Enhanced superconductivity through virtual tunneling in Bernal bilayer graphene coupled to WSe$_2$

Yang-Zhi Chou,1,∗ Fengcheng Wu,2,3 and Sankar Das Sarma1

1Condensed Matter Theory Center and Joint Quantum Institute, Department of Physics, University of Maryland, College Park, Maryland 20742, USA
2School of Physics and Technology, Wuhan University, Wuhan 430072, China
3Wuhan Institute of Quantum Technology, Wuhan 430206, China

(Dated: November 4, 2022)

Motivated by a recent experiment [arXiv:2205.05087], we investigate a possible mechanism that enhances superconductivity in hole-doped Bernal bilayer graphene due to a proximate WSe$_2$ monolayer. We show that the virtual tunneling between WSe$_2$ and Bernal bilayer graphene, which is known to induce Ising spin-orbit coupling, can generate an additional attraction between two holes, providing a potential explanation for enhancing superconductivity in Bernal bilayer graphene. Using microscopic interlayer tunneling, we derive the Ising spin-orbit coupling and the effective attraction as functions of the twist angle between Bernal bilayer graphene and the WSe$_2$ monolayer. Our theory provides an intuitive and physical explanation for the intertwined relation between Ising spin-orbit coupling and superconductivity enhancement, which should motivate future studies.

Introduction.— Recent experiments on Bernal bilayer graphene (BBG) [1] and rhombohedral trilayer graphene (RTG) [2] reveal multiple symmetry broken phases and provide a new understanding for superconductivity in general graphene systems (i.e., moiré [3–13] or moiréless systems [1, 2, 14]). In these non-moiré crystalline graphene multilayers, superconducting states with $T_c \leq 0.1$K are found in narrow regions close to interaction-driven “isospin” polarized phases [1, 2, 15]. Theoretically, the acoustic-phonon-mediated pairing can provide a consistent resolution for superconductivity in BBG and RTG [16–18], while interaction-driven mechanisms are also proposed [19–28].

A new experiment on superconductivity in BBG demonstrates that superconductivity can be significantly enhanced with a proximate WSe$_2$ [14]. The system consists of a WSe$_2$ monolayer on top of a BBG, and a displacement field ($D$) is used to control the layer-polarization of the low-energy bands. For a sufficiently large $D > 0$, the carriers of hole-doped BBG reside entirely on the top layer, and superconductivity is observed around 0.3K without a magnetic field. The superconducting state shows a nontrivial response to an in-plane magnetic field – a Pauli-limit violation at lower doping and Pauli-limited behavior at higher doping. For $D < 0$, no superconductivity is observed for $T_c \geq 30$mK, but the normal states are essentially consistent with the previous experiment without a WSe$_2$ layer [1]. The strikingly different results for $D > 0$ and $D < 0$ suggest the significance of the proximate WSe$_2$ layer. It is important to emphasize that WSe$_2$ enhances superconductivity quite substantially – the superconducting temperature is enhanced by an order of magnitude (from 30mK to 300mK), the region of the superconducting phase also becomes wider, and an in-plane magnetic field is no longer required to induce superconductivity.

The key task is to identify the physical origin of superconductivity enhancement. Since a small observable $T_c$ has been found in BBG [1], any additional pairing glue or reduction of Coulomb repulsion can result in a noticeable enhancement in superconductivity. However, such a cooperative enhancing mechanism must be absent without a nearby WSe$_2$ layer, manifesting an asymmetric effect in $D > 0$ and $D < 0$. The main goal of the current work is to provide a potential explanation for the WSe$_2$-enhanced superconductivity in BBG [14].

In this Letter, we propose a novel mechanism that enhances pairings in a BBG-WSe$_2$ system based on the interlayer tunneling between WSe$_2$ and BBG. Such a tunneling process is believed to induce Ising spin-orbit coupling (ISOC) in BBG [14, 29–32], implying significant interlayer tunneling. We develop a minimal theory that produces an effective attraction between two holes in the slightly hole-doped BBG via a virtual interlayer tunneling process in combination with an interaction between hole carriers and the virtual electron. Furthermore, we

FIG. 1. Setup of BBG-WSe$_2$ and band structure. (a) Side view of the BBG-WSe$_2$ system. The WSe$_2$ monolayer is on top of the BBG. A displacement field along the $z$-direction is exerted. (b) The schematic band structures of WSe$_2$ and BBG. The green line indicates the Fermi energy, which is on the band edge of the first BBG valence band ($E_2$). We ignore the spin splitting of the WSe$_2$ conduction bands in this illustration. We use $E_1$, $E_2$, $E_3$, and $E_4$ to label the BBG bands in ascending order in energy.
as well as the interlayer tunneling between WSe$_2$ and BBG. A site (blue dots) and B site (blue open circles) correspond to the positions of 1A and 1B sites in Fig. 1(a), respectively; the red dots indicate the 2B sites in the bottom graphene layer; 2A sites are right below the 1B sites. (b) ISOC due to virtual tunneling. (c) Attraction due to virtual tunneling.

Derive the ISOC and effective attraction as functions of the relative angle between WSe$_2$ and BBG, incorporating microscopic tunneling at extended Brillouin zones [33, 34]. Our results suggest that the enhanced superconductivity can be explained by the virtual tunneling from the WSe$_2$ layer in cooperation with the electron-phonon interaction, paving the way for higher-$T_c$ superconducting states in graphene systems.

Model. — We are interested in a BBG-WSe$_2$ system as depicted in Fig. 1. In the presence of a sufficiently large displacement field along $z$ direction ($D > 0$), the low-energy valence band of BBG is polarized at the 1A site [illustrated in Fig. 1(a)] on the top graphene layer. It was shown theoretically [31, 32] and experimentally [14, 29] that ISOC is induced primarily on the layer proximate to WSe$_2$, suggesting that tunneling between WSe$_2$ and the top graphene layer is essential. Thus, a minimal model shown theoretically [31, 32] and experimentally [14, 29] is inside the $2\Delta$ gap of Fig. 1(b) corresponds to a state with completely filled A sites and empty B sites. In our case with $\mathcal{E}_F$ at the $E_2$ band edge, the system is slightly hole-doped, and we can consider ground states with dilute holes on the A sites of the effective honeycomb lattice model [given by Eq. (1)]. Again, the effective description here is valid when $\mathcal{E}_F$ is at the $E_2$ band edge.

In addition to the onsite potential, we consider electron-electron interactions given by

$$
\hat{H}_I = \frac{U_0}{2} \sum_r \delta n(r) \{ \delta n(r) - 1 \} + U_1 \sum_{(r_A,r_B)} \delta n(r_A) \delta n(r_B),
$$

where $\delta n(r) = n(r) - \langle 0 | n(r) | 0 \rangle$, $| 0 \rangle$ is the state with a charge neutral configuration (i.e., filled A sites and empty B sites), $U_0 > 0$ ($U_1 > 0$) is the onsite (nearest-neighbor) Coulomb interaction, and $\langle r_A,r_B \rangle$ denotes the nearest-neighbor pair. We consider a sufficiently large $U_0$ such that at most one hole (electrons) can be created on sublattice A (B). The $U_1$ term describes the interaction between nearest-neighbor sites, and $U_1 < E_B$ is assumed (as the spontaneous formation of dipoles is forbidden). We will focus on the electron-hole attraction between electron on the B site and a hole on the nearest-neighbor A site in the virtual process.

The WSe$_2$ layer can be described by a semiconductor bandstructure with spin split valence bands [37] as illustrated in Fig. 1(b). Specifically, the energy splittings can be described by an ISOC, $\lambda \tau_z s_z$, with $\tau_z$ ($s_z$) being the $z$-component Pauli matrix for valley (spin). The interlayer tunneling between WSe$_2$ and BBG can facilitate spin-orbit splitting in BBG valence bands. To provide an intuitive understanding, we treat WSe$_2$ valence bands as a few representative energy levels described by a simplified Hamiltonian,

$$
\hat{H}_d = -W \left( d_{\tau\tau,s_\uparrow}^+ d_{\tau\tau,s_\downarrow} + d_{\tau\tau,s_\downarrow}^+ d_{\tau\tau,s_\uparrow} \right) - (W + \delta) \left( d_{\tau\tau,s_\uparrow}^+ d_{\tau\tau,s_\downarrow} + d_{\tau\tau,s_\downarrow}^+ d_{\tau\tau,s_\uparrow} \right),
$$

where $d_{\tau,s}$ denotes the fermionic annihilation operator with valley $\tau K$ and spin $s$ in the WSe$_2$ valence bands. $W$ and $\delta$ are the parameters for the WSe$_2$ valence bands. Finally, we consider a tunneling Hamiltonian between WSe$_2$ and BBG given by

$$
\hat{H}_V = \sum_{\tau,s,A,B} \left\{ c_{\tau A s A}^+ (r_A) \left[ V_{\tau A s A}^* d_{\tau A s A} + \bar{V}_{\tau A s A}^* d_{\bar{\tau} A s A} \right] + H.c. \right\} + \sum_{\tau,s,B} \left\{ c_{\tau B s B}^+ (r_B) \left[ V_{\tau B s B}^* d_{\tau B s B} + \bar{V}_{\tau B s B}^* d_{\bar{\tau} B s B} \right] + H.c. \right\},
$$

where $V_{\tau s}^\tau$ is the tunneling strength for the intravalley process (i.e., $\tau K$ to $\bar{\tau} K$), and $\bar{V}_{\tau s}^\tau$ is the tunneling strength for the intervalley process (i.e., $-\tau K$ to $\tau K$).
ergy levels on the A sites, realizing an ISOC with strength $Eqs. (1), (2), (3), and (4)$ can straightforwardly produce interlayer tunneling between WSe$_2$ and BBG. The main idea is to capture the full microscopic detail. We will later use a detailed approach incorporating WSe$_2$ and the reciprocal lattice vectors of WSe$_2$ and BBG, respectively.

