)\,Z_{\overline{p},\overline{k}}\,e^{i\ell\theta_{\overline{k}}}\,d\theta_{\overline{k}}. \tag{19}$$

Note, $V^{\ell}_{\overline{p},\overline{k}}$ is independent of $\theta_{\overline{p}}$, since the integrand in the second line of (19) is a function of $\theta_{\overline{k}} - \theta_{\overline{p}}$.

Brute force numerical solution of (19) is straightforward. We consider three cases: (i) suspended BG with the dielectric constant $\epsilon_r = 1$; (ii) single-sided hBN substrate with effective $\epsilon_r = (3.9+1)/2 = 2.45$; and (iii) double-sided hBN substrate with effective $\epsilon_r = 3.9$. The binding energy of the s-wave ($\ell = 0$) ground state of intervalley exciton versus Δ is plotted in Fig. 4(a)(i). There are nine lines corresponding to the three values of the dielectric constant and to the three values of the distance to metallic gates, $d = 20, 100, 1000$ nm. As expected, there is a significant dependence of the binding energy on the dielectric constant ϵ_r. However, given that the characteristic exciton radius is $r \sim 10$ nm, the strong dependence on the gate distance $d > 20$ nm is somewhat unexpected. Ultimately, this is because the Coulomb interaction, at zero momentum, is $V_{p\to 0} = 2\pi e^2 d$. Another surprising observation is practical independence of the wave function on the binding energy. The wave functions corresponding to very different binding energies are plotted in Fig. 4(a)(ii); we see that, in dimensionless momenta, the wave fuctions are insensitive to the system parameters. In the original units they are of course different: $\overline{p} = 1$ corresponds to $p = 0.130\ \text{nm}^{-1}$ for $\Delta = 20$ meV and to $p = 0.206\ \text{nm}^{-1}$ for $\Delta = 50$ meV. We attribute this universal behavior to the shape of the polarization operator, which is small $\propto \overline{p}^2$ up to a scale $\overline{p} = 1$.

Due to the difference of the Z factors, Eqs. (14) and (15), binding energies of intervalley and intravally excitons are slightly different. The intervalley exciton has a stronger binding. The difference of binding energies,

$$\Delta\epsilon^s_b = \epsilon^{(0)}_b(\text{inter}) - \epsilon^{(0)}_b(\text{intra}), \tag{20}$$

is plotted versus Δ in Fig. 4(a)(iii) for $\epsilon_r = 1$ and $d = 1000$ nm. The difference in binding energies is less than 1%. It is even smaller for higher ϵ_r and lower d.

For the p-wave exciton we again start from the intervalley case. In this case the energies of $l = \pm 1$ states are degenerate. The binding energy of the lowest p-wave state of the intervalley exciton is plotted in Fig. 4(b)(i) versus Δ for the same values of ϵ_r and d as that for the s wave in Fig. 4(a)(i). The wave functions corresponding to different binding energies are plotted in Fig. 4(b)(ii).

We already pointed out that the inervalley p-wave exciton states with $l = \pm 1$ are degenerate. At the same time the intravalley p-wave states with $l = \pm 1$ are nondegenerate, and the following relation is valid,

$$\epsilon^{(0)}_b(\text{intra}) = \epsilon^{(0)}_b(\text{inter}) + \tau l\,\Delta\epsilon^p_b, \tag{21}$$

where τ indicates the valley. The splitting $\Delta\epsilon^p_b$ is plotted in Fig. 4(b)(iii) versus Δ for $\epsilon_r = 1$ and $d = 1000$ nm. The splitting is again less than 1%. It is even smaller for higher ϵ_r and lower d.

IV. RETARDATION AND BETHE-SALPETER EQUATION

Now we proceed to the central message of this work. In the analysis of the previous section, an important effect was neglected—retardation of the screened Coulomb interaction. According to Fig. 2, screening of the Coulomb

