



This is the accepted manuscript made available via CHORUS, the article has been published as:

Origin of the Variation of Exciton Binding Energy in Semiconductors

Marc Dvorak, Su-Huai Wei, and Zhigang Wu

Phys. Rev. Lett. **110**, 016402 — Published 2 January 2013

DOI: [10.1103/PhysRevLett.110.016402](https://doi.org/10.1103/PhysRevLett.110.016402)

Origin of the Variation of Exciton Binding Energy in Semiconductors

Marc Dvorak,¹ Su-Huai Wei,² and Zhigang Wu^{1,*}

¹*Department of Physics, Colorado School of Mines, Golden, CO 80401*

²*National Renewable Energy Laboratory, Golden, CO 80401*

Abstract

Excitonic effects are crucial to optical properties, and the exciton binding energy, E_b , in technologically important semiconductors varies from merely a few meV to about 100 meV. This large variation, however, is not well understood. We investigate the relationship between the electronic band structures and exciton binding energies in semiconductors, employing first-principles calculations based on the density functional theory (DFT) and the many-body perturbation theory using Green's functions (GW/BSE). Our results clearly show that E_b increases as the localization of valence electrons increases due to the reduced electronic screening. Furthermore, E_b increases in ionic semiconductors such as ZnO because, contrary to the simple two-level coupling model, it has both conduction and valence band edge states strongly localized on anion sites, leading to an enhanced electron-hole interaction. These trends are quantized by electronic structures obtained from the DFT; thus, our approach can be applied to understand the excitonic effects in complex semiconducting materials.

PACS numbers: 71.35.-y, 71.15.Qe, 71.15.Mb,

Excitonic effects in semiconductors, which are determined by not only the quasiparticle (QP) energies but also the electron-hole interaction, play a critical role in optoelectronic devices, such as photovoltaic cells and light emitting diodes (LEDs) [1, 2]. Without considering the electron-hole interaction, the independent-QP description of optical spectra is often significantly deviated from experimental data [3–10]. The strength of electron-hole interaction is characterized by the exciton binding energy, E_b , which varies from a few meV to about 100 meV in technologically important inorganic semiconductors. This rather large variation in exciton binding energy in bulk semiconductors is not well understood from fundamental principles. Even structurally and/or electronically similar materials can have surprisingly different E_b . For example, both GaN and ZnO have the Wurtzite structure with similar lattice constants, band gaps ($E_g = 3.44$ eV), and effective masses [11, 12], yet the E_b for GaN of 28 meV [13] is much smaller than that for ZnO of 59 meV [14], despite the fact that ZnO is more ionic. Based on the simple tight-bonding model [15], the overlap between electron and hole in the more ionic ZnO is expected to be less and thus ZnO should have smaller E_b than GaN.

Understanding the variation in E_b among common semiconductors is, therefore, not only practically important, but also of fundamental interest. The state-of-the-art first-principles method for electronic excitations is the many-body perturbation theory using Green functions, and accurate E_b can be obtained by solving the Bethe-Salpeter equation (BSE) [4, 16]. However, this approach is computationally extremely demanding; furthermore, the calculated results do not provide a simple and clear explanation to the large variation in exciton binding energy. Instead, electronic structures obtained from the Kohn-Sham (KS) density functional theory (DFT) [17, 18] often give good insights to many electronic and optical properties. Prediction of the trend in excitonic properties using DFT is thus particularly advantageous and convenient due to its simplicity and the tremendously reduced computational efforts compared to the many-body techniques.

In this Letter, we explore the relationship between exciton binding energies and the localization of the KS eigenstates. In the Tamm-Dancoff approximation to the BSE, the exciton is a superposition of quasi-electron/quasi-hole pairs, with each pair interacting by the screened direct Coulomb and bare exchange interactions. Thus, the localization and overlap of electron and hole states is critical because of the $1/r$ nature of the interaction. Additionally, valence states not explicitly participating in the exciton formation can contribute to the

screening of the interaction, and it is a commonly held tenet that localized charge is poorer at screening than delocalized charge. Thus, we investigate separately the localization of the band edge states and valence charge density, using the fact that in bulk semiconductors DFT valence and conduction states are good approximation to quasi-hole and quasi-electron states, respectively [19, 20], with QP energy corrections applied. By solving the BSE, we rigorously study the effect of wave function localization on exciton binding energies from a purely *ab-initio* framework.

