Dynamical screening and excitonic bound states in biased bilayer graphene

Harley D. Scannell$^1$ and Oleg P. Sushkov$^1$

$^1$School of Physics, the University of New South Wales, Sydney, NSW, 2052, Australia

(Dated: July 8, 2022)

Excitonic bound states are characterised by a binding energy $\epsilon_b$ and a single-particle band gap $\Delta_b$. 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 (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.

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.

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 wavefunction description, whereas for the BSE a wavefunction is not possible, and instead the correct object is 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 $\epsilon_b$ and single-particle band gap $\Delta_b$ 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 bottle-neck. 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 to the case of BBG, our analysis demonstrates that the strong binding regime in BBG is possible and we specify conditions in which it can can be achieved. The conditions concern the bias, the configuration of gates, and the substrate material.

In BBG, the single-particle band gap $\Delta_b$ 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, $\epsilon_b$. There are three sources of screening: (i) dielectric due to a material between BBG and the gates, (ii) metallic gates, (iii) in-plane 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. 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 5nm. Therefore, the continuum approximation is sufficient for analysis of the problem. The low-energy single particle Hamiltonian of BBG is\(^{[19]}\)

\[
H_0 = \begin{pmatrix}
\Delta & -\frac{p^2}{2m} \\
-\frac{p^2}{2m} & -\Delta
\end{pmatrix} .
\]  

(1)

It 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 \( \Pi_{\pm} = \tau(p_x - eA_x) \pm i(p_y - eA_y) \equiv \tau \Pi_x \pm i \Pi_y \), \( \mathbf{p} \) is the in-plane momentum, \( \mathbf{A} \) the magnetic vector potential, \( \tau = \pm 1 \) 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,

\[
\delta H = \begin{pmatrix}
\Pi_{\pm} & -\frac{\Pi_0 \Pi_{\pm}}{2M} \left(1 - \frac{\Delta_b}{\Delta}\right) \\
\Pi_0 \Pi_{\pm} & \Pi_{\pm} \frac{\Pi_0}{2M} \left(1 + \frac{\Delta_b}{\Delta}\right)
\end{pmatrix} .
\]

(2)

This captures three types of perturbations, \( \frac{\Pi_0 \Pi_{\pm}}{2M} \), \( \frac{\Pi^2}{2M} \), and \( \frac{\Pi^2}{2M} \Delta_b \). 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, with \( eD/(2\Delta) \approx 10.4 \text{nm}^{-1} \) taken from experiment\(^{[16]}\). Meanwhile, all other parameters of the single particle Hamiltonian can be related to the standard Slonczewski-Weiss-McClure (SWM) parameters of BG\(^{[20]}\)

\[
m = \frac{2\gamma_1}{3e^2\gamma_0} \approx 0.032m_e ,
\]

\[
M = \frac{m}{2\gamma_1/m + \Delta/\gamma_1} \approx \frac{m}{0.146} \approx 0.22m_e ,
\]

\[
\Delta_0 = \frac{\gamma_1 m}{2M} \approx 28\text{meV} ,
\]

\[
\Pi_0 = \frac{2\gamma_1 \gamma_3}{\sqrt{3}\gamma_0^2 a} \approx 0.068\text{nm}^{-1} .
\]

(3)

To obtain the numerical values, we have taken the SWM parameters established by the experiment and analysis of Ref.\(^{[21]}\),

\[
a = 2.46\text{Å} ,
\]

\[
\gamma_0 = 3.16\text{eV} ,
\]

\[
\gamma_1 = 0.381\text{eV} ,
\]

\[
\gamma_3 = 0.38\text{eV} ,
\]

\[
\gamma_4 = 0.14\text{eV} ,
\]

\[
\Delta' = 0.022\text{eV} .
\]

(4)