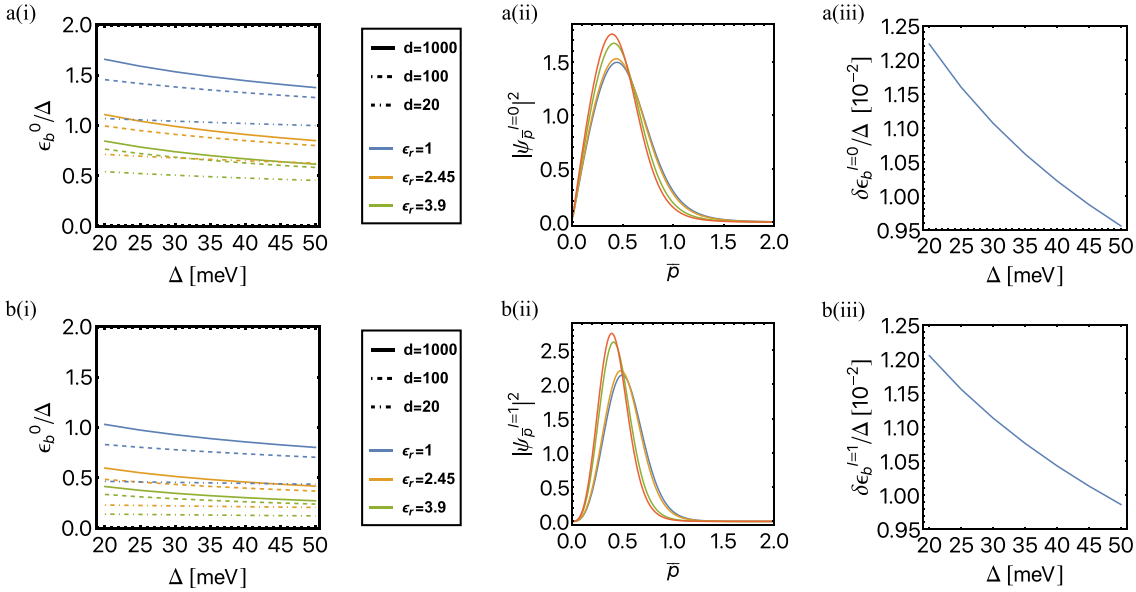

FIG. 4. (a) Exciton s-wave ($\ell = 0$) ground state, and (b) exciton p-wave ($\ell = 1$) ground state. (a) and (b) (i) Binding energies of intervalley exciton in the LSE approximation versus Δ. The plots are presented for three values of the dielectric constant $\epsilon_r = \{1, 2.45, 3.9\}$ and three values of the distance to gates $d = \{20, 100, 1000\}$ nm. (a) and (b) (ii) Wave functions versus dimensionless momentum $\bar{p}$ for different sets of parameters. The blue line corresponds to $\Delta = 20$ meV, $\epsilon_r = 1$, $d = 1000$ nm; the orange line corresponds to $\Delta = 50$ meV, $\epsilon_r = 1$, $d = 1000$ nm; the green line corresponds to $\Delta = 20$ meV, $\epsilon_r = 3.9$, $d = 20$ nm; the red line corresponds to $\Delta = 50$ meV, $\epsilon_r = 3.9$, $d = 20$ nm. (a) and (b) (iii) The energy splitting, Eq. (20), between the intervalley and intravalley excitons versus Δ for $\epsilon_r = 1$ and $d = 1000$ nm.

attraction is monotonically reduced with frequency. We already pointed out, and we would like to stress again, that this is true only for imaginary frequency. Analytic continuation to real frequency obscures this simple behavior. It is therefore convenient and physically transparent to work with imaginary frequency. Reduced screening enhances the binding energy. Heuristically, the typical frequency is set by the binding energy, and therefore when the binding energy is much smaller than the band gap, $\epsilon_b \ll 2\Delta$, the effect of screening reduction is small. However, at strong binding, $\epsilon_b \sim 2\Delta$, the effect of frequency dependence (retardation) becomes significant.

With account of retardation, the electron and hole interact at different times and the bound state cannot be described by a wave function. Instead of the wave function, the correct object is the amputated two-particle Green's function $\chi_{\xi,p}$—written here in terms of the relative momentum $\boldsymbol{p}$ and the relative frequency ξ. In our analysis, the total momentum of the electron and hole is encoded in the valley indices, and is either zero (intravalley pairing) or $2\boldsymbol{K}$ (intervalley pairing). The imaginary frequency ξ is Fourier conjugated to the retardation time (plus a Wick rotation). BSE for $\chi_{\xi,p}$ reads [14]

$$\chi_{\xi,p} = -\frac{1}{(E/2-\epsilon_p)^2+\xi^2}\int V_{p,k}(\xi-\lambda)\chi_{\lambda,k}Z_{p,k}D_\lambda D_k,$$
$$D_\lambda = \frac{d\lambda}{2\pi}, \quad D_k = \frac{d^2k}{(2\pi)^2}. \tag{22}$$

Here $E = 2\Delta - \epsilon_b$, and ξ, λ are imaginary frequencies. If the interaction is independent of frequency, $V_{p,k}(\xi-\lambda) \to V_{p,k}(0)$, BSE (22) is equivalent to LSE (12), and χ is related to the usual wave function Ψ_p as

$$\chi^{(0)}_{\xi,p} = \frac{2a_p}{a_p^2+\xi^2}\Psi_p,$$
$$a_p = -E_0/2 + \epsilon_p. \tag{23}$$

The superscript/subscript "0" indicates that this is the solution without retardation. Note that a_p is always positive.

A. Perturbation theory for the retardation effect

Equation (22) is not a linear eigenvalue problem, and a direct numerical solution of Eq. (22) is an involved calculation which we detail in Sec. IV B. In this subsection we develop a perturbative method, and explicitly consider the first-order term in perturbation series. This technique is numerically efficient to implement and will be seen to be sufficient to quantitatively describe the regimes of weak and intermediate binding, while providing qualitative information on the strong-binding regime. Let us first replace the interaction in (22) by a frequency-independent interaction, $V_{p,k}(\xi) \to V^{(0)}_{p,k}$. It can be the interaction at zero frequency, $V^{(0)}_{p,k} = V_{p,k}(\xi = 0)$, or interaction at some typical frequency $V^{(0)}_{p,k} = V_{p,k}(\xi = \xi_{\text{typical}})$, or something else. Here we exclusively consider $V^{(0)}_{p,k} = V_{p,k}(\xi = 0)$. BSE with $V^{(0)}_{p,k}$ is reduced to LSE which is a linear eigenvalue problem and can be solved numerically with ease. The solution is given by Eq. (23) where Ψ_p and E_0 is the eigenfunction and the eigenenergy of LSE. Next, let us consider

$$\delta V_{p,k}(\xi) = V_{p,k}(\xi) - V^{(0)}_{p,k} \tag{24}$$

as a perturbation. We obtain the following expression for the first-order retardation correction to the binding energy,

$$\delta E = \int \left(\chi^{(0)}_{\xi,p}\right)^* \delta V_{p,k}(\xi-\lambda)\chi^{(0)}_{\lambda,\mathbf{k}} Z_{p,k} D_\xi D_p D_\lambda D_{\mathbf{k}}. \quad (25)$$