Our analysis has three parts, studying separately the localization of excitons, valence electrons, and band edge electron and hole states. Comparison of the relative exciton distribution, fixing hole or electron, to the binding energy indicates that compact excitons have higher binding energies than delocalized excitons. Next, we compare localization of excitons to localization of the valence electrons. We expect that a localized charge density is relatively poor at screening, thus charge localization should correspond to high E_b . Determining localization of electrons in the Bloch wave picture is tricky since the electrons are distributed periodically over all unit cells; instead, we construct maximally localized Wannier functions (MLWFs [21]) for each valence band. The total variance of these localized orbitals is a quantifiable measure of valence electron localization. We find that localized electrons do, in fact, correlate with high E_b because of their reduced screening to the electron-hole interaction. Finally, we examine the localization of the band-edge states near ions. In contrast to the simple two-level tight-binding picture of ionic bonding, our DFT results indicate that ionic materials have a significant anion contribution to the conduction band minimum (CBM) electron state, resulting in significant overlap between the CBM electron state and the valence band maximum (VBM) hole state on the anion sites. The strong Coulomb interaction between these states near the anion sites is responsible for the tightly bound excitons for systems such as ZnO.

Because excitons are intrinsically two-particle objects, common independent electron methods in electronic structure calculations are not adequate to describe their behavior, hence a more rigorous level of theory must be employed. We perform electronic ground-state calculations with DFT, and then use the many-body perturbation theory with Green's functions (GW/BSE) on top of the KS eigenstates. The conceptual advantage of the BSE is that of a two-particle equation, which is necessary for describing the excitonic effects. We diagonalize the two-particle Hamiltonian to explicitly obtain the exciton eigenfunctions and

eigenvalues, and then the exciton binding energy - the difference between the QP gap and the optical gap. The intermediate GW step essentially results in an adjustment of the KS eigenvalues towards QP energy levels, which can be matched to experiment. It has been shown that the combined GW/BSE method provides significantly better agreement with experimental optical properties than single-particle calculations alone [3–10].

Density functional calculations are performed with the open-source code QUANTUM ESPRESSO [22]. We use a planewave basis expanded to 45 Ha and norm-conserving pseudopotentials to perform non-spin polarized valence electron calculations. The Perdew-Burke-Ernzerhof (PBE) parametrization of the generalized gradient approximation (GGA) [23] is adopted for the exchange-correlation functional, with the Brillouin zone sampling on a $12 \times 12 \times 12$ Monkhorst-Pack grid. Wannier functions are constructed using the WANNIER90 code [24, 25]. To better describe the strong Coulomb and exchange interactions of the localized Zn d states, the GGA+ U method [26, 27] is used for ZnO, with the screened Coulomb and exchange parameters $U = 8.0$ eV and $J = 0.9$ eV, respectively, which were calculated from first-principles [28]. GW and BSE calculations employ the BerkeleyGW package [29], and we perform the single shot G_0W_0 calculations using the plasmon-pole approximation [20] to the dynamical screening. The BSE kernel is constructed using the four highest valence bands and four lowest conduction bands of the system on the same $12 \times 12 \times 12$ \mathbf{k} -point grid; however, accurate representations of the band edge excitons require an enormously dense grid of \mathbf{k} -points. Fortunately, BerkeleyGW can interpolate the interaction kernel onto a much finer grid of \mathbf{k} -points, eliminating the formidable need to compute a huge number of matrix elements. To keep the matrix diagonalization tractable, we interpolate the kernel onto a very dense grid in roughly 2% of the full Brillouin zone, centered at the direct gap of the material.

For completeness in our analysis, we first confirm the intuitive picture of more localized excitons having higher E_b . Before comparing the localization of the exciton to the binding energy, we first define our measure of localization in three dimensions. To obtain a three-dimensional distribution from the six-dimensional exciton wave function $\Psi(\mathbf{r}_1, \mathbf{r}_2)$ for electron and hole coordinates \mathbf{r}_1 and \mathbf{r}_2 , respectively, we examine the relative distribution $\phi(\mathbf{r})$ of the electron around the hole:

$$\phi(\mathbf{r}) = \int \Psi(\mathbf{r}', \mathbf{r}' + \mathbf{r}) d\mathbf{r}'. \quad (1)$$

