14
Electrostatic Interactions between Charged Defects in Supercells

Christoph Freysoldt, Jörg Neugebauer, and Chris G. Van de Walle

14.1
Introduction

Theoretical calculations have revolutionized our understanding of the doping behavior in semiconductor materials for electronic and optoelectronic devices [1, 2]. The advent of density-functional theory (DFT), and, more recently, *ab initio* approaches beyond it such as many-body perturbation theory in the *GW* approximation [3] or quantum Monte-Carlo methods [4], has enabled us to study the microscopic details of point defects with almost no *a priori* assumptions. These calculations complement experiment in various ways: they not only help to interpret experimental findings and link them to an atomistic model of the relevant defects, but also provide additional data such as formation energies, geometric structures, or the character of the wavefunctions that cannot be obtained with present-day experimental methods.

The work-horse of these calculations has been DFT with local or semilocal functionals [1, 2]. The defect is usually modeled in a supercell, consisting of the defect surrounded by a few dozen to a few 100 atoms of the host material, which is then repeated periodically throughout space. This allows to employ the highly efficient and thoroughly tested computer codes developed for periodic solids [5]. Recent advances in the theoretical framework tend to maintain the supercell models for the same reason [3, 4]. However, it must be kept in mind that the use of supercells implies that the isolated defect is replaced by a periodic array of defects. Such a periodic array contains unrealistically large defect concentrations, resulting in artificial interactions between the defects that cannot be neglected. These interactions include overlap of the wavefunctions, elastic interactions, and – in the case of charged defects – electrostatic interaction. The focus of this contribution are the electrostatic interactions which typically dominate.

A large variety of approaches to control electrostatic artifacts exists in the literature, as has been reviewed recently by Nieminen [6]. Our aim is not to compare the different approaches, but to work out explicitly all assumptions that are made to estimate defect–defect interactions. This allows – at least in principle – to verify each

of them for any defect or material under consideration, thereby greatly enhancing the reliability of the method.

Supercell calculations for formally charged systems must always include a compensating background charge, since the electrostatic energy of a system with a net charge in the unit cell diverges [7, 8]. It is most common to include a homogeneous background, which is equivalent to setting the average electrostatic potential to zero. By increasing the supercell lattice constant L, the isolated defect limit can be recovered in principle for $L \rightarrow \infty$. Nevertheless, the defect energy converges only slowly with respect to L. The origin of this effect lies in the unphysical electrostatic interaction of the defect with its periodic images and the background, decaying asymptotically as q^2/L, where q is the defect charge [7, 8]. Its magnitude can be estimated from the Madelung energy of an array of point-charges with neutralizing background [7]. Makov and Payne [8] proved for isolated ions that the quadrupole moment of the charge distribution gives rise to a further term scaling like L^{-3}. For realistic defects in condensed systems, however, such corrections, scaled by the macroscopic dielectric constant ε to account for screening, do not always improve the convergence [9–11]. Therefore, the prefactors have often been regarded as parameters to be obtained from fitting a series of supercell calculations [6, 9–12]. Unfortunately, such "scaling laws" require large supercells and may include higher order L^{-n} terms with not well defined physical significance. While it is relatively straightforward to determine the most slowly decaying terms of the relevant interactions (which give the leading terms in a $1/L$ expansion), higher-order terms can have a variety of functional forms. Focussing on electrostatic interactions only, contributions beyond the $1/L$ asymptotic limit arise from finite overlap of the defect charge densities (which have an asymptotically exponential decay), details of the microscopic screening (which decay faster than $1/L^2$, but exhibit oscillatory behavior), non-linear effects, and higher-order moments of the charge distributions. They can however not be separated from the remaining wavefunction overlap errors (asymptotic exponential decay) and the cell-shape variations usually present in "scaling law" approaches to increase the number of available supercells. Representing all these contributions by one L^{-n} term (or a few of them) is clearly a very strong reduction of the underlying complexity and and removes any physical meaning from the L^{-n} prefactor even if individual contributions exhibit such an asymptotic limit. The loss of significance is apparent already for the L^{-n} term in the standard scaling law approach: the fitted prefactor usually deviates from the predictions of macroscopic electrostatic theory, which – at least in the limit of sufficiently large supercells – is the physically correct limit for any localized defect.

Recently, a modified version of the Makov–Payne corrections has been proposed by Lany and Zunger [13, 14]. The approach has been employed very successful in practice. A potential drawback is that the approach proposed in Ref. 14 does not always recover the asymptotic $1/L$ limit of Makov–Payne theory. Significant efforts have been undertaken to assess the applicability of the existing correction schemes, but no clear picture regarding applicability and limitations has emerged so far [9–11, 15, 16].

As an alternative to the homogeneous-background approaches, several authors suggested to modify the computation of the electrostatic potential in the DFT

calculation itself to remove the unwanted and unphysical interactions [17–20]. These approaches can also be regarded as the introduction of a neutralizing surface charge at the boundary of the supercell instead of the homogeneous background when the potential is calculated. Complications arise if the boundary cuts through the material, since the material will try to screen away the electric-field discontinuities.

Recently, we have proposed a scheme that accounts for the electrostatic screening of the defect right from the beginning. A major advantage is that all approximations are well defined and easily verified by the actual calculations. The aim of this paper is to provide a careful discussion of the underlying assumptions, provide examples how the approach can be used in practice, and analyze its performance with respect to supercell size convergence. The aim is not to give a detailed overview over all the alternative approaches, since this would go well beyond the scope of the present paper. A recent overview on alternatives can be found in Ref. [6].

The remainder of this chapter is organized as follows. In Section 14.2, electrostatics and screening in real materials is discussed. We will then summarize in Section 14.2.1 the key steps to derive an explicit and exact expression for the electrostatic artifacts in the supercell approach and discuss tractable approximations that yield a parameter-free correction scheme for these artifacts. Results for two model point defects that have been used already by other groups to study supercell size convergence, namely the Ga vacancy in GaAs and the vacancy in diamond will be discussed in Section 14.3.