To derive (25) let us represent the Green's function as $\chi_{\xi,p} = \chi^{(0)}_{\xi,p} + \delta\chi_{\xi,p}$. Hence, BSE (22) can be rewritten as

$$\begin{aligned}
&\chi^{(0)}_{\xi,p} + \delta\chi_{\xi,p} \\
&\quad = -\frac{1}{(E/2-\epsilon_p)^2+\xi^2}\int \left[V^{(0)}_{p,k} + \delta V_{p,k}(\xi-\lambda)\right] \\
&\qquad \times \left[\chi^{(0)}_{\lambda,\mathbf{k}} + \delta\chi_{\lambda,\mathbf{k}}\right] Z_{p,k} D_\lambda D_k \\
&\quad \approx -\frac{1}{(E/2-\epsilon_p)^2+\xi^2}\int \left[V^{(0)}_{p,k}\chi^{(0)}_{\lambda,\mathbf{k}} + \delta V_{p,k}(\xi-\lambda)\chi^{(0)}_{\lambda,\mathbf{k}}\right. \\
&\qquad \left. + V^{(0)}_{p,k}\delta\chi_{\lambda,\mathbf{k}}\right] Z_{p,k} D_\lambda D_k. \quad (26)
\end{aligned}$$

The second-order term $\delta V \delta\chi$ has been neglected in the last line. Let us denote

$$\delta\Psi_p = \int \delta\chi_{\xi,p} D_\xi. \quad (27)$$

Integrating (26) over ξ we get

$$\begin{aligned}
\Psi_p + \delta\Psi_p = &+\frac{1}{E-2\epsilon_p}\int V^{(0)}_{p,k} Z_{p,k}\Psi_k D_k \\
&- \int \frac{D_\xi}{(E/2-\epsilon_p)^2+\xi^2}\delta V_{p,k} \\
&\times (\xi-\lambda)\chi^{(0)}_{\lambda,\mathbf{k}} Z_{p,k} D_\lambda D_k \\
&+ \frac{1}{E-2\epsilon_p}\int V^{(0)}_{p,k} Z_{p,k}\delta\Psi_k D_k,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
(E-2\epsilon_p)(\Psi_p+\delta\Psi_p) = &\int V^{(0)}_{p,k} Z_{p,k}\Psi_k D_k \\
&+ \int \frac{2(-E/2+\epsilon_p)}{(E/2-\epsilon_p)^2+\xi^2}\delta V_{p,k} \\
&\times (\xi-\lambda)\chi^{(0)}_{\lambda,\mathbf{k}} Z_{p,k} D_\xi D_\lambda D_k \\
&+ \int V^{(0)}_{p,k} Z_{p,k}\delta\Psi_k D_k.
\end{aligned}$$

Representing $E = E_0 + \delta E$ and neglecting all the second-order terms, i.e., $O(\delta\times\delta)$, this is transformed to

$$\begin{aligned}
\delta E\Psi_p + (E_0-2\epsilon_p)\delta\Psi_p = &\int \frac{2(-E_0/2+\epsilon_p)}{(E_0/2-\epsilon_p)^2+\xi^2}\delta V_{p,k} \\
&\times (\xi-\lambda)\chi^{(0)}_{\lambda,\mathbf{k}} Z_{p,k} D_\xi D_\lambda D_k \\
&+ \int V^{(0)}_{p,k} Z_{p,k}\delta\Psi_k D_k.
\end{aligned}$$

Finally, multiplying this equation by Ψ^*_p and integrating by $\boldsymbol{p}$ we arrive at Eq. (25). Using (23), one frequency integration in (25) can be performed analytically and the retardation correction to the energy reduces to

$$\delta E = \int \Psi^*_p \frac{4(a_p+a_k)\delta V_{p,k}(\nu)}{(a_p+a_k)^2+\nu^2}\Psi_{\mathbf{k}} Z_{p,k} D_\nu D_p D_{\mathbf{k}}. \quad (28)$$

Here the ν integration goes from 0 to ∞. Since δV is negative the retardation correction to the total energy is negative, $\delta E < 0$, and thereby increases the binding energy. If δV is independent of frequency, $\delta V_{p,k}(\nu) \to \delta V_{p,k}$, the ν integration in (28) is trivial and the energy correction is reduced to the familiar expression from quantum mechanics, $\delta E = \int \Psi^*_p \delta V_{p,k}\Psi_{\mathbf{k}} Z_{p,k} D_p D_{\mathbf{k}}$.