The model designed here is physically motivated for understanding ISOC and effective attraction via interlayer tunneling. However, such a simplified model cannot capture the full microscopic detail. We will later use a detailed approach incorporating WSe$_2$ band structures and the microscopic interlayer tunneling matrix elements.

**Ising spin-orbit coupling and effective attraction.**— The minimal model $H = H_{IG} + H_I + H_d + H_V$ (given by Eqs. (1), (2), (3), and (4)) can straightforwardly produce ISOC in the first valence band of BBG. The main idea is that the second-order perturbation in $H_V$ as sketched in Fig. 2(b) (see also [35]) generates spin-valley-splitting energy levels on the A sites, realizing an ISOC with strength

$$
\lambda_I = \frac{|V_{A+\uparrow}|^2 - |\tilde{V}_{A+\uparrow}|^2}{W} - \frac{|V_{A+\downarrow}|^2 - |\tilde{V}_{A+\downarrow}|^2}{W + \delta}.
$$

The main goal of this work is to investigate if the interlayer tunneling between WSe$_2$ and BBG can generate an effective attractive interaction. Specifically, we consider two holes that contain a common nearest-neighbor B site at $x_B$ as illustrated in Fig. 2(c). At the second-order perturbation of $H_V$, the interplay between virtual tunneling and the interaction $H_I$ [Eq. (2)] generates an effective attraction between the holes, described by [35]

$$
\hat{H}_\text{att} = -\mathcal{U}_{\text{eff}} \sum_{\langle \ell A, A' \ell' \rangle} \delta n(\ell A) \delta n(\ell' A'),
$$

where the sum runs over the nearby pairs on A sites and

$$
\mathcal{U}_{\text{eff}} = \frac{V_{s}^2}{W + E_B} + \frac{V_{t}^2}{W + E_B - 2U_1} - \frac{2V_{t}^2}{W + E_B - U_1}
$$

$$
+ \frac{V_{t}^2}{W + \delta + E_B} + \frac{V_{t}^2}{W + \delta + E_B - 2U_1} - \frac{2V_{t}^2}{W + \delta + E_B - U_1}
$$

with $V_{s}^2 = 2 |V_{s+\uparrow}|^2 + 2 |V_{s+\downarrow}|^2$. $\mathcal{U}_{\text{eff}}$ vanishes as $U_1$ is absent. While the electron on a B site has a large local energy $E_B$, the nearest-neighbor electron-hole attraction can lower the total energy in a virtual state. In Eq. (7), those terms with $-2U_1$, corresponding to virtual tunneling to $x_B$, yield the dominant contributions as long as $U_1 > 0$. We assume that $W + E_B > 2U_1$ so that the any charge transfer between WSe$_2$ and BBG should be absent. Since $|V_{A+\uparrow}|$ and $|V_{A+\downarrow}|$ are of the same order of magnitude [35], we expect that the virtual tunneling generates a sizable effective attraction $\mathcal{U}_{\text{eff}}$. Our theory therefore provides an intuitive understanding of the effective attraction due to a proximate WSe$_2$ layer.

The proposed mechanism here is conceptually related to the “polarizer” idea [39, 40] and the repulsion-induced attraction in models on the honeycomb lattice [41–44]. We discuss a few differences between our work and Refs. [41–44]— (i) The virtual process is due to interlayer tunneling rather than intralayer hopping, and (ii) the electron-hole attraction in the virtual process rather than electron-electron repulsion. Point (i) is crucial as our mechanism describes a possible enhanced attraction from WSe$_2$ rather than pairings due to intralayer processes. In addition, point (ii) allows for a wider parameter range for a sizable effective attraction because the large onsite energy $E_B$ can be compensated by a nearest neighbor attraction.

**Interlayer tunneling and twist angle.**— The interlayer tunneling between WSe$_2$ and BBG crucially determines ISOC as well as the virtual-tunneling-induced effective attraction. The interlayer tunneling preserves crystal momentum as the matrix element primarily depends on the distance between the sites. Within the two-center approximation scheme, the tunneling between the layers can be described by [33]

$$
\mathcal{T}_{k, p}^{\alpha, \sigma; \beta, \sigma'} = \frac{1}{\sqrt{A_0 A_1}} \left( a_{k, \alpha, \sigma}^{(1)} \right)^* \left( a_{p, \beta, \sigma'}^{(0)} \right) \times \sum_{Q^{(1)}, Q^{(0)}} t_{k+Q^{(1)}, p+Q^{(0)}}^{\alpha, \sigma} q^{\alpha, \sigma}_{k+Q^{(1)}, p+Q^{(0)}} q^{\beta, \sigma'}_{k+Q^{(1)}, p+Q^{(0)}},
$$

where $A_0$ ($A_1$) is the unit-cell area of WSe$_2$ (BBG), $\alpha$ and $\beta$ are the band indexes, $k$ of BBG and $p$ of WSe$_2$ are the momentum relative to $\Gamma$ point (Brillouin zone center), $\sigma$ and $\sigma'$ are the sublattice (orbital) indexes, $a_{k, \alpha, \sigma}^{(1)}$ is the sublattice (orbital) projection of a Bloch state $|k, \alpha, \sigma\rangle$ at layer $l$, $Q^{(0)}$, $Q^{(1)}$ are the reciprocal lattice vectors in $\tau K$) [38]. Since the entire system preserves the (spinful) time-reversal symmetry, the tunneling terms obey $|V_{s+\uparrow}| = |\tilde{V}_{s+\uparrow}|$ and $|V_{s+\downarrow}| = |\tilde{V}_{s+\downarrow}|$, where $\tilde{s} (\bar{s})$ means the time-reversal partner of $s$ (\bar{s}).
WSe$_2$ and BBG, respectively, and $\phi^Q_{\sigma',\sigma}(0)$ is a phase factor depending on sublattice. In the above expression, we use index 0 for the WSe$_2$ layer and index 1 for the top graphene layer of BBG. $I_{k}$ is the 2D Fourier transform of the interlayer tunneling amplitude with a finite range, and we use a stretched exponential ansatz form [33],

$$|I_{k}| = t_0 e^{-\chi(|k| z_\perp)^\gamma},$$

where $t_0$ is an overall constant, $\chi$ is an order 1 numerical constant, $\gamma$ is the exponent of the stretched exponential, and $z_\perp$ is the distance between the WSe$_2$ monolayer and the top graphene layer of BBG [45]. Note that Eq. (9) is an empirical expression for a finite $k$, and the potential complications for $k \to 0$ are not relevant to our problem.

The interlayer tunneling described by Eq. (8) is a highly nontrivial single-particle process. Microscopically, WSe$_2$ and BBG have a relative angle that is tunable experimentally as well as different lattice constants ($d = 3.31\text{Å}$ for WSe$_2$, $a = 2.46\text{Å}$ for BBG), resulting in Brillouin zones illustrated in Fig. 3. Note that $T^{\alpha,\sigma,\sigma'}_{k,p}$ is nonzero as long as $k + Q^{(1)} = p + Q^{(0)}$ for some reciprocal lattice vectors $Q^{(0)}$ and $Q^{(1)}$. Thus, a full calculation must incorporate a sufficiently large number of extended Brillouin zones. Using the virtual-tunneling ideas discussed previously, we can compute the $\lambda_I$ and $U_{\text{eff}}$ incorporating the microscopic interlayer tunneling matrix elements between the $k \cdot p$ WSe$_2$ bands [37] and our honeycomb model. We summarize the main results, and the complete derivations are provided in [35].

In Fig. 4, we plot the twist-angle ($\theta$) dependence of $\lambda_I$ with a few representative values of $\chi$ and $\gamma$ [46]. $\lambda_I$ is generally positive for small positive $\theta$ and becomes negative slightly above $\theta = 15^\circ$. Moreover, nonmonotonic behavior can manifest near $\theta = 5^\circ$ and $\theta = 20^\circ$ for smaller $\gamma$ and $\chi$. These results can be understood by the geometry of momentum at $\theta = 5^\circ$ and $\theta \geq 20^\circ$ [35]. At $\theta = 30^\circ$, $\lambda_I$ vanishes because the intravalley and intravalley tunnelings have exactly the same contributions. Similar nonmonotonic behavior near $\theta = 20^\circ$ was also theoretically reported in graphene coupled to transition metal dichalcogenide [34, 47, 48]. In addition, the sign changing in $\lambda_I$ (curves with $\gamma = 1$ in Fig. 4) was also obtained in Ref. [34]. In Fig. 5, we plot the $\theta$ dependence of $U_{\text{eff}}$ with $U_1 = 0.3eV$ and $U_{1} = 0.43eV$ (the largest possible value in our theory) and a few different values of $\gamma$ and $\chi$. $U_{\text{eff}}$ is generally larger at smaller $\theta$, but nonmonotonic features might develop for smaller $\gamma$ and $\chi$. While the qualitative results are insensitive to $U_1$, the quantitative values depend on $U_1$ [35]. An important implication here is that ISOC strength and the effective attraction are not directly related to each other. The full results depend a lot on the details of $T^{\alpha,\sigma,\sigma'}_{k,p}$.