14.2
Electrostatics in Real Materials

When a localized charge $q(\boldsymbol{r})$ (total charge q) is introduced into condensed matter, it attracts a screening charge of opposite sign that reduces its long-range potential to $\frac{q}{\varepsilon r}$ where r is the distance from the localized charge.[1] It immediately follows that the amount of the screening charge is $(1-\frac{1}{\varepsilon})q$. For a finite system, the total charge q is conserved, but the screening charge is expelled to the surface. In an infinite system, on the other hand, the screening charge is almost homogeneously taken from the host material, modulated only by the underlying atomic structure. It is worthwhile to note that in the idealized case of an isolated charge in an infinite covalently bound semiconductor, the total screening charge inside any *finite* distance is non-zero, or in other words the screening process does not conserve charge in any *finite* region. This surprising result is a direct consequence of the quantum-mechanical, truly non-local nature of screening in condensed matter [21, 22]. In an ideal ionic material with separated polarizable ions, the screening is effectuated by induced dipoles. Even though the charge on each ion is conserved, there is a net flow of charge along the radial axis. For large distances, the spherically averaged charge distribution, which is relevant for the distance-dependence of screening, approaches the one of an ideal

---

1) In the following, we will define the center of the charge such that the dipole moment of $(q(\boldsymbol{r})-q\delta(\boldsymbol{r}))$ becomes zero.

homogeneous (jellium-like) semiconductor. The reason is that no matter how well the ions are separated in three-dimensional space, since the radial spacing between the ionic shells becomes arbitrarily small with increasing distance, they will necessarily overlap when projected onto the radial coordinate.

In a periodic array of defects the screening charge is taken from the supercell. In consequence, the average electron density far from the defect is shifted from its bulk value by $\frac{q(\varepsilon-1)}{\varepsilon\Omega}$ where Ω denotes the volume of the supercell. This effect is indeed visible in defect calculations in the framework of DFT. We illustrate the difference in the defect charge before and after screening in Figure 14.1 for the vacancy in diamond in the $2-$ charge state. The unscreened charge density

$$q(\mathbf{r}) = 2|\psi_d(\mathbf{r})|^2 \tag{14.1}$$

results from filling the sixfold degenerate defect state ψ_d with two additional electrons. It is clearly localized. The screened charge density is obtained from the change in the self-consistent charge density with respect to the neutral state, and includes all screening effects. It is completely delocalized and indeed approaches the homogeneous limit $\frac{q(\varepsilon-1)}{\varepsilon\Omega}$ far from the defect.

In order to understand the quantum-mechanical nature of screening in a real material in more detail, it is illuminative to decompose the supercell error in the defect formation energy into the energy terms of the underlying electronic-structure calculation such as Hartree energy, kinetic energy, exchange–correlation energy, etc. We report such an analysis in Appendix A for a model system that avoids the electronic-structure complications of real defects. It reveals that the *electrostatic effects are not restricted to the energy contributions formally associated with electrostatics*, but are distributed to all parts of the total energy due to self-consistency. Recovering the electrostatic energy from the self-consistent electron density employing density-based expressions for electrostatic interactions might therefore be very difficult if not
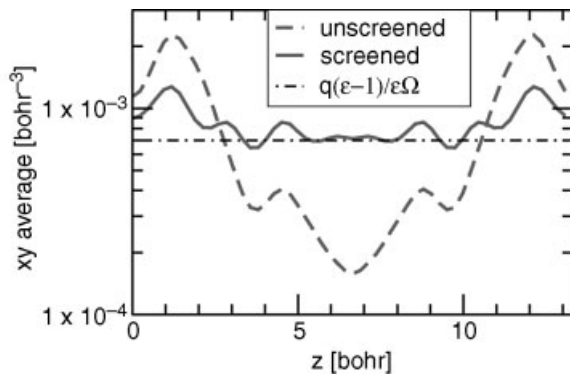

Figure 14.1 (online color at: www.pss-b.com) Comparison of defect charges before and after screening (see text) for the 2− vacancy in diamond in a 64-atom supercell. The defect is located at $z = 0$ with a periodic image at $z = 13.3$ bohr. The limit of a homogeneously distributed screening charge is indicated by the black dash-dotted line.

even impossible. We therefore provide in the following a rigorous and computationally highly efficient scheme that is based on a recently developed approach [23].

14.2.1
Potential-based Formulation of Electrostatics

In view of the aforementioned difficulties when working with the electron density, it turns out to be advantageous to express electrostatic interactions in terms of the unscreened charge density $q(\mathbf{r})$ and the electrostatic potential $V(\mathbf{r})$. To be more specific, let us consider the interaction of a single defect $q(\mathbf{r})$ with one of its periodic images $q_{\mathbf{R}}(\mathbf{r}) = q(\mathbf{r}-\mathbf{R})$. Figure 14.2 depicts this situation graphically. The electrostatic interaction is given by

$$E = \int d^3\mathbf{r}\, q(\mathbf{r}) V[q_{\mathbf{R}}](\mathbf{r}). \tag{14.2}$$

Here, $V[q_{\mathbf{R}}]$ is the potential due to the presence of the charge $q_{\mathbf{R}}$ including all the screening effects. At a sufficient distance from the defect center, the potential approaches its macroscopic value of $1/\varepsilon r$. At which length scale the macroscopic behavior is reached, depends on the localization of the charge and the characteristic screening length of the material under consideration. The potential is further modulated by the effects of microscopic screening (local field effects), in particular due to the underlying atomic structure. However, for the electrostatic energy according to Eq. (14.2) the microscopic details tend to average out. The same is true for the details of the charge distribution if the potential is sufficiently smooth. Note that the modulation amplitude decays faster than $1/r$. The potential therefore becomes smoother as the distance is increased. The electrostatic energy due to the unphysical interaction of the defect charge with its periodic image can then be estimated to good accuracy from a simplified model of the charge density and the long-range potential V^{lr}. The key advantage of this view on electrostatic interactions is that reasonable approximations to the charge distribution and the long-range
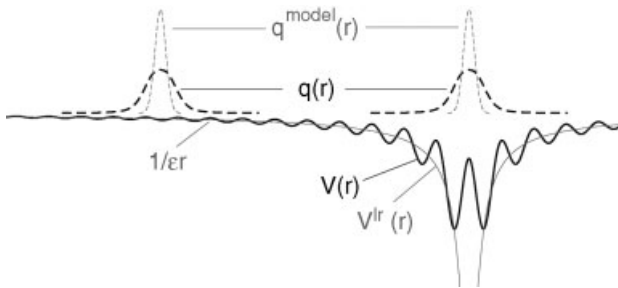