Momenta integrations in (28) are well convergent. On the other hand, the perturbation $\delta V(\nu)$ increases with frequency, and we find that at intermediate-to-large ν the integrand in (28) decays as $1/\nu$ leading to a logarithmic divergence of the integral. However, ultimately the $\delta V(\nu\to\infty) \to$ constant; the integrand is therefore $\propto\nu^{-2}$ and hence the integral over ν converges. (By direct numerical evaluation, we find that the integral converges at $\nu \lesssim 3\,\mathrm{eV}$.) The validity of the screened Coulomb interaction at large frequencies is discussed in the Appendix.

B. Exact BSE

To complement the perturbative treatment of the retardation effect, we appeal now to the numerically exact solution of the BSE. Starting from the BSE (22), let us perform the following replacement of variables,

$$\begin{aligned}
\mathcal{X}_{\xi,p} &= \sqrt{a_p^2+\xi^2}\,\chi_{\xi,p}, \\
\mathcal{V}_{p,k}(\xi-\lambda) &= \frac{V_{p,k}(\xi-\lambda)Z_{p,k}}{\sqrt{\left(a_p^2+\xi^2\right)\left(a_{\mathbf{k}}^2+\lambda^2\right)}},
\end{aligned} \quad (29)$$

with $a_p = -E/2+\epsilon_p$. In these variables BSE reads

$$\mathcal{X}_{\xi,p} = -\int \mathcal{V}_{p,k}(\xi-\lambda)\mathcal{X}_{\lambda,k} D_\lambda D_k. \quad (30)$$

To efficiently solve Eq. (30), we fix the energy E, e.g., guided by the LSE solution, and introduce a new variable R, such that

$$R\mathcal{X}_{\xi,p} = -\int \mathcal{V}_{p,k}(\xi-\lambda)\mathcal{X}_{\lambda,k} D_\lambda D_k, \quad (31)$$

which is a standard eigenvalue problem for R and can be solved numerically. We need only the maximal eigenvalue $R_{\max}$. The energy E that satisfies $R_{\max}(E) = 1$ gives the solution of the original BSE (22). We will denote the corresponding binding energy $\epsilon_b^{(\infty)} = 2\Delta - E$, since in the language of perturbation theory, this corresponds to the infinite summation of corrections in all orders.

C. Comparison of techniques

The intravalley s-wave exciton binding energies vs Δ are plotted in Fig. 5, with the various results reflecting the solution using the LSE $\epsilon_b^{(0)}$, the perturbative correction $\epsilon_b^{(1)} = \epsilon_b^{(0)} - \delta E$, and the numerically exact BSE $\epsilon_b^{(\infty)}$. Moreover, results for various combinations of gate distance d and dielectric ϵ_r are shown. To discuss the key findings, we distinguish the three regimes:

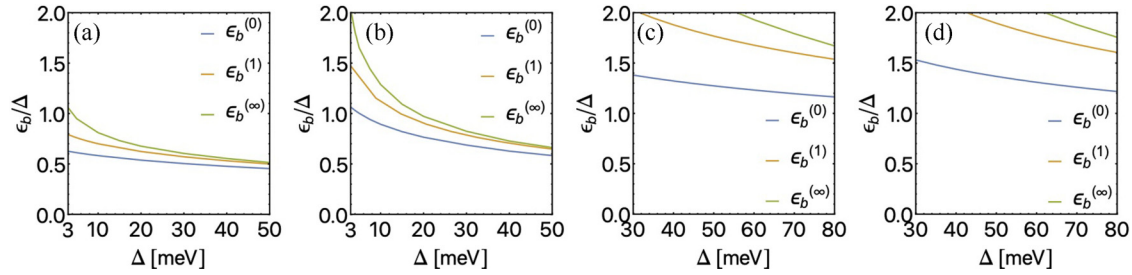

FIG. 5. Binding energies (divided by Δ) vs Δ. The LSE, perturbative BSE, and exact BSE are denoted $\epsilon_b^{(0)}$, $\epsilon_b^{(1)}$, and $\epsilon_b^{(\infty)}$. Screening parameters: (a) $d = 20$ nm, $\epsilon_r = 3.9$; (b) $d = 100$ nm, $\epsilon_r = 3.9$; (c) $d = 100$ nm, $\epsilon_r = 1$; (d) $d = 1000$ nm, $\epsilon_r = 1$.

(1) Weak binding $\epsilon_b \ll 2\Delta$, e.g., for the case of $\epsilon_r = 3.9$ and $d = 20$ nm [Fig. 5(a)]. We find $\epsilon_b^{(0)} \approx \epsilon_b^{(1)} \approx \epsilon_b^{(\infty)}$, demonstrating that retardation corrections are negligible.

(2) Intermediate binding $\epsilon_b \sim \Delta$, e.g., for the case of $\epsilon_r = 3.9$, $d = 100$ nm [Fig. 5(b)] and for $\Delta \gtrsim 15$ meV. We find $\epsilon_b^{(0)} < \epsilon_b^{(1)} \approx \epsilon_b^{(\infty)}$, demonstrating that retardation corrections are non-negligible and that the first-order perturbation theory quantitatively captures the influence of the retardation.

(3) Strong binding $\epsilon_b \sim 2\Delta$, e.g., for the cases of $\epsilon_r = 1$, $d = 100$ and 1000 nm [Figs. 5(c) and 5(d)]. We see that the LSE is completely insufficient and that the first-order perturbation theory captures approximately 50%, or more, of the retardation correction.

