Difficulty in predicting shallow defects with hybrid functionals: Implication of the long-range exchange interaction

Junhyeok Bang, Y. Y. Sun, Tesfaye A. Abtew, Amit Samanta, Peihong Zhang, and S. B. Zhang

Phys. Rev. B **88**, 035134 — Published 25 July 2013

DOI: 10.1103/PhysRevB.88.035134

Can Hybrid Functionals Predict Shallow Defects? Implication of the Long-Range Exchange Interaction

Junhyeok Bang,[1] Y. Y. Sun,[1] Tesfaye A. Abtew,[2] Amit Samanta,[3] Peihong Zhang,[2,*] and S. B. Zhang[1,†]

[1]*Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, Troy, New York 12180, USA*

[2]*Department of Physics, University at Buffalo, State University of New York, Buffalo, New York 14260, USA*

[3]*Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey, USA*

We identify an important issue in defect studies using hybrid functionals. When modeling a defect, which is supposedly an isolated system, with a finite-size supercell, the inclusion of a fraction of the Hartree-Fock interaction results in a strong cell-size dependence and an extremely slow convergence of the calculated defect properties, especially for shallow defect. These behaviors may give rise to a number of errors in calculated defect properties, including the deepening of transition level and over-stabilization of shallow defects. Numerical results from hybrid functional calculations for a diverse array of systems can be understood within the Hartree-Fock theory of an electron-gas model, indicating that the long-range exchange is the main cause for the errors in the calculated defect properties within hybrid functionals.

PACS numbers: 71.55.-i, 71.15.-m, 71.15.Mb, 61.72.Bb

I. INTRODUCTION

Density-functional theory[1,2] with a local or semi-local exchange-correlation functional, such as the local density approximation (LDA)[3] or the generalized gradient approximation (GGA),[4] has been successfully applied to predict atomic structures and ground-state properties of condensed matter systems, including many defect-related properties using the supercell approach.[5] However, for many semiconductors and insulators the band gap is underestimated and the electronic states tend to be delocalized.[6] This deficiency often leads to a large uncertainty in the calculated defect properties, a problem that has plagued the electronic structure and materials theory community for decades.

It has been demonstrated that hybrid functionals such as PBE0[7] and HSE,[8] which incorporate a fraction of the nonlocal Hartree-Fock (HF) exchange energy, improve the calculated band gap of a broad range of semiconductors at a much lower computational cost[9] compared with more sophisticated many-body methods, such as the *GW* method.[10] The improvement of the calculated band gap, although empirical in nature and highly parameter and system dependent, has generated significant interest in applying these new functionals to the study of defects in semiconductors.[11–14] Another reason for the popularity of the hybrid functionals is that some of the classic defect related problems, for which the LDA and GGA fails, can now be treated within hybrid functionals, at least qualitatively. Examples include the V_k center[15] and other polaronic defects,[16,17] which typically involve localization of a charge carrier at the defect site.

Even with these successes, one has to be cautious in interpreting results obtained with hybrid functionals. It is unclear how a mixing of a fraction of the HF exchange is able to correctly and accurately account for the exchange-correlation energy of a many-electron system. Different hybrid schemes may give different results for a given physical system. Recently, it was reported that current hybrid functionals do not give reliable band gaps, especially for nanostructures and surfaces.[18]

In this paper, we show that the long-range nature of the HF exchange leads to consequences that have not been fully understood. The effect could become particularly serious when applied to isolated systems, such as defects, modeled by the supercell approach. We show that defect transition levels calculated using hybrid functionals such as PBE0 are very sensitive to the size of the supercell and the convergence is extremely slow. As a result, many known shallow defects would be classified as deep centers in practical calculations with the PBE0 functional. In addition, the thermodynamic properties of defects, such as the shallow-to-deep transition, could also be affected because of the over-stabilization of the shallow state. We show that all these problems arise from the long-range nature of the HF exchange interaction. This raises a serious question regarding the applicability of some hybrid functionals (such as PBE0) for defect calculations.

