Scaling Universality between Band Gap and Exciton Binding Energy of Two-Dimensional Semiconductors

Zeyu Jiang,[1] Zhirong Liu,[2,*] Yuanchang Li,[3,†,‡] and Wenhui Duan[1,4,§]
[1]State Key Laboratory of Low-Dimensional Quantum Physics and Collaborative Innovation Center of Quantum Matter, Department of Physics, Tsinghua University, Beijing 100084, China
[2]College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China
[3]CAS Center for Excellence in Nanoscience, National Center for Nanoscience and Technology, Beijing 100190, China
[4]Institute for Advanced Study, Tsinghua University, Beijing 100084, China
(Received 10 November 2016; revised manuscript received 12 April 2017; published 27 June 2017)

Using first-principles GW Bethe-Salpeter equation calculations and the $\mathbf{k}\cdot\mathbf{p}$ theory, we unambiguously show that for two-dimensional (2D) semiconductors, there exists a robust linear scaling law between the quasiparticle band gap (E_g) and the exciton binding energy (E_b), namely, $E_b \approx E_g/4$, regardless of their lattice configuration, bonding characteristic, as well as the topological property. Such a parameter-free universality is never observed in their three-dimensional counterparts. By deriving a simple expression for the 2D polarizability merely with respect to E_g, and adopting the screened hydrogen model for E_b, the linear scaling law can be deduced analytically. This work provides an opportunity to better understand the fantastic consequence of the 2D nature for materials, and thus offers valuable guidance for their property modulation and performance control.

DOI: 10.1103/PhysRevLett.118.266401

Parameter-free universal phenomena are amazing while rare in the material science, which are usually associated with profound physics behind, e.g., the topology for quantized conductance [1,2], the fine structure constant for opacity [3], and the self-organized criticality for $1/f$ noise [4]. Ever since the discovery of graphene, more and more two-dimensional (2D) materials have been fabricated, and they exhibit a lot of novel properties that are distinctly different from those of their three-dimensional (3D) counterparts, potentially acting as the host for the parameter-free universality. An example is the recently reported linear scaling law [5] between the band gap (E_g) and the exciton binding energy (E_b). However, a nonzero intercept up to 0.4 eV was revealed there which significantly shook the validity of the finding because it means an E_b of 0.4 eV for a gapless system. A following study [6] thus claimed that such a linear scaling law only holds for 2D materials with large E_g and breaks down when $E_g < 2$ eV.

For semiconductors, E_g and E_b are two fundamental parameters critical to their applications in electronics, optics, and energy harvest. Generally, these two quantities are considered as a function of the carrier effective mass (m^*) and dielectric constant (ε). Because of the large screening effect in the 3D semiconductors, it appears that the m^* is more crucial for the material properties, and the universal linear relation between m^* and E_g is experimentally established [7], regardless of the ε. By contrast, the role of ε becomes dominant in the 2D cases, which itself solely determines the E_b, independent of the m^* over a wide range of polarizabilities [8]. Then a key to answer whether the linear scaling law holds for 2D materials lies in the relationship between E_g and ε. Moreover, the interplay between m^* and ε is of special interest for the understanding of the essential difference between 2D and 3D nature.

In this work, we first identify the linear scaling law between E_g and E_b by conducting a series of first-principles GW Bethe-Salpeter equation (BSE) calculations on a variety of 2D semiconductors, paying particular attention to those with the E_g almost uniformly distributed in the range 0–8 eV. We find a robust linear relation between them with an intercept of exact zero, thus resolving the aforementioned disagreement between Refs. [5,6]. Based on the $\mathbf{k}\cdot\mathbf{p}$ approach, we then establish a universal expression of E_g solely in terms of 2D polarizability, and therefore analytically obtain the linear relation between E_b and E_g. Finally, we show that the m^* exhibits neither the linear nor any clear relation with E_g or E_b in the 2D situation, revealing a substantial difference from the 3D materials.