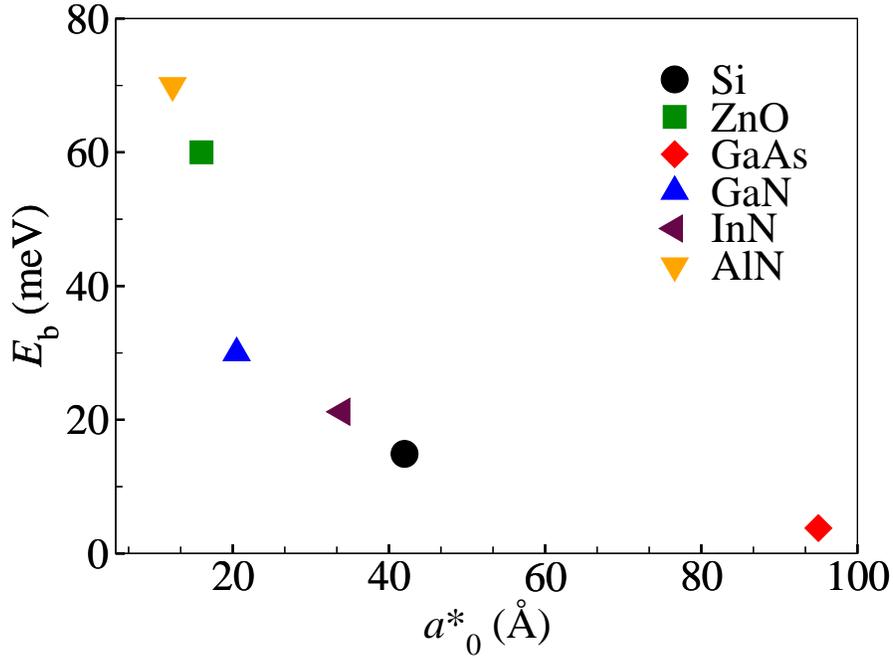


FIG. 1: (color online) Calculated exciton binding energy (E_b) vs exciton Bohr radius (a_0^*). These theoretical values obtained from solving the BSE agree very well with experimental data. The curve represents E_b fitted to $C/(a_0^*)^2$, with C a constant. We also computed a very strongly bound and localized exciton in NaCl, which is off the scale of the plot but in agreement with the overall trend.

Then one can compute the spread of $\phi(\mathbf{r})$, defined as

$$\Omega = \langle r^2 \rangle - \langle \mathbf{r} \rangle^2, \quad (2)$$

and the localization length as $\sqrt{\Omega}$.

Assuming $\phi(\mathbf{r})$ of the (lowest-energy) exciton has the 1s hydrogen-like wave function, then its Bohr radius $a_0^* \approx \sqrt{\Omega}/\sqrt{3}$, and we can simply compute a_0^* to compare the localization levels of excitons. Fig. 1 summarizes the computed exciton binding energies in Si, GaAs, AlN, GaN, InN, and ZnO, as a function of exciton Bohr radius. As expected, E_b increases

monotonically as a_0^* decreases. Using the hydrogenic model,

$$a_0^* = \frac{m}{m^*} \epsilon a_0, \quad (3)$$

$$E_b = \frac{e^2}{\epsilon a_0^*} = \frac{e^2 a_0}{(m^*/m)(a_0^*)^2}, \quad (4)$$

where m (m^*) is electron (effective reduced) mass, a_0 is the Bohr radius for hydrogen atom, e is the electron charge, and ϵ is the dielectric constant. Because the variation of m^*/m is relatively small in the considered semiconductors except for GaAs, $E_b \propto 1/(a_0^*)^2$, as demonstrated in Fig. 1.

In both the simple hydrogenic model of excitons and the BSE description, screening, predominantly by the valence electrons, leads to a weaker Coulomb interaction between electron and hole, and hence a reduced exciton binding energy. Although the actual behavior of electronic screening in real materials is rather complicated, qualitatively, a more extended (delocalized) valence electron distribution tends to screen more strongly than less extended (localized) distribution. Stronger screening leads to higher E_b and more localized excitons as is evident in Fig. 2, in which both valence electrons and exciton distributions are plotted for Si and ZnO. The more localized valence electrons in ZnO lead to smaller a_0^* and more compact exciton than those in Si.