II. CALCULATIONAL METHODS

Our calculations were carried out using the VASP package[19] with projector augmented wave method.[20] For PBE0 calculations, we used mixing parameter $\alpha = 0.25$, while we used $\alpha = 0.375$ and screening parameter $\mu = 0.2$ Å^{-1} for HSE calculations. We used a large HF mixing parameter in the HSE calculations so that the two functionals (HSE and PBE0) give a similar band gap for GaAs. Wave functions were expanded in a plane-wave basis with an energy cutoff of 277, 318, and 306 eV for GaAs, Si, and ZnO, respectively. We used the experimental lattice parameters for all systems studied. For

defect systems, atomic positions were fully relaxed until the residual forces were less than 0.03 eV/Å. Γ-point is used for the Brillouin zone integration. The singularity term in the HF-exchange energy is calculated using the method described in Ref. 21.

III. RESULTS: EFFECT ON TRANSITION LEVELS

Figure 1 shows the defect transition level $\epsilon(+/0)$ [using the conduction band minimum (CBM) as a reference] of two textbook examples of shallow donors, namely, GaAs:Si_{Ga} and Si:P_{Si}, as a function of the supercell size. The transition levels are calculated using the well-esbalished technique:[22] $\epsilon(q/q') = (E^q_{\rm tot} - E^{q'}_{\rm tot})/(q'-q) - E_{\rm CBM}$. For comparison, results obtained using the PBE functional[4] are also shown. It can be seen that the $\epsilon(+/0)$ transition level calculated using the PBE0 functional are very sensitive to the supercell size, and the convergence is extremely slow. This is in sharp contrast to the PBE results, which quickly converge with a reasonably large (e.g., 100-atom) supercell. Because the PBE0 results do not converge with any practically affordable supercell size, the $\epsilon(+/0)$ levels for such shallow defects would always appear to be deep in practical calculations using PBE0. For example, with a supercell size of 250 atoms, the $\epsilon(+/0)$ level calculated using the PBE0 functional is 0.30 eV for GaAs:Si_{Ga}, and 0.32 eV for Si:P_{Si}, to be compared with the experimental values of 6 meV for GaAs:Si_{Ga} and 45 meV for Si:P_{Si}.[23,24] By inspecting Fig. 1, one may consider an extrapolation scheme to obtain a converged result. An appropriate extrapolation, however, requires an analytical understanding of the scaling of the PBE0 results, as we will discuss later.

It should be mentioned that there are two corrections, i.e., the electrostatic correction and the alignment of the potential, involved in the calculations of charged systems using the supercell technique. Both of these corrections will not affect the main results presented here. This is because these corrections exist in calculations with any energy functional, and our PBE results converge to within 0.01 eV with a 128-atom cell as it is shown in Fig. 1.

IV. RESULTS: EFFECT ON STABILITY

The error demonstrated in Fig. 1 could lead to serious consequences in the study of other defect-related physical processes using hybrid functionals. For example, many defects exist in two different configurations (one stable and the other metastable). Often one of these configurations has a shallow transition level, while the other exhibits a deep level. The transition between these two states could determine the dopability and doping limit of a semiconductor.[25] A classic example is the DX center in III-V semiconductors, where a donor atom (e.g., Si on a Ga site in GaAs) could undergo a displacement from