Discussion.— The virtual-tunneling-induced attraction might explain the enhanced superconductivity in BBG-WSe$_2$ [14] as it enters the theory nonperturbatively in an exponential function determining $T_c$ and the effect is primarily on $D > 0$. In the BBG experiment without WSe$_2$ [1], superconductivity with $T_c \sim 30\text{mK}$ was reported at the doping density $n_c = -6 \times 10^{12}\text{cm}^{-2}$ with a Fermi-surface degeneracy factor of 2, implying that the pairing interaction between holes overcomes Coulomb suppression. Since the enhanced superconductivity is found near the same doping density with a very similar Fermi-surface degeneracy factor, we anticipate that the same pairing mechanism manifests in BBG with or without the proximate WSe$_2$. According to Ref. [17], the acoustic-phonon-mediated attraction is slightly stronger than the Coulomb suppression, so any additional pairing glue due to presence of the WSe$_2$, albeit small compared to the Coulomb repulsion, might considerably enhance superconductivity because $T_c$ is exponentially related to the coupling constant. As such, our theory based on the interplay between virtual tunneling and interaction is a possible explanation for the BBG-WSe$_2$ experiment [14]. We also derive the twist-angle dependence of the ISOC parameter and the effective attraction, providing guidance for future BBG-WSe$_2$ systems.

It is worth mentioning that our theory most likely underestimates $U_{\text{eff}}$ because we consider only the nearest-neighbor Coulomb interaction and only valley momenta (i.e., $K$ and $-K$ points) in the BBG bands. With a
long-range Coulomb interaction, a long-range effective attraction can also arise from virtual tunneling processes, resulting in an overall stronger pairing. Moreover, incorporating all the allowed momenta in the BBG $E_4$ band also enhances $U_{\text{eff}}$ because there are more states to be tunneled into.

Now we discuss the pairing symmetry and the normal state in the BBG-WSe$_2$ system. We anticipate that inter-valley intrasublattice pairings dominate [16, 49, 50] because of the layer-sublattice polarization in BBG. Based on the previous work on acoustic phonons [17] and the ideas discussed in this Letter, we assume that the dominant pairing is due to the acoustic phonons, and the virtual-tunneling-induced attraction is the subleading pairing glue. Note that the normal states for superconductivity in Refs. [1] and [14] are qualitatively different due to the differences in spin-orbit coupling and the in-plane magnetic field. While our theory does not determine the normal state properties, the major pairing glue, the acoustic-phonon-mediated pairing, has a $SU(2) \times SU(2)$ symmetry [16, 17, 49] (allowing for singlet, triplet, and singlet–triplet mixing). Thus, the superconductivity should be enhanced regardless of the details of the spin symmetry in the normal state or in the subleading pairing [51]. Due to the induced spin-orbit coupling, we anticipate an admixture of singlet and triplet pairings [52, 53], which can produce a beyond-Pauli-limit response to the in-plane magnetic field, consistent with the lower-doping superconductivity in the BBG-WSe$_2$ experiment [14].

While our theory provides a potentially consistent resolution to the BBG-WSe$_2$ experiment, there are a few issues that require further investigations. First, the effective attraction $U_{\text{eff}}$ depends on the value of $U_1$, which we treat as a parameter. This makes a quantitative estimate of $U_{\text{eff}}$ difficult in our theory. The other issue is related to the differences in the normal states between Refs. [1] and [14]. In our work, we simply assume that such differences are not crucial to the superconductivity enhancement. However, as pointed out in Ref. [14], a nontrivial normal state due to an interplay between the interaction and spin orbit coupling might explain the enhancement of superconductivity. To fully understand the factor-of-10 enhancement of $T_c$ in the BBG-WSe$_2$ experiment [14], one needs to incorporate both the additional pairing glue and the change in normal states.

Finally, we discuss implications for experiments. First, the twist-angle dependence of ISOC [Fig. 4] and effective attraction [Fig. 5] might provide insights on why enhanced superconductivity is not always achieved in BBG-WSe$_2$ experiments [14]. Based on our theory, applying a pressure to the system should considerably raise $T_c$ since an applied pressure should enhance tunneling. Furthermore, one can use the ideas of this work to look for the optimal proximate layer for enhancing superconductivity. An important message of this Letter is that superconductivity can still be enhanced even if the induced spin-orbit coupling is very small, because of the exponential nature in $T_c$ and the complex relation between effective attraction and spin-orbit coupling.

Acknowledgments.— We thank Étienne Lantagne-Hurtubise, Alex Thomson, and Cyprian Lewandowski for pointing out an issue in an earlier version of the manuscript. We are also grateful to Seth Davis, Jiabin Yu, and Ming Xie for useful discussions. This work is supported by the Laboratory for Physical Sciences (Y.-Z.C. and S.D.S.), by JQI-NSF-PFC (Y.-Z.C.), and by ARO W911NF2010232 (Y.-Z.C.). F.W. is supported by National Key R&D Program of China 2021YFA1401300 and start-up funding of Wuhan University.

* yzhou@umd.edu

[1] H. Zhou, L. Holleis, Y. Saito, L. Cohen, W. Huynh, C. L. Patterson, F. Yang, T. Taniguchi, K. Watanabe, and A. F. Young, Science 375, 774 (2022).
[2] H. Zhou, T. Xie, T. Taniguchi, K. Watanabe, and A. F. Young, Nature 598, 434 (2021).
[3] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Nature 556, 43 (2018), URL http://dx.doi.org/10.1038/nature26160.
[4] M. Yankowitz, S. Chen, H. Polshyn, Y. Zhang, K. Watanabe, T. Taniguchi, D. Graf, A. F. Young, and C. R. Dean, Science 363, 1059 (2019).
[5] X. Lu, P. Stepantov, W. Yang, M. Xie, M. A. Aamir, I. Das, C. Urgell, K. Watanabe, T. Taniguchi, G. Zhang, et al., Nature 574, 653 (2019).
[6] Z. Hao, A. Zimmerman, P. Ledwith, E. Khalaf, D. H. Najafabadi, K. Watanabe, T. Taniguchi, A. Vishwanath, and P. Kim, Science 371, 1133 (2021).
[7] J. M. Park, Y. Cao, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, Nature 590, 249 (2021).
[8] Y. Cao, J. M. Park, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, Nature 595, 526 (2021).
[9] X. Liu, K. Watanabe, T. Taniguchi, and J. Li, arXiv preprint arXiv:2108.03338 (2021).
[10] Y. Zhang, R. Polski, C. Lewandowski, A. Thomson, Y. Peng, Y. Choi, H. Kim, K. Watanabe, T. Taniguchi, J. Alicea, et al., arXiv preprint arXiv:2112.09270 (2021).
[11] J. M. Park, Y. Cao, L. Xia, S. Sun, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, arXiv preprint arXiv:2112.10760 (2021).
[12] G. W. Burg, E. Khalaf, Y. Wang, K. Watanabe, T. Taniguchi, and E. Tutuc, arXiv preprint arXiv:2201.01637 (2022).
[13] P. Siriviboon, J. Lin, X. Liu, H. Scammell, S. Liu, D. Rhodes, K. Watanabe, T. Taniguchi, J. Hone, M. Scheurer, et al., arXiv preprint arXiv:2112.07127 (2021).
[14] Y. Zhang, R. Polski, A. Thomson, É. Lantagne-Hurtubise, C. Lewandowski, H. Zhou, K. Watanabe, T. Taniguchi, J. Alicea, and S. Nadj-Perge, arXiv preprint arXiv:2205.05087 (2022).
[15] H. Zhou, T. Xie, A. Ghazaryan, T. Holder, J. R. Ehrets, E. M. Spanton, T. Taniguchi, K. Watanabe, E. Berg,
In this work, we use $E_B = \sqrt{t_1^2 + \Delta^2 + \Delta + \Delta'}$, where $t_1$ denotes the interlayer dimerization energy between the 1B and 2A sites of BBG, $\Delta$ is the potential energy induced by the displacement field, $\Delta'$ is the local potential shift at 1B and 2A sites of BBG. (See Ref. [54] and Supplementary Material for the values used in this work [55].)

In principle, the tunneling strength should vary in space because the WSe$_2$ layer and top graphene layer of BBG can form a moiré pattern. However, the BBG-WSe$_2$ system has a mean free path of order of sample size [14], suggesting that the spatial dependence is not essential. Thus, we ignore the spatial dependence in $V_{rs}$ and $V_{rs}^\prime$.

[54] J. Jung and A. H. MacDonald, Phys. Rev. B 89, 035405 (2014), URL https://link.aps.org/doi/10.1103/PhysRevB.89.035405.