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, \( \epsilon_p^{(-)} \), and conduction, \( \epsilon_p^{(+)} \), bands, as well as the corresponding wave functions \( \psi_{p,\tau}^{(\pm)} \) (spin index is idle, and so is not included). We will perform calculations for Hamiltonian (1) both 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. Excluding the small corrections (2), it is convenient to perform the analysis in rescaled units, whereby energy is measured in units of \( \Delta \) and momentum in units of \( p_0 = \sqrt{2m\Delta} \),

\[
\tau = \epsilon/\Delta ,
\]

\[
\bar{p} = p/p_0 = p/\sqrt{2m\Delta} .
\]

(5)

In these units, and in zero magnetic field, the Hamiltonian (1) reads

\[
H_0 = \begin{pmatrix}
1 & -\bar{p}_p^2 \\
-\bar{p}_p^2 & -1
\end{pmatrix} .
\]

(6)

The eigenenergies are \( \epsilon_q^{(\pm)} = \pm \sqrt{1 + \bar{p}_p^4} \).

B. Screened Coulomb interaction

The many-body RPA screening of Coulomb interaction in BG is significant. We recall the essential details. The polarisation operator at an imaginary frequency \( \xi \), as given by the diagram in Fig. 1, reads
\( \Pi(p, i\xi) = 2 \times 2 \times \int Dq \frac{2}{\epsilon_q - \epsilon_{q+p}} (\epsilon_q^{(-)} - \epsilon_{q+p}^{(+)}) F_{q,q+p}^{+-} (\epsilon_q^{(-)} - \epsilon_{q+p}^{(+)})^2 + \xi^2 \)  

(7)

Throughout the paper we use the notation

\[ Dq = \frac{d^2q}{(2\pi)^2}. \]

(8)

The prefactor \( 2 \times 2 \) in (7) comes from spin and valley degeneracy, the vertex factor is the overlap of the conduction and valence single particle wave functions, \( F_{q,q+p}^{+-} = |\langle \psi_{q+p}^{(-)} | \psi_q^{(-)} \rangle|^2 \). 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 \( q = 0 \). The polarisation operator has dimension of mass, rewriting Eq. (7) in dimensionless units (5) gives,

\[ \Pi(p, i\xi) = 2m\Pi(p, i\xi) \]

(9)

where \( \Pi \) is given by the same Eq. (7), but all the variables replaced to those with bars. At \( \bar{q} \gg 1 \) the polarisation operator is \( \Pi = \frac{\bar{q}^2}{4\pi} \approx 0.441 \), which reduces to the \( \Delta = 0 \) case considered in Ref. [22]. In the general case of arbitrary \( \Delta \), the operator is straightforwardly computed numerically. Plots of \( -\Pi \) versus \( \bar{p} \) for different values of \( \xi \) are presented in Fig. 2a. The polarisation operator is zero at \( \bar{p} = 0 \) and it is approaching \( (\ln 4)/\pi \) at large \( \bar{p} \).

The frequency dependence of the polarisation operator becomes significant when the frequency is comparable and larger than the band gap, \( \xi \gg \Delta_b = 2\Delta \).

There are two points to note on the behaviour of the polarisation operator: (i) The polarisation screening becomes significant at \( \bar{p} > 1 \). The scale \( \bar{p} \approx 1 \) will determine the size of the exciton, \( 1/r \approx p_0 = \sqrt{2m\Delta} \); and (ii) the (imaginary) frequency monotonically reduces screening, i.e. \( \Pi(\bar{p}, i\xi) < \Pi(\bar{p}, 0) \) 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.