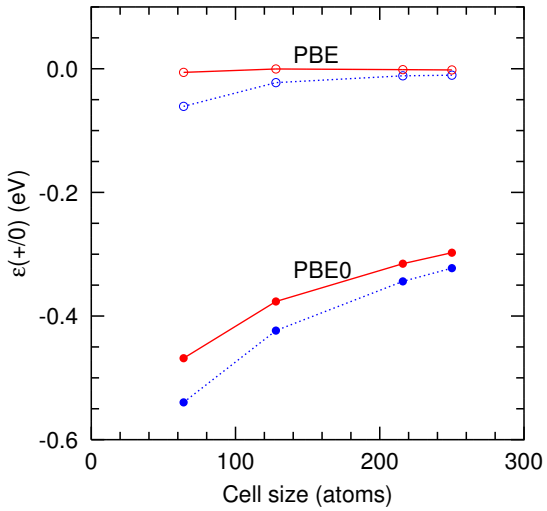

FIG. 1: (Color online) The defect transition level $\epsilon(+/0)$ of GaAs:Si_{Ga} (red solid) and Si:P_{Si} (blue dotted) as a function of the supercell size. Open and filled dots show the PBE and PBE0 results, respectively. Energy is referenced to the CBM level.

the substitutional lattice site (a shallow-donor state) to a neighboring interstitial site (a deep-donor state).[26–28] Another example, which has recently attracted significant attention,[16,29,30] is the so-called small-polaron defects in wide-band-gap semiconductors, such as a Li atom on a Zn site in ZnO and, similarly, a Mg atom on a Ga site in GaN. To describe such systems, not only the shallow transition levels, but also the relative thermodynamic stability of the two states, need to be correctly predicted. However, as we show below, shallow defects can be unphysically over-stabilized in hybrid functional calculations.

Figure 2 shows the schematic energy diagrams and structures of the negatively charged Si_{Ga} in GaAs, which has two states denoted by Si^{-1}_{d} and Si^{-1}_{DX}, as shown in Figs. 2(a) and 2(b), respectively. The Si^{-1}_{d} is four-fold coordinated with a T_d symmetry, while the Si^{-1}_{DX} is displaced to an interstitial site (C_{3v} symmetry) by breaking one of the Si-As bonds, leaving the Si dangling bond (DB) level deep inside the band gap. The bond breaking tends to increase the total energy of the system, but the Si^{-1}_{DX} state is stabilized by the lowering of the doubly occupied DB level. The relative stability of the two states is determined by the amount of the lowering of the DB level. Experimentally, the Si^{-1}_{d} state is known to be a hydrogenic defect, with a shallow transition level of about 6 meV.[23] Therefore, in this state the doubly occupied defect level should be located nearly at the CBM, which is indeed the case using the LDA or GGA functional. However, the use of the PBE0 functional results in defect states dropped deep into the band gap (shown in Fig. 2), which is apparently at odds with experiment. With such an unphysical lowering of the hydrogenic defect state (about −0.29 eV calculated using a 250-atom

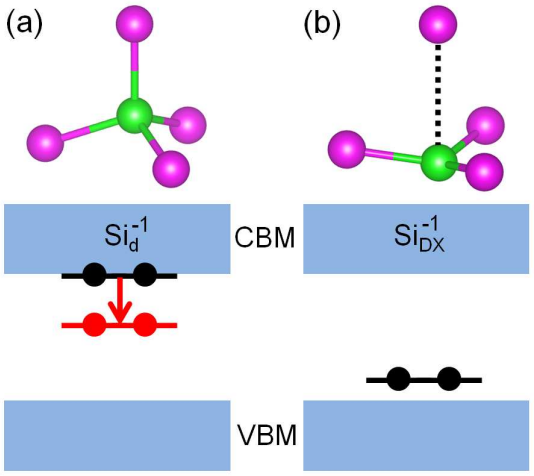

FIG. 2: (Color online) Atomic structure and schematic energy level of (a) Si_d^{-1} and (b) Si_{DX}^{-1} in GaAs. Green and violet balls stand for Si and As atoms. In hybrid functional (PBE0) calculations, the Si_d^{-1} level is significantly lowered (−0.29 eV calculated using a 250-atom supercell) as represented by the red arrow.