To quantitatively demonstrate the relationship between the localization of valence electrons and exciton binding energy, however, we need to use Wannier functions for extended systems whose localization is otherwise not obvious in the Bloch-wave picture. Wannier functions are expected to display localization characteristic of the total charge density. An ionic crystal, for instance, would likely have Wannier functions localized to the anions of the system; while in GaAs Wannier functions are expanded to most space.

As introduced by Marzari and Vanderbilt [25], the spread functional Ω of an N -band crystal in real space is

$$\Omega = \sum_{n=1}^N [\langle r^2 \rangle_n - \langle \mathbf{r} \rangle_n^2], \quad (5)$$

which is minimized with respect to unitary transformations, and the localization length of the valence electrons $\sigma_{\text{el}} = \sqrt{\Omega}/N$. Ω can be decomposed into one gauge invariant term Ω_{I} and a variant term $\tilde{\Omega}$. The WANNIER90 code [24, 25] is employed, which searches over a range of unitary transformations to the wave functions to identify the Wannier orbitals (MLWFs) that minimize $\tilde{\Omega}$.

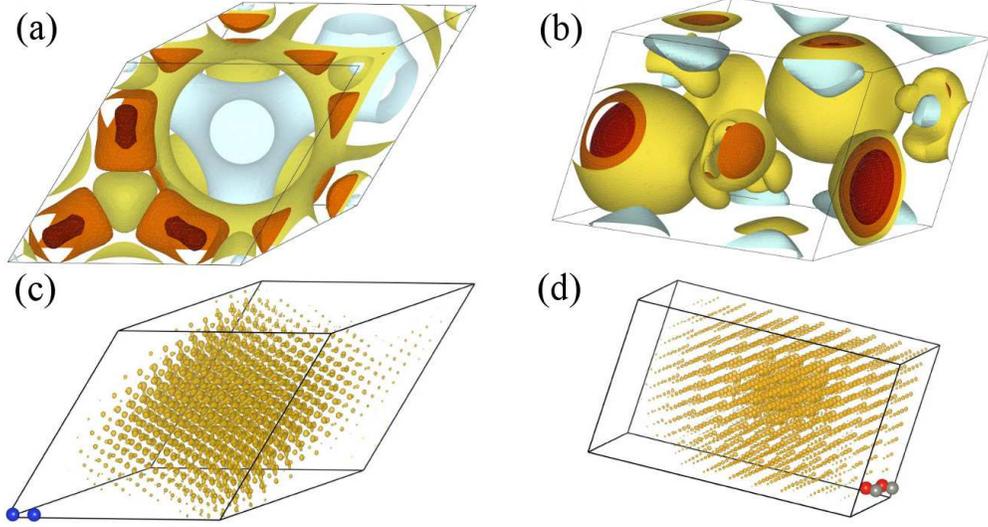


FIG. 2: (color online) Isosurfaces of valence charge density (upper panels) and exciton distribution (lower panels), $|\phi(\mathbf{r})|^2$, in Si (left panels) and in ZnO (right panels). Here four valence bands for Si and six for ZnO are included, and the corresponding isosurface values are set to be identical in panels (a) and (b), with high to low values ranging from red (dark) to yellow (light) colors. Charge density is periodic, plotted in a single unit cell, while exciton is non-periodic, extended to many unit cells. Atoms in a unit cell of Si and ZnO are illustrated in panel (c) and (d), respectively.

Since the induced change in macroscopic polarization ΔP_{mac} depends linearly on the localization length σ_{el} [30], dielectric constant is also expected to depend on σ_{el} linearly,

$$\varepsilon = C_1 + C_2\sigma_{\text{el}}, \quad (6)$$

as clearly demonstrated in Fig. 3(a), with C_1 and C_2 as two fitting parameters. Combining Eqs. (3) and (6), we find that

$$a_0^* = \frac{m}{m^*}a_0(C_1 + C_2\sigma_{\text{el}}), \quad (7)$$

i.e., the exciton Bohr radius depends linearly on valence electron localization length as well, if the variation of m/m^* in semiconductors is omitted. Fig. 3(b) compares a_0^* with σ_{el} ; although the linear dependence is not quantitatively accurate, qualitatively, a low level of valence electron localization leads to strong screening and large value of ε , which cause the exciton to be loosely bound with low E_b and large a_0^* . While these evident trends are not particularly surprising, our *ab initio* calculations carried out to the Bethe-Salpeter