Our calculations are carried out in the framework of DFT-GW-BSE [9,10] scheme by VASP package [11] with the projector-augmented wave method (PAW) and the Perdew-Burke-Ernzerhof (PBE) [12] exchange correlation functional. A vacuum of 25 Å is used to minimize the interlayer coupling. The geometric structures are fully relaxed until forces on atoms are less than 0.005 eV/Å, and the k-point meshes and cutoff energy for plane wave basis are tested until the energy change is less than 0.001 eV. In the GW step, a single-shot G_0W_0 approach is implemented to obtain the band gaps; the number of empty bands and the cutoff energy for response function are tested to ensure that the gaps are converged with an accuracy of 0.01 eV. The details of parameters used in the *GW*-BSE calculations can be found in Table S1 of the Supplemental Material [13], given their critical role in reaching convergence in *GW* and BSE calculations [14].

We perform the calculations on 23 2D semiconductors, whose E_b is plotted as a function of the corresponding quasiparticle E_g in Fig. 1. Remarkably, these two distinct physical quantities, which were seldom thought to be relevant in the 3D materials, exhibit a robust linear scaling law in a wide range of E_g (varying between 0 and 8 eV). Through the numerical fitting, we obtain a slope of 0.27 and an intercept of zero that is consistent with the conventional knowledge on exciton formation. It is worth emphasizing that these 23 materials cover almost all kinds of popular 2D monolayer semiconductors including transition metal dichalcogenides (TMDs), graphene derivatives, III–V compounds, black phosphorus, transition metal carbides, and nitrides (MXenes)[15] and even topological materials [16–18]. They are essentially distinct in various attributes, ranging from the lattice configuration to the bonding characteristic and to the gap origin. Note that the

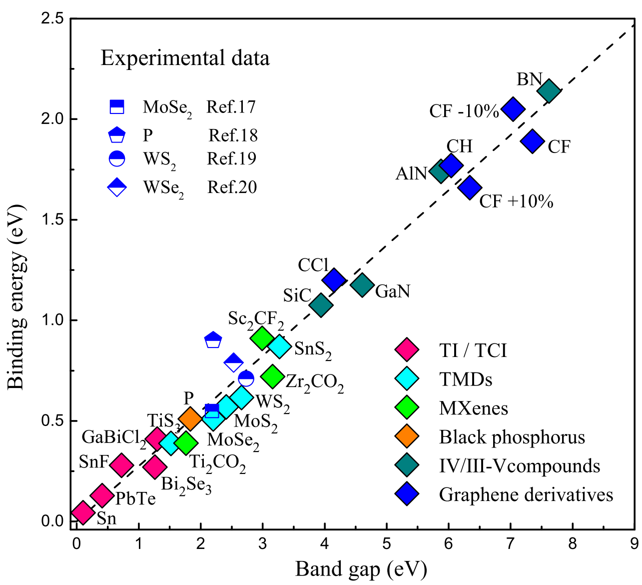

FIG. 1. Linear relationship between E_g and E_b. Calculations cover almost all the popular 2D monolayer semiconductors including TMDs [monolayer MoS_2, $MoSe_2$, WS_2, SnS_2, and TiS_3], graphene derivatives [graphane (CH), chlorinated graphene (CCl), fluorinated graphene (CF), and fluorinated graphene under external field with 10% pressure (CF − 10%) or 10% tension (CF + 10%)], IV/III–V compounds [SiC, GaN, AlN, and BN in the honeycomb lattice], black phosphorus, functionalized MXenes [15] [Ti_2CO_2, Zr_2CO_2, and Sc_2CF_2] and topological materials Sn [16], SnF [16], PbTe [17], single-quintuple layer Bi_2Se_3, and $GaBiCl_2$ [18]. The exciton is determined by the lowest peak in absorption spectrum with the only exception of SnS_2 where the first dark exciton is used. The "minimum direct band gap" is used for the indirect gap materials.

linear law is not only applicable for the ground state of materials but also valid under external field (such as strain, either a compression or a tension type), as demonstrated by an example of the fluorinated graphene. More importantly, this finding is experimentally supported. In Fig. 1, we also incorporate the experimental results [19–22] reported so far. It can be seen that all of them are well located around the fitted line. Therefore, such an exotic linear scaling law is likely universal for all the 2D monolayer semiconductors, neither dependent upon any material parameter nor the fundamental constants.