[55] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao, Phys. Rev. Lett. 108, 196802 (2012), URL https://link.aps.org/doi/10.1103/PhysRevLett.108.196802.
Enhanced superconductivity through virtual tunneling in Bernal bilayer graphene coupled to WSe$_2$

SUPPLEMENTAL MATERIAL

In this supplemental material, we provide some technical details for the main results in the main text.

**SINGLE-PARTICLE BANDS AND EFFECTIVE HONEYCOMB LATTICE MODEL**

The \( \mathbf{k} \cdot \mathbf{p} \) Hamiltonian for BBG is based on Ref. [1]. The \( \hat{h}_r(\mathbf{k}) \) in the main text is given by

\[
\hat{h}_r(\mathbf{k}) = \begin{pmatrix}
-\Delta & v_0 \Pi(\mathbf{k}) & -v_4 \Pi(\mathbf{k}) & -v_3 \Pi(\mathbf{k}) \\
v_0 \Pi(\mathbf{k}) & \Delta' - \Delta & t_1 & -v_4 \Pi(\mathbf{k}) \\
-v_4 \Pi(\mathbf{k}) & t_1 & \Delta' + \Delta & v_0 \Pi(\mathbf{k}) \\
-v_3 \Pi(\mathbf{k}) & -v_4 \Pi(\mathbf{k}) & v_0 \Pi(\mathbf{k}) & \Delta
\end{pmatrix},
\]

where \( \Pi(\mathbf{k}) = \tau \kappa_x a + i \kappa_y a \), \( a = 2.46 \) Å is the lattice constant of graphene and \( \Delta \) encodes the electric potential difference from the displacement field. In this work, we take \( \Delta = 50 \) meV. Note that \( \mathbf{k} \) is relative to \( \tau K \). Other parameters are given by \( v_0 = 2.261 \) eV, \( v_3 = 0.245 \) eV, \( v_4 = 0.12 \) eV, \( t_1 = 0.361 \) eV, and \( \Delta' = 0.015 \) eV. The basis of the matrix is \((1A, 1B, 2A, 2B)\).

The low-energy bands manifest layer-sublattice polarization, so that only the A sites of the top layer (1A) and B sites of the bottom layer (2B) are essential for slightly doped BBG. This property arises from the interlayer nearest-neighbor tunnelings which tend to form dimerized bonds between 1B and 2A sites. For \( \Delta > 0 \), the low-energy valence band can be well described by 1A sites alone. To see this, we examine \( \mathbf{k} = 0 \) case in the following. Right at \( \tau K \),

\[
\hat{h}_r(0) = \begin{pmatrix}
-\Delta & 0 & 0 & 0 \\
0 & \Delta' - \Delta & t_1 & 0 \\
0 & t_1 & \Delta' + \Delta & 0 \\
0 & 0 & 0 & \Delta
\end{pmatrix}.
\]

The eigenvalues of \( \hat{h}_r(0) \) are

\[
E_1 = -\sqrt{\Delta^2 + t_1^2} + \Delta', \ E_2 = -\Delta, \ E_3 = \Delta, \ E_4 = \sqrt{\Delta^2 + t_1^2} + \Delta',
\]

corresponding to the following eigenstates

\[
\psi_1 \approx \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \ \psi_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ \psi_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \ \psi_4 \approx \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix},
\]

where we have used \( t_1 \gg \Delta \). \( E_2 \) and \( E_3 \) correspond to the low-energy valence band edge and the low-energy conduction band edge respectively, while \( E_1 \) and \( E_4 \) are the high energy band edges.

We are interested in a slightly hole-doped BBG with a sufficiently large \( D > 0 \). Ignoring the trigonal wrapping terms, the zero energy states are precisely at \( K \) and \(-K\) valleys, and \( \hat{h}_r(\mathbf{k} \to 0) \) corresponds to a collection of decoupled atomic sites in the position space. Since \( \hat{h}_r(0) = \hat{h}_{-}(0) \), we can treat the valley as an internal quantum number, and we construct an effective honeycomb lattice. In such a model, we concentrate only on 1A and 1B sites with \( \mathbf{k} \approx 0 \). The minimal model for BBG contains only the \( E_2 \) and \( E_4 \) bands because contributions form other bands are much weaker. We explain the relevant band degrees of freedom in the following.

First, \( \psi_2 \) has weights only on the 1A sites, so the \( E_2 \) band must be included for the interlayer tunneling. On the contrary, the \( E_3 \) band can be ignored because \( \psi_3 \) has weights only on the 2B sites. Since both \( \psi_1 \) and \( \psi_4 \) have finite weights on the 1B sites, we need to investigate the virtual processes involving both the \( E_1 \) and \( E_4 \) bands. The primary interlayer process involving the \( E_1 \) band corresponds to tunneling an electron from the BBG \( E_1 \) band to the WSe$_2$ conduction bands; the primary interlayer process involving the \( E_4 \) band corresponds to tunneling an electron from the
WSe\textsubscript{2} valence bands to the BBG \textit{E}_{1} band. For the induced ISOC, the spin-splitting in the WSe\textsubscript{2} conduction bands is much smaller than the valence bands, so we can ignore the contributions from the WSe\textsubscript{2} conduction bands. For the effective attraction in hole-doped BBG, the virtual process \textit{(i)}, involving the BBG \textit{E}_{1} band and the WSe\textsubscript{2} conduction bands, is a much weaker than the virtual process \textit{(ii)}, involving the BBG \textit{E}_{\text{4\textsubscript{1}}} band and the WSe\textsubscript{2} valence bands. The results are due to the Coulomb interaction in the virtual states such that the energy is increased by \(2U_1\) in \textit{(i)} and the energy is decreased by \(2U_1\) in \textit{(ii)}. Thus, the process \textit{(i)} yields a much smaller contribution, and we focus only on the process \textit{(ii)}. Therefore, the interlayer tunneling processes with the \textit{E}_{\text{2\textsubscript{1}}} and \textit{E}_{\text{4\textsubscript{1}}} bands are the dominating contributions.

The effective honeycomb lattice in the main text is related to the top graphene layer but with some differences. The B sites of the effective honeycomb model are \textit{not} the microscopic BBG 1B sites because we ignore the \textit{E}_{\text{1\textsubscript{1}}} band completely. In the main text, we consider \(E_A = 0\) and \(E_B = \sqrt{t^2 + \Delta^2 + \Delta + \Delta'}\) (the energy difference between \textit{E}_{\text{4\textsubscript{1}}} and \textit{E}_{\text{2\textsubscript{1}}} bands). In addition, the charge neutrality configuration in the effective model is described by completely filled A sites and empty B sites, which is quite different from the microscopic graphene where both sites have similar electron density. This effective honeycomb lattice description allows us to present the effective attraction in an intuitive and transparent way.

**MONOLAYER TUNGSTEN DISELENIDE**

The WSe\textsubscript{2} monolayer can be modeled by a massive Dirac model [2] given by

\[
\hat{h}_{\text{WSe}_{2}}(k) = v(\tau \hat{\sigma}_x k_x + \hat{\sigma}_y k_y) + \frac{\Omega}{2} \hat{\sigma}_z + \lambda \frac{\tau - 1}{2} \hat{s}_z = \begin{bmatrix}
\frac{\Omega}{2} & 0 & v(\tau k_x - ik_y) & 0 \\
0 & \frac{\Omega}{2} & 0 & v(\tau k_x - ik_y) \\
v(\tau k_x - ik_y) & 0 & -\Omega/2 + \lambda & 0 \\
0 & v(\tau k_x - ik_y) & 0 & -\Omega/2 - \lambda
\end{bmatrix}
\]

(S5)

where \(v = td, d = 3.31\,\text{Å}, t = 1.19\,\text{eV}, \Omega = 1.6\,\text{eV}\) and \(2\lambda = 0.46\,\text{eV}\). The basis of \(\hat{h}_{\text{WSe}_{2}}(k)\) is \((C^\uparrow, C^\downarrow, V^\uparrow, V^\downarrow)\), where The label \(C\) corresponds to the \(d_{z^2}\) orbital function and the label \(V\) corresponds to the \(d_{x^2-y^2} + i\tau d_{xy}\) orbital function [2]. The eigenvalues of \(\hat{h}_{\text{WSe}_{2}}(k)\) are given by

\[
E_{1\text{WSe}_{2}}(k) = -\lambda - \sqrt{4v^2k^2 + (\Omega + \lambda)^2}, \quad \text{(S6a)}
\]
\[
E_{2\text{WSe}_{2}}(k) = \lambda - \sqrt{4v^2k^2 + (\Omega - \lambda)^2}, \quad \text{(S6b)}
\]
\[
E_{3\text{WSe}_{2}}(k) = -\lambda + \sqrt{4v^2k^2 + (\Omega + \lambda)^2}, \quad \text{(S6c)}
\]
\[
E_{4\text{WSe}_{2}}(k) = \lambda + \sqrt{4v^2k^2 + (\Omega - \lambda)^2}. \quad \text{(S6d)}
\]
corresponding to

\[
\psi_1^{(\text{WS}e_2)}(\mathbf{k}) = \frac{1}{\sqrt{1 + \frac{2v|k|}{\Omega + \lambda + \sqrt{4v^2k^2 + (\Omega + \lambda)^2}}}} \begin{pmatrix} 0 \\ \frac{2v(-k_x+ik_y)}{\Omega + \lambda + \sqrt{4v^2k^2 + (\Omega + \lambda)^2}} \\ 0 \\ 1 \end{pmatrix} \approx \frac{1}{\sqrt{1 + \frac{v^2k^2}{(\Omega + \lambda)^2}}} \begin{pmatrix} 0 \\ \frac{v(-k_x+ik_y)}{\Omega + \lambda} \\ 0 \\ 1 \end{pmatrix},
\]