Using notation \( \epsilon^2 = \epsilon_0^2/\epsilon_r \), where \( \epsilon_0 \) is the bare charge and \( \epsilon_r \) is the dielectric constant, the screened Coulomb interaction is

\[ V_p(\xi) = -\frac{2\pi\epsilon^2}{\bar{p} - 2\pi\epsilon^2\Pi(p, i\xi)} = \frac{1}{2m}V_p(\xi) \]

\[ \nabla_p(\xi) = -\frac{4\pi}{s\bar{p} - 4\pi\Pi(p, i\xi)} \]

\[ s = \sqrt{\frac{\epsilon_r^2\Delta}{me^2/2}} \]

(10)

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

\[ \chi_{scr} = \frac{-\nabla_p(\xi)}{(4\pi/s\bar{p})} = \frac{s\bar{p}}{s\bar{p} - 4\pi\Pi(p, i\xi)} \]

(11)

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

III. LIPPMANN-SCHWINGER EQUATION (LSE)

We define quasi-momentum, \( p \), with respect to the valley minimum (K-point) and only consider a bound states with total quasi-momentum zero, i.e. electron in valley \( \tau \), with momentum \( p + \tau K \) pairing with a hole in valley \( \tau' \), with total momentum \( -p - \tau' K \). This means that the total momentum, i.e. as defined with respect to the Γ-point, is zero if electron-hole pair in the same valley (intravalley exciton) and it is equal to \( \pm 2K \) 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 \( \epsilon^{(+)}_p = -\epsilon^{(-)}_p \equiv \epsilon_p = \sqrt{\Delta^2 + \frac{p^2}{4m^*}} \). For some of our numerics, we account for the corrections (2). The bound state equation reads

\[
(E_0 - 2\epsilon_p)\Psi_p = \int V_{p-k}(0)Z_{p,k}\Psi_k D_k
\]

\[
E_0 = 2\Delta - \epsilon^{(0)}_b
\]

Here \( \Psi_p \) is the exciton wavefunction, and \( \epsilon^{(0)}_b \) is the exciton binding energy (the subscript/superscript ‘0’ is used to distinguish from the case with account of retardation, to be discussed in Section IV). The vertex form factors are given by

\[
Z_{p,k}^{\tau',\tau} = \langle \psi^{(\tau')}_{p,k} | \psi^{(\tau)}_{k} \rangle \langle \psi^{(\tau)}_{k} | \psi^{(\tau')}_{p,k} \rangle
\]

The form factors do not distinguish spin, yet they weakly distinguish between intra- and inter-valley excitons. Note that \( \tau \) corresponds to the valley where the electron is located and \( \tau' \) corresponds to the valley where the hole is located. Explicitly, the expression is

\[
Z_{p,k}^{\tau',\tau} = \frac{(1 + \Omega_p \Omega_k e^{2i\tau'\theta}) (1 + \Omega_p \Omega_k e^{2i\tau\theta})}{1 + \Omega^2_p} \frac{1 + \Omega^2_k}{1 + \Omega^2_k}.
\]

Here \( \theta = \theta_k - \theta_p \) and \( \Omega_p = 2m(\epsilon_p - \Delta)/p^2 \). Hence, for the intervalley exciton, \( \tau' = -\tau \), the Z-factor is real

\[
Z_{p,k}^{\tau',\tau} = \frac{1 + \Omega^2_p \Omega^2_k + 2 \Omega_p \Omega_k \cos(2\theta)}{(1 + \Omega^2_p)(1 + \Omega^2_k)}.
\]

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

\[
Z_{p,k}^{\tau',\tau} = \frac{1 + \Omega^2_p \Omega^2_k e^{4i\tau\theta} + 2 \Omega_p \Omega_k e^{2i\tau\theta}}{(1 + \Omega^2_p)(1 + \Omega^2_k)}.
\]

In dimensionless units (5), Eq.(12) reads

\[
(E_0 - 2\epsilon_p)\Psi_p = \int \nabla^\tau \Psi_p D_{\xi}\nabla^\tau \Psi_p D_{\xi}
\]

The interaction \( \nabla^\tau \) 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 accounted 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 distance from the BG plane. Accounting for gate-screening, via the method of images, the interaction in Eq.(10) is replaced by,

\[
\nabla^\tau \Psi_p = -\frac{4\pi}{s \pi / \tau_p - 4\pi \Pi(p, r \xi)}
\]

\[
\forall \tau_p = \tanh(pd) = \tanh(p_0pd).
\]