Figure 14.2 (online color at: www.pss-b.com) Potential-based formulation of defect–defect interactions (schematically). Dashed lines indicate the unscreened defect density of a defect (left) and its periodic image (right). The true potential (black solid line) of the image includes local-field effects (wiggles). The major part of the interaction can be captured by simplified models (thin red lines).

potential are easily found. Moreover, the electrostatic potential is a byproduct of a DFT calculation and therefore available at no additional computational cost.

Neglecting the influence of microscopic screening, which typically takes place at a length scale of a few bond lengths, the long-range potential is given by

$$V^{\mathrm{lr}}(\boldsymbol{r}) = \int d^3\boldsymbol{r}\, \frac{q^{\mathrm{model}}(\boldsymbol{r}')}{\varepsilon|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{14.3}$$

It remains to choose an appropriate model charge density q^{model}. For strongly localized defects, a point charge or a Gaussian with a small width (≈ 1 bohr) is for most cases a reasonable choice. For very delocalized states, a more advanced model is needed and will be discussed in Section 14.3.2.

In addition to the interaction of the defect with its periodic images, also the interaction with the homogeneous background must be taken into account. For this, the potential-based expressions for the electrostatic energy are ideally suited. There is no need to introduce approximations here since the homogeneous background density and the defect-induced potential are known exactly. However, the defect images and the background cannot be considered separately as the individual energy contributions would diverge. Instead, we will separate all interactions into a long-range part, that is treated at the model level, and a remaining short-range part. The corresponding expressions will be derived in the following.

14.2.2
Derivation of the Correction Scheme

In this section, tractable expressions are derived for the electrostatic interactions introduced by the supercell approximation. The key idea is to exploit that long-range interactions can be captured with a simplified model as described in Section 14.2.1, and correct for the short-range interactions beyond this in a consistent way.

We start from the defect-induced potential

$$V = V^{\mathrm{els}}(\text{charged defect}) - V^{\mathrm{els}}(\text{reference}). \tag{14.4}$$

Here, V^{els} denotes the DFT electrostatic potential, i.e., the sum of the external (ionic) local potential and the Hartree potential. The scheme can of course be applied to any other electronic-structure method that provides the electrostatic potential. We will start with the neutral defect as reference and discuss the transition to the bulk reference below.

V can be formally split into a long-range and short-range part, i.e.,

$$V = V^{\mathrm{lr}} + V^{\mathrm{sr}}, \tag{14.5}$$

which implicitly defines V^{sr} from V and V^{lr}. The long-range potential V^{lr} is obtained from the model charge density q^{model} via Eq. (14.3). If q^{model} is well chosen and the supercell is large enough, V^{sr} decays to zero within the supercell. In this case, the short-range energy of the defect charge does not differ between the isolated defect and the periodic array of defects. For strongly localized defects, a Gaussian model charge is

usually sufficient to ensure a fast decay of V^{sr}. However, the periodic array also includes a neutralizing background $n = -q/\Omega$, with a short-range interaction energy

$$\int d^3 r n V^{sr}(r) = -q \left[\frac{1}{\Omega} \int d^3 r V^{sr}(r) \right], \tag{14.6}$$

between the background and the defect.

The long-range potential for the periodic array (including the background) is obtained from the Fourier transform of Eq. (14.3) as

$$\tilde{V}^{lr}(G \neq 0) = \frac{4\pi q^{model}(G)}{\varepsilon |G|^2}, \qquad\qquad \tilde{V}^{lr}(0) = 0. \tag{14.7}$$

Note that the homogeneous background does not induce local-field effects; its only role is to cancel the divergence of the $G = 0$ term [23]. The long-range interaction energy can then be estimated from the screened lattice energy (Madelung energy) of the model charge [23]. For spherical charge densities, it can be easily computed in reciprocal space via

$$E^{lat}[q^{model}] = \frac{2\pi}{\varepsilon\Omega} \sum_{\substack{G \neq 0}}^{|G| \leq G_{cut}} \frac{\{q^{model}(|G|)\}^2}{|G|^2} - \frac{1}{\pi\varepsilon} \int_0^{G_{cut}} dg \{q^{model}(g)\}^2, \tag{14.8}$$

where G runs over the reciprocal lattice vectors. The first term in Eq. (14.8) is the energy of the periodic array in its own potential, with a prefactor of $\frac{1}{2}$ to account for double counting. The second term removes the electrostatic interaction energy of the model charge with itself, that is contained in the first term.

With these ingredients, the formation energy of a charged defect in a supercell can be expressed as [23]

$$E_f^q = E_f^0 + \Delta E^{iso}(q) + E^{lat}[q^{model}] - q\Delta, \tag{14.9}$$

where E_f^0 denotes the formation energy of the neutral defect in the same supercell and $\Delta E^{iso}(q)$ the difference in formation energy between an isolated charged defect and the neutral one. The alignment-like term

$$\Delta = \frac{1}{\Omega} \int d^3 r V^{sr}(r) \tag{14.10}$$

is obtained from the short-range potential

$$V^{sr} = \tilde{V}^{els}(charged) - \tilde{V}^{els}(neutral) - \tilde{V}^{lr} + \Delta V, \tag{14.11}$$