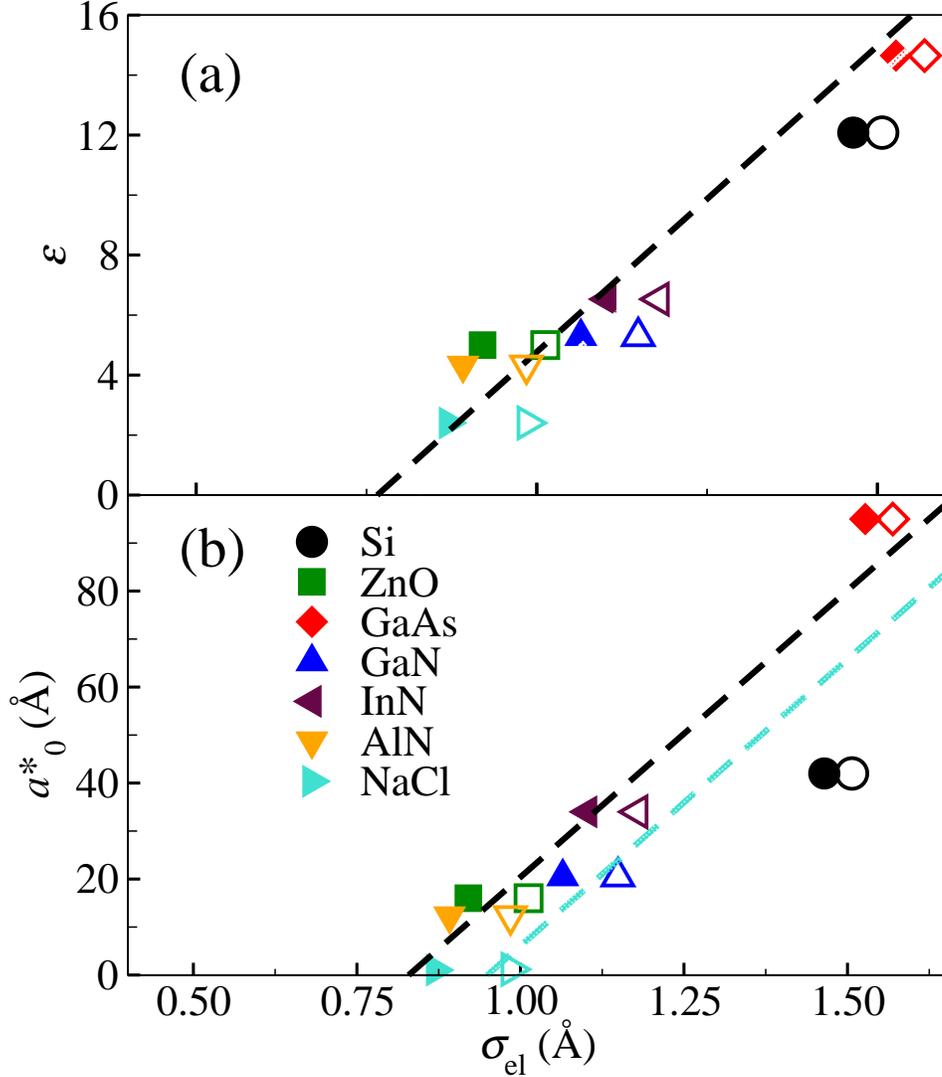


FIG. 3: (color online) Calculated dielectric constant (ϵ) in panel (a) and exciton Bohr radius (a_0^*) in panel (b) as functions of valence electron localization length (σ_{el}). Here the dashed lines are linear fittings, and the black and green (grey) symbols correspond to σ_{el} computed from the invariant (Ω_I) and total (Ω) spread, respectively.

level provide rigorous confirmation of the relationship between localized charge, electron screening, and exciton binding energies.