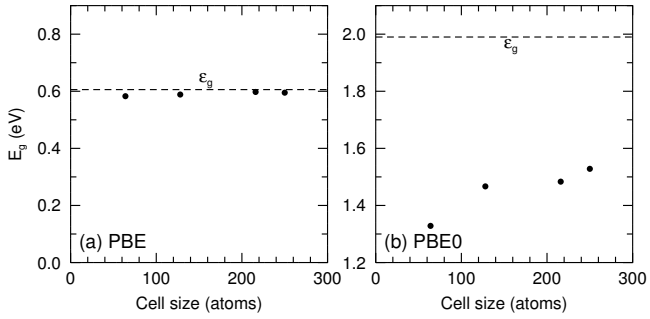

FIG. 3: The calculated E_g of GaAs as a function of supercell size using the PBE (a) and PBE0 (b) functionals. The dashed lines indicates the Kohn-Sham band gap ϵ_g.

supercell), one cannot expect an appropriate description on the thermodynamic stability of shallow defects and the transition between the shallow and deep states.

V. ORIGIN OF THE ISSUES

Since hybrid functionals have been shown to give improved results for many systems compared with the LDA or GGA, the results shown above are rather intriguing and it is imperative to understand the origin of these problems with an aim to eventually correct them. To this end, we find it is beneficial to investigate the bulk energy gap E_g first. Here we define $E_g = E_{CBM} - E_{VBM}$, where $E_{VBM} = E^0_{tot} - E^{+1}_{tot}$, $E_{CBM} = E^{-1}_{tot} - E^0_{tot}$, and E^q_{tot} is the total energy of a supercell of a pure system carrying a charge q. Within the LDA or GGA, the additional charge q has little effect on the band edge levels for a reasonably large supercell (e.g., 100-atom cell). Therefore, as shown in Fig. 3(a), even using a relatively small supercell, E_g calculated with the PBE functional approaches the Kohn-Sham (KS) band gap, $\epsilon_g = \epsilon_{CBM} - \epsilon_{VBM}$, where ϵ_{CBM} and ϵ_{VBM} are the KS eigenvalues of the CBM and VBM, respectively. However, this is not the case at all for results calculated using hybrid functionals such as PBE0.

Figure 3(b) shows the calculated E_g for bulk GaAs as a function of cell size using the PBE0 functional. Again, the strong size dependence and slow convergence are clearly seen, suggesting that these issues arise from the intrinsic properties of the hybrid functionals. We note that similar behavior has been observed in the previous study on α-quartz.[31] The underestimation of E_g can be understood by inspecting the one-electron levels involved in calculating E_g, as shown in Fig. 4. An intriguing finding is that, upon introducing an extra electron, the occupied CBM state (labeled as spin-up in the figure) shifts downward by 0.75 eV relative to the original CBM state (calculated with 128-atom supercell). Clearly, even without a defect, the addition of an electron (or hole) in an otherwise perfect crystal already produces a state deep inside the band gap in a relatively large supercell (in this case, a 128-atom cell). Results for the shallow donor GaAs:Si_{Ga} (also shown in Fig. 4), an isoelectronic system to the defect-free $(Ga_{64}As_{64})^-$, shows a very similar behavior: The occupied (supposedly shallow) donor level is lower in energy than the unoccupied level by about 0.78 eV. Therefore, it is this unphysical over-binding of the additional electron (or hole) within PBE0 that is responsible for all of the above described results. It is worth noting that the results for charged, defect-free supercells can be important on their own because studies of defects such as the V_k center[15] and other self-traps of carriers[17,32,33] involve such calculations using supercell techniques.