(S7a)

\[
\psi_2^{(\text{WS}e_2)}(\mathbf{k}) = \frac{1}{\sqrt{1 + \frac{2v|k|}{\Omega - \lambda + \sqrt{4v^2k^2 + (\Omega - \lambda)^2}}}} \begin{pmatrix} 0 \\ \frac{2v(-k_x+ik_y)}{\Omega - \lambda + \sqrt{4v^2k^2 + (\Omega - \lambda)^2}} \\ 0 \\ 1 \end{pmatrix} \approx \frac{1}{\sqrt{1 + \frac{v^2k^2}{(\Omega - \lambda)^2}}} \begin{pmatrix} 0 \\ \frac{v(-k_x+ik_y)}{\Omega - \lambda} \\ 0 \\ 1 \end{pmatrix},
\]

(S7b)

\[
\psi_3^{(\text{WS}e_2)}(\mathbf{k}) = \frac{1}{\sqrt{1 + \frac{2v|k|}{\Omega + \lambda + \sqrt{4v^2k^2 + (\Omega + \lambda)^2}}}} \begin{pmatrix} 0 \\ \frac{2vk_x}{\Omega + \lambda + \sqrt{4v^2k^2 + (\Omega + \lambda)^2}} \\ 0 \\ 1 \end{pmatrix} \approx \frac{1}{\sqrt{1 + \frac{v^2k^2}{(\Omega + \lambda)^2}}} \begin{pmatrix} 0 \\ \frac{v(k_x+ik_y)}{\Omega + \lambda} \\ 0 \\ 1 \end{pmatrix},
\]

(S7c)

\[
\psi_4^{(\text{WS}e_2)}(\mathbf{k}) = \frac{1}{\sqrt{1 + \frac{2v|k|}{\Omega - \lambda + \sqrt{4v^2k^2 + (\Omega - \lambda)^2}}}} \begin{pmatrix} 0 \\ \frac{2vk_x}{\Omega - \lambda + \sqrt{4v^2k^2 + (\Omega - \lambda)^2}} \\ 0 \\ 1 \end{pmatrix} \approx \frac{1}{\sqrt{1 + \frac{v^2k^2}{(\Omega - \lambda)^2}}} \begin{pmatrix} 0 \\ \frac{v(k_x+ik_y)}{\Omega - \lambda} \\ 0 \\ 1 \end{pmatrix}.
\]

(S7d)

In the above expressions, we have assumed that \(v|k| \ll \Omega - \lambda < \Omega + \lambda\). At \(\mathbf{k} = 0\), \(E_1^{\text{WS}e_2} = -\Omega/2 - \lambda\), \(E_2^{\text{WS}e_2} = -\Omega/2 + \lambda\), and \(E_3^{\text{WS}e_3} = E_4^{\text{WS}e_3} = \Omega/2\). Assuming that WSe\(_2\) layer has a potential energy lower than BBG top layer by \(\Delta\) (tuned by the displacement field), we can identify that \(W = \Delta + \Omega/2 - \lambda\) and \(\delta = 2\lambda\).

**DERIVATION OF ISOC AND VIRTUAL-TUNNELING-INDUCED ATTRACTION**

The effective model in the main text allows for an intuitive understanding on ISOC as well as virtual-tunneling-induced attraction by treating interlayer tunneling using second-order perturbation theory. The unperturbed Hamiltonian is described by \(\hat{H}_0 = \hat{H}_{1G} + \hat{H}_I + \hat{H}_d\), where \(\hat{H}_{1G}\) is the effective honeycomb lattice, \(\hat{H}_I\) encodes the short-range Coulomb interaction, and \(\hat{H}_d\) represents the effective energy levels for WSe\(_2\). The perturbed Hamiltonian is the interlayer tunneling given by \(\hat{H}_V\). Note that the ground state of the unperturbed Hamiltonian is described by filled WSe\(_2\) levels and dilute (singly occupied) holes on A sites of the effective honeycomb lattice model. The correction of energy in second-order perturbation theory is given by

\[
\delta E_g = \sum_e \langle g | \hat{H}_V | e \rangle \langle e | \frac{1}{E_g - \hat{H}_0} | e \rangle \langle e | \hat{H}_V \hat{P} | g \rangle,
\]

(S8)

where \(E_g\) and \(|g\rangle\) are the energy and the wavefunction of the ground state of \(\hat{H}_0\), \(|e\rangle\) denotes the excited state of \(\hat{H}_0\), and the sum runs over the excited states of \(\hat{H}_0\). The ground state is described by dilute holes on A sites, and no more than two holes are occupied at the same position. Since the Fermi energy is at the band edge of the valence band, the ground states consists a few holes. Specifically, we focus on one-hole states and two-hole states (with two holes connected by a nearest-neighbor common B site).
Ising spin-orbit coupling

To derive ISOC, we consider single hole ground states and compute the correction of energies based of the valley and spin quantum numbers. The energy correction to a hole with valley $\tau$ and spin $s$ is given by $E_{\tau s}$, where

$$
\delta E_{\tau s}^{(1)} = \frac{|V_{\tau s}|^2}{W + E_B} - \frac{|\tilde{V}_{\tau s}|^2}{W + \delta + E_B - U_1},
$$

(S14)

$$
\delta E_{\tau s}^{(2)} = \frac{|V_{\tau s}|^2}{W + E_B - 2U_1} - \frac{|\tilde{V}_{\tau s}|^2}{W + \delta + E_B - 2U_1},
$$

(S15)

The energy difference between the nearby-holes state and the remote-holes state is given by

$$
\Delta E = E_{\tau s}^{(0)} + E_{\tau s}^{(2)} - E_{\tau s}^{(1)};
$$

(S17)

Virtual-tunneling-induced attraction

Now, we discuss the effective attraction induced by virtual tunnelings. We consider a state with two holes that are located at two nearest-neighbor A sites as illustrated in Fig. 2(c) in the main text. Two A sites are connected by a common B site (marked by $x_B$), which represents a high energy conduction band of BBG. We compare the energy difference between the state with two nearby holes and the state with two far separated holes. Note that we need to take into account the entire energy renormalization due to the virtual interlayer tunneling. In Fig. S1, we plot two configurations of states, and we use different colors to specify the different virtual tunneling contributions. Specifically, the white circles correspond to the on-site energy correction $E_v^{(0)}$ (without any nearest-neighbor hole), the green circles correspond to the on-site energy correction $E_v^{(1)}$ (with one nearest-neighbor hole), the red circles correspond to the on-site energy correction $E_v^{(2)}$ (with two nearest-neighbor holes). These onsite energy correction at the second order are given by

$$
E_v^{(0)} = \frac{|V_{\tau s}|^2}{W + E_B} - \frac{|V_{\tau s}|^2}{W + \delta + E_B - U_1},
$$

(S18)
where we have used $|V_{rs}^{B}| = |V_{rs}^{B}|$ and $|\bar{V}_{rs}^{B}| = |\bar{V}_{rs}^{B}|$. Thus, the effective attraction in the main text is expressed by

$$U_{\text{eff}} = \left(2 |V_{s+t}^{B}|^2 + 2 |\bar{V}_{s+t}^{B}|^2\right) \left[\frac{1}{W + E_B} + \frac{1}{W + E_B - 2U_1} - \frac{2}{W + E_B - U_1}\right]$$

$$+ \left(2 |V_{s+t}^{B}|^2 + 2 |\bar{V}_{s+t}^{B}|^2\right) \left[\frac{1}{W + \delta + E_B} + \frac{1}{W + \delta + E_B - 2U_1} - \frac{2}{W + \delta + E_B - U_1}\right].$$

(S19)

We note that $U_{\text{eff}} \propto U_1^2$ for an infinitesimal $U_1$, indicating that the effective attraction is absent in the noninteracting limit.

### Stability of second-order perturbation theory

Since our analysis is based on the second-order perturbation, it is important to justify that the higher-order contributions are small. The expansion parameters in our theory are $|V_A|/W$ and $|V_B|/(W + E_B - 2U_1)$. Using $|V_A| \approx 100\text{meV}$ (based on Ref. [3]), we obtain $|V_A|/W \approx 0.16$ and $|V_B|/(W + E_B - 2U_1) \approx 0.37$ for $U_1 = 0.43\text{eV}$ ($V_B \approx V_A/\sqrt{2}$). The higher-order contributions start at the fourth order in tunneling, and there are two distinct contributions – (a) fourth order in $V_B$ and (b) second order in $V_A$ and second order in $V_B$. The former case results in a quantitative change in $U_{\text{eff}}$, and the latter case generates an interaction that breaks spin symmetry. Using the naive counting estimate, the case (a) gives a correction of order $U_{\text{eff}}|V_B|^2/(W + E_B - 2U_1)^2 \approx 0.14U_{\text{eff}}$, and case (b) generates a spin-dependent interaction of order $U_{\text{eff}}|V_A|^2/W^2 \approx 0.03U_{\text{eff}}$. Both contributions are small compared to $U_{\text{eff}}$, suggesting that the results based on second-order perturbation theory are reasonable.

