Density Functional Theory Calculations of Defect Energies Using Supercells.

Christopher WM Castleton

School of Science and Technology, Nottingham Trent University, NG11 8NS, UK

E-mail: `Christopher.Castleton@ntu.ac.uk`

A. Höglund and S. Mirbt

Department of Physics, Uppsala University, Box 530, SE-75121, Uppsala, Sweden

Abstract.   Reliable calculations of defect properties may be obtained with density functional theory using the supercell approximation. We systematically review the known sources of error and suggest how to perform calculations of defect properties in order to minimize errors. We argue that any analytical error-correction scheme relying on electrostatic considerations is not appropriate to derive reliable defect formation energies, especially not for relaxed geometries. Instead we propose finite size scaling of the calculated defect formation energies, and compare the use of this with both fully converged and "Gamma" (Γ) point only k-point integration. We give a recipe for practical DFT calculations which will help to obtain reliable defect formation energies and demonstrate it using examples from III-V semiconductors.

1. Introduction

The prediction of defect properties is still a significant challenge within condensed matter theory and computation. Although defect properties have been studied for the last 70 years, the reliability of calculations has always been limited by the state-of-the-art in the approximations used. For many years defects with shallow levels where modeled as hydrogen atoms and solved within the effective mass approximation,[1] but with the advent of faster computers it has become possible to perform first principles predictions of the individual properties of a wide variety of defects in many different materials.

In this paper we focus on the use of density functional theory in predicting defect properties. Density Functional Theory[2] (DFT) is one of the most versatile atomic scale tools available for the investigation of defect properties in semiconductors and insulators. The key quantity to calculate is the defect formation energy

$$E_{\mathrm{d}}^{\mathrm{C}} = E_{\mathrm{T}}^{\mathrm{C}}(\text{defect}^q) - E_{\mathrm{T}}^{\mathrm{C}}(\text{no defect}) + \sum_i \mu_i n_i - q\left(\epsilon_v + \epsilon_F\right) \tag{1}$$

where $E_{\mathrm{T}}^{\mathrm{C}}(\text{defect})$ and $E_{\mathrm{T}}^{\mathrm{C}}(\text{no defect})$ are the total energy of the supercell "C" with and without the defect, (of charge q,) calculated using the same basis set or planewave cutoff, k-point grid, etc, to make use of the cancellation of errors. The defect is formed by adding/removing n_i atoms of chemical potential μ_i. ϵ_F is the Fermi level, measured from ϵ_v, the valence band edge (VBE).
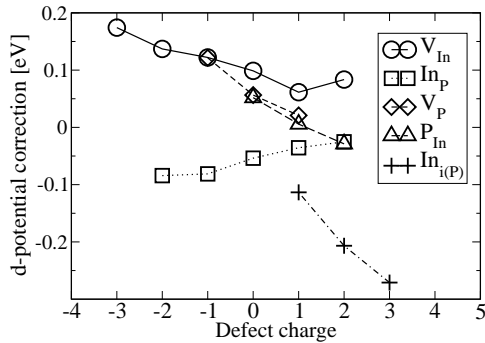
Figure 1. Corrections to the defect formation energy arising from the treatment of the In-$4d$-electrons, as a function of the defect's charge. On the y-axis we plot the difference in formation energy between using an US-PP with In-$4d$ treated as valence and that obtained using an US-PP with In-$4d$ treated as core. The lines are a guide to the eye.

Almost all properties of a defect can be derived from variations in and differences between formation energies.

One major drawback of DFT is that the standard tractable non-empirical approximations to it - the Local Density Approximation (LDA) and the Generalized Gradient Approximation (GGA) strongly underestimate HOMO-LUMO and bandgaps, usually by around 50%. This complicates the prediction of defect properties because many of them are influenced by the energetic position of the defect level in relation to the band edges.

Tractable DFT calculations are limited to clusters of relatively small numbers of atoms - at present 100s or 1000s. This means that calculated defect concentrations are much higher than those found in experiment. For example, to calculate a realistic dopant concentration on the order of 10^{18} cm^{-3}, a supercell size of close to 20,000 atoms would have to be used. The limited number of atoms has another, even more serious effect: the boundary conditions become all-important. One of the most common approaches is to use Periodic Boundary Conditions (PBCs) together with a plane wave basis set.[3] For defects this means use of the supercell approximation: a cluster of atoms containing the defect, is periodically repeated throughout space. The cell boundary thus looks bulk-like, and the computationally helpful translational symmetry is restored. However, it also means that the defect interacts with an infinite array of images of itself seen in the PBCs. These defect/image interactions alter the calculated $E_{\mathrm{d}}^{\mathrm{C}}$, making them (and most other defect properties) supercell size dependent.

From our studies[4, 5, 6, 7] of defect properties in InP and GaP we have concluded that it is possible to use DFT with PBCs and still obtain reliable results for defect properties. We have found that finite size scaling calculated defect formation energies as a function of supercell size improves the agreement with experiment tremendously. It also allows us to estimate an error bar on our corrected energies. Finite size scaling means the following: We calculate the same defect formation energy in at least three supercells with different numbers of atoms but with the same symmetry. Performing a fit as a function of the inverse supercell size we then extract the defect formation energy for a supercell of infinite size. One example of the importance of finite size scaling is the substitutional Zn defect in InP. In industrial applications, Zn doping of InP gives rise to p-type doping. But in calculations in supercells of up to at least 512 atoms (and probably up to several 10,000s), the most stable Zn defect has a charge state of +2, which would mean that Zn gives rise to n-type doping. Only after finite size scaling does the -1 charged substitutional site becomes preferred, predicting p-type doping in agreement with experiment.

In addition to finite size scaling, other corrections have to be taken in account. We divide the errors inherent in calculations using DFT-LDA with PBCs in three categories:

Firstly we have the errors which are not dependent on the size of the supercell used, such as errors due to the pseudopotential approximation or to the bandgap underestimation. These

errors can be calculated once for every defect in a supercell of moderate size and then used to correct the results in all other supercells. We have concluded (and show later) that the bandgap underestimation is best corrected for by applying the scissors scheme.

Secondly we have the errors which depend in a regular and often monotonic way on the supercell size. These errors may be referred to as "rational" in the sense that a functional form should in principle exist for them. They are principally those errors which are due to the interaction of the defect with it's images. These errors can be eliminated (up to a certain accuracy) by finite size scaling the DFT results from different supercells.

Thirdly we have errors which vary in an irregular "non-rational" way on the supercell size - in particular the errors due to the choice of k-point mesh. These need to be treated in each supercell individually using, for example, k-point convergence enhancement recipes.

This paper is organized as follows: In section II we list the computational details. In section III we describe the sources and types of errors. In section IV we describe the main proposed solutions to treat the supercell approximation errors and discuss their advantages and disadvantages. In section V we use finite size scaling to examine the bandgap problem. In section VI we give a recipe for practical calculations of defect properties and in section VII we conclude with a summary.