where the alignment constant ΔV is chosen such that V^{sr} decays to zero in between the defects [23]. We demonstrate the alignment in Figure 14.3 for the Ga 3 – vacancy in GaAs referenced to the bulk potential (the change of reference from the neutral defect to the bulk is discussed below). For this, the potentials were averaged over the xy plane and plotted as a function of z. The defect-induced potential shows a parabolic shape in between the two defects [23]. This shape is well reproduced by the long-range
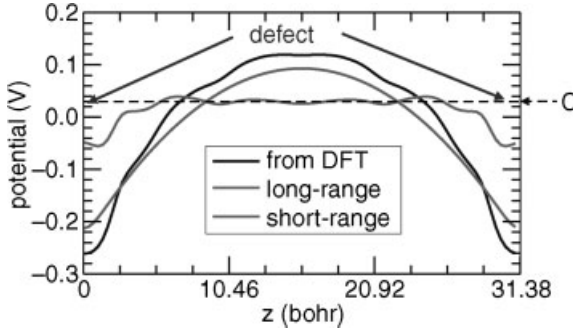
Figure 14.3 (online color at: www.pss-b.com) Potentials (see text, averaged along x, y) for a V_{Ga}^{3-} in a $3 \times 3 \times 3$ cubic GaAs supercell. The defect is located at $z = 0$ bohr with a periodic image at $z = 31.38$ bohr.

model, in this case a 1 bohr wide Gaussian. The difference between the two reaches a plateau at $C = -\Delta V = 0.03$eV. The appearance of the plateau clearly demonstrates that the model has correctly reproduced the long-range tail of the defect potential. Any remaining curvature (apart from the unavoidable modulation due to the underlying atomic structure and/or the screening response) would indicate that the long-range modeling should be improved, e.g., by choosing a model charge that is derived from the actual defect wave function. It should be emphasized that the main approximation in the scheme lies in the neglect of the details of microscopic screening and their coupling to the details of the actual unscreened charge distribution beyond q^{model} for the long-range interactions. These neglected details are by definition cast into V^{sr}. The validity of the central approximation can therefore be easily controlled and checked by verifying that V^{sr} is well behaved.

We now proceed to replacing the neutral defect reference by the bulk, which constitutes the reference of interest in most practical applications. As shown in Appendix B, the potential alignment constant ΔV and the energy alignment constant Δ are related via

$$\Delta = \Delta V. \tag{14.12}$$

For the neutral defect reference discussed up to now, this can be seen as follows: the inclusion of the homogeneous background in the Hartree energy by setting $V(\boldsymbol{G} = 0)$ to zero implies that the average potential does not change between the neutral and the charged defect. Likewise, the average V^{lr} vanishes in this alignment convention. Eq. (14.12) then immediately follows from Eq. (14.11). Equation (14.12) remains valid if the neutral defect reference is replaced by another one, notably the bulk, even if the average alignment changes. The reason is that the alignment of the potentials reflects the dependence of the total energy (and derived quantities) when the formal charge is changed. In contrast to the original formulation [23], where it was incorrectly stated that Eq. (14.10) was to be used for any reference potential, using Eq. (14.12) for the alignment guarantees full internal consistency as outlined in Appendix B.

The formation energy of an isolated defect with charge q then becomes

$$E_f^{\text{iso}} = E^{\text{DFT}}(\text{defect} + \text{bulk}) - E^{\text{DFT}}(\text{bulk}) - E^{\text{lat}}[q^{\text{model}}]$$

$$+ q\Delta V - \sum n_s\mu_s + q(E^{\text{Fermi}} + \varepsilon^{\text{vbm}}), \tag{14.13}$$

where we also included the reference chemical potentials for the chemical species s added ($n_s > 0$) or removed ($n_s < 0$) to form the defect. E^{Fermi} is the Fermi energy relative to the valence band maximum (vbm), and ε^{vbm} is the valence band maximum as obtained from the bulk reference calculation.

14.2.3
Dielectric Constants

The long-range modeling requires the dielectric constant ε of the material at hand. To be consistent with the theoretical framework, the dielectric constant must be computed. This can be done by perturbation theory [24, 25], or by a direct approach [22, 26]. The direct approach is straightforward to apply. For this purpose, a sawtooth potential V^{saw} is applied to an elongated cell [22, 26], typically a $1 \times 1 \times 6$ supercell of the simple-cubic bulk cell. The change in the effective potential ΔV^{SCF} then also shows a sawtooth-like shape, however reduced by the dielectric constant, see Figure 14.4. By comparing the slope of the applied and effective potentials between the turning points, the dielectric constant can be determined as

$$\varepsilon = \frac{\partial V^{\text{saw}}/\partial z}{\partial \Delta V^{\text{SCF}}/\partial z}. \tag{14.14}$$

The direct approach has the advantage to provide immediate insight into the linearity and the screening length of the perturbation applied. It yields very accurate
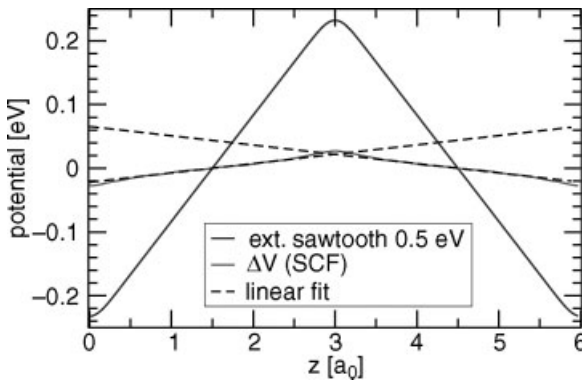

Figure 14.4 (online color at: www.pss-b.com) Determination of the dielectric constant of GaAs. Applying a (smoothened) sawtooth potential V^{saw} (blue) to the material induces a change ΔV^{SCF} in the self-consistent potential (red). The dielectric constant is given by $\frac{\partial V^{\text{saw}}}{\partial z} / \frac{\partial \Delta V^{\text{SCF}}}{\partial z} \approx 12.7$.