We conclude that the perturbative approach accounts for the majority of the retardation correction, except for deep in the strong-binding regime, $\epsilon_b \to 2\Delta$.

D. Comparison with existing data

Assuming hBN encapsulation, giving dielectric enhancement $\bar{\epsilon} = 3.9$, and taking the metallic gates to be at a distance $d = 20$ nm, we can directly compare with the experimental measurements of Ref. [9]. We compare the exciton energies, $E = 2\Delta - \epsilon_b$ vs displacement field (D) obtained using LSE and BSE, to those measured experimentally. We use the linear conversion between parameter Δ and field D, provided in Eq. (3). The experiment measures both s- and p-wave intravalley excitons, so it is necessary to use form factor (15). The comparison is provided in Fig. 6, from which we see that both LSE and BSE techniques provide quantitative agreement,

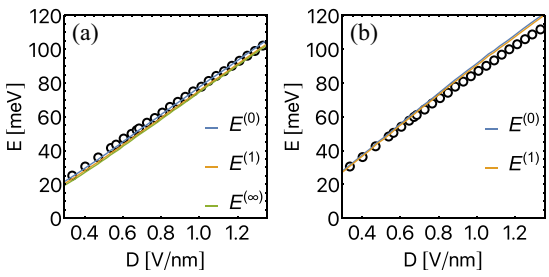

FIG. 6. Exciton energies vs displacement field D for (a) s- and (b) p-wave states. Experiment [9] for s and p wave are shown by black circles, while the LSE, perturbative BSE, and exact BSE are shown as the $E^{(0)}$, $E^{(1)}$, and $E^{(\infty)}$ curves. In (a) the $E^{(1)}$ and $E^{(\infty)}$ curves are essentially indistinguishable, while in (b) just $E^{(0)}$ and $E^{(1)}$ are shown.

without fitting parameters. We reiterate that this is the case of weak binding; i.e., retardation corrections are small, and therefore the LSE and BSE eigenenergies show very little difference.

V. DISCUSSION

In this work we considered the influence of dynamical screening (retardation) on exciton binding—i.e., the binding energy ϵ_b. We considered the particular example of biased bilayer graphene, whereby the bias field induces single-particle band gap Δ_b. However, the techniques developed are applicable to many other 2D materials.

We found that for $\epsilon_b \ll \Delta_b$, retardation effects can be safely neglected, and the properties of the exciton bound states are very well captured by the Lippmann-Schwinger equation (a two-body Hamiltonian approach). However, in the strong-binding regime $\epsilon_b \sim \Delta_b$ retardation is significant and therefore a Hamiltonian approach is insufficient. Instead the correct formalism is the Bethe-Salpeter equation (BSE). The BSE is costly to numerically implement and thereby presents a bottleneck. To handle this situation, we develop a simple perturbative expansion of the BSE, which allows us to systematically compute corrections to the binding energy relative to the static case. The perturbative approach provides a numerically simple means to compute and assess the importance of retardation. By comparison with the exact solution, we found that for strong binding the first-order perturbative approach captured approximately 50% of the retardation correction; for intermediate binding the first-order perturbative approach captured most of the effect; and for weak binding the retardation is negligible. The first-order correction therefore provides an efficient means to understand the role of retardation. The quantitative accuracy of our methods are checked via directly fitting to available experimental data in this regime [9]; crucially, we take parameters established elsewhere, and as such do not use any fitting parameters.

Screening significantly influences the excitonic binding energy. We argue that to probe the critical regime of $\epsilon_b \sim \Delta_b$, one must reduce screening from the environment, both dielectric and gate. We propose suspended bilayer graphene (i.e., dielectric $\epsilon_r = 1$) with placement of metallic gates $d > 20$ nm above the plane. Counterintuitively, even though the characteristic radius of excitons considered here is $r \lesssim 10$ nm, the difference in binding energies for metallic gates at $d = 20$ and 100 nm is significant, Fig. 4. We exam-

ine exciton condensation, i.e. $\epsilon_b \geqslant \Delta_b$, in a companion work [24].

A closely related system, rhombohedral (ABC-stacked) trilayer graphene, also hosts an electric field tunable single particle band gap [25]. A recent theoretical study [26] suggests that excitons in this system can reach strong binding; that work does not include dynamical screening. We speculate that the retardation correction to binding energy, detailed in the present work, will be significant in biased ABC trilayer graphene.

ACKNOWLEDGMENTS

This work benefited from a range of fruitful discussions with Alex Hamilton, Oleh Klochan, Dmitry Efimkin, and Mike Zhitomirsky. We acknowledge funding support from the Australian Research Council Centre of Excellence in Future Low-Energy Electronics Technology (FLEET) (Grant No. CE170100039).

APPENDIX

In this Appendix we compute the *exact* polarization operator, i.e., using the Hamiltonian valid to energy scales $\sim$10 eV and for all momenta, and compare with the polarization operator computed in the main text, which relied on the quadratic band Hamiltonian (1).