2. Computational Details

We use planewave ab initio DFT [2] within the LDA together with ultrasoft pseudopotentials [8] (US-PP) using the VASP code.[9] The choice of LDA rather than GGA is because it produces slightly better lattice parameters and bandgaps than GGA for the materials considered here. For InP (GaP) with LDA, a planewave cutoff energy of 200 eV (250 eV) was previously found [10, 6] sufficient to restrict errors to O(0.01 eV) or less. For InP (GaP) our optimized LDA lattice constant using these chosen pseudopotentials[10, 6] is 5.827 Å (5.39 Å) and the bandgap is 0.667 eV (1.42 eV), compared to 5.869 Å (5.45Å) and 1.344 eV (2.35 eV) in experiment. We chose chemical potentials corresponding to stoichiometric conditions, which for InP means μ_P = 6.243 eV and μ_{In} = 3.485 eV and for GaP means μ_P = 6.47 eV and μ_{Ga} = 4.06 eV.

We report both non-relaxed (ions at ideal lattice sites) and relaxed calculations. Except where otherwise stated, no restrictions are placed upon the symmetry of relaxations, but we do not allow atoms located on the boundary of the cell to relax. Charged defects were calculated with a uniformly charged jellium background to ensure convergence of the total energy.

3. Sources of Errors

3.1. Errors Not Dependent on Supercell Size

3.1.1. Pseudopotential Using a plane wave basis code, the accuracy of the results obtained is, of course, limited by the accuracy of the treatment used for the core electrons. We generally use Ultra-Soft Pseudopotentials (US-PPs), but the Projector Augmented Wave method[11] (PAW) produces more accurate results than any pseudopotential method, since it reconstructs the exact valence wavefunctions, with all nodes in the core region. The replacement of US-PP by PAW produces only small (O(0.1 eV) or less) changes in the formation energies of the native defects in InP, with virtually no supercell size dependence. The largest difference we have seen between the correction in the 8 atom cell and that in the 64 atom cell is only 0.01 eV [5]. This demonstrates that the widely used US-PP[8] are perfectly reasonable for this type of calculation.

For InP an additional question arises regarding the treatment of the $4d$-electrons of In. Treatment as core is computational preferable, allowing us to tackle larger supercells, (and hence get further with the supercell approximation errors,) but it does introduce additional errors of it's own. However, we find that these are also largely independent of the supercell size[5]. In Fig.1 we show the difference in defect formation energy between In $4d$-electrons treated as valence and treated as core (d-potential correction) as a function of the defect's charge. The d-potential
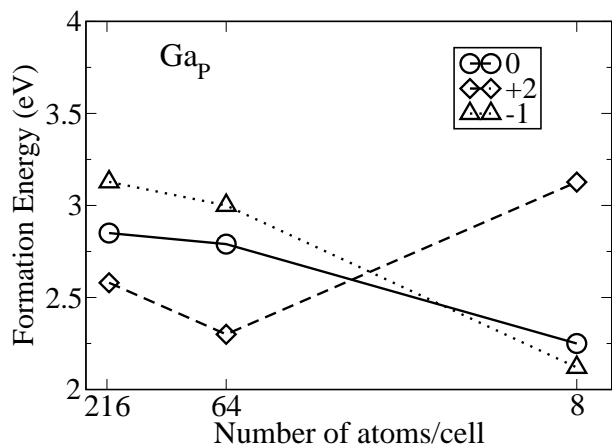
Figure 2. Calculated formation energy of the Gallium antisite in GaP, Ga_P, as a function of the number of atoms per supercell for three different charge states. The lines are a guide to the eye.

correction does vary with the charge, which gives rise to a correction for the transfer levels as well. The d-potential correction changes monotonically with the defect's charge state, but comparing different defects there is no general pattern. In addition, of course, the d-potential correction differs between a calculation allowing for relaxations and a calculation assuming ideal atomic positions. In general we find the correction to become smaller after relaxations are included.

3.1.2. Affect of the choice of Exchange-Correlation potential on formation energies and bandgaps
The errors arising from the exchange-correlation functional have two main forms. Firstly, the bandgap is (usually) strongly reduced compared to experiment. This reduction leads to ambiguities in the definition of the formation energy for charged defects, in turn leading to large uncertainties in predictions. For some semiconductors the bandgap can even be reduced to zero making the material appear metallic and dramatically altering the properties of many defects. The influence of the underestimation of the bandgap on the prediction of defect properties is further discussed in section V.

A second exchange-correlation related error is also present: LDA overbinds all bonds, while GGA underbinds them. An exact solution to the DFT formation energies would lie somewhere in between, but any tractable solution will do one or the other, moving some defect formation energies up and others down. To quantify the effect of this, we have compared some of our LDA results with results from GGA, finding errors of O(0.1 eV), with no significant supercell size dependence.

3.1.3. Spin-polarization We have not allowed spin polarization in our calculations, using LDA instead of the Local Spin Density Approximation (LSDA). However, for most of our defects the use of LSDA would simply cause further finite size errors since it would introduce spurious magnetic interactions between the defect and it's PBC images. The exceptions occur for the Jahn-Teller active defects[12] (In_P, V_P and V_{In} in InP, for example,) where the degeneracies can in certain cases be lifted by Hund's rule couplings. However, this is generally only important for materials[13] with more tightly localized bonds or dangling bonds than we have in InP and GaP.
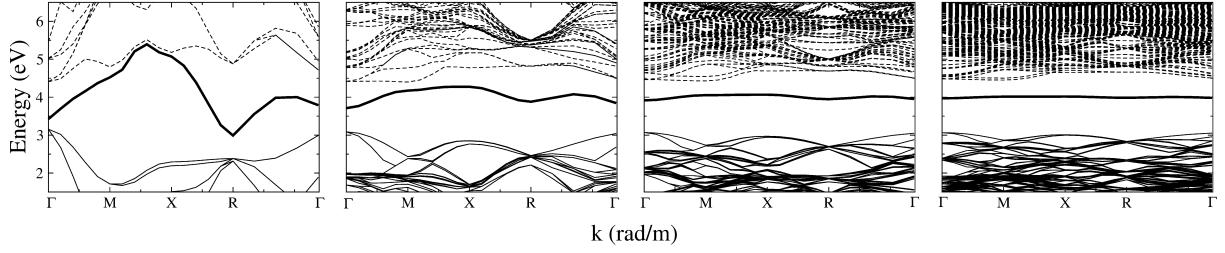
Figure 3. Band structure of the neutral P antisite in GaP. The deep defect level is shown as the thick black line and its dispersion is presented for the 8, 64, 216, and 512 atom supercells.

3.2. Rational Supercell Size Dependent Errors

As described in the introduction, one of the largest problems with planewave DFT studies of defects is that the formation energies, E_d^C, and indeed most other properties, become dependent upon the size of the supercell used. The resulting errors (compared to the results for genuinely single defects) can also be referred to as supercell approximation errors, or as finite size errors. In Fig. 2 we show the effect of the supercell size on the defect formation energy of the Gallium antisite in GaP, for three different charge states, $0, +2$, and -1. The "true" defect properties are only recovered in the limit of an infinitely large supercell, equivalent to the limit of an isolated defect. This problem is particularly severe in the case of charged defects, where the Madelung energy becomes infinite if the charge is not neutralized using a uniform jellium background[14]. Even with jellium, the calculated formation energies can be wrong by several eV in supercells of the order of 10s or 100s of atoms, and we have previously shown[4, 5] that finite size errors on this scale can even arise for neutral defects.