It is worth noting that an earlier work [5] also demonstrated the existence of a linear law between E_b and E_g. However, the nonzero intercept seems physically unreasonable [6]. In the study presented here, by examining more diverse systems, especially those with a small gap, we achieve a zero intercept, which is no longer in conflict with the fundamental knowledge of exciton. By careful comparison, we find the data presented in Ref. [5] are actually fully compatible with ours, and the observed discrepancy is just due to the lack of enough data therein.

There are rare universal phenomena in condensed matter physics and materials science that do not depend on material parameters. A well-known example in the new era of 2D materials is the opacity of graphene and other Dirac materials [3], which is defined only by the fine structure constant and does not depend on material parameters or external conditions. The revealed linear relation between E_b and E_g here, which is not only independent of parameters of material, but even independent of any fundamental constants, adds a new example to the universal phenomena.

We now turn to the origin of this universal linear behavior. Considering the complicacy of *GW*-BSE method, it seems impossible to straightforwardly obtain a relation between E_g and E_b. With an alternative strategy, we seek answer in simplified analytical models, i.e., the conventional $\mathbf{k}\cdot\mathbf{p}$ perturbation theory and hydrogenlike model which are frequently used to describe E_g and E_b in semiconductors.

The key to understand the observed linear law lies in the dielectric constant ε [5,6]. For a hydrogen atom without dielectric screening ($\varepsilon = 1$), the characteristic size (Bohr radius) is about 0.53 Å. Nevertheless, previous studies [19,23–25] indicated that the magnitude of a 2D exciton radius is typically at the level of several angstroms or even larger. This reveals that the screening effect still plays a role in 2D materials even if they have ultrathin thickness. Different from that in 3D cases, the dielectric environment in 2D semiconductors is highly inhomogeneous because the charge polarization is confined solely in the material plane [26]. As a consequence, only the in-plane electric field distribution in material is affected [27], leading to the strong in-plane screening while no screening exists in vacuum (outside the plane).

In principle, the 2D static dielectric function takes the form of $\varepsilon(\mathbf{q}) = 1 + 2\pi\alpha_{2D}|\mathbf{q}|$, where α_{2D} is the 2D polarizability that can be calculated from first principles [28]. A handy definition of α_{2D} is given as

$$\alpha_{2D} = L\frac{\varepsilon_{3D} - 1}{4\pi}, \tag{1}$$

where L and ε_{3D} are, respectively, the supercell thickness and macroscopic dielectric constant. In the random-phase approximation (RPA) approach, the microscopic frequency-dependent dielectric function is given as [28]

$$\varepsilon_{\mathbf{GG}'}(\mathbf{q},\omega) = \delta_{\mathbf{GG}'} + 2\frac{4\pi e^2}{V}\frac{1}{|\mathbf{q}+\mathbf{G}|}\frac{1}{|\mathbf{q}+\mathbf{G}'|}\sum_{\mathbf{k}}\sum_{c,v}\langle v,\mathbf{k}|e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|c,\mathbf{k}+\mathbf{q}\rangle\langle v,\mathbf{k}|e^{-i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}}|c,\mathbf{k}+\mathbf{q}\rangle^* \times\left(\frac{1}{E_{c,\mathbf{k}+\mathbf{q}} - E_{v,\mathbf{k}} - \omega + i0^+} + \frac{1}{E_{c,\mathbf{k}+\mathbf{q}} - E_{v,\mathbf{k}} + \omega + i0^+}\right), \tag{2}$$