### MICROSCOPIC INTERLAYER TUNNELING

In the BBG-WSe$_2$ experiment, there is a relative angle $\theta$ between WSe$_2$ and BBG. In addition, the lattice constant of WSe$_2$ is $d = 0.331\text{nm}$ which is larger than the lattice constant of graphene, $a = 0.246\text{nm}$. These two factors are crucial for the microscopic interlayer tunneling, which are discussed in the section.

First of all, the interlayer tunneling preserves the crystal momentum because we assume that the tunneling between two positions depends only on their relative displacement. In our case, the Brillouin zones of WSe$_2$ and BBG are in different sizes, and WSe$_2$ is rotated by an angle $\theta$. In Fig. S2, the Brillouin zones for both the WSe$_2$ and BBG are illustrated. Within the two-center approximation scheme, the spin-preserving tunneling between two layers (from the $\beta$th band with sublattice $\sigma'$ to the $\alpha$ band with sublattice $\sigma$) can be described by [4]

$$T_{k,p}^{\alpha,\sigma;\beta,\sigma'} \equiv T(\alpha, \sigma, k; \beta, \sigma', p) = \frac{1}{\sqrt{\mathcal{A}_{0,A1}}} \left(\tilde{\epsilon}_p^{(1)}\right)^* \sum_{Q^{(0)},Q^{(1)}} \tilde{\epsilon}_{k+Q^{(1)}} e^{iQ^{(1)} \cdot p} e^{-iQ^{(0)} \cdot p} R_{\sigma}^{(1)} e^{-iQ^{(0)} \cdot p} \delta_{k+Q^{(1)},p+Q^{(0)}}$$

(S20)
relative to \( \Gamma \) point, the blue lattice is rotated by an angle \( \theta \), and \( k \) in infinitely many crystal momenta, but the contributions from the large crystal momentum \( k \) which can be modeled as a stretched exponentially decaying function in \( Q \) at layer \( l \) where \( A \). Figure S2. Brillouin zones of WSe\(_2\) and BBG. The green (blue) lines indicate the reciprocal lattice of WSe\(_2\) (BBG). The blue lattice is rotated by an angle \( \theta = 10^\circ \) in this illustration. \( G^{(0)}_1 = \left(-\frac{2\pi}{\sqrt{3d}}, \frac{2\pi}{\sqrt{3d}} \right) \), \( G^{(0)}_2 = \left(-\frac{2\pi}{\sqrt{3d}}, -\frac{2\pi}{\sqrt{3d}} \right) \), \( G^{(1)}_1 = \frac{2\pi}{\sqrt{3d}} \left(-\cos \theta - \frac{\sin \theta}{\sqrt{d}}, \frac{\cos \theta}{\sqrt{d}} - \sin \theta \right) \), \( G^{(1)}_2 = \frac{2\pi}{\sqrt{3d}} \left(-\cos \theta + \frac{\sin \theta}{\sqrt{d}}, -\frac{\cos \theta}{\sqrt{d}} - \sin \theta \right) \).

where \( A_0 \) (\( A_1 \)) is unit cell area of the WSe\(_2\) (graphene), \( \alpha \) and \( \beta \) are the band indexes, \( k \) and \( p' \) are the momentum relative to \( \Gamma \) point, and \( \sigma \) and \( \sigma' \) are the sublattice indexes, \( a^{(l)}_{k,\alpha,\sigma} \) is the sublattice projection of a Bloch state \( |k,\alpha\rangle \) at layer \( l \), \( R^{(l)}_\sigma \) is the sublattice vector of \( l \)th layer, \( Q^{(0)} \) and \( Q^{(1)} \) represent the reciprocal lattice vectors in WSe\(_2\) and BBG, respectively, and \( \tilde{t}_k \) is the 2D Fourier transform of the interlayer tunneling amplitude with a finite range. \( T(\alpha,\sigma,k;\beta,\sigma',p) \) indicates the components of both bands and sublattices. In the above expression, we use index \( 0 \) as the WSe\(_2\) layer and index \( 1 \) as the top graphene layer of BBG. We have ignored relative layer shift \( d \) for simplicity. \( Q^{(0)} = n_0 G^{(0)}_1 + m_0 G^{(0)}_2 \) and \( Q^{(1)} = n_1 G^{(1)}_1 + m_1 G^{(1)}_2 \) for integer values of \( n_0, m_0, n_1, m_1, \) and \( G^{(0)}_1, G^{(0)}_2, G^{(1)}_1, G^{(1)}_2 \) are the primitive lattice vectors for the reciprocal lattices as illustrated in Fig. S2. \( R^{(0)}_A = (0,0) \), \( R^{(0)}_B = (0,d/\sqrt{3}) \), \( R^{(1)}_A = (0,0) \), and \( R^{(1)}_B = (0,a/\sqrt{3}) \).

Equation (S20) describes the microscopic interlayer tunneling between two layers. The contributions come from infinitely many crystal momenta, but the contributions from the large crystal momentum \( k \) is suppressed by \( \tilde{t}_k \), which can be modeled as a stretched exponentially decaying function in \( k \) [4] (will be discussed later). In addition, the contribution is small if the momentum \( k \) is too far away from the valleys due to the structure of \( a^{(l)}_{k,\alpha,\sigma,\sigma'} \).

Our main purpose is to connect the microscopic calculations to the minimal model approach. We focus only on the tunneling between the \( d_{\pi,-\pi} \pm \pi d_{\pi} \) orbitals of WSe\(_2\) and the top graphene layer of BBG. Since our theory is based on electrons tunneling from WSe\(_2\) valence bands, only the 1st and 2nd bands in \( \tilde{h}^{(\text{WSe}_2)}(\mathbf{k}) \) [given by Eq. (S5)] are considered. With respect to BBG, only the 2nd and 4th bands in \( \tilde{h}^{(\text{BBG})}(\mathbf{k}) \) [given by Eq. (S1)] are considered because these two bands are relevant to the intervalley tunneling in our model. To simplify our calculations, we keep only \( K \) and \( -K \) momenta in BBG bands, and we consider a momentum cutoff \( \Lambda \) (relative to \( K \) or \( -K \)) for WSe\(_2\) bands. We define \( \mathbf{K}_0 = \frac{4\pi}{3d}(1,0) \) and \( \mathbf{K}_1 = \frac{4\pi}{3d}(\cos \theta, \sin \theta) \), where \( a = 0.246 \text{nm} \) (lattice constant of graphene) and \( d = 0.331 \text{nm} \) (lattice constant of WSe\(_2\)).

**ISOC and virtual-tunneling-induced attraction**

In the main text, we discuss ideas of deriving ISOC and the virtual-hopping-induced attraction. The same procedures can be done with the microscopic interlayer tunneling. The main differences is that multiple momenta states in WSe\(_2\) can contribute to the virtual processes instead of one. This is a consequence of the moiré structure of the BBG-WSe\(_2\) such that crystal momentum conservation can allow for tunnelings from extended zones.

We can formally compute ISOC based on second order perturbation theory of the microscopic interlayer tunneling.
and derive that
\[
\lambda_I = \sum_{q,|q|<\Lambda} \left[ \frac{|T(2,A,K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta} - \frac{|T(2,A,K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta} - \frac{|T(2,A,-K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta} + \frac{|T(2,A,-K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta} \right],
\]
where we have incorporated the potential energy difference \((2\Delta)\) between WSe\(_2\) and the top graphene layer (assuming the same distance as the top and bottom graphene layers) and \(\Lambda\) is the cutoff for the momentum sum. \(\Lambda \leq \frac{2\pi}{a}\) is required for consistently defining valleys. In this work, we use \(\Lambda = \frac{2\pi}{a} = 0.633\text{Å}^{-1}\). With the above expression and Eq. (S13), we identify that
\[
\frac{|V_{A1}|^2}{W} = \sum_{q,|q|<\Lambda} \frac{|T(2,A,K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta},
\]
\[
\frac{|V_{A1}|^2}{W + \delta} = \sum_{q,|q|<\Lambda} \frac{|T(2,A,K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta},
\]
\[
\frac{|V_{A1}|^2}{W} = \sum_{q,|q|<\Lambda} \frac{|T(2,A,-K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta},
\]
\[
\frac{|V_{A1}|^2}{W + \delta} = \sum_{q,|q|<\Lambda} \frac{|T(2,A,-K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta}.
\]
Similarly, we can compute the virtual-tunneling-induced attraction and derive that
\[
U_{\text{eff}} = 2 \sum_{q,|q|<\Lambda} \left[ \frac{|T(4,B,K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta+E_B} + \frac{|T(4,B,-K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta+E_B} + \frac{|T(4,B,K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta+E_B} + \frac{|T(4,B,-K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta+E_B} \right],
\]
With the above expression and Eq (S19), we identify that
\[
\frac{|V_{B1}|^2}{W + E_B - xU_1} = \sum_{q,|q|<\Lambda} \frac{|T(4,B,K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta+E_B - xU_1},
\]
\[
\frac{|V_{B1}|^2}{W + \delta + E_B - xU_1} = \sum_{q,|q|<\Lambda} \frac{|T(4,B,K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta+E_B - xU_1},
\]
\[
\frac{|V_{B1}|^2}{W + E_B - xU_1} = \sum_{q,|q|<\Lambda} \frac{|T(4,B,-K_1;2,B,K_0+q)|^2}{-E_2^{(\text{WSe}_2)}(q)+2\Delta+E_B - xU_1},
\]
\[
\frac{|V_{B1}|^2}{W + \delta + E_B - xU_1} = \sum_{q,|q|<\Lambda} \frac{|T(4,B,-K_1;1,B,K_0+q)|^2}{-E_1^{(\text{WSe}_2)}(q)+2\Delta+E_B - xU_1},
\]
where \(x = 0, 1, 2\).
Evaluating interlayer tunneling