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

\[
\Psi_p = \sum_\ell \frac{1}{\sqrt{p}} \psi^\ell_p e^{i\ell\tau},
\]

where \( \psi^\ell_p \) depends only on the absolute value of momentum. In a channel with a given orbital momentum \( l \), Eq.(16) is reduced to

\[
(E_0 - 2\epsilon_p)\psi^\ell_p = \int \nabla_{\xi} \Psi_p D_{\xi}
\]

\[
V^\ell_{p,\xi} = \frac{\sqrt{p}}{(2\pi)^2} \int \nabla_{\xi} e^{-i\ell\theta_p} \nabla_{\xi} \Psi_p D_{\xi}
\]

where \( \psi^\ell_p \) is independent of \( \theta_p \), since the integral in the second line of (19) is a function of \( \theta_p - \theta_p \).

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 (\( l = 0 \)) ground state of intervalley exciton versus \( \Delta \) is plotted in Fig. 4a. 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 \( \epsilon_r \). 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 \( V_{p=0} = 2\pi e^2d \). 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.4b, and (in dimensionless momenta) are insensitive. In the original units they are of course different: \( p = 1 \) corresponds to \( p = 0.130nm^{-1} \) for \( \Delta = 20meV \) and to \( p = 0.206nm^{-1} \) for \( \Delta = 50meV \). We attribute this universal behaviour to the shape of the polarisation operator, which is small \( \propto \xi \) up to a scale \( \xi = 1 \).

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

\[
\Delta \epsilon^\ell_p = \epsilon^{(0)}_b (\text{inter}) - \epsilon^{(0)}_b (\text{intra}),
\]

is plotted versus \( \Delta \) in Fig.4c for \( \epsilon_r = 1 \) and \( d = 1000 \) nm. The difference in binding energies is less than 1%. It is even smaller for higher \( \epsilon_r \) and lower \( d \).
FIG. 4. Exciton s-wave ground state, l=0. (a) Binding energies of intervalley exciton in the LSE approximation versus ∆. The plots are presented for three values of the dielectric constant, ε_r = 1 (blue curves), ε_r = 2.45 (orange curves), ε_r = 3.9 (green curves), and three values of the distance to gates d = 1000nm (solid), d = 100nm (dashed), d = 20nm (dotted). (b) Wave functions versus dimensionless momentum p for different sets of parameters. The blue line corresponds to ∆ = 20meV, ε_r = 1, d = 1000nm; the orange line corresponds to ∆ = 50meV, ε_r = 1, d = 1000nm; the green line corresponds to ∆ = 50meV, ε_r = 3.9, d = 20nm; the red line corresponds to ∆ = 50meV, ε_r = 3.9, d = 20nm. (c) The energy splitting, Eq.(20), between the intervalley and intravalley excitons versus ∆ for ε_r = 1 and d = 1000nm.

FIG. 5. Exciton lowest p-wave state. (a) Binding energies of intervalley exciton in the LSE versus ∆. The plots are presented for three values of the dielectric constant, ε_r = 1 (blue curves), ε_r = 2.45 (orange curves), ε_r = 3.9 (green curves), and three values of the distance to gates d = 1000nm (solid), d = 100nm (dashed), d = 20nm (dashed-dotted). (b) Wave functions versus dimensionless momentum p for different sets of parameters. The blue line corresponds to ∆ = 20meV, ε_r = 1, d = 1000nm; the orange line corresponds to ∆ = 50meV, ε_r = 1, d = 1000nm; the green line corresponds to ∆ = 20meV, ε_r = 3.9, d = 20nm; the red line corresponds to ∆ = 50meV, ε_r = 3.9, d = 20nm. (c) The l-dependent, Eq.(21) energy splitting between the intervalley and intravalley excitons versus ∆ for ε_r = 1 and d = 1000nm.

For the p-wave exciton we again start from the intervalley case. In this case the energies of l = ±1 states are degenerate. The binding energy of lowest p-wave state of the intervalley exciton is plotted in Fig.5a versus ∆ for the same values of ε_r and d as that for the s-wave in Fig.4. The wave functions corresponding to different binding energies are plotted in Fig.5b.