results (only 1–2% scatter between different cells and different amplitudes) if the following points are observed: (i) The total height of the sawtooth potential must not exceed the bandgap to avoid a dielectric breakdown, i.e., transfer of valence electrons from near the top of the potential to conduction band states near the bottom of the potential. (ii) The sawtooth potential must be adapted to the symmetry of the system. In particular, turning points should lie on symmetry planes to avoid the induction of additional dipoles at the turning points. If this is not observed, the rising and falling parts of the sawtooth potential may yield different apparent dielectric constants (even negative ones). (iii) The induced potential fluctuations are very soft modes. Therefore, the convergence criteria for the self-consistent field calculations must be very tight to yield accurate results. A helpful check is to compare the dielectric constants of the rising and falling parts of the potential. (iv) The induced potentials are modulated by the underlying atomic structure. In order to average out the modulations (also known as local-field effects), the slopes must be determined over full periods of these modulations. We also note from Figure 14.4 that there are deviations from the macroscopic screening behavior close to the turning points which extend over a range of $a_0/2$. The deviations reflect the finite screening length, and must of course be excluded from the slope fitting.

If ionic relaxations are taken into account for the defect calculation, the dielectric constant employed for the correction scheme must also reflect ionic screening. In the sawtooth approach, this implies that ionic relaxation must be included [27].

14.3
Practical Examples

In the following, we discuss the application and performance of the correction scheme for two representative examples. Specifically, we will consider the Ga vacancy in GaAs as a deep, well localized defect, and the carbon vacancy as a shallow, rather delocalized defect state to test and discuss the limits of the corrections. The calculations were performed in the framework of DFT in the local-density approximation, employing plane waves and normconserving pseudopotentials as implemented in the SPHInX code [28]. A plane-wave cutoff of 20 Ry for GaAs and 40 Ry for C was found to yield sufficient accuracy. In order to disentangle the electrostatic from strain effects, the ions were not relaxed. The supercell artifacts due to wavefunction overlap, on the other hand, cannot be avoided, but were minimized by a constant occupation scheme [2].

14.3.1
Ga Vacancy in GaAs

The first example is the Ga vacancy in GaAs. It shows a deep level in the band gap that supports – in its unrelaxed geometry – charge states between 0 and 3–. The corrections are obtained from a 1 bohr wide Gaussian model charge. The calculated

formation energies for the four charge states are displayed in Figure 14.5 for a series of supercells. The uncorrected formation energies show a strong dependence on the supercell size. Since electrostatic interactions scale with q^2, it is most prominent for the $q = -3$ case shown in Figure 14.5a. This dependence is largely removed by including the correction energies according to Eq. (14.13). The remaining scatter in the data is ~0.1 eV, see Figure 14.5b. It does not scale with the charge state and is present even in the neutral state. This suggests that other than electrostatic effects are responsible. Since strain effects have been excluded by not considering ionic relaxation, it is probably caused by the overlap of the wavefunctions and the resulting Pauli repulsion. Clearly, corrections aiming at electrostatic interactions cannot and should not capture such effects.
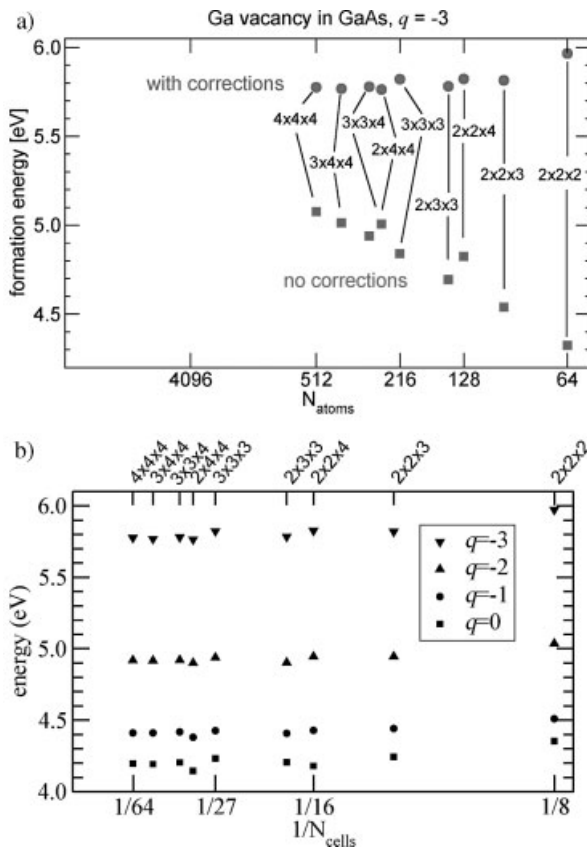

Figure 14.5 (online color at: www.pss-b.com) Formation energies of the Ga vacancy for a variety of supercells. The Fermi energy is set to the valence band maximum. (a) Comparison of uncorrected and corrected value for $q = -3$. (b) Corrected values for $q = 0 \ldots -3$.
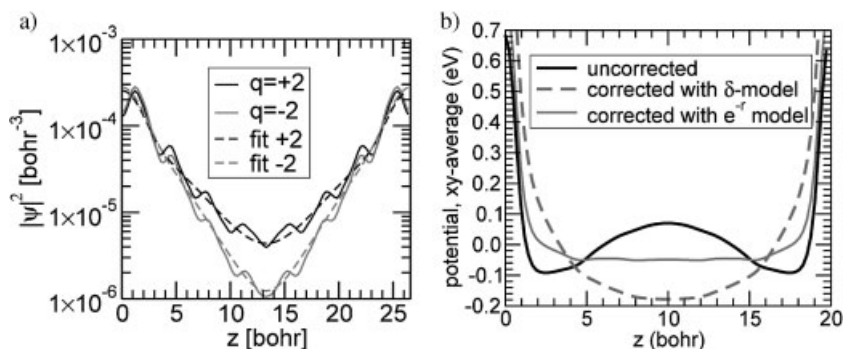
Figure 14.6 (online color at: www.pss-b.com) Localization analysis of the carbon vacancy. Shown is the *xy*-plane average of (a) the defect wavefunction, and (b) the defect-induced potentials. The defect is located at $z=0$ with a periodic image at $z=26.6$.(a) Shape of the defect state $|\psi(r)|^2$ for two different charge states averaged over the Brillouin zone and the three defect bands (solid lines). (b) Comparison of the defect-induced potential for the $2+$ case without long-range corrections (black line), with long-range corrections from a 1 bohr wide Gaussian (red dashed), and from the fitted exponential-tail model (green).