The four-orbital single-particle Hamiltonian for bilayer graphene reads [20]

$$H = \begin{pmatrix} -\Delta & -\gamma_0 f_{\boldsymbol{p}} & \gamma_4 f_{\boldsymbol{p}} & -\gamma_3 f_{\boldsymbol{p}}^* \\ -\gamma_0 f_{\boldsymbol{p}}^* & -\Delta + \Delta' & \gamma_1 & \gamma_4 f_{\boldsymbol{p}} \\ \gamma_4 f_{\boldsymbol{p}}^* & \gamma_1 & \Delta + \Delta' & -\gamma_0 f_{\boldsymbol{p}} \\ -\gamma_3 f_{\boldsymbol{p}} & \gamma_4 f_{\boldsymbol{p}}^* & -\gamma_0 f_{\boldsymbol{p}}^* & \Delta \end{pmatrix}, \quad \text{(A1)}$$

where $f_{\boldsymbol{p}} = e^{ip_y a/\sqrt{3}} + e^{-ip_y a/(2\sqrt{3})}\cos(p_x a/2)$. Values of the parameters a, γ_0, γ_1, γ_3, γ_4, Δ', are presented in Eq. (5) of the main text. The four branches of the dispersion near the K point are shown in Fig. 7. By black lines we also show the dispersion in quadratic approximation considered in the main text. The quadratic approximation fails at energy higher than 200–300 meV, which corresponds to $a(p-K) \approx 0.15$. The size of the bound state in the momentum space is about $\sim p_0/2$, where $p_0 = \sqrt{2m\Delta}$; see Figs. 4(a)(ii) and 4(b)(ii) of the main text. This is a very small momentum; e.g., at $\Delta = 15$ meV the characteristic momentum is just $ap_0 = 0.028$. Hence, the

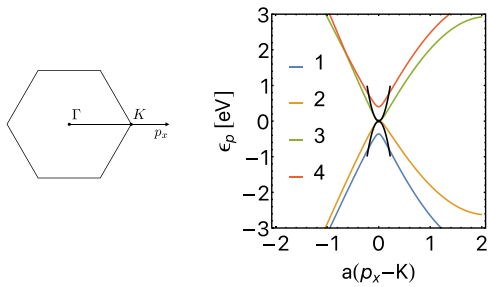

FIG. 7. Left: Brillouin zone. Right: Four branches of the dispersion near the K point, $K = \frac{4\pi}{3a}$, at zero band gap $\Delta = 0$. Black lines show the dispersion in quadratic approximation of Eq. (1).

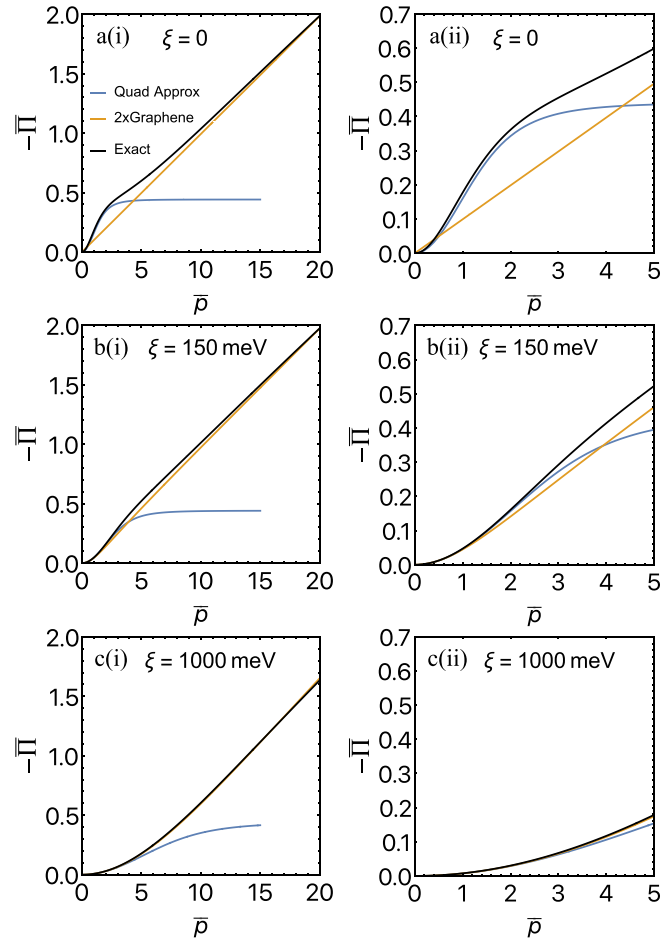

FIG. 8. The polarization operator versus dimensionless momentum for band gap $2\Delta = 30$ meV and frequency (a) $\xi = 0$, (b) $\xi = 150$ meV, and (c) $\xi = 1000$ meV. The black line shows the exact value (A2), the blue line corresponds to the quadratic approximation given by Eq. (8) in the main text, and the orange line corresponds to two decoupled graphene planes given by Eq. (A3). The left panels, (i), correspond to a broader momentum range and the right panels, (ii), are a zoom-in of the momentum range relevant to the exciton problem.

quadratic approximation in BSE is justified. However, the energy integration related to the retardation in BSE, i.e., Eq. (25), goes up to 2–3 eV, and we therefore need to know the polarization operator $\Pi(q, i\xi)$ in the range $0 < q < 5p_0$, $0 < \xi < 3$ eV. It seems that the quadratic approximation at such high energy should not be valid. We therefore calculate, in this Appendix, the polarization operator directly from the four-band Hamiltonian (A1).