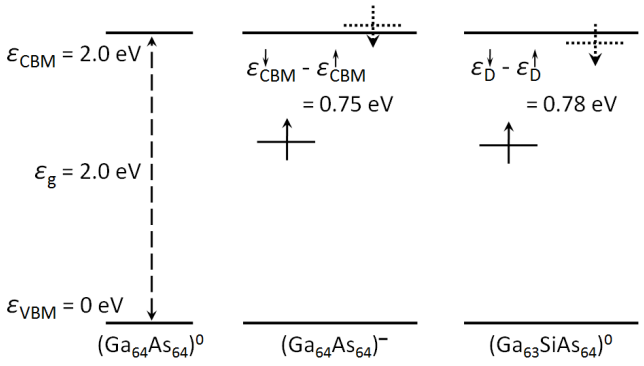

FIG. 4: One-electron levels of neutral (left) and negatively charged (middle) 128-atom GaAs supercell without defect and with the shallow donor GaAs:Si_{Ga} (right) calculated using the PBE0 functional, showing surprisingly large level splittings.

Next, we show that all the issues discussed above are a direct consequence of the HF exchange potential included in the PBE0 functional. The wave function of a shallow defect state is substantially delocalized well beyond the

size of supercells used in first-principles calculations, and the defect state forms a band. This, when coupled with the nonlocal and long-range nature of the HF exchange potential, results in a significant intra-band exchange energy for the (supposedly) isolated defect state represented by a supercell model. It is this intra-band exchange that is responsible for the splitting between the occupied and the unoccupied one-electron levels discussed above. Note that this intra-band exchange shall not be confused with the self-exchange since within the HF theory, the self-exchange and self-Coulomb potentials cancel each other exactly. Of course, other factors, such as correlation and the difference between up-spin and down-spin wave functions, will also contribute to the energy difference. But as we will show below, the intra-band exchange is the major cause for the unphysical results calculated using the PBE0 functional.

We now focus on the additional electron at the CBM. The behavior of this additional electron resembles a fully spin-polarized uniform electron gas (EG), i.e., all up-spin states are occupied whereas all down-spin states are unoccupied. Within the uniform EG model, the exact HF exchange energy is known. If we neglect the subtle difference between the wave functions of the occupied and that of the unoccupied state, within the Hartree-Fock theory, the one-electron energy difference between the unoccupied down-spin state and the occupied up-spin state is

$$\Delta\epsilon_{\rm CBM} = \epsilon^{\downarrow}_{\rm CBM} - \epsilon^{\uparrow}_{\rm CBM} = \alpha\frac{2k_{\rm F}}{\pi}F(\frac{k}{k_{\rm F}}), \qquad (1)$$

where k is the wave vector, $k_{\rm F}$ is the Fermi wave vector, α is the scaling parameter of the HF potential used in the hybrid functional ($\alpha = 0.25$ in PBE0), and $F(y) = \frac{1-y^2}{4y}\ln|\frac{1+y}{1-y}|$. At the Γ point, $\Delta\epsilon_{\rm CBM} = \alpha\frac{2k_{\rm F}}{\pi} = \alpha\frac{41.9}{r_s}$(eV), where r_s (in a.u.) is defined by considering only the additional electron (or hole) and is related to the volume of the supercell via $r_s = (\frac{3V_{\rm cell}}{4\pi})^{\frac{1}{3}}$.

Figure 5 compares the calculated results using hybrid functionals and the result from the EG model for a range of difference systems including Si, GaAs, and ZnO. Let us focus on the PBE0 results (filled symbols) first. It is clear that PBE0 results follow closely the scaling behavior of the EG model discussed above, regardless of if the system is an ideal system with an additional electron or a defective system. Considering the diverse systems that we studied, the quantitative agreement between the PBE0 and EG model is rather impressive. The average difference between the PBE0 result and that from the EG model is about 0.15 eV, which indicates that the long-range HF exchange is indeed the root of the various issues discussed above. The cell-size sensitivity and slow convergence are originated from the scaling behavior of the intra-band exchange energy, which scales as $1/r_s$ or $(1/N_{\rm atom})^{\frac{1}{3}}$ and converges slowly with respect to the cell size. Since the computational cost of hybrid functional calculations (e.g., PBE0) scales roughly as $O(N^4)$, it is extremely difficult to carry out a converged calculation