where c and v refer to empty and occupied bands, respectively. Without the local field effect, ε_{3D} can be calculated by taking the macroscopic and static limit of Eq. (2),

$$\begin{aligned}\varepsilon_{3D} &= \lim_{\mathbf{q}\to 0}\varepsilon_{00}(\mathbf{q},\omega=0)\\ &= 1 + \lim_{\mathbf{q}\to 0}\frac{16\pi e^2}{V}\frac{1}{|\mathbf{q}|^2}\times\sum_{\mathbf{k}}\sum_{c,v}|\langle v,\mathbf{k}|e^{-i\mathbf{q}\cdot\mathbf{r}}|c,\mathbf{k}+\mathbf{q}\rangle|^2\\ &\quad\times\frac{1}{E_{c,\mathbf{k}+\mathbf{q}} - E_{v,\mathbf{k}}},\end{aligned} \tag{3}$$

where V is the volume of the 3D supercell. The correction caused by nonzero ω is small (see more details in the Supplemental Material Fig. S1 [13]), not affecting the conclusion of the paper.

In the Bloch theory, the eigenfunctions are written as periodically modulated plane waves, leading to

$$\begin{aligned}&\langle v,\mathbf{k}|e^{-i\mathbf{q}\cdot\mathbf{r}}|c,\mathbf{k}+\mathbf{q}\rangle\\ &\quad= \langle u_{v,\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}}|e^{-i\mathbf{q}\cdot\mathbf{r}}|u_{c,\mathbf{k}+\mathbf{q}}e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}}\rangle = \langle u_{v,\mathbf{k}}|u_{c,\mathbf{k}+\mathbf{q}}\rangle.\end{aligned} \tag{4}$$

And in the $\mathbf{q}\to 0$ limit, by means of Taylor series over $\mathbf{q}$, we have

$$\langle u_{v,\mathbf{k}}|u_{c,\mathbf{k}+\mathbf{q}}\rangle = \mathbf{q}\cdot\langle u_{v,\mathbf{k}}|\nabla_{\mathbf{k}}|u_{c,\mathbf{k}}\rangle. \tag{5}$$

So α_{2D} is finally expressed as (a $1/2$ factor is introduced given the tensor nature of α_{2D})

$$\alpha_{2D} = \frac{2e^2}{S}\sum_{\mathbf{k}}\sum_{c,v}\frac{|\langle u_{c,\mathbf{k}}|\nabla_{\mathbf{k}}|u_{v,\mathbf{k}}\rangle|^2}{E_{c,\mathbf{k}} - E_{v,\mathbf{k}}}, \tag{6}$$

where S is the surface area of the 2D supercell, and $\mathbf{k}$ is summed over the 2D reciprocal plane. Equation (6) can also be rewritten in an integral form,

$$\alpha_{2D} = \frac{2e^2}{(2\pi)^2}\sum_{c,v}\int_{\mathbf{k}}\frac{|\langle u_{c,\mathbf{k}}|\nabla_{\mathbf{k}}|u_{v,\mathbf{k}}\rangle|^2}{E_{c,\mathbf{k}} - E_{v,\mathbf{k}}}d^2\mathbf{k}. \tag{7}$$