The polarization operator reads

$$\Pi(\boldsymbol{p}, i\xi) = 2\sum_{\substack{i=1,2 \\ j=3,4}} \int_{q\in BZ} D_{\boldsymbol{q}} \frac{2\left(\epsilon_{\boldsymbol{q}}^{(i)} - \epsilon_{\boldsymbol{q}+\boldsymbol{p}}^{(j)}\right) F_{\boldsymbol{q},\boldsymbol{q}+\boldsymbol{p}}^{ij}}{\left(\epsilon_{\boldsymbol{q}}^{(i)} - \epsilon_{\boldsymbol{q}+\boldsymbol{p}}^{(j)}\right)^2 + \xi^2}, \quad \text{(A2)}$$

where the band indices i, j are as defined in Fig. 7; $\epsilon_{\boldsymbol{q}}^{(i)}$, $\psi_{\boldsymbol{q}}^{(i)}$ are the eigenenergies and eigenfunctions of Eq. (A1); $F_{\boldsymbol{q},\boldsymbol{q}+\boldsymbol{p}}^{ij} = |\langle\psi_{\boldsymbol{q}+\boldsymbol{p}}^{(j)}|\psi_{\boldsymbol{q}}^{(i)}\rangle|^2$ is the vertex factor accounting for the

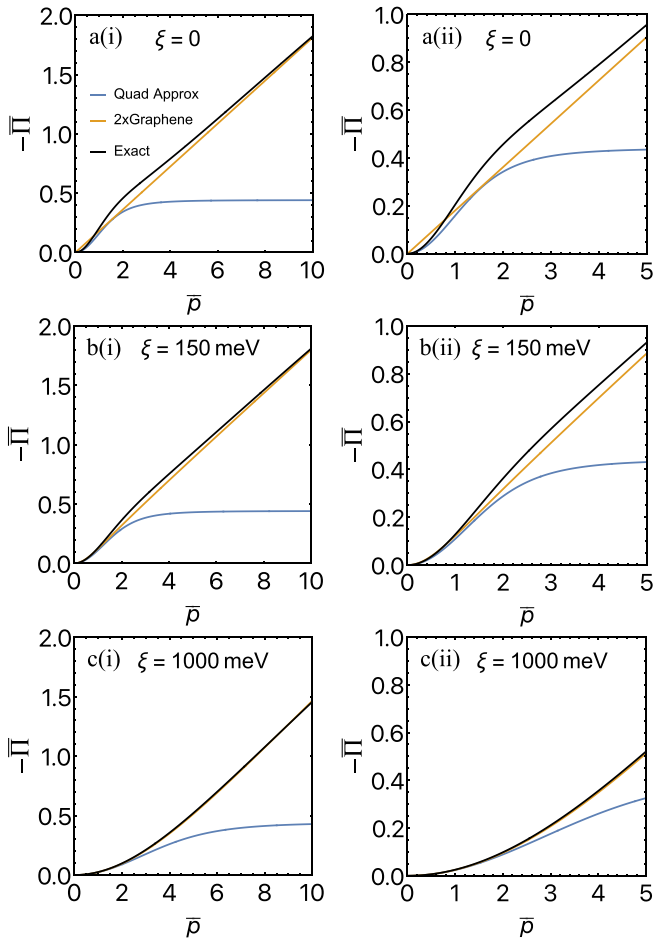

FIG. 9. The polarisation operator versus dimensionless momentum for bandgap $2\Delta = 100$ meV and frequency: (a) $\xi = 0$; (b) $\xi = 150$ meV; and (c) $\xi = 1000$ meV. The black line shows the exact value (A2), the blue line corresponds to the quadratic approximation given by Eq. (8) in the main text, the orange line corresponds to two decoupled graphene planes given by Eq. (A3). The left panels, (i), corresponds to a broader momentum range and the right panels, (ii), is zoom in of the momentum range relevant to the exciton problem.

overlap of the conduction and valence single-particle wave functions; and the prefactor 2 in (A2) comes from spin degeneracy. Similarly to the main text we define the dimensionless polarization operator as $\overline{\Pi} = \Pi/(2m)$.

Figures 8 and 9 compares the $\bar{\Pi}(\bar{q}, \xi)$ at $\Delta = 15$ and 50 meV respectively using the quadratic approximation [Eq. (8)]; the exact esxpression [Eq. (A2)]; and the polarization operator of two decoupled graphene planes, given by

$$\overline{\Pi}_D(\boldsymbol{p}, i\xi) = -\frac{p^2}{4m\sqrt{v_F^2 p^2 + \xi^2}},$$
$$v_F = \frac{\sqrt{3}a\gamma_0}{2\hbar}. \tag{A3}$$

Considering the polarization operator at $\xi = 0$ [Figs. 8(a) and 9(a)]: As one should expect at small $\bar{p}$ the exact calculation is consistent with that of the quadratic approximation, while at large $\bar{p}$ it is inconsistent with the quadratic approximation but instead becomes consistent with two decoupled graphene planes. Considering now the polarization operator at large frequencies, $\xi = 150$ meV and $\xi = 1000$ meV [Figs. 8 and 9 panels b and c]: Again, at large $\bar{p}$ the exact calculation is consistent with two decoupled graphene planes; this is expected. Quite surprisingly, at small $\bar{p}$ the exact calculation is consistent with the parabolic approximation up to $\xi \approx 3$ eV. Less surprising is consistency with the two decoupled graphene planes at small $\bar{p}$.