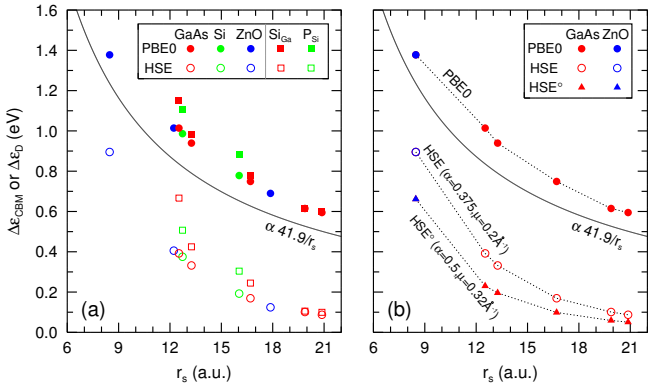

FIG. 5: (Color online) Comparison between results calculated using EG model (solid lines) and PBE0 and HSE functional (symbols) for a range of difference systems including GaAs, Si, and ZnO with respect to r_s. The Si_{Ga} and P_{Si} in figure stand for shallow donors Si in GaAs and P in Si, respectively. The HSE results with different mixing and screening parameters are shown in (b) as discussed in the text.

using the PBE0 functional within current computational capabilities. As demonstrated earlier, unconverged calculations will lead to artificially deep defect transition levels and over-stabilize shallow defects.

Ideally, defects in semiconductors are isolated systems. When they are modeled using the supercell technique, the long-range exchange interaction included in the hybrid functionals poses a serious problem since this fictitious interaction converges extremely slowly with respect to the cell-size, and it is nearly impossible to achieve a converged result within current computational capabilities. Screened hybrid functionals such as HSE may alleviate this problem if the screening length is appropriately chosen. However, one still has to be cautious in interpreting the results obtained from these functionals with adjustable parameters.

Within the HSE functional, an error function is introduced to separate out the long-range part of the HF potential, which is then replaced by the corresponding PBE component. This scheme was originally proposed to improve the efficiency of hybrid functional based calculations for periodic systems. A direct consequence of the removal of the long-range exchange is the reduction in the intra-band exchange energy and a faster convergence with respect to the supercell size for defect-related calculation. Fig. 5 also shows the results calculated using the HSE functional (open symbols), which indeed show that $\Delta\epsilon_{\rm CBM}$ and $\Delta\epsilon_{\rm D}$ are reduced significantly compared with PBE0. Since we use a screening parameter of 0.2 Å^{-1}, the Coulomb potential is effectively cut off at $r \sim$10 Å^{-1} ($\sim$19 a.u.). Therefore, the HSE results should converge at around $r_s = 19$ a.u., as shown in Fig. 5(a).

It would be interesting to investigate the convergence behavior of the HSE functional with respect to the screening parameters. Figure 5(b) compares the results calculated with the PBE0 functional and two versions of

HSE functional with different mixing and screening parameters. The parameters are chosen such that all three functionals give a similar band gap for a given system. It is clear that HSE functional with a larger $\mu = 0.32$ Å^{-1} converges faster.

Since deep centers are usually more localized and one would expect that the results for deep centers converges faster compared with shallow defects. We have tested supercell size dependence for deep defects, e.g., the $(0/-)$ transition level of Cr in GaAs, which is about 0.8 eV from VBM in experiment. The calculated $(0/-)$ level using the PBE0 functional is 1.01 eV with a 64-atom supercell, and 0.95 eV with a 250-atom cell. The size dependence is indeed smaller than those for shallow defects as shown in Fig. 1.

Our results imply that 1) for shallow defects, the PBE or a highly screened HSE functional is preferred over PBE0 functional and 2) the EG model can be used to extrapolate the PBE0 results calculated with finite cell size. It seems that using a screened hybrid functional is a more practical solution for defect calculations. However, one has to be cautious in choosing the mixing and screening parameters since different defects might require different screening parameters. In addition, defect systems are composite systems in which the defect states and the host states may experience different screening. Whether or not the *ad hoc* truncation can yield more reliable result is still subject to further study.