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

$$\epsilon_b^{(0)}(\text{intra}) = \epsilon_b^{(0)}(\text{inter}) + \tau I \Delta \epsilon_b^P$$  \hspace{1cm} (21)

where τ indicates the valley. The splitting Δϵ_b^P is plotted in Fig.5c versus ∆ for ε_r = 1 and d = 1000nm. The splitting is again less than 1%. It is even smaller for higher ε_r and lower d.

IV. RETARDATION AND BETHE-SALPETER EQUATION (BSE)

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 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 behaviour. 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, ϵ_b ≪ 2Δ, the effect of screening reduction is small. However, at strong binding, ϵ_b ∼ 2Δ, 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 $p$ and the relative frequency $\xi$. In our analysis, the total momentum of the electron and hole is encoded in the valley indices, and is either zero (intravalley pairing) or $2K$ (intervalley pairing). The imaginary frequency $\xi$ is Fourier conjugated to the retardation time (plus a Wick rotation). BSE for $\chi_{\xi,p}$ reads [14],

$$\chi_{\xi,p} = -\frac{1}{(E/2 - \omega_p)^2 + \xi^2} \int V_{p,k}(\xi - \lambda)\chi_{\lambda,k} Z_{p,k} D_\lambda D_k,$$

$$D_\lambda = \frac{d\lambda}{2\pi}, \quad D_k = \frac{d^2k}{(2\pi)^2}. \quad (22)$$

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

$$\chi_{\xi,p}^{(0)} = \frac{2a_p}{a_p^2 + \xi^2} \Psi_p,$$

$$a_p = -E_0/2 + \omega_p. \quad (23)$$

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

Eq.(22) is not a linear eigenvalue problem, and a direct numerical solution of Eq.(22) is an involved calculation. Here we develop a perturbation theory method that is sufficient for our purposes. This is the regime when the retardation correction while being important is still relatively small. Let us first replace the interaction in (22) by a frequency independent interaction, $V_{p,k}(\xi \rightarrow V_{p,k}(0)$. It can be the interaction at zero frequency, $V_{p,k}^{(0)}(0) = V_{p,k}(\xi = 0)$, or interaction at some typical frequency $V_{p,k}^{(0)} = V_{p,k}(\xi = \xi_{\text{typical}})$, or something else. We will discuss specific possibilities later. BSE with $V_{p,k}^{(0)}$ is reduced to LSE which is a linear eigenvalue problem and can solved numerically with ease. The solution is given by Eq.(23) where $\Psi_p$ and $E_0$ is the eigenfunction and the eigenenergy of LSE. Next, let us consider

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

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

$$\delta E = \int (\chi_{\xi,p}^{(0)})^* \delta V_{p,k}(\xi - \lambda)\chi_{\lambda,k} Z_{p,k} D_\lambda D_p D_\lambda D_k \quad (25)$$

To derive (25) let us represent the Green’s function as

$$\chi_{\xi,p} = \chi_{\xi,p}^{(0)} + \delta \chi_{\xi,p}. \quad \text{Hence, BSE (22) can be rewritten as}$$

$$\chi_{\xi,p}^{(0)} + \delta \chi_{\xi,p} = -\frac{1}{(E/2 - \omega_p)^2 + \xi^2} \int [\delta V_{p,k}(\xi - \lambda)\chi_{\lambda,k}^{(0)} Z_{p,k} D_\lambda D_k,$$

$$\approx -\frac{1}{(E/2 - \omega_p)^2 + \xi^2} \int [\delta V_{p,k}(\xi - \lambda)\chi_{\lambda,k}^{(0)} + \delta V_{p,k}(\xi - \lambda)\chi_{\lambda,k}^{(0)} Z_{p,k} D_\lambda D_k].$$