Rational supercell size errors arise for several different reasons:

3.2.1. Elastic Effects

Elastic errors arise because the supercell is simply not large enough to contain all of the local relaxation around the defect, leaving calculated formation energies too high. In addition, there are direct defect/image elastic interactions, which can affect both energies and the size and symmetry of structural relaxations. These direct elastic interactions can easily be truncated by freezing all atoms lying on the surface of the cell at their ideal lattice positions. The indirect elastic errors on the other hand can not be avoided. We have shown that the indirect elastic errors, even for Jahn-Teller active defects, are essentially inverse-linear in supercell size.[5, 7]

3.2.2. Electrostatic Interactions

The most important defect/image interactions are electrostatic. In contrast to the direct elastic interactions, the electrostatic interactions cannot be truncated or removed. They result in errors in the calculated formation, binding and migration energies which can be on the same order as the energies themselves. For practical supercell sizes they need not even be negligible for neutral defects, since dipolar and quadrupolar interactions can remain significant. These latter can result in errors in the calculated structures, since they favor certain symmetries and local relaxation modes over others [4]. It is now known [5, 7, 15, 16] that the leading electrostatic errors in E_d^C scale as L^{-1} and L^{-3} (where L is the supercell size).

3.2.3. Dispersion

Additional errors come from the spurious dispersion of the defect levels introduced by overlap between the defect state wavefunctions and their PBC images. The errors related to this dispersion (or tunneling) are expected to have only a fairly small and rather short ranged (exponential) effect.

In Fig.3 we show the bandstructure of the neutral phosphorus antisite in GaP as a function of the supercell size. It is clearly seen that the dispersion is a short ranged effect, though significant in the smallest cells.

3.2.4. Choice of Potential Zero Since the electrostatic potential in a supercell with PBCs is only defined up to a constant, the zero on the energy scale must be chosen arbitrarily in each calculation. In the case of the pseudopotential codes we use this occurs as an implicit average over values appropriate to each atom species in the supercell, weighted by the number of atoms of each species. This means that the numerical value of ϵ_F entering Eq. 1 changes with the contents of the cell, leading to an additional finite size error. If the number of defects per supercell is constant then this error decreases with the number of atoms in the cell - essentially with the volume of the cell, L^3. This error is completely taken care of in the infinite supercell limit of the finite size scaling scheme described below. For individual supercells, Van de Walle and Neugebauer[17] suggest correcting the error by re-aligning the potential in the defect cell to that of the bulk, using its real-space value measured in a suitable bulk-like region far from the defect.

3.3. Non-Rational Supercell Size Dependent Errors

Errors from k-point integration are not in themselves caused by the supercell approximation, but the effects of incomplete k-point convergence do depend upon the specific supercell in question. As a result they can potentially add additional terms in the supercell approximation errors which will not vary with supercell size according to any rational function, with the probable exception of calculations using only the Gamma (Γ)-point for k-point integrations. Hence we have two choices: Either full k-point convergence or Gamma Point only scaling.

3.3.1. Full k-point convergence Full k-point convergence implies using strong k-point convergence (preferably to at least +-0.001 eV) in EACH cell individually. To get errors safely below $\mathcal{O}(0.005$ eV) a k-point grid [18] of at least 18x18x18 is needed for the 8 atom cell [5]. This is costly, so to improve the convergence from smaller grids we use weighted averages over $E_{\mathrm{d}}^{\mathrm{C}}$ values calculated using a series of grids. Taking a weighted average over all the $E_{\mathrm{d}}^{\mathrm{C}}$ values up to a particular NxNxN

$$\overline{E_{\mathrm{d}}^{\mathrm{C}}} = \frac{\sum_N w_N E_d^C(N)}{\sum_N w_N} \tag{2}$$

helps dramatically, as it effectively increases the k-point density: the points in a 4x4x4 grid are not contained in a 6x6x6 grid, for example. The best choice of the weight we find[5] to be $w_N = N^3$, i.e. the number of k-points in the full Brillouin zone.

In reference [10] we also showed that the relaxation energy

$$\epsilon_{\mathrm{Relax}}(N) = E_{\mathrm{d}}^{\mathrm{C:Rx}}(N) - E_{\mathrm{d}}^{\mathrm{C:Id}}(N) \tag{3}$$

(where $E_{\mathrm{d}}^{\mathrm{C:Rx}}(N)$ and $E_{\mathrm{d}}^{\mathrm{C:Id}}(N)$ are $E_{\mathrm{d}}^{\mathrm{C}}(N)$ with atoms at relaxed and ideal positions respectively) converges faster with k-point grid than $E_{\mathrm{d}}^{\mathrm{C:Rx}}(N)$ and $E_{\mathrm{d}}^{\mathrm{C:Id}}(N)$ themselves. The relaxed formation energies $E_{\mathrm{d}}^{\mathrm{C:Rx}}$ are then approximated by

$$\overline{E_{\mathrm{d}}^{\mathrm{C:Rx}}} \approx \overline{E_{\mathrm{d}}^{\mathrm{C:Id}}} - \epsilon_{\mathrm{Relax}}(N) = \overline{E_{\mathrm{d}}^{\mathrm{C:Id}}} + E_{\mathrm{d}}^{C:Rx}(N) - E_{\mathrm{d}}^{C:Id}(N) \tag{4}$$

The relaxation energies used in the examples here are weighted averages using 6x6x6 and 8x8x8 Monkhorst-Pack k-point grids in the 8 atom cell, 2x2x2 and (if the convergence is uncertain) 4x4x4 grids in the 64 atom cell and 2x2x2 in the 216 and 512 atom supercells. For the latter two we usually restrict the k-point grid to the irreducible Brillouin zone of the undisturbed

bulk lattice. In other words, we use just the special k-point (0.25,0.25,0.25). This amounts to assuming that the distortion in the bandstructure due to the presence of the defect is either important only away from this point Γ or is symmetric. It introduces a small error whose significance disappears in the large supercell limit.

3.3.2. Γ Point only   In the literature it has been customary to assume that for large supercells it is sufficient to conduct the defect calculation with a k point grid consisting of only the Γ-point. There is a strong advantage to this as not only is there just a single k-point to consider, but it also displays time-reversal symmetry, allowing computational times to be halved. However, it has also been shown that if the average over the Brillouin zone is to be represented by a single k-point then the Γ-point is never the best one. [19, 20] To see this, assume one uses a set of special k points for some small cell size. An increase of the cell size corresponds to a folding back of the special k-points. One can show that the folded special k-points never coincide with the Γ-point, only approach it more and more closely. Therefore the use of only the Γ-point reduces the computation time, but at the expense of accuracy. According to Remeler and Madden,[21] k-points away from the Γ-point are needed to describe wavefunctions with components that do not fit in the supercell. Certainly, we find errors in the 512 atom InP supercell defect energy for the unrelaxed neutral phosporus (indium) vacancy which are 0.36 eV (0.06 eV) worse when using Γ-point only calculations than when using converged k-point grids.[5] We conclude that a 1x1x1 Monkhort-Pack grid (the Γ-point) is in general not sufficient, at least for the 512 atom cell. More detailed studies of the convergence of supercell calculations of defect properties as a function of the Brilloun-Zone sampling has been performed by, for example, Puska et al.[22] and Probert and Payne.[23]

In section 4.4.3 we will show the results of finite size scaling using calculated formation energies obtained through Γ-point only calculations.