VI. SUMMARY

We have shown that applying hybrid functionals to defect study using the supercell approach is prone to systematic errors: (1) The transition levels for hydrogenic shallow defects can be as deep as several tenths of an eV. (2) Adding an electron (or a hole) to an otherwise ideal semiconductor without defects can introduce an unphysically deep defect-like state in the band gap. (3) Shallow defects can be significantly over-stabilized. (4) Bulk band gap can be severely underestimated if evaluated using total-energy difference and may deviate significantly from the KS gap.

This work is supported by the US Department of Energy under Grant No. DE-SC0002623. The supercomputer time was provided by NERSC under the Grant No. DE-AC02-05CH11231, the CCNI at RPI, and the Center for Computational Research at the University at Buffalo, SUNY.

* Email: pzhang3@buffalo.edu
† Email: zhangs9@rpi.edu

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[4] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[5] Y. Bar-Yam and J. D. Joannopoulos, Phys. Rev. B **30**, 1844 (1984).
[6] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Science **321**, 792 (2008).
[7] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).
[8] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).
[9] T. M. Henderson, J. Paier, and G. E. Scuseria, Phys. Status Solidi B **248**, 767 (2011).
[10] L. Hedin, Phys. Rev. **139**, A796 (1965).
[11] J. L. Lyons, A. Janotti, and C. G. Van de Walle, Phys. Rev. Lett. **108**, 156403 (2012).
[12] P. Ágoston, K. Albe, R. M. Nieminen, and M. J. Puska, Phys. Rev. Lett. **103**, 245501 (2009).
[13] D. O. Scanlon, B. J. Morgan, G. W. Watson, and A. Walsh, Phys. Rev. Lett. **103**, 096405 (2009).
[14] D. O. Scanlon and G. W. Watson, Phys. Rev. Lett. **106**, 186403 (2011).
[15] A. Yu. Kuznetsov, A. B. Sobolev, A. S. Makarov, and M. A. Botov, J. Luminescence **129**, 1937 (2009).
[16] M.-H. Du and S. B. Zhang, Phys. Rev. B **80**, 115217 (2009).
[17] C. Franchini, G. Kresse, and R. Podloucky, Phys. Rev. Lett. **102**, 256402 (2009).
[18] M. Jain, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. Lett. **107**, 216806 (2011).
[19] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[20] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[21] J. Paier, R. Hirschl, M. Marsman, and G. Kresse, J. Chem. Phys. **122**, 234102 (2005).
[22] S. B. Zhang and John E. Northrup, Phys. Rev. Lett. **67**, 2339 (1991).
[23] M. D. McCluskey and E. E. Haller, *Dopants and Defects in Semiconductors* (CRC press, Boca Raton, 2012)
[24] J.-W. Chen and A. G. Milnes, Ann. Rev. Mater. Sci. **10**, 157 (1980)
[25] S. B. Zhang, S.-H. Wei, and A. Zunger, Phys. Rev. Lett. **84**, 1232 (2000).
[26] D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).
[27] S. B. Zhang and D. J. Chadi, Phys. Rev. B **42**, 7174 (1990).
[28] P. M. Mooney, J. Appl. Phys. **67**, R1 (1990).
[29] S. Lany and A. Zunger, Phys. Rev. B **80**, 085202 (2009).
[30] S. Lany and A. Zunger, Appl. Phys. Lett. **96**, 142114 (2010).
[31] P. Broqvist, A. Alkauskas, and A. Pasquarello, Phys. Rev. B **80**, 085114 (2009).
[32] N. Sai, P. F. Barbara, and K. Leung, Phys. Rev. Lett. **106**, 226403 (2011).
[33] D. M. Ramo, A. L. Shluger, J. L. Gavartin, and G. Bersuker, Phys. Rev. Lett. **99**, 155504 (2007).