Borrowing the results from Ref. [29],

$$|\langle u_{c,\mathbf{k}}|\nabla_{\mathbf{k}}|u_{u,\mathbf{k}}\rangle|^2 = \frac{1}{(E_{v,\mathbf{k}} - E_{c,\mathbf{k}})^2}\left|\langle u_{c,\mathbf{k}}|\frac{\partial H_{\mathbf{k}}}{\partial\mathbf{k}}|u_{v,\mathbf{k}}\rangle\right|^2, \tag{8}$$

we have

$$|\langle u_{c,\mathbf{k}}|\nabla_{\mathbf{k}}|u_{v,\mathbf{k}}\rangle|^2 = \frac{\hbar^2}{2\mu}\frac{1}{(E_g + \frac{\hbar^2k^2}{2\mu})}, \tag{9}$$

under the $\mathbf{k}\cdot\mathbf{p}$ perturbation theory and an effective mass approximation, when considering only the contributions from the highest valence band and lowest conduction band. It is noted that the deduction does not depend on any specific Hamiltonian, and the details can be found in the Supplemental Material [13]. Here, μ is the reduced mass which is assumed to be independent of $\mathbf{k}$. Substituting Eq. (9) into Eq. (7), we get

$$\alpha_{2D} = \frac{e^2}{2\pi}\left(-\frac{1}{E_g + x}\right)\Bigg|_0^{x_M=\frac{\hbar^2K_M^2}{2\mu}}, \tag{10}$$

where K_M denotes the maximal k at the boundary of the first Brillouin zone. Usually x_M is much larger than E_g (herein the typical value of x_M is around several tens of eV). This implies that we may take the upper limit in Eq. (10) to be infinite, which leads to a compact formula for α_{2D},

$$\alpha_{2D} = \frac{e^2}{2\pi E_g}. \tag{11}$$

Equation (11) only accounts for the contribution from states near one gap. If there are more than one equivalent gaps, e.g., for MoS_2 or Sn, there are two gaps located at K and K', respectively, the results should be multiplied by the number of band gaps (N_g), i.e.,

$$\alpha_{2D} = \frac{N_g e^2}{2\pi E_g}. \tag{12}$$

For Sn system, which has a direct gap as small as 0.1 eV with $N_g = 2$, Eq. (12) predicts $\alpha_{2D} = 46$ Å, close to the numerical GW calculation of 56 Å. If the contribution from an extra larger gap of 0.3 eV at Γ is included, Eq. (12) increases the value to 53 Å, nearly identical to the GW result.

To relate α_{2D} to E_b, we adopted the screened hydrogen model by Olsen *et al.* [8] and obtained E_b as

$$E_b = \frac{8\mu e^4}{\hbar^2(1 + \sqrt{1 + \frac{32\pi e^2}{3\hbar^2}\alpha_{2D}\mu})^2}, \tag{13}$$

where the nonlocal effect of the 2D dielectric constant in the standard hydrogen model is described by averaging the screening over the extension of the exciton. When α_{2D} is large enough, it can be further simplified into

$$E_b = \frac{3e^2}{4\pi\alpha_{2D}}, \tag{14}$$

which is also independent of the reduced mass. Combining Eq. (12) with (14), we immediately obtain a simple linear relation between E_g and E_b,

$$E_b = \frac{3}{2N_g}E_g. \tag{15}$$

The value of N_g is essential for the linear slope between E_b and E_g. For another small E_g system, SnF, the band gaps are located along $\Gamma - K$, and $N_g = 6$ due to the hexagonal symmetry. Inserting the value of N_g into Eq. (15) gives a simple expression of