3.3.3. Defect structures   The symmetry of the most stable relaxed structures can vary with supercell size. As an example we discuss here the Jahn-Teller active defect, In_P. For a Jahn-Teller active defect there is a partially filled degenerate state at the Fermi level, which the Jahn-Teller theorem [12] says will be lifted by symmetry reducing relaxations (if no other effect achieves this first). We have done a search for the various stable and metastable structures in four supercells of varying size [5].

In the 8 atom cell the lowest lying structure has the full T_d symmetry of the unrelaxed antisite and it is the only stable structure we find in this cell. In the larger cells relaxation breaks the T_d symmetry to give (primarily) C_{3v}, D_{2h} or C_{2v} point groups at the defect site. Overall, these results suggest that a lone In_P in an infinite supercell would have a structure of D_{2h} symmetry.

The changes in relative stability of the different structures are due to one or a combination of two things: a) stabilizing/destabilizing dipolar, quadrupolar or higher order interactions, which can in certain cases lift the symmetry without distortion, (in the 8 atom cell for example,) or favour certain Jahn-Teller structures over others. These effects become weaker as the cells grow. b) The lack of shells of atoms in the smaller cells to absorb the elastic strain, which favours more symmetric structures. In section IV.E we will continue to discuss the stability of different structures.

4. Proposed Solutions to Supercell Size Errors
Various authors,[15, 16, 24] have attempted to create corrections schemes to estimate and remove the supercell size errors, and we will describe some of these here. We will assess them by comparison with the scheme we currently consider the most pragmatic and reliable, namely finite size scaling, which is described in detail later, in subsection 4.4.

Table 1. Assessment of Makov-Payne and potential realignment as corrections methods: Residual errors in Makov-Payne corrected defect formation energies, (δ_{MP}), (Eq.5) for the 64 atom supercell relative to the finite size scaling results, ideal and relaxed. Error of 'Potential realignment' corrected defect formation energies, (δ_ϕ), (section IV.C) for the 64 atom supercell relative to the finite size scaling results, ideal and relaxed. Finite size scaled relaxed and unrelaxed (ideal lattice sites) formation energies (E_{d}^∞) including the calculated error bars for various example defects in InP. All energies in eV.

	Ideal structures			Relaxed structures		
Defect	δ_ϕ	δ_{MP}	E_{d}^∞	δ_ϕ	δ_{MP}	E_{d}^∞
V_{P}^{+1}	0.02	0.18	1.95±0.09	0.11	0.30	1.77±0.14
V_{In}^{-3}	-0.13	-0.51	6.52±0.06	-0.01	-0.08	4.95±0.05
P_{In}^{+2}	-0.02	0.11	2.36±0.04	0.04	0.23	1.07±0.04
In_{P}^{-2}	-0.03	0.08	4.25±0.08	0.05	0.30	3.85±0.13
$P_{\mathrm{i(P)}}^{+3}$	-0.08	-0.28	4.05±0.07	-0.03	0.05	2.43±0.11
$In_{\mathrm{i(P)}}^{+3}$	-0.21	-0.50	3.67±0.08	-0.19	3.48	1.85±0.04
$Zn_{\mathrm{i(P)}}^{+2}$	-0.13	-0.16	1.28±0.01	-0.12	-0.03	0.50±0.02
Zn_{In}^{-1}	-0.04	-0.16	0.98±0.01	0.01	-0.03	0.48±0.01
Si_{P}^{-1}	-0.04	-0.03	1.82±0.03	-0.03	0.03	1.71±0.04
S_{P}^{+1}	-0-06	0.00	-1.17±0.02	-0.07	0.05	-1.34±0.01
Si_{In}^{+1}	-0.02	-0.04	0.62±0.01	0.07	0.06	-0.36±0.03

4.1. Makov-Payne Corrections

The most widely known correction scheme is that of Makov and Payne[15]. Although this scheme is often used, its accuracy has been strongly questioned, with several studies suggesting that it is not reliably enough for regular use[24, 25, 17, 5], though it helps form the conceptual basis for much that has followed.

Makov and Payne predicted that the two leading terms in the errors should be linear and cubic. Their "corrected" formation energy takes the form

$$E_{\mathrm{d:MP}}^{\mathrm{C}}(L) = E_{\mathrm{d:MP}}^\infty - k_1(\epsilon L)^{-1} - k_3(\epsilon L)^{-3} \tag{5}$$

where ϵ is the dielectric constant, $k_1 = q^2\alpha/2$ and $k_3 = 2\pi qQ/3$. (q is the charge of the defect, α is the Madelung constant for the supercell and Q is the quadrupole moment of the defect.)

We have evaluated this scheme in Table 1 by comparing the "corrected" formation energies with the finite size scaled results to be described later. When the Makov Payne corrections are used the correct formation energies are obtained in some cases, such as V_{In}^{-3}, but in other cases, such as P_{In}^{+2}, they help but are not sufficient. For other cases, such as V_{P}^{+1}, the corrections actually move the formation energies in the wrong direction.

We have previously shown that the corrections are generally more likely to succeed for unrelaxed formation energies[7]. This is as expected, since the electrostatic monopole terms are not the only ones to scale as L^{-1}: the elastic errors do too. This means that even in principle the Makov-Payne corrections are only useful for non-relaxed formation energies, which are rarely the interesting ones. Besides this, the corrections also do better for more highly charged defects. This confirms that one of the problems is that they do not take into account the various other error terms which depend upon supercell size but not on charge state. Overall, even the leading linear error term may be very different from that predicted by Makov and Payne's corrections. Unfortunately, beyond noting that things get better on average for larger charge states and for non-relaxed calculations there seems to be no *a priori* method for determining whether the

corrections will make things better or worse in a specific case. They are thus of little practical help, since they do not take into account enough of the specific behaviour of individual defects and materials.