To compute $T(\alpha, \sigma, \mathbf{k}; \beta, \sigma', \mathbf{p})$, $a^{(l)}_{\mathbf{k}_0, \alpha, \sigma}$ and $\tilde{t}_k$ are needed. $a^{(l)}_{\mathbf{k}_0, \alpha, \sigma}$ corresponds to the wavefunctions of $\hat{h}_{\text{WS}2}(\mathbf{k})$ and $\hat{h}_r(\mathbf{k})$, given by Eqs. (S5) and (S4). We find

$$
 a^{(0)}_{\mathbf{k}_0 + \mathbf{q}, 1, B} = \frac{1}{\sqrt{1 + \left[\frac{2v|q|}{\Omega + \lambda + \sqrt{4v^2|q|^2 + (\Omega + \lambda)^2}}\right]^2}}, \quad a^{(0)}_{\mathbf{k}_0 + \mathbf{q}, 2, B} = \frac{1}{\sqrt{1 + \left[\frac{2v|q|}{\Omega - \lambda + \sqrt{4v^2|q|^2 + (\Omega - \lambda)^2}}\right]^2}}, \quad a^{(1)}_{\pm, 1, 2, A} = 1, \quad a^{(1)}_{\pm, 1, 4, B} = \frac{1}{\sqrt{2}}.
$$

(S25)

Based on the wavefunction amplitudes above, we conclude that $|T(2, A, \mathbf{K}_1; 1, B, \mathbf{K}_0 + \mathbf{q})| < |T(2, A, \mathbf{K}_1; 2, B, \mathbf{K}_0 + \mathbf{q})|$ and $|T(4, B, \mathbf{K}_1; 2, B, \mathbf{K}_0 + \mathbf{q})| < |T(4, B, \mathbf{K}_1; 1, B, \mathbf{K}_0 + \mathbf{q})|$, but the difference might be small. In addition, $|T(4, B, \mathbf{K}_1; 1, B, \mathbf{K}_0 + \mathbf{q})| = \frac{1}{\sqrt{2}} |T(2, A, \mathbf{K}_1; 1, B, \mathbf{K}_0 + \mathbf{q})|$ and $|T(4, B, \mathbf{K}_1; 2, B, \mathbf{K}_0 + \mathbf{q})| = \frac{1}{\sqrt{2}} |T(2, A, \mathbf{K}_1; 2, B, \mathbf{K}_0 + \mathbf{q})|$. 

Regarding $\tilde{t}_k$, we follow Ref. [4] and adopt a stretched exponential ansatz as follows:

$$
|\tilde{t}_k| = t_0 e^{-\chi(|k|z_\perp)^\gamma},
$$

(S26)

where $t_0$ is an overall constant, $\chi$ is an order 1 numerical constant, $\gamma$ is the exponent of stretched exponential, and $z_\perp$ is the distance between WS2 monolayer and the top graphene layer of BBG. We use $z_\perp = 3.34\text{Å}$, the same as the distance between two graphene layers. $\chi$ and $\gamma$ control the decay of $\tilde{t}_k$ for a sufficiently large $k$, and we will discuss the results with a few representative values of $\chi$ and $\gamma$ in addition to the twist angle $\theta$. Since $\tilde{t}_k$ becomes small for a sufficiently large $k$, we can ignore contributions such that $k + \mathbf{Q}^{(1)} > \Lambda$ in Eq. (S20), where we use $\Lambda = \frac{2\pi}{a} = 5.11\text{Å}^{-1}$.

In Fig. S3, the kinematics of momenta involved in tunneling is plotted for a few values of $\theta$. The big black dashed circle draws the region of $k + \mathbf{Q}^{(1)} < \Lambda$ considered in the tunneling. We also use the small gray dashed circles to mark the momentum cutoff $\Lambda$ for momentum relative to $\bar{K}$ valley. The blue circles indicate finite contributions.
of intravalley processes; the yellow circles indicate finite contributions of intervalley processes. Now, we are in the position to compute interlayer tunneling between WSe$_2$ and BBG.

$$T(\alpha, \sigma, k; \beta, \sigma', p) \text{ at different angles}$$

The interlayer tunneling matrix element crucially depends on the twist angle $\theta$. In Fig. S3, we plot crystal momenta involved in the interlayer tunneling with different $\theta$. We will list all the needed tunneling matrix elements for $0 \leq \theta \leq 30^\circ$. In our model, angle $60^\circ - \theta$ is related to $\theta$, but the intravalley and intervalley contributions are interchanged. See the kinematics in Figs. S3(f) and S3(h) for an example.

$\theta = 0$

For $\theta = 0$, there are three momenta contributing to the intravalley tunnelings and six momenta contributing to the intervalley tunnelings. Since the three intravalley tunnelings are related by the $C_{3z}$ rotation, we only need to evaluate one of them. The six intervalley tunneling momenta are also related by symmetry operation, so only one of them needs to be evaluated. The spin-dependent interlayer tunneling terms are as follows:

$$\left| T(2, A, K_1; 2, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega - \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right]^2},$$  \hspace{1cm} (S27a)

$$\left| T(2, A, K_1; 1, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega + \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega + \lambda)^2}} \right]^2},$$  \hspace{1cm} (S27b)

$$\left| T(2, A, -K_1; 2, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega - \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right]^2},$$  \hspace{1cm} (S27c)

$$\left| T(2, A, -K_1; 1, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega + \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega + \lambda)^2}} \right]^2},$$  \hspace{1cm} (S27d)

where $|K_1 - K_0| = 0.437\AA^{-1}$, $\Delta k = |K_1 - 2G_2^{(1)} - K_0 - G_1^{(0)} + 2G_2^{(0)}| = 0.416\AA^{-1}$ and $K = |K_1 - 2G_2^{(1)}| = 4.51\AA^{-1}$.

$\theta = 5^\circ$

For $\theta = 5^\circ$, there are only two distinct momenta. The spin-dependent interlayer tunneling terms are as follows:

$$\left| T(2, A, K_1; 2, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega - \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right]^2},$$  \hspace{1cm} (S28a)

$$\left| T(2, A, K_1; 1, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega + \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega + \lambda)^2}} \right]^2},$$  \hspace{1cm} (S28b)

$$\left| T(2, A, -K_1; 2, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega - \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right]^2},$$  \hspace{1cm} (S28c)

$$\left| T(2, A, -K_1; 1, B, K_0 + q) \right|^2 \approx \frac{t_0^2}{A_0 A_1} \frac{e^{-2\chi(|K_1|z_\perp)}}{1 + \left[ \frac{2\pi|K_1 - K_0|}{\Omega + \lambda + \sqrt{4\pi^2|K_1 - K_0|^2 + (\Omega + \lambda)^2}} \right]^2},$$  \hspace{1cm} (S28d)
where $|K_1 - K_0| = 0.456\,\text{Å}^{-1}$, $\Delta k = | - K_1 - 2G_2^{(1)} - K_0 - G_2^{(0)} + 2G_1^{(0)} | = 0.06\,\text{Å}^{-1}$, and $\mathcal{K} = | - K_1 - 2G_2^{(1)} | = 4.51\,\text{Å}^{-1}$. Since the intervalley processes have a very small $\Delta k$, the intervalley contribution can be large. This explains the weak nonmonotonicity in ISOC an effective attraction for $\gamma = 1$ and $\chi = 0.2$.

$\theta = 10^\circ$

For $\theta = 10^\circ$, there are three distinct momenta. The spin-dependent interlayer tunneling terms are as follows:

$$
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S29a)

$$
|T(2, A, K_1; 1, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S29b)

$$
|T(2, A, -K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S29c)

$$
|T(2, A, -K_1; 1, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S29d)

$$
|T(2, A, -K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S29e)

$$
|T(2, A, -K_1; 1, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S29f)

where $|K_1 - K_0| = 0.507\,\text{Å}^{-1}$, $\Delta k = | - K_1 - G_2^{(1)} - G_2^{(2)} - K_0 + G_2^{(0)} | = 0.539\,\text{Å}^{-1}$, $\mathcal{K} = | - K_1 - G_1^{(1)} - G_2^{(1)} | = 3.41\,\text{Å}^{-1}$, $\Delta k' = | - K_1 - 2G_2^{(1)} - K_0 - G_1^{(0)} + 2G_2^{(0)} | = 0.383\,\text{Å}^{-1}$, and $\mathcal{K'} = | - K_1 - 2G_2^{(1)} | = 4.51\,\text{Å}^{-1}$.