$$E_b = \frac{E_g}{4}. \tag{16}$$

However, for general cases, the situation gets complicated due to the two approximations adopted during the calculation of Eq. (11). Firstly, we employ the two-band model which omits the contributions from the other bands, hence usually yielding the underestimation of α_{2D} (see more details in the Supplemental Material [13]). For systems with small E_g, such as Sn and SnF, the screening is indeed contributed dominantly from the band edge bands, while for systems with moderate or large gaps, the contributions from other bands are nonignorable (see Fig. S2). Secondly, the μ is assumed to be independent of k, which neglects the complicated band dispersion. For example, a gapped Dirac Hamiltonian (where the μ implicitly depends on k) gives a result slightly different from Eq. (11) (see the Supplemental Material [13]). Taking the upper limit in Eq. (10) to be infinite also introduces discrepancy (see the Supplemental Material [13]). Therefore, the observed simple linear relations of E_b vs E_g and α_{2D} vs $1/E_g$ are consequence of various factors. Actually, our numerical GW-BSE results suggest that the influence can be equivalently represented by $N_g = 6$ in the simplified analysis. To further clarify this point, we present E_g and E_b [from the GW-BSE calculations and theoretical Eq. (12) with $N_g = 6$ and Eq. (14)] as a function of α_{2D} in Fig. 2. It can be seen that the theoretical predictions are in excellent agreement with GW-BSE calculations when E_g is smaller than 4 eV. However, both E_g and E_b deviate from the predictions under large gaps, although their deviations cancel out and the linear relation between E_g and E_b is well kept in the whole energy range considered (as shown in Fig. 1). The origin of the mysterious cancellation and $N_g = 6$ is not well understood at this stage, which awaits for further research.

Physically, the larger the energy gap is, the weaker the screening becomes. Weaker screening naturally corresponds to a smaller exciton radius and hence the higher binding energy as schematically illustrated in Fig. 3. In this sense, a proportional relationship between E_g and E_b is taken for granted. Nevertheless, what is much more interesting here is the universal parameter-free slope as well as the formulas of E_g and E_b as functions of α_{2D}, which is clearly not valid for the 3D semiconductors. Such a difference essentially arises from the nonlocal *vs* local screening effect between the 2D and 3D materials. For the 2D case, the electron-hole interaction is long-ranged and thus notably affects the exciton radius. In contrast, for the 3D case, the strong screening weakens the electron-hole interaction rapidly, generally leading to a larger exciton radius, and hence a smaller E_b.

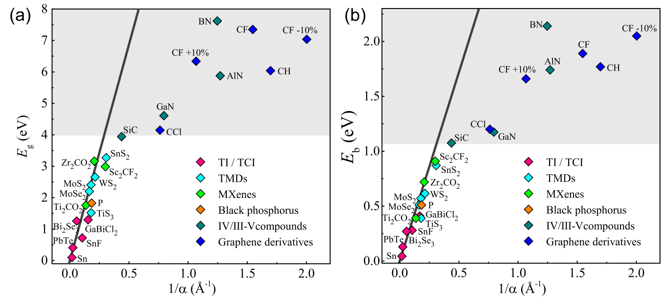

FIG. 2. Comparison of the results between analytical model and first principles calculations. E_g (left) and E_b (right) as a function of α_{2D}. The brown lines are from analytical model of Eqs. (12) and (14) while the scattered points are from the GW-BSE calculations. They agree rather well when $E_g < 4$ eV.

Figure 2 also shows an increasing deviation of Eq. (14) from the GW-BSE calculating results as the α_{2D} becomes smaller than 2 angstroms. Let us briefly discuss their origins. For the determination of E_b, the averaging scheme [8] is used to handle the nonlocal screening. Intrinsically, the validity of such a treatment lies in the fact that the exciton extends over a wide range under small E_g. However, the exciton will be compressed within a small space with the enlargement of E_g as schematically demonstrated in Fig. 3. Averaging then becomes no longer valid for a local quantity, thus causing the deviation. In this sense, the observed linear relationship between E_b and E_g in the whole range (see Fig. 1) is surprising. The underlying mechanism under small E_g is understood in terms of α_{2D}, but it keeps unclear under large E_g.