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

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

Hence, integrating (26) over $\xi$ we get

$$\Psi_p + \delta \Psi_p = + \frac{1}{E - 2\omega_p} \int V_{p,k}^{(0)} Z_{p,k} \Psi_k D_k$$

$$- \int \frac{D_\xi}{(E/2 - \omega_p)^2 + \xi^2} \delta V_{p,k}(\xi - \lambda)\chi_{\lambda,k}^{(0)} Z_{p,k} D_\lambda D_k$$

$$+ \frac{1}{E - 2\omega_p} \int V_{p,k}^{(0)} Z_{p,k} \delta \Psi_k D_k.$$

This is equivalent to

$$(E - 2\omega_p)(\Psi_p + \delta \Psi_p) = \int V_{p,k}^{(0)} Z_{p,k} \Psi_k D_k$$

$$+ \int \frac{2(-E/2 + \omega_p)}{(E/2 - \omega_p)^2 + \xi^2} \delta V_{p,k}(\xi - \lambda)\chi_{\lambda,k}^{(0)} Z_{p,k} D_\lambda D_k$$

$$+ \int V_{p,k}^{(0)} Z_{p,k} \delta \Psi_k D_k.$$

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

$$\delta E \Psi_p + (E_0 - 2\omega_p) \delta \Psi_p$$

$$= \int \frac{2(-E_0/2 + \omega_p)}{(E_0/2 - \omega_p)^2 + \xi^2} \delta V_{p,k}(\xi - \lambda)\chi_{\lambda,k}^{(0)} Z_{p,k} D_\lambda D_k$$

$$+ \int V_{p,k}^{(0)} Z_{p,k} \delta \Psi_k D_k.$$

Finally, multiplying this Eq. by $\Psi_p$ and integrating by $p$ we arrive to Eq.(25). Using (23), one frequency integration in (25) can be performed analytically and the retardation correction to the energy reduces to

$$\delta E = \int \Psi_p^* \frac{4(a_p + a_k)\delta V_{p,k}(\mu)(a_p + a_k)^2 + \mu^2}{(a_p + a_k)^2 + \mu^2} \Psi_k Z_{p,k} D_\mu D_p D_k. \quad (28)$$

Here the $\mu$-integration goes from $0$ to $\infty$. Since $\delta V$ is negative the retardation correction to the total energy is negative, $\delta E < 0$, and thereby increases the binding energy. If $\delta V$ is independent of frequency, $\delta V_{p,k}(\mu) \rightarrow \delta V_{p,k}$ the $\mu$-integration in (28) is trivial and the energy correction is reduced to the familiar expression from quantum mechanics, $\delta E = \int \Psi_p^* \delta V_{p,k} \Psi_k Z_{p,k} D_p D_k.$
The exciton condensation regime will be considered in a separate publication. The BSE solution includes the retardation correction (28).

Momenta integrations in (28) are well convergent. However, the perturbation $\delta V(\mu)$ increases with frequency, and we find that at large $\mu$ the integrand in (28) decays as $1/\mu$ leading to a logarithmic divergence of the integral. To remedy, we impose an ultraviolet cutoff $\Lambda = 0.5eV$, which corresponds to the energy scale that the single particle Hamiltonian (1) becomes invalid; one must account for additional bands (i.e. consider the basis $\{A_1, A_2, B\}$) see e.g. [20]. A change of the cutoff to, say, 1 eV does not significantly influence the retardation correction.

B. Zero frequency RPA potential as starting approximation

Let us consider first the solution of LSE (12) obtained in Section III as the zeroth approximation. So, we set $V_{p,k}^{(0)} = V_{p,k}(\xi = 0)$. In Fig.6a we present binding energies of the s-wave ground state for two sets of parameters: (i) $d = 1000nm$, $\epsilon_r = 1$; and (ii) $d = 20nm$, $\epsilon_r = 3.9$. The blue solid and dashed-dotted lines show the LSE solutions. These lines are identical to that in Fig.4a. Orange lines, solid and dashed-dotted, show the same energies with account of the retardation correction (28). The retardation correction is significant, especially for the $d = 1000nm$, $\epsilon_r = 1$. Moreover, it is qualitatively significant because for sufficiently small $\Delta$ it brings the system to the exciton condensation regime, $\epsilon_b > 2\Delta$. The exciton condensation regime will be considered in a separate publication.