14.3.2
Vacancy in Diamond

The second example is the vacancy in diamond. This defect has been used previously to study finite-size effects in supercells by Shim *et al.* [10]. The authors found that the $2-$ charge state is easily corrected by a Makov–Payne-like scheme, while the $2+$ charge state shows a completely different behavior. This discrepancy was tentatively attributed to a qualitatively different screening for the two cases [10]. Such defect-induced changes in the electrostatic screening, however, are expected to converge proportional to the inverse volume of the supercell if the defect can be regarded as an impurity with a changed polarizability compared to the host material.

When the correction scheme of Section 14.2.2 is applied to the $2+$ case, the alignment procedure immediately indicates problems to reproduce the long-range potential, see Figure 14.6b. The "short-range" potential shows a significant over-correction when a Gaussian charge model is employed. In order to derive an improved charge model, let us analyze the defect states. In the long-distance limit, every defect state shows an exponential decay behavior with increasing distance of the defect center. Indeed, as shown in Figure 14.6a, the shape of the $2+$ and $2-$ defect states of the diamond vacancy reveals a significant contribution of this exponential tail to the overall wavefunction. While the states show a close match near the defect center, the exponential decay differs significantly for the two states. This can be attributed to the position of the levels in the band gap. The $2-$ Kohn–Sham level is located around midgap (cf. Figure 14.7, the Kohn–Sham level lies half-way between the $-/2-$ and $2-/3-$ transition levels of the corresponding supercell). The state thus decays very quickly. The $2+$ level is close to the valence band, and the associated state then starts to hybridize with the valence band states, leading to a delocalization that is sizeable and cannot be neglected even at the boundaries of the supercell.
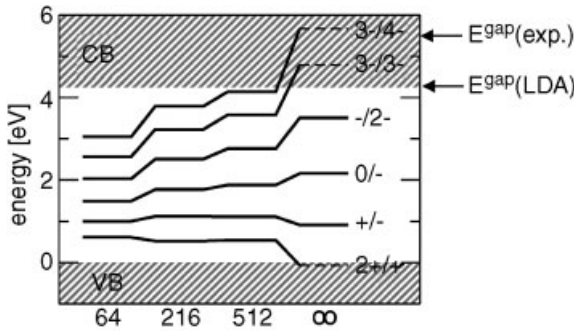
Figure 14.7 (online color at: www.pss-b.com) Charge transition levels for the unrelaxed vacancy in diamond as obtained from different supercells (size of 64, 216, 512 atoms) without correction compared to the corrected ones (∞).

To take the exponential tail into account, the defect charge can be modeled by the radial ansatz

$$q(r) = qxN_\gamma e^{-r/\gamma} + q(1-x)N_\beta e^{-x^2/\beta^2}. \tag{14.15}$$

N_β and N_γ denote the normalization constants for the exponential and the Gaussian, respectively. The decay constant γ and the tail weight x is obtained by fitting the wavefunctions of the $4 \times 4 \times 4$ (512 atom) cell. The exact value of β turns out to be relatively unimportant as long as the Gaussian stays localized; here, a value of 2 bohr was used. To ensure that this ansatz can be generally applied, we used it for all charge states, even though a Gaussian-only model works reasonably well for the more localized mid-gap states $1 \geq q \geq -2$. Interestingly, all states have essentially the same tail weight ($x = 54$–60%). The decay constants, on the other hand, sensitively depend on the energetic distance to the valence or conduction band edge. They are listed in Table 14.1 along with the uncorrected and corrected values for cubic supercells of 64, 216, and 512 atoms. It is apparent that the exponential-tail model can be equally applied to all charge states. The supercell corrections reduce the errors in the calculated formation energies from up to 7 eV (4−, 64 atoms) to ∼0.1 eV. The charge transition levels derived from the supercells without the corrections as well as from the extrapolated values for the isolated defect are visualized in Figure 14.7. It is noteworthy that the corrections push the transition levels to high charge states into the valence band (2 + / +) and the conduction band (2−/3− and 3−/4−), respectively. The occurence of high charge states thus turns out to be an artifact of the small supercells. This tendency to push the charged levels away from the neutral one is a general feature of supercell corrections whenever the Madelung term dominates.

The examples shown here illustrate the importance of supercell corrections for predicting the correct position of defect levels, sometimes even qualitatively. The charge correction scheme presented here is able to remove the dominating electrostatic artifacts from finite-size supercell calculations. In combination with the constant-occupation scheme, the agreement between different supercells is typically on the order of 0.1 eV, and thus considerably smaller than the errors that arise from the DFT framework employed. We therefore expect that major improvements over

Table 14.1 Formation energies of a vacancy in diamond for N × N × N supercells in different charge states q with and without supercell corrections. Column 2 lists the defect states' decay constants γ (see text). The Fermi energy is set to the valence band maximum.

q	γ	uncorrected			corrected		
		$N=2$	$N=3$	$N=4$	$N=2$	$N=3$	$N=4$
$2+$	2.41	6.65	6.49	6.63	7.69	7.45	7.44
$+$	2.06	7.26	7.11	7.17	7.43	7.33	7.37
0	1.85	8.26	8.23	8.28	8.26	8.23	8.28
$-$	1.77	9.75	10.00	10.16	10.40	10.38	10.44
$2-$	1.81	11.78	12.51	12.92	13.93	13.91	13.95
$3-$	2.03	14.35	15.74	16.50	18.77	18.73	18.74
$4-$	2.56	17.40	19.55	20.64	24.55	24.54	24.41

the present state-of-the-art in defect calculations will come from the application of advanced electronic-structure methods in the supercell approach.