Finally, we investigate the relationship between the carrier effective mass m^* and E_g in the 2D materials. It is worth emphasizing that a scalar m^* usually cannot be defined in the 2D case because of emerging anisotropy and degeneracy at the band edges. Among 23 2D materials considered, the electron effective mass (m_e^*) is only well-defined in 14 materials, and the corresponding first-principles results are given in Fig. 4(a). While in the calculation of the reduced mass μ, the presence of

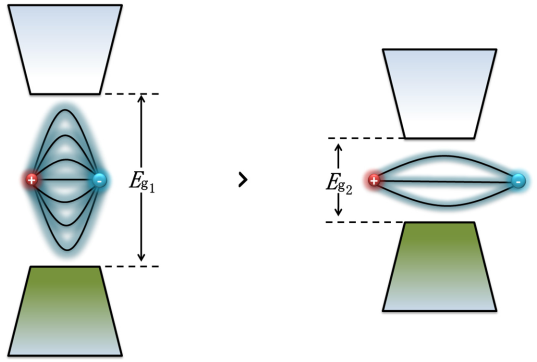

FIG. 3. Illustration of the relationship between the energy gap and exciton. For a system with a large E_g, the screening is weak and the exciton binds strongly, giving rise to a relatively narrow spatial extension and high E_b. In contrast, for a system with a small E_g, the screening is strong and the exciton binds loosely, giving rise to a relatively wide spatial extension and a low E_b.

anisotropy and degeneracy at the valence band maximum as well as the indirect gap character further decreases the number of suitable materials to five (Sn, PbTe, MoS_2, WS_2, and $MoSe_2$). In this regard, we pick out the materials with well-defined m_e^* and, respectively, calculate the μ with respect to heavy and light holes when the valence band maximum is degenerate. Figure 4(b) summarizes the results, where the filled and half filled symbols, respectively, correspond to the values from heavy and light holes. Obviously, no clear correlation is observed between effective mass (m_e^* or μ) and E_g. This is remarkably distinct from the 3D materials in which a linear relation exists between E_g and m_e^* [7], manifesting an essential difference of 2D and 3D materials.

In conclusion, we confirm the parameter-free linear scaling law between E_g and E_b in the whole band gap range for 2D materials by using first-principles *GW*

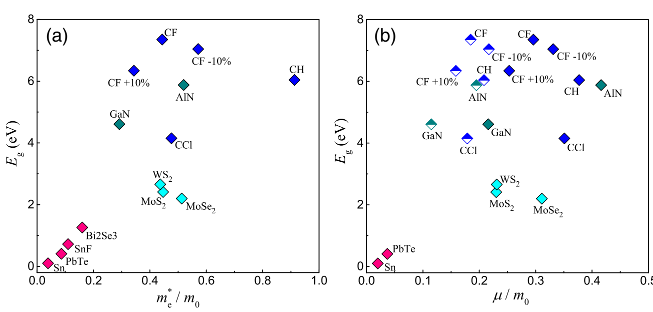

FIG. 4. Dependence of E_g as the functions of (a) electron effective mass m_e^* and (b) reduced mass μ. Only the data of 14 2D materials are presented in (a) due to that the scalar m_e^* is ill-defined for the cases with emerging anisotropy and degeneracy at the conduction band minimum. In (b), apart from the well-defined μ for Sn, PbTe, MoS_2, WS_2, and $MoSe_2$, the filled and half filled diamonds, respectively, correspond to the μ calculated from heavy and light holes owing to the twofold degeneracy at the valence band maximum.

Bethe-Salpeter equation calculations and the $\mathbf{k} \cdot \mathbf{p}$ method. We deduce a universal expression for the quasiparticle band gap of 2D materials which is solely dependent upon the static 2D polarizability but has nothing to do with the effective mass. It is expected that there exists more generalized E_g expression suitable for all the 2D semiconductors, and here obtained linear scaling law may provide an alternative way to realize the E_g expression from the knowledge of E_b, or vice versa. We also find that the effective mass does not play an essential role in the universal behavior of 2D materials, unlike the case in their 3D counterparts. Our study offers not only a new perspective to distinguish the 3D and 2D nature but also some insights for the band gap engineering in the optimal design of functionalized 2D materials.

This work was supported by the Ministry of Science and Technology of China (Grant No. 2016YFA0301001), the National Natural Science Foundation of China (Grants No. 11674071, No. 21373015, No. 11674188, and No. 11334006).