4.2. Correction of Defect Level Dispersion

It has been suggested[26] that the defect level dispersion artificially raises $E_{\mathrm{d}}^{\mathrm{C}}$ when k-points other than just the Γ point are used. It is suggested[26] that $E_{\mathrm{d}}^{\mathrm{C}}$ should then be shifted by $q \times \left(\epsilon_D^{\Gamma} - \epsilon_D^{ks} \right)$, where ϵ_D^{Γ} and ϵ_D^{ks} are the values of the Kohn-Sham level corresponding to the defect state calculated in the defect cell at the Γ point and averaged over the sampled k-points, respectively. The assumption is that the value of the defect level is correct at the Γ point, so the difference between that and the k-point averaged value should be removed. It is clear[6] from Fig. 3 that this is completely incorrect for the example of the phosphorus antisite in GaP. By plotting the "bandstructure" of the defect level in different sized supercells (Fig3) it can be seen that the defect level in the smaller cells is more-or-less correct when averaged over the sampled k-points, but much too low at the Γ point. The same is also true for the As vacancy on the GaAs(110) surface, for example[27]. Van de Walle and Neugebauer[17] instead point out[17] that in this respect there is a fundamental difference between deep levels such as these and shallow defect levels. They suggest that the correction should only be applied when evaluating transfer levels for shallow donors and acceptors. However, we have found[7] that even for shallow defect levels, which follow[17] closely the VBE or conduction band edge (CBE), ϵ_D^{Γ} still produces worse formation energies than ϵ_D^{ks}.

4.3. Potential Realignment

The potential realignment scheme was described earlier in Section 3.2.4. Although it was not originally intended for the purpose, we have found that it can help with the supercell size errors as well, as shown in Table 1. In this table we compare formation energies of InP defects in the 64 atom supercell obtained using the potential realignment to those obtained later using finite size scaling. (We use the potential in the "corner" of the simple cubic supercell furthest from the defect.) The errors after potential realignment are very small indeed.

Although it works well for individual cells, we find[7] that finite size scaling the potential realignment corrected defect formation energies leads to wide error bars (see Section 4.4 for a description of the method). We find error bars of up to ± 0.78 eV, average ± 0.24 eV, compared to error bars of up to ± 0.14 eV, average ± 0.05 eV when non-realigned formation energies are scaled. This means that potential realignment is a useful correction for the results from individual supercells, but should *not* be used if more accurate results or defined error bars on results are required. In that case non-realigned values should be scaled. The reason is that the correction scheme, good though it is, is not actually complete or correct. Even in the largest supercells, the region furthest from the defect is *not* bulk-like, as the scheme assumes, resulting in either an over estimate or an under estimate, depending upon the specific conditions.

This obviously raises the question of why the potential re-alignment scheme is *so* successful, when it does not set out to correct defect-image interaction errors at all! The fact that it produce similar (but more reliable) corrections to the Makov-Payne scheme suggests that it is some how dealing with the electrostatic errors anyway. It assumes that the real-space potential at some point in the cell far from the defect is bulk-like, even though for practical cell sizes it is not bulk-like at all. The resulting additional shift in this local real-space potential reflects the effects of the electrostatic defect-image interactions. Doing the potential re-alignment in this way therefore fails to properly correct the mismatch in the zeros of the energy scales between the bulk and defect cells, but the "error" in the re-alignment roughly corrects for the electrostatic errors arising from the PBCs, at least for the defects considered here.

4.4. Finite Size Scaling

We previously[4, 5, 7] proposed that the supercell size errors can instead be eliminated by calculating the same defect properties in a series of supercells of different sizes but the same symmetry and then finite size scaling the results to recover those of the infinite supercell.

4.4.1. Possible power laws for scaling

We found that $E_{\mathrm{d}}^{\mathrm{C}}$ varies with the supercell size as a power law: L, as

$$E_{\mathrm{d}}^{\mathrm{C}}(L) = E_{\mathrm{d}}^{\infty} + a_1 L^{-1} + a_n L^{-n} \tag{6}$$

where a_1, a_n and E_{d}^{∞} are fitting parameters, E_{d}^{∞} being the finite size scaled formation energy corresponding to an infinitely large supercell. The linear term has been discussed many times previously, first by Leslie and Gillian [14]. We tried fitting the scaling curves of various defects in InP, both neutral and charged defects, relaxed and nonrelaxed, using $n = 2$, 3 and 4 and assessing the fits using χ^2 tests. We found[5] the best value of n to be 3, with $n = 2$ on average 2.9 times worse and $n = 4$ on average 2.2 times worse. Computational limitations at the time forced us to use the 8 atom cell in our analysis. This introduced some additional short ranged errors, contributing to the error bars obtained and occasionally leading to either $n = 2$ or $n = 4$ providing a slightly better fit than $n = 3$, but in combination the conclusion was that finite size errors vary with L^{-1} and L^{-3}.

This conclusion has been confirmed and extended by others in ever larger cells. Levasseur et al used the finite size scaling scheme in an *ab initio* study of the diffusion mechanism of gallium in a silicon matrix [28, 29]. Finite size scaling was used for the first time by Lento et al [30] for the study of the unrelaxed +2 Si self-interstitial, with 1/L dependence. Wright et al. [31] use a mixture of Makov-Payne and finite size scaling, whereas Wilson et al. [32] use finite size scaling but in addition to the 1/L and 1/L³ dependence they add a third term with a 1/L² dependence. Hine et al [33] extended the finite size scaling scheme by varying not only the size of the supercell but also its shape. They have conducted a 1/L fit for supercells of varying shape and size where they assume the 1/L³ dependence to be negligible. Obviously, for sufficiently large cells this must be true, and for certain defects in certain materials it is even true for smaller supercells. However, it is clear from our data that at least for defects in InP it cannot be reliably assumed until the smallest cell included is by today's standards large, perhaps even a ∼1000 atoms or more. Paudel et al [34] on the other hand assume the 1/L dependence to be negligible and use only scaling with a 1/L³ term. Again, while this will occasionally be true for some individual defects, it is clearly not true in the majority of cases. Lany and Zunger [35] have conducted a systematic study of finite size effects in supercell defect calculations and propose a Makov-Payne fit after finite size effects not related to electrostatic interactions are eliminated. This is achieved by performing the potential-alignment method in parallel with the Makov-Payne correction method. In contrast to their results on GaAs and ZnO, we do *not* find that this works for defects in InP, nor do we expect it be to true generally.

Indeed, it must be pointed out that while various combinations of Makov-Payne (or other) electrostatic schemes and/or potential realignment etc will always for some defects in some materials, such schemes can only *ever* be generally reliable for *un*relaxed defect structures. Such a scheme may well emerge at some point, indeed, Freysoldt et al [36] recently proposed a promising new analytical correction scheme for unrelaxed defect geometries which they show to be more reliable than the Makov Payne scheme. However, for the more practical case of relaxed defects, the supercell error in the relaxation energy also scales with 1/L, and no combination of electrostatic based analytic corrections will account for this. Whether or not it is possible to find an analytic method of predicting these relaxation energy errors is unclear at present.