In Fig.6b we present the binding energy of the lowest $p$-wave state for $d = 1000nm$, $\epsilon_r = 1$. The blue solid line shows the LSE solution. This line is identical to that in Fig.5a. The orange line show the same energy with account of the retardation correction (28). For the $p$-wave the retardation is less important. This is natural, as we already pointed out the retardation is more important for a larger binding energy.

C. Averaged over frequency RPA potential as starting approximation

We can improve accuracy of the calculation of the retardation effect. To do so let us change the zero approximation potential that enters LSE. Instead of the zero frequency, $V_{p,k}^{(0)} = V_{p,k}(\xi = 0)$, we take the frequency averaged potential

$$V_{p,k}^{(0)} = \frac{2}{\pi} \int_0^\Lambda d\xi \frac{\Gamma}{\xi^2 + \Gamma^2} V_{p,k}(\xi)$$

(29)

Where the frequency dependent potential is given by Eq.(10). Here $\Lambda = 0.5eV$ is the ultraviolet cutoff. The case $\Gamma = 0$ corresponds to the static screening considered in the previous subsection. From discussion in previous sections we suggest that the optimal value of $\Gamma$ is $\Gamma \approx \epsilon_b$. It is worth noting that the averaging (29) makes sense only in imaginary frequency. This is because the dependence of $V_{p,k}(\xi)$ on $\xi$ is monotonic. A similar averaging in real frequency would have limited meaning. In Fig.7a we present plots of the screening factors, Eq.(11), for averaged potentials with $\Gamma = 0, \Delta, 2\Delta$ at $\Delta = 40meV$. The averaging captures some of the physics or retardation, and thereby reduces screening in the zeroth approximation (i.e. in the LSE approach).

In Fig.7b we present binding energies calculated using the averaged potential method. Blue lines present LSE binding energies calculated with averaged potentials with $\Gamma = 0$ (solid), $\Gamma = \Delta$ (dashed), and $\Gamma = 2\Delta$ (dashed-dotted). The orange lines are the BSE binding energies (LSE with added retardation correction (28)).

---

FIG. 6. Exciton binding energies. (a) s-wave ground state for two sets of parameters, $d = 1000nm$, $\epsilon_r = 1$ (two upper solid lines) and $d = 20nm$, $\epsilon_r = 3.9$ (two lower dashed lines). (b) Lowest p-wave state for $d = 1000nm$, $\epsilon_r = 1$. In both panels LSE (blue curves) is solved with zero frequency RPA screened potential. The BSE solution includes the retardation correction (28).

FIG. 7. (a) Screening factors defined by Eq.(11) for frequency averaged potentials, Eq.(29), with $\Gamma = 0, \Delta, 2\Delta$ at $\Delta = 40meV$. (b) Binding energies calculated using the averaged potential method. Blue lines present LSE binding energies calculated with averaged potentials with $\Gamma = 0$ (solid), $\Gamma = \Delta$ (dashed), and $\Gamma = 2\Delta$ (dashed-dotted). The orange lines are the BSE binding energies (LSE with added retardation correction (28)).
frequency with subsequent usage of instantaneous LSE can account up to 50% of the retardation correction. For the accurate result one should combine the potential averaging, Eq.(29), and the explicit retardation correction (28). We reiterate again, this analysis is important for large binding energy, $\Delta \lesssim \epsilon_b \lesssim 2\Delta$. For weak binding the retardation correction is small and the method of the correction calculation is not very important.