Finally, in the simplest and widely accepted picture of atomic bonding in solids [15], the anion states and cation states form bonding and anti-bonding bands. The bonding VBM state is mainly composed of anion atomic orbitals while the anti-bonding CBM state mainly consists of cation atomic orbitals. Our DFT results indicate that this description is not

completely correct, especially for ionic materials; instead, a significant anion contribution to the CBM for AlN and ZnO is found. The deviation is because these ionic compounds have some high-lying unoccupied anion s -orbitals that can hybridize strongly with the cation s -orbitals, making the CBM contain an appreciable amount of anion orbitals. Furthermore, we find that, for all materials in our study, deviations from this model are critical to explain trends in E_b . In an exciton, the hole is derived mostly from the states close to VBM, while the electron is mostly derived from the states close to CBM; therefore, the contribution of anion orbitals near the CBM would strongly enhance the Coulomb and exchange interactions, which are important matrix elements in the electron-hole interaction kernel K for BSE. As a result, materials with a significant amount of CBM and VBM states *both* localized at the same atomic site are expected to have a high exciton binding energy, instead of CBM and VBM states being on the anions and cations, separately.

The BSE kernel K has screened direct and exchange interactions, as constructed by terms such as $\psi_c(\mathbf{r}_1)\psi_v^*(\mathbf{r}_2)$. This suggests that N_cN_v , the product of the numbers of *localized* valence and conduction states near band edges, could be used to indicate the strength of K and therefore the strength of the electron-hole interaction. We compute the product N_cN_v at each atomic site to show that this onsite localization at anions in ionic semiconductors is a good indicator of the strength of the exciton binding. Here N_c and N_v are the numbers of near gap states within a cutoff energy, which in real space are near ions within a cutoff radius. In reasonable ranges, the trend of E_b as a function of N_cN_v is not sensitive to the choices of both cutoffs, and N_cN_v is a practical way to account for CBM and VBM states localized on the same atomic sites. Fig. 4 shows that N_cN_v for cations is negligible compared to that for anions. Most importantly, E_b increases as N_cN_v (at the anion sites) increases, and a good numerical fitting is $E_b \propto \sqrt{N_cN_v}$, as plotted in Fig. 4. This cannot be derived from the simple effective mass model, and a high value of total density of states at the band edges, by itself, is not a strict enough criterion to indicate the strength of excitonic binding.

In particular, returning to the comparison of GaN and ZnO, their difference in E_b can be explained by their different numbers of near-gap states at the anion site. ZnO has a large CBM component on oxygen sites because of strong hybridization of the $4s$ states of Zn with the high lying $3s$ states of oxygen. GaN does not have as much hybridization for the CBM state because the N $3s$ level has a higher energy than the Ga $4s$ level, and, as a result, has a reduced value of N_c at the N site.

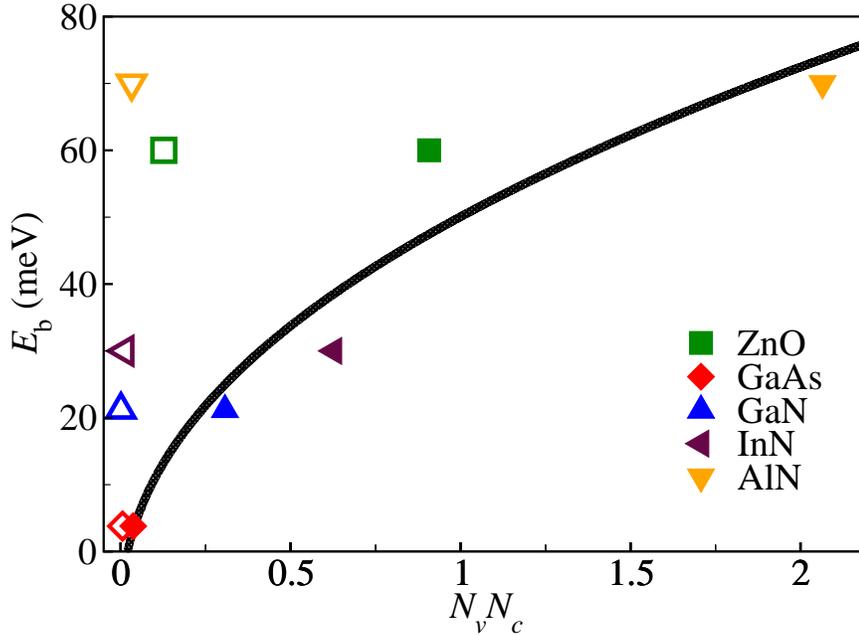


FIG. 4: (color online) Exciton binding energy (E_b) as a function of $N_c N_v$. See the text for the definition of N_v and N_c . The solid (open) symbols are $N_c N_v$ for the anion (cation) sites, and the curve is a guide to the eye. The data point for NaCl is off the scale off the plot to the top-right, but agrees with the overall trend.