14.4
Conclusions

In this paper, we have presented an analysis of supercell artifacts in charged point defect calculations arising from electrostatic interactions. For the electrostatics in real materials, an exact, potential-based formulation overcomes the limitation of previous correction schemes that rely on *a priori* simplifications such as the restriction to macroscopic screening and point-like charge densities. A new correction scheme emerges from this analysis. The scheme itself requires no empirical parameters or fitting procedures, and requires only a single supercell calculation. It employs certain simplifications to model the long-range interaction, but in contrast to other approaches, the validity of these approximations can be verified immediately by visually inspecting the short-range potential. If needed, the underlying charge model can be refined in a straightforward manner. It should be emphasized that the correction scheme does not rely on DFT, but can be applied to any electronic-structure theory that provides the electrostatic potential. We believe that this approach may help to reduce the errors induced by approximate correction schemes and thus extend the applicability of advanced methods to charged defects, even if the supercells that can be afforded with these methods introduce significant electrostatic interactions.

Acknowledgements

Appendix

(A) Energy Decomposition of Electrostatic Artifacts in DFT

In order to understand the screening behavior in more detail, let us consider a model system that avoids the electronic-structure complications of real defects. For this purpose, a Gaussian charge is placed in an interstitial site in GaAs. (As anti-bonding, tetrahedrally surrounded by As.) A charge of -1 ensures that this model charge repels the surrounding electrons, preventing the formation of any localized electronic bounding states. We then calculated the self-consistent total energy within DFT in 22 different supercells using norm-conserving, non-local pseudopotentials. The calculated defect formation energy

$$E_d = E_{tot}(\text{defect}) - E_{tot}(\text{bulk}) \qquad (14.16)$$

depends on the supercell. Figure 14.8 shows a decomposition of how the error in the formation energy is related to the various contributions to the Kohn–Sham functional. For this, we plot the change in each energy contribution with respect to the corresponding bulk as a function of the error of the defect formation energy with respect to the isolated case, i.e., the supercell error. We note that positive supercell errors occur for elongated cells, negative for cubic supercells. The energy of the isolated case was obtained with the defect correction scheme described in Section 14.2.2. If a single energy contribution were responsible for the supercell error, it would appear as a straight line of slope $m = 1$. It becomes apparent that the energy contributions vary non-linearly with the supercell error. The electrostatic part
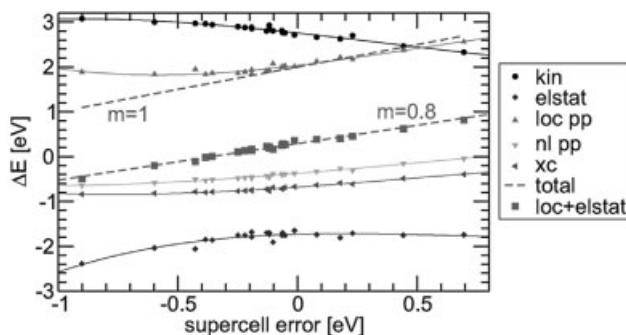

Figure 14.8 (online color at: www.pss-b.com) Decomposition of the defect formation energy into the various energy contributions to the Kohn–Sham functional for different supercells: kinetic (kin), Hartree + long-range pseudopotential (elstat), short-range local pseudopotential (loc pp), non-local pseudopotential (nl pp), exchange–correlation (xc). The supercell error corresponds to the sum of all contributions (red line), shifted by the isolated defect limit. The solid curves are polynomial fits to highlight the trends. Perfect correlation between an energy contribution and the supercell error corresponds to a line of slope $m = 1$.

of the Kohn–Sham functional can be identified with the sum of Hartree and local pseudopotential contributions. This sum varies almost linearly with the supercell, but with a slope of only $m = 0.8$. In other words, the electrostatic part of the functional accounts for only 80% of the electrostatic supercell error. This highlights that the *electrostatic effects are not restricted to the energy contributions formally associated with electrostatics*, but are distributed to all parts of the total energy due to self-consistency.

(B) Alignment Issues in Supercell Calculations

Depending on the computer code implementation at hand, the treatment of the local pseudopotential may influence the potential alignment and thus the total energy for charged systems. It is common practice to split off the long-range part of the pseudopotential by subtracting the potential of a compensation charge (e.g., a Gaussian), adding the compensation charge to the charge density used for the Hartree energy, and correcting for the added self-energy of the compensation charges [29]. The remaining short-range pseudopotential may now be used for the electron density only [29] (which is also the case for the SPHInX pseudopotential code [28] employed for the examples in Section 14.3) or for the neutral "Hartree density" including the compensation charges, with a corresponding correction term for each atom [30]. In the latter case, one may alter the alignment of the periodic superposition of the short-range potentials, and in particular set its average to zero and thus include the neutralizing background. Only in this alignment convention, the energy of a charged system becomes independent of the choice of the compensation charge. Otherwise the potential shift acting on the electrons is only compensated in part by the self-energy of the compensation charges.

In this general case, Δ must be determined as follows. The alignment convention of setting the average of the periodic Hartree potential to zero is the consistent way to include a neutralizing background in the Hartree energy. For the neutral reference, Δ then becomes identical to ΔV: the Hartree alignment convention implies that the average potential does not change between the neutral and charged defect (the pseudopotentials and compensation charges are the same in both cases). Likewise, the average $\tilde{V}^{lr}$ vanishes in this alignment convention. The identity $\Delta = \Delta V$ (Eq. (14.12)) then immediately follows from Eq. (14.11). Using Eq. (14.12), we may now replace the neutral defect reference by any other, notably the bulk, even if the alignment changes. The crucial point is that the alignment is not changed arbitrarily, but consistent with the total-energy expressions and quantities derived from it. For instance, a change in the bulk alignment by an amount A will change ΔV by $+A$, which compensates the shift in the valence band energy $\varepsilon^{vbm} = \frac{\partial E}{\partial f_{vbm}}$ (f_{vbm} is the occupation number of the valence band maximum) appearing in Eq. (14.13) by $q^{electron} A = -A$. On the other hand, a shift of B in the defect potential (due to altered compensation charges) will change ΔV by $-B$, but also the total energy of the charged defect calculation by qB.