$\theta = 15^\circ$

For $\theta = 15^\circ$, there are two distinct momenta. The spin-dependent interlayer tunneling terms are as follows:

$$
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S30a)

$$
|T(2, A, K_1; 1, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S30b)

$$
|T(2, A, -K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S30c)

$$
|T(2, A, -K_1; 1, B, K_0 + q)|^2 \approx \frac{t_0^2}{\mathcal{A}_0 \mathcal{A}_1} \frac{e^{-2\chi(|K_1| z_\perp)}\gamma}{1 + \left( \frac{2\chi|K_1 - K_0|}{\Omega - \lambda + \gamma + \sqrt{4\chi|K_1 - K_0|^2 + (\Omega - \lambda)^2}} \right)^2},
$$

(S30d)

where $|K_1 - K_0| = 0.581\,\text{Å}^{-1}$, $\Delta k = | - K_1 - G_1^{(1)} - G_2^{(2)} - K_0 + G_2^{(0)} | = 0.249\,\text{Å}^{-1}$, and $\mathcal{K} = | - K_1 - G_1^{(1)} - G_2^{(1)} | = 3.41\,\text{Å}^{-1}$.
\( \theta = 20^\circ \)

For \( \theta = 20^\circ \), there are three distinct momenta. The spin-dependent interlayer tunneling terms are as follows:

\[
|T(2, A, -K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k}{\Omega - \lambda + \sqrt{4\epsilon^2 \Delta k^2 + (\Omega - \lambda)^2}}} 2^1. \tag{S31a}
\]

\[
|T(2, A, -K_1; 1, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k'}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k'^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S31b}
\]

\[
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k'}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k'^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S31c}
\]

\[
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k'}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k'^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S31d}
\]

\[
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k''}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k''^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S31e}
\]

\[
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k''}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k''^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S31f}
\]

where \( \Delta k = |K_1 - G_1^{(1)} - G_2^{(1)} - K_0 + G_2^{(0)}| = 0.7788 \text{Å}^{-1} \), \( K = |K_1 - G_1^{(1)} - G_2^{(1)}| = 3.41 \text{Å}^{-1} \), \( \Delta k' = |K_1 - G_1^{(1)} - K_0 + G_1^{(0)} + G_2^{(0)}| = 0.562 \text{Å}^{-1} \), \( K' = |K_1 - G_1^{(1)}| = 4.51 \text{Å}^{-1} \), \( \Delta k'' = |K_1 - G_2^{(1)} - K_0 - G_1^{(0)} + 2G_2^{(0)}| = 0.556 \text{Å}^{-1} \), \( K'' = |K_1 - G_2^{(1)}| = 4.51 \text{Å}^{-1} \). At \( \theta = 20^\circ \), the intervalley processes have smaller crystal momentum than the intravalley processes. In addition, \( \Delta k \) for the interlayer tunneling is quite small. This is why we find \( \lambda_1 < 0 \) at \( \theta = 20^\circ \). Another important point is that the tunneling crystal momentum is no longer at the first Brillouin zone corner (i.e., \( K \) or \( -K \)). This explains the small effective attraction for \( \theta \geq 20^\circ \).

\( \theta = 25^\circ \)

For \( \theta = 25^\circ \), there are two distinct momenta. The spin-dependent interlayer tunneling terms are as follows:

\[
|T(2, A, -K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k}{\Omega - \lambda + \sqrt{4\epsilon^2 \Delta k^2 + (\Omega - \lambda)^2}}} 2^1. \tag{S32a}
\]

\[
|T(2, A, -K_1; 1, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k'}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k'^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S32b}
\]

\[
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k'}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k'^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S32c}
\]

\[
|T(2, A, K_1; 2, B, K_0 + q)|^2 \approx \frac{t_0^2}{A_0 A_1} 1 + \frac{e^{-2\chi(K_{z+}|)|}}{1 + \frac{2\Delta k''}{\Omega + \lambda + \sqrt{4\epsilon^2 \Delta k''^2 + (\Omega + \lambda)^2}}} 2^1. \tag{S32d}
\]
λ contributions are stronger than intravalley contributions, and we plot \( \tilde{\mathcal{G}} \) for different values of \( \theta \) in Fig. S4. Effective attraction for different values of \( U \). We use \( \gamma = 1.25 \) and \( \chi = 0.2 \) in all the curves.

where \( \Delta k = \mid K_1 - G_1^{(1)} - G_2^{(1)} - K_0 + G_2^{(0)} \mid = 0.352 \text{Å}^{-1}, \ K = \mid K_1 - G_1^{(1)} - G_2^{(2)} \mid = 3.41 \text{Å}^{-1}, \ \Delta k' = \mid K_1 - G_2^{(1)} - K_0 - G_1^{(0)} + 2G_2^{(0)} \mid = 0.168 \text{Å}^{-1}, \ K'' = \mid K_1 - G_2^{(2)} \mid = 4.51 \text{Å}^{-1}. \) Similar to \( \theta = 20^\circ \), the intervalley contributions are stronger than intravalley contributions, and \( \lambda_I < 0 \) (albeit vanishingly small in some cases).

\[
\theta = 30^\circ
\]

For \( \theta = 30^\circ \), there is actually only one distinct momentum, but we check both intravalley and intervalley processes. The spin-dependent interlayer tunneling terms are as follows:

\[
\begin{align*}
&T(2, A, -K_1; 2, B, K_0 + q) \approx \frac{t_0^2 A_0 A_1}{1 + e^{-2\lambda (K_2)}} \left( 1 - \frac{2v\Delta k}{\Omega + \lambda + \sqrt{4v^2\Delta k^2 + (\Omega - \lambda)^2}} \right)^2, \\
&T(2, A, -K_1; 1, B, K_0) \approx \frac{t_0^2 A_0 A_1}{1 + e^{-2\lambda (K_2)}} \left( 1 - \frac{2v\Delta k}{\Omega + \lambda + \sqrt{4v^2\Delta k^2 + (\Omega - \lambda)^2}} \right)^2, \\
&T(2, A, K_1; 2, B, K_0 + q) \approx \frac{t_0^2 A_0 A_1}{1 + e^{-2\lambda (K_2)}} \left( 1 - \frac{2v\Delta k' \gamma}{\Omega + \lambda + \sqrt{4v^2\Delta k'^2 + (\Omega - \lambda)^2}} \right)^2, \\
&T(2, A, K_1; 2, B, K_0) \approx \frac{t_0^2 A_0 A_1}{1 + e^{-2\lambda (K_2)}} \left( 1 - \frac{2v\Delta k' \gamma}{\Omega + \lambda + \sqrt{4v^2\Delta k'^2 + (\Omega - \lambda)^2}} \right)^2,
\end{align*}
\]

where \( \Delta k = \mid q = -K_1 - 2G_2^{(1)} - K_0 - 2G_2^{(0)} \mid = 0.245 \text{Å}^{-1}, \ K = \mid K_1 - 2G_2^{(1)} \mid = 4.51 \text{Å}^{-1}, \ \Delta k' = \mid K_1 - G_2^{(1)} - K_0 - G_1^{(0)} + 2G_2^{(0)} \mid = 0.245 \text{Å}^{-1}, \ K'' = \mid K_1 - G_2^{(2)} \mid = 4.51 \text{Å}^{-1}. \)

Effective attraction versus \( U \)

Here, we discuss \( U \) (short-range Coulomb interaction) dependence in the effective attraction strength \( \mathcal{U}_{\text{eff}} \). In Fig. S4, we plot \( \mathcal{U}_{\text{eff}} = \mathcal{U}_{\text{eff}} A_0 A_1 eV/t_0^2 \) (the dimensionless attraction) as a function of \( \theta \) with (a) \( \gamma = 1, \ \chi = 0.2 \) and (b) \( \gamma = 1.25, \ \chi = 0.2 \), and different values of \( U \) ranging from 0.2eV to 0.43eV. \( U \leq E_B \) in our theory, where \( E_B \approx 0.43eV. \) The qualitative trends are very similar because the values of \( U \) do not generate extremely small denominators in second order perturbation theory. However, the results show quantitative dependence in \( U \), and a larger \( U \) generally enhances \( \mathcal{U}_{\text{eff}} \).

* yzhou@umd.edu
[1] J. Jung and A. H. MacDonald, Phys. Rev. B 89, 035405 (2014), URL https://link.aps.org/doi/10.1103/PhysRevB.89.035405.

[2] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao, Phys. Rev. Lett. 108, 196802 (2012), URL https://link.aps.org/doi/10.1103/PhysRevLett.108.196802.

[3] A. David, P. Rakyta, A. Kormányos, and G. Burkard, Phys. Rev. B 100, 085412 (2019), URL https://link.aps.org/doi/10.1103/PhysRevB.100.085412.

[4] R. Bistritzer and A. H. MacDonald, Phys. Rev. B 81, 245412 (2010), URL https://link.aps.org/doi/10.1103/PhysRevB.81.245412.