So, to summarize, the elastic errors are essentially inverse-linear in supercell size, while the total formation energy errors (relaxed or unrelaxed) scale with the inverse-linear dimension and the inverse volume of the supercell. This result is physically very intuitive: most sources of
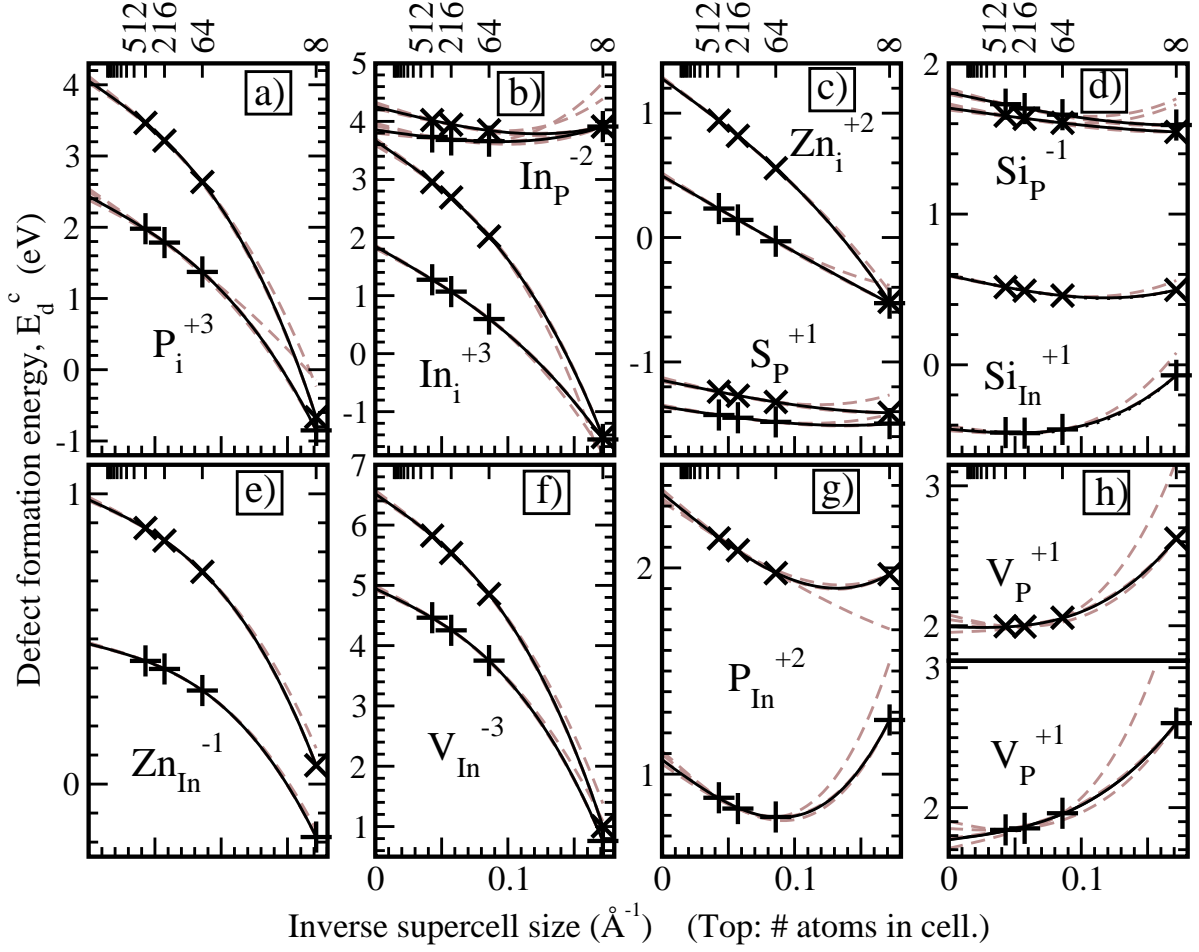
Figure 4. (Colour online.) Scaling of unrelaxed ($\times$) and relaxed ($+$) formation energies. Curves are fits to Eq 6 with $n=3$. Solid (black) curves are fits to the four points as calculated (no corrections.) Dashed (grey) lines each have one cell omitted for accuracy assessment.

error should vary with either the supercell size L (the defect-defect image distance) or with the cell volume L^3 (proportional to the jellium charge density, the number of atoms, the number of electrons, etc). Terms scaling with the surface area, $6L^2$, seem unlikely to be dominant.

4.4.2. Examples of finite size scalings Fig.4 shows relaxed formation energy scaling for 11 example defects in InP. The scaling curves using the uncorrected, as-calculated values are shown as solid lines in the figures (black in the online colour version). Their y-axis intersects give the E_d^∞ values listed in Table 1. The curves also serve to predict the formation energy which would be expected in any finite sized supercell: for example the formation energies in the 8000 atom supercell are those at $1/L = 0.1$ in the figures. We can estimate how accurate the E_d^∞ values are by adding the four dashed (grey) curves shown for each example in Fig 4, in each of which one of the four data points has been omitted. (Note that for some cases the errors are so small that the dotted lines are hard to pick out, but they are still present in the figure.) The spread in the resulting y-axis intersects gives us the error bars for the scaled formation energies E_d^∞, which are also listed in Table 1. This is one of the particular advantages of using finite size scaling: it is possible not only to correct for the finite size errors themselves, but also to obtain
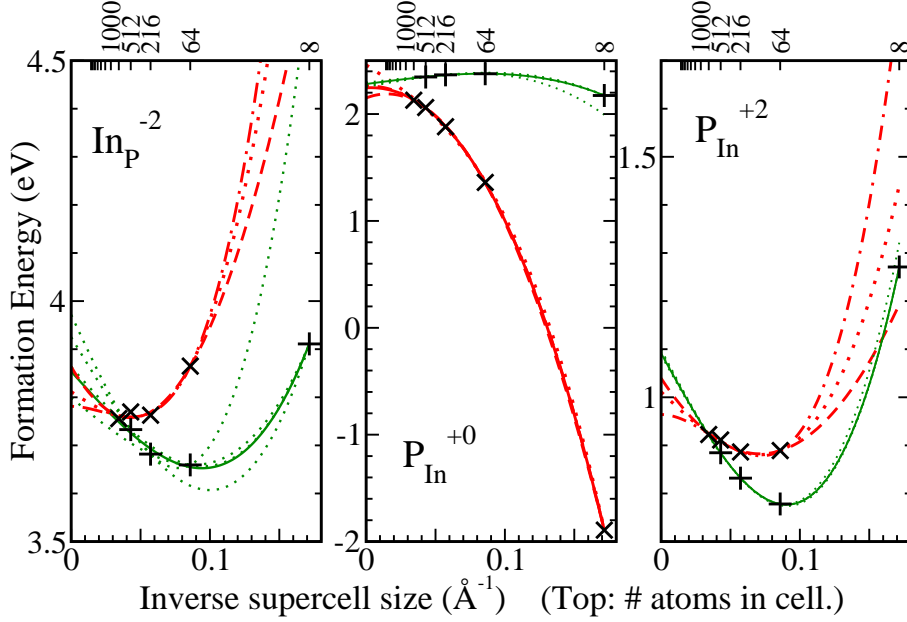
Figure 5. (Colour online.) Defect formation energies of In_P^{-2}, P_{In}^{+0} and P_{In}^{+2} calculated using only the Γ-point (+) and using converged k-point grids ($\times$) versus inverse supercell size. For the converged k-point grids the solid lines are fits using the form $1/L + 1/L^3$, while the dotted lines are error estimates for these fits as described in the text. For the Γ-point only calculations, the dotted lines are again fitted assuming $1/L + 1/L^3$, while the dashed lines assume $1/L + 1/L^2$, and the dot-dashed lines assume $1/L^2 + 1/L^3$. The solid line (P_{In}^{+0} only) is a fit assuming $1/L + 1/L^2 + 1/L^3$.

a well defined uncertainty on the resulting energies - something other correction schemes can not provide.