*LiuZhiRong@pku.edu.cn
†liyc@nanoctr.cn
‡Present address: Advanced Research Institute for Multidisciplinary Science, Beijing Institute of Technology, Beijing 100081, China.
§dwh@phys.tsinghua.edu.cn

[1] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
[2] X. L. Qi and S. C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[3] R. R. Nair, P. Blake, A. N. Grigorenko, K. S. Novoselov, T. J. Booth, T. Stauber, N. M. R. Peres, and A. K. Geim, Science **320**, 1308 (2008).
[4] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
[5] J. H. Choi, P. Cui, H. P. Lan, and Z. Y. Zhang, Phys. Rev. Lett. **115**, 066403 (2015).
[6] M. L. Zhang, L. Y. Huang, X. Zhang, and G. Lu, Phys. Rev. Lett. **118**, 209701 (2017).
[7] J. Singh, *Semiconductor Devices: Basic Principles* (Wiley, New York, 2001).
[8] T. Olsen, S. Latini, F. Rasmussen, and K. S. Thygesen, Phys. Rev. Lett. **116**, 056401 (2016).
[9] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
[10] M. Rohlfing and S. G. Louie, Phys. Rev. B **62**, 4927 (2000).
[11] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[12] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[13] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.118.266401 for the details of computational parameters and theoretical deduction.
[14] D. Y. Qiu, F. H. da Jornada, and Steven G. Louie, Phys. Rev. B **93**, 235435 (2016).

[15] M. Khazaei, M. Arai, T. Sasaki, C. Y. Chung, N. S. Venkataramanan, M. Estili, Y. Sakka, and Y. Kawazoe, Adv. Funct. Mater. **23**, 2185 (2013).
[16] Y. Xu, B. H. Yan, H. J. Zhang, J. Wang, G. Xu, P. Z. Tang, W. H. Duan, and S. C. Zhang, Phys. Rev. Lett. **111**, 136804 (2013).
[17] J. W. Liu, X. F. Qian, and L. Fu, Nano Lett. **15**, 2657 (2015).
[18] L. Y. Li, X. M. Zhang, X. Chen, and M. W. Zhao, Nano Lett. **15**, 1296 (2015).
[19] M. M. Ugeda, A. J. Bradley, S.-F. Shi, F. H. da Jornada, Y. Zhang, D. Y. Qiu, W. Ruan, S.-K. Mo, Z. Hussain, Z.-X Shen, F. Wang, S. G. Louie, and M. F. Crommie, Nat. Mater. **13**, 1091 (2014).
[20] X. M. Wang, A. M. Jones, K. L. Seyler, V. Tran, Y. C. Jia, H. Zhao, H. Wang, L. Yang, X. D. Xu, and F. N. Xia, Nat. Nanotechnol. **10**, 517 (2015).
[21] B. R. Zhu, X. Chen, and X. D. Cui, Sci. Rep. **5**, 9218 (2015).
[22] A. T. Hanbicki, M. Currie, G. Kioseoglou, A. L. Friedman, and B. T. Jonker, Solid State Commun. **203**, 16 (2015).
[23] P. Cudazzo, C. Attacalite, I. V. Tokatly, and A. Rubio, Phys. Rev. Lett. **104**, 226804 (2010).
[24] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (2013).
[25] V. Tran, R. Soklaski, Y. F. Liang, and L. Yang, Phys. Rev. B **89**, 235319 (2014).
[26] P. Cudazzo, I. V. Tokatly, and A. Rubio, Phys. Rev. B **84**, 085406 (2011).
[27] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. L. Li, O. B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. Lett. **113**, 076802 (2014).
[28] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **48**, 17791 (1993).
[29] Z. R. Liu, J. Wu, and W. H. Duan, Phys. Rev. B **69**, 085117 (2004).