V. COMPARISON WITH EXISTING DATA

Assuming hBN encapsulation, giving dielectric enhancement $\bar{\epsilon} = 3.9$, and taking the metallic gates to be at a distance $d = 20\text{nm}$, we can directly compare with the experimental measurements of Ref. [9]. We compare the exciton energies, $2\Delta - \epsilon_b$, vs $\Delta$ obtained using LSE and BSE, to those measured experimentally. 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. 8, from which we see that our techniques provide quantitative agreement. We stress that we have not introduced fitting parameters. We reiterate that this is the case of weak binding.

VI. DISCUSSION

Summary. In this work we considered the influence of dynamical screening (retardation) on exciton binding – i.e the binding energy $\epsilon_b$. We considered the particular example of biased bilayer graphene, whereby the bias field induces single-particle band gap $\Delta_b$. 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.

Screening significantly influences the excitonic binding energy. We argue that to probe 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 $d > 20\text{nm}$ above the plane. Counterintuitively, even though the characteristic radius of excitons considered here is $r \lesssim 10\text{nm}$, the difference in binding energies for metallic gates at $d = 20$ and 100nm is significant, Fig. 4.

We verify the quantitative accuracy of the methods 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.

Outlook. The crucial finding is the role of dynamical screening, which becomes significant in the regime $\epsilon_b \sim \Delta_b$. To probe this regime, we needed to consider the case of suspended BG with well separated metallic gates to reduce environment screening and thereby maximise the Coulomb interaction. Additionally, one could consider engineering of BG so as to generate a larger effective mass $m$ (smaller bandwidth). All things equal, enhancing the effective mass reduces the kinetic energy and therefore helps to promote condensation. This situation could be achieved via e.g. modulated electrostatic gating or via relative twist of the layers, or other van der Waals engineering. Application of the present techniques to modified BG is an important line of inquiry left for future work.

Finally, we leave it for future work to probe the possibility of exciton condensation, $\epsilon_b \geq \Delta_b$, and characterise the subsequent condensate.

VII. ACKNOWLEDGEMENTS

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) (CE170100039).
[1] L. V. Keldysh and Y. V. Kopaev, Possible instability of the semimetallic state against coulomb interaction, Fiz. Tverd. Tela., 6, 2791 (1964) [Sov. Phys. Solid State 6, 2219 (1965)] (1964).
[2] D. Jérôme, 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 (Academic Press, 1968) pp. 115–192.
[4] Y. E. Lozovik and V. I. Yudson, Feasibility of superfluidity of paired spatially separated electrons and holes; a new superconductivity mechanism, JETP Lett. (USSR) (Engl. Transl.); (United States).
[5] M. Pogrebinskii, Mutual drag of carriers in a semiconductor-insulator-semiconductor system, Soviet Physics-Semiconductors 11, 372 (1977).
[6] J. M. Blatt, K. Böer, and W. Brandt, Bose-einstein condensation of excitons, Physical Review 126, 1691 (1962).
[7] M. Kellogg, J. Eisenstein, L. Pfeiffer, and K. West, Vanishing hall resistance at high magnetic field in a double-layer two-dimensional electron system, Physical review letters 93, 036801 (2004).
[8] J.-J. Su and A. MacDonald, How to make a bilayer exciton condensate flow, Nature Physics 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), https://www.science.org/doi/pdf/10.1126/science.aam9175.
[10] C.-H. Park and S. G. Louie, Tunable excitons in biased bilayer graphene, Nano Letters 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. Berestetskii, E. Lifshitz, and L. Pitaevskii, Quantum Electrodynamics: Volume 4, Course of theoretical physics (Elsevier Science, 1982).
[15] M. M. Glazov and A. Chernikov, Breakdown of the static approximation for free carrier screening of excitons in monolayer semiconductors, Physica Status Solidi B 255, 180021 (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 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), https://www.science.org/doi/pdf/10.1126/science.1194988.
[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, Reports on Progress in Physics 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] E. H. Hwang and S. Das Sarma, Screening, kohn anomaly, friedel oscillation, and rkky interaction in bilayer graphene, Phys. Rev. Lett. 101, 156802 (2008).