The errors obtained are on the 0.01-0.1 eV range or below, (smaller errors are here rounded to 0.01 eV) and can be further improved when still larger supercells are used[28, 31, 35]. We note that, by construction, the errors which arise if only the 8, 64, and 216 atom supercells are used for the scaling are also on this 0.01-0.1 eV level [7]. The fact that such small error bars can be obtained indicates that scaling is a viable and practical approach to supercell approximation errors. (It also indicates that our enforced use of the 8 atom supercell was also reasonable.)

In the following we list various examples confirming the importance of finite size scaling: Experimentally, the $Zn_{In}^{0/-}$ acceptor level is found to be 35 meV above the VBE. Even for the 512 atom supercell this level incorrectly lies within the conduction band. Only after finite size scaling does it enter the bandgap citeManage (see also Fig. 7). The formation energy of the V_{Ga}^{-3} in GaP calculated in the 216 atom supercell is raised by about 0.8 eV in the infinite limit [6]. In the 216 atom supercell of GaP, Ga_P^{+2} is the most stable charge state of the Ga antisite defect in p-type GaP. After finite size scaling the neutral state becomes the lowest state and the +2 state is never stable [6]. Even for neutral defects we find that scaling can help: E_d^∞ for the unrelaxed neutral In vacancy, V_{In}^0, is about 0.2eV higher than E_d^C for the 512 atom cell [4]. These examples together with the results shown in Fig. 4 demonstrate that finite size scaling is very important for obtaining even qualitative results. Moreover, finite size scaling can help even when the largest cells we can actually calculate are too small to be even partially converged.

4.4.3. Finite size scaling using the Γ point only An interesting variant of the finite size scaling method is to use only the Γ point for k-point integrations when calculating $E_{\mathrm{d}}^{\mathrm{C}}$ in Eq. 6. We gain in computation time, but considerable additional finite size errors are introduced. In principle these should scale to zero in the large supercell limit, so the technique is promising, particularly for Jahn-Teller active defects, as we will discuss below. Unfortunately these additional errors do not necessarily vary as $1/L + 1/L^3$, but scaling may still work using some (probably other) power law dependence, at least for larger supercells. In Fig. 5 we show the scaled results using only the Γ point (+ in Fig.5) for three defects in InP: $\mathrm{In_P^{-2}}$, $\mathrm{P_{In}^{+0}}$, and $\mathrm{P_{In}^{+2}}$. For comparison we show also the scaled results for the same defects using fully k-point converged $E_{\mathrm{d}}^{\mathrm{C}}$ ($\times$ in Fig.5). For $\mathrm{In_P^{-2}}$ and $\mathrm{P_{In}^{+2}}$, the Γ point calculation is not stable in the 8 atom cell, but we are able to include the 1000 atom simple cubic cell instead. For the k-point converged results we show the scaling and error bars in the figure as before. For the Γ point scaling we try three alternative power law forms: $1/L + 1/L^2$ (dashed lines), $1/L^2 + 1/L^3$ (dotted lines), $1/L^2 + 1/L^3$ (dot-dashed) and for $\mathrm{P_{In}^{+0}}$only we also show a four parameter fit to $1/L + 1/L^2 + 1/L^3$. In the infinite limit the scaled energies are rather similar. We generally find good agreement between Γ point scaling and full k-point converged scaling, the best being when using a fit including the square term, i.e. $1/L + 1/L^2$.

The problem with the Γ point scaling is that we have no real *a priori* idea about what the scaling *ought* to be. There are certainly going to be components which scale as $1/L$ and $1/L^3$, as there are for the k-point converged. However, there are also going to be other terms associated with the dispersion of both the defect levels and the host bands. We would guess that these are primarily exponential, but with a rather slow exponent, so that they are significant for all of our supercell sizes. We know that the dispersion is significant for the 8 and 64 atom cells at least (see Fig. 3). So far we have not tried to include an exponential, but to do such a complicated fitting would probably require rather more supercells than we have data for at the moment. Since we do not know the functional form, nor have a physical argument to suggest what it could be, we can have little confidence in the results obtained, and we cannot estimate the errors either. Hence we do not use it at present.

Never-the-less, finite size scaling using Γ point only k-point integration is worthy of further consideration in the future, as computers become more powerful. In particular, it is sometimes pointed out that the most correct way to treat Jahn-Teller active defects is to use k-point integration at the Γ point only. For Jahn-Teller active defects any other k-point grid can artificially lift degeneracies, and (worse) it may even lead to a "ground state" structure which is actually of MIXED symmetry. For example, performing the scaling for $\mathrm{In_P^{+0}}$ using the Γ point only, we get a ground state of pure D_{2h} point group, with 2 lengthened neighbour-neighbour bonds. In other words it becomes a double anti-dimer structure (DAD). On the other hand, performing the scaling using fully converged k-point grids we actually get a small additional DX-like symmetry breaking with point group C_{3v}, in which the antisite atom moving towards one neighbour, reducing the final structure to C_1 symmetry.

The scaling of both the ground state and various excited state structures of $\mathrm{In_P^{+0}}$ using Γ point only is shown in Fig 6. We have specifically looked at a full range of possible symmetry breakings, finding formation energies of each stable structure in each supercell. (Again, symmetry broken relaxed structures are not stable in the 8 atom cell, probably due to the very limited scope for relaxation.) Besides the DX and DAD structures mentioned above, we have found another metastable D_{2h} structure in which two pairs of neighbours move towards each other (double dimer: DDM) and a structure in which the antisite again shifts off centre, towards the middle of one of the neighbour-neighbour bonds (hence in the direction of a bond centre BCR interstitial site). Other structures were tried but were not stable. Finally we also plot the formation energies with a forced unbroken tetrahedral (Td) symmetry with breathing mode relaxations only.
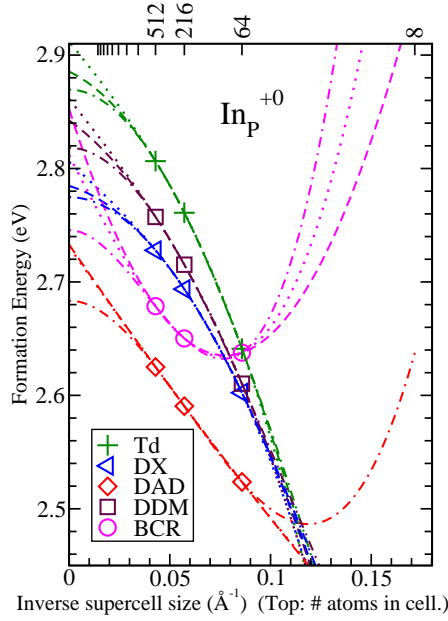
Figure 6. Defect formation energy for In_P^{+0} in InP for different structures calculated using only the Γ-point. The Td structure has an unbroken tetrahedral symmetry (T_d) with breathing mode relaxation only. The DX structure is a C_{3v} symmetry distortion in which the antisite moves towards one of its neighbours (and away from the others). DAD and DDM are both D_{2h} symmetry distortions in which the antisite stays put, but the four neighbours rotate around it in pairs, forming either two dimers or two anti-dimers. In the BCR structure the antisite again shifts off centre, moving towards the middle of one neighbour-neighbour bond (hence in the direction of a bond centre interstitial site). The finite size fittings assume scaling with the powers $1/L+1/L^2$ (dashed), $1/L+1/L^3$ (dotted) and $1/L^2+1/L^3$ (dot-dashed).