The message of this Appendix is the following: For the exciton problem with the band gap less that 100 meV ($\Delta < 50$ meV) one needs the polarization operator in the range $0 < q < 5\sqrt{2m\Delta}$, $0 < \xi < 3$ eV. Within this range the exact calculation of the polarization operator and a similar calculation within the quadratic approximation give good agreement.

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

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

[3] B. Halperin and T. Rice, The excitonic state at the semiconductor-semimetal transition, *Solid State Physics*, Vol. 21 (Academic Press, 1968), pp. 115–192.

[4] Yu. E. Lozovik and V. I. Yudson, Feasibility of superfluidity of paired spatially separated electrons and holes: A new superconductivity mechanism, Pis'ma Zh. Eksp. Teor. Fiz. **22**, 556 (1975) [JETP Lett. **22**, 274 (1975)].

[5] M. B. Pogrebinskii, Mutual drag of carriers in a semiconductor-insulator-semiconductor system, Fiz. Tekh. Poluprovodn. **11**, 637 (1977) [Sov. Phys. Semicond. **11**, 372 (1977)].

[6] J. M. Blatt, K. Böer, and W. Brandt, Bose-Einstein condensation of excitons, Phys. Rev. **126**, 1691 (1962).

[7] M. Kellogg, J. Eisenstein, L. N. Pfeiffer, and K. W. West, Vanishing Hall Resistance at High Magnetic Field in a Double-Layer Two-Dimensional Electron System, Phys. Rev. Lett. **93**, 036801 (2004).

[8] J.-J. Su and A. MacDonald, How to make a bilayer exciton condensate flow, Nat. Phys. **4**, 799 (2008).

[9] L. Ju, L. Wang, T. Cao, T. Taniguchi, K. Watanabe, S. G. Louie, F. Rana, J. Park, J. Hone, F. Wang, and P. L. McEuen, Tunable excitons in bilayer graphene, Science **358**, 907 (2017).

[10] C.-H. Park and S. G. Louie, Tunable excitons in biased bilayer graphene, Nano Lett. **10**, 426 (2010).

[11] P. Li and I. Appelbaum, Excitons without effective mass: Biased bilayer graphene, Phys. Rev. B **99**, 035429 (2019).

[12] M. O. Sauer and T. G. Pedersen, Exciton absorption, band structure, and optical emission in biased bilayer graphene, Phys. Rev. B **105**, 115416 (2022).

[13] J. C. G. Henriques, I. Epstein, and N. M. R. Peres, Absorption and optical selection rules of tunable excitons in biased bilayer graphene, Phys. Rev. B **105**, 045411 (2022).

[14] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, Vol. 4, 2nd ed. (Butterworth Heinemann, Oxford, 1982), p. 389.

[15] M. M. Glazov and A. Chernikov, Breakdown of the static approximation for free carrier screening of excitons in monolayer semiconductors, Phys. Status Solidi B **255**, 1800216 (2018).

[16] Y. Zhang, T.-T. Tang, C. Girit, Z. Hao, M. C. Martin, A. Zettl, M. F. Crommie, Y. R. Shen, and F. Wang, Direct observation of a widely tunable bandgap in bilayer graphene, Nature (London) **459**, 820 (2009).

[17] R. T. Weitz, M. T. Allen, B. E. Feldman, J. Martin, and A. Yacoby, Broken-symmetry states in doubly gated suspended bilayer graphene, Science **330**, 812 (2010).

[18] F. Freitag, J. Trbovic, M. Weiss, and C. Schönenberger, Spontaneously Gapped Ground State in Suspended Bilayer Graphene, Phys. Rev. Lett. **108**, 076602 (2012).

[19] E. McCann and V. I. Fal'ko, Landau-Level Degeneracy and Quantum Hall Effect in a Graphite Bilayer, Phys. Rev. Lett. **96**, 086805 (2006).

[20] E. McCann and M. Koshino, The electronic properties of bilayer graphene, Rep. Prog. Phys. **76**, 056503 (2013).

[21] A. B. Kuzmenko, I. Crassee, D. van der Marel, P. Blake, and K. S. Novoselov, Determination of the gate-tunable band gap and tight-binding parameters in bilayer graphene using infrared spectroscopy, Phys. Rev. B **80**, 165406 (2009).

[22] G. Mahan, *Many-Particle Physics* (Plenum Press, New York, 1990).

[23] E. H. Hwang and S. Das Sarma, Screening, Kohn Anomaly, Friedel Oscillation, and RKKY Interaction in Bilayer Graphene, Phys. Rev. Lett. **101**, 156802 (2008).

[24] H. D. Scammell and O. P. Sushkov, Exciton condensation in biased bilayer graphene, arXiv:2301.07864.

[25] C. H. Lui, Z. Li, K. F. Mak, E. Cappelluti, and T. F. Heinz, Observation of an electrically tunable band gap in trilayer graphene, Nat. Phys. **7**, 944 (2011).

[26] M. F. C. M. Quintela and N. M. R. Peres, Tunable excitons in rhombohedral trilayer graphene, Phys. Rev. B **105**, 205417 (2022).