Moreover, Eq. (14.12) as the general definition of Δ allows for changing the $\tilde{V}^{lr}$ alignment convention defined by Eq. (14.7). In Eq. (14.7), the $\boldsymbol{G} = \boldsymbol{0}$ component was set to zero, consistent with excluding the $\boldsymbol{G} = \boldsymbol{0}$ component in Eq. (14.8). This choice

corresponds to including the interaction of the model charge with the compensating background in the lattice energy. The lattice energies of non-overlapping charge densities therefore depend on the actual shape of the charge distribution. This is compensated for by a corresponding change in the depth of the short-range potential according to Eq. (14.11), and thus the $-q\Delta$ term in Eq. (14.9). An alternative is to modify the definition of the Madelung energy [Eq. (14.8)] and the associated potential [Eq. (14.7)] consistently. Using the analytic limit

$$V_0 = \lim_{\boldsymbol{G} \to 0} \; V(\boldsymbol{G}) = \frac{2\pi}{\varepsilon} \frac{\partial^2 q^{\mathrm{model}}(g)}{\partial g^2}\bigg|_{g=0},\tag{14.17}$$

for the alignment convention, the modified definitions are

$$\tilde{V}^{\mathrm{lr}}(\boldsymbol{G} = 0) := V_0,\tag{14.18}$$

$$E^{\mathrm{lat}} := E^{\mathrm{lat}}[\mathrm{Eq.}(15.8)] + qV_0.\tag{14.19}$$

The lattice energy of Eq. (14.19) is corrected for the interaction with the background density and coincides with the lattice energy of an array of point charges (provided that the model charges do not overlap in the periodic array). Using Eqs. (14.18) and (14.19) instead of Eqs. (14.7) and (14.8) removes the shape dependence of the individual contributions (lattice energy and alignment term) to the defect energy correction in the non-overlapping case. The realignment nicely highlights the need for a consistent treatment of energies and potentials in this context, but is not needed in the practical approach.

References

1 Seebauer, E.G. and Kratzer M.C. (2006) *Mater. Sci. Eng. R,* **55**, 57.

2 Van de Walle C.G. and Neugebauer J. (2004) *J. Appl. Phys.,* **95**, 3851.

3 Rinke, P., de Walle, C.G.V., and Scheffler, M. (2009) *Phys. Rev. Lett.,* **102**, 026402.

4 Needs, R.J. (2007) in: *Theory of Defects in Semiconductors, Topics in Applied Physics,* Vol. 104, (eds D.A. Drabold and S.K. Estreicher, (Springer, Berlin, Heidelberg, New York), p. 141.

5 http://www.dft.sandia.gov/Quest/DFT_codes.html.

6 Nieminen, R. (2009) *Modell. Simul. Mater. Sci. Eng.,* **17**, 084001.

7 Leslie M. and Gillan M.J. (1985) *J. Phys. C, Solid State,* **18**, 973.

8 Makov G. and Payne M.C. (1995) *Phys. Rev. B* **51**, 4014.

9 Wright A.F. and Modine N.A. (2006) *Phys. Rev. B* **74**, 235209.

10 Shim J., Lee E.K., Lee Y.J., and Nieminen R.M., (2005) *Phys. Rev. B,* **71**, 035206.

11 Castleton, C.W.M., Höglund, A., and Mirbt, S. (2006) *Phys. Rev. B,* **73**, 035215.

12 Castleton, C.W.M., Höglund, A., and Mirbt, S. (2009) *Modell. Simul. Mater. Sci. Eng.,* **17**, 084003.

13 Lany, S. and Zunger, A. (2008) *Phys. Rev. B,* **78**, 235104.

14 Lany, S. and Zunger, A. (2009) *Modell. Simul. Mater. Sci. Eng.,* **17**, 084002.

15 Lento, J., Mozos, J.L., and Nieminen, R.M. (2002) *J. Phys.: Condens. Matter* **14**, 2637.

16 Gerstmann, U., Deák, P., Rurali, R., Aradi, B., Frauenheim, T., and Overhof, H. (2003) *Physica B,* **340–342**, 190.

17 Carloni, P., Blöchl, P., and Parinello, M. (1995) *J. Phys. Chem.,* **99**, 1338.

18 Schultz, P.A. (2000) *Phys. Rev. Lett.,* **84**, 1942.

19 Rozzi C.A., Varsano D., Marini A., Gross E.K.U., and Rubio A. (2006) *Phys. Rev. B,* **73**, 205119.

20 Schultz, P.A. (2006) *Phys. Rev. Lett.,* **96**, 246401.

21 Resta, R. (1994) *Rev. Mod. Phys.,* **66**, 899.

22 Resta, R. and Kunc, K. (1986) *Phys. Rev. B,* **34**, 7146.

23 Freysoldt, C., Neugebauer, J., and Van de Walle, C.G. (2009) *Phys. Rev. Lett.,* **102**, 016402.

24 Baroni, S. and Resta, R. (1986) *Phys. Rev. B,* **33**, 7017.

25 Gonze, X. and Lee, C. (1997) *Phys. Rev. B,* **55**, 10355.

26 Kunc, K. and Resta, R. (1983) *Phys. Rev. Lett.,* **51**, 686.

27 Pham, T.A., Li T., Shankar, S., Gygi, F., and Galli, G. (2010) *Appl. Phys. Lett.,* **96**, 062902.

28 Boeck, S., Freysoldt, C., Ismer, L., Dick, A., and Neugebauer, J. (2011) *Comput. phys. commun.*, **182**, 543, Web page: http://www.sphinxlib.de.

29 Bockstedte, M., Kley, A., Neugebauer, J., and Scheffler, M. (1997) *Comput. Phys. Commun.,* **107**, 187.

30 Kresse, G. and Joubert, D. (1999) *Phys. Rev. B,* **59**, 1758.