In conclusion, the variation in exciton binding energy of semiconductors manifests itself in the real space localization of the exciton, which in turn is determined by the strength of electron screening in semiconductors. The more localized valence electrons, which we quantify by the spread of the Wannier functions, are found to be poorer at screening, leading to a smaller dielectric constant. Furthermore, E_b is large for ionic semiconductors because, in contrast to the simple two-level coupling model, the conduction band edge states have significant overlap with valence band edge states localized to anion sites, increasing the electron-hole interaction. All these trends can be qualitatively fitted to simple scaling rules based on essential physical pictures of exciton, valence electrons and near gap states, and they pave the way for understanding and predicting the excitonic effects in more complicated semiconductors without resorting to computationally very demanding many-particle perturbation theory.

This work at CSM was supported by DOE Early Career Award (No. DE-SC0006433) and the startup fund from Colorado School of Mines (CSM). The work at NREL was supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308. Computations were carried out at the Golden Energy Computing Organization at the CSM and National Energy Research Scientific Computing Center (NERSC). ZW thanks insightful discussions with M. Lusk and D. Wood.

* Electronic address: `zhiwu@mines.edu`

- [1] R. D. Schaller and V. I. Klimov, *Phys. Rev. Lett.* **92**, 186601 (2004).
- [2] A. Khan, K. Balakrishnan, and T. Katona, *Nature Photon.* **2**, 77 (2008).
- [3] M. Rohlfing and S. G. Louie, *Phys. Rev. Lett.* **81**, 2312 (1998).
- [4] M. Rohlfing and S. G. Louie, *Phys. Rev. B* **62**, 4927 (2000).
- [5] S. Albrecht, G. Onida, and L. Reining, *Phys. Rev. B* **55**, 10278 (1997).
- [6] R. Laskowski, N. E. Christensen, G. Santi, and C. Ambrosch-Draxl, *Phys. Rev. B* **72**, 035204 (2005).
- [7] R. Laskowski and N. E. Christensen, *Phys. Rev. B* **73**, 045201 (2006).
- [8] A. Marini and R. Del Sole, *Phys. Rev. Lett.* **91**, 176402 (2003).
- [9] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, *Phys. Rev. Lett.* **80**, 4510 (1998).
- [10] L. X. Benedict, E. L. Shirley, and R. B. Bohn, *Phys. Rev. Lett.* **80**, 4514 (1998).
- [11] M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur, eds., *Properties of Advanced Semiconductor Materials GaN, AlN, InN, BN, SiC, SiGe* (John Wiley and Sons, Inc., New York, 2001).
- [12] S. Pearton, D. Norton, K. Ip, Y. Heo, and T. Steiner, *Prog. in Mat. Sci.* **50**, 293 (2005).
- [13] W. Shan, B. D. Little, A. J. Fischer, J. J. Song, B. Goldedberg, W. G. Perry, M. D. Bremser, and R. F. Davis, *Phys. Rev. B* **54**, 16369 (1996).
- [14] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors* (Springer, 2005), 3rd ed.
- [15] W. A. Harrison, *Electronic Structure and the properties of solids* (W. H. Freeman, San Francisco, 1980).
- [16] G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002).
- [17] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).

- [18] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [19] M. S. Hybertsen and S. G. Louie, *Phys. Rev. Lett.* **55**, 1418 (1985).
- [20] M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* **34**, 5390 (1986).
- [21] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, *Rev. Mod. Phys.* **84**, 1419 (2012).
- [22] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, et al., *J. Phys.: Cond. Matt.* **21**, 395502 (2009).
- [23] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [24] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, *Compu. Phys. Commu.* **178**, 685 (2008).
- [25] N. Marzari and D. Vanderbilt, *Phys. Rev. B* **56**, 12847 (1997).
- [26] V. I. Anisimov, J. Zaanen, and O. K. Andersen, *Phys. Rev. B* **44**, 943 (1991).
- [27] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, *Phys. Rev. B* **52**, R5467 (1995).
- [28] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, *Phys. Rev. Lett.* **105**, 146401 (2010).
- [29] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, *Compu. Phys. Commu.* **183**, 1269 (2012).
- [30] I. Souza, T. Wilkens, and R. M. Martin, *Phys. Rev. Lett.* **62**, 1666 (2000).