5. Using Finite Size Scaling to examine the bandgap Problem

We now turn to the bandgap problem and the issue of how to map calculated transfer levels onto the experimental gap. In practise several alternative - and essentially incompatible - methods are used.

1) The Extended Gap Scheme Align the theoretical and experimental VBEs and start plotting defect transfer levels from there, continuing *past* the theoretical conduction band edge (CBE) until one reaches the experimental one. In the section of the thus plotted "bandgap" which lies above the theoretical CBE one automatically includes calculations in which supposedly localized, defect-bound electrons are in reality located in delocalized conduction band states. The properties of the defect itself (transfer levels and local relaxed structure etc) re-enter primarily via hybridization of the conduction band states with the localized defect states, though this hybridization becomes smaller as the supercell size grows.

2) The Scissors Scheme Align both the theoretical VBE *and* CBE with their experimental counterparts, performing a "scissors" operation to stretch out the theoretical gap states over the experimental gap. The manner in which this scissors operation should be done is not uniquely defined. A common option is to place acceptor levels the same distance above the experimental VBE that they appear above the theoretical VBE in calculations, and donor levels the same distance below the experimental CBE that they appear below the theoretical CBE in calculations. Deeper levels are harder: our practical suggestions for this are given in section 6.

3) The Reference Level Scheme. The basis of this scheme is rather different: the transfer level for the defect of interest is calculated, together with that of a similar reference defect for which the experimental value of the transfer level is well known, both done to the same level of accuracy. The difference between the experimental and calculated levels of the known defect is subtracted from the calculated value of the new defect, so that the new level is only found relative to the old one. This idea is not without practical merit, but is very empirical. Its accuracy depends critically upon the choice of an appropriate reference defect, which must be as similar to the new one as possible, so it will not be discussed further here. This scheme also has an occasionally used *ab initio* variant, namely:

4) The Charged Bulk Reference Scheme Here the reference state is not that of another defect, but is either one or more holes at the VBE or electrons at the CBE. In other words, a
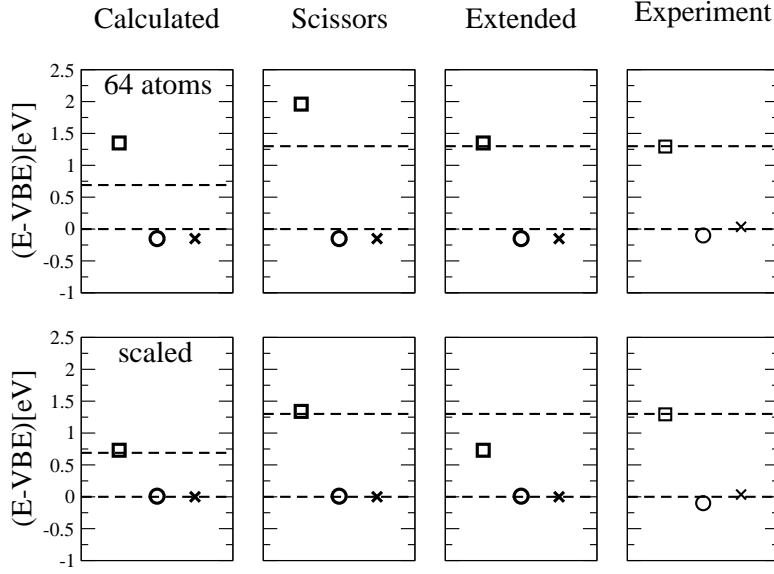
Figure 7. Transfer levels for $S_P^{+/0}$ ($\square$), $S_P^{+2/+}$ ($\circ$), and $Zn_{In}^{0/-}$ ($\times$) in InP calculated using LDA. In the top panel we show the calculated results for the 64 atom cell without any corrections. In the bottom panel we show the finite size scaled results. In the second (third) column we show the transfer levels according to the scissors scheme (extended scheme) and in the last column we show as reference the experimentally determined transfer levels.
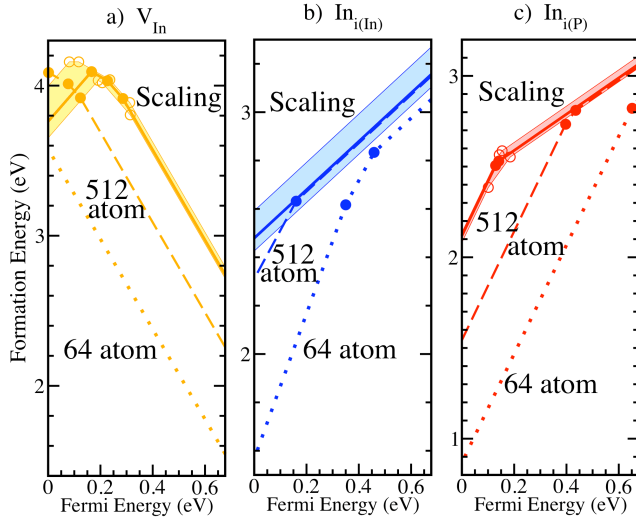

Figure 8. Plotting fully scaled and corrected formation energies (lines) and transfer levels (circles) versus Fermi level across the LDA bandgap, for defects in InP. Solid lines indicate the scaled formation energies and shaded areas indicate the error bars on those scaled results. Dotted and dashed lines indicate the un-scaled formation energies in the 64 and 512 atom supercells, respectively.

charged bulk total energy appears in Eq.1, rather than a neutral one. In principle this provides an alternative route around the bandgap problem. However, we have earlier shown[7] that this approach is not even internally consistent and is fundamentally incorrect.

Obviously, none of these schemes is fully correct, since the LDA/GGA bandgap problem is a fundamental one, but the important practical question of which approach comes closer to giving the correct physical picture must still be considered. In principle it can be decided by examining

various experimentally well known defect levels. The exact location of most native defect levels is rather hard to measure to a sufficiently high accuracy to answer this question, but many simple donor and acceptor levels are known very accurately. We use the 0/- acceptor level of Zn_{In}, which in experiment lies 0.035 eV from the VBE, and the +/0 donor level of S_P, which experiment finds about 0.006 eV from the CBE (see Fig 7 last column). Unfortunately, calculations of these levels in finite sized supercells in the 100-200 atom range have never produced a clear answer to the question, so we use finite size scaling to correct for the supercell approximation errors. In addition, in Fig.7 we also show the second donor level, $S_P^{+2/+}$, calculated using the same scheme. Since this level is never observed experimentally it must lie outside the bandgap. Hence the VBE should lie between the second donor level (∘ in Fig.7) and the single acceptor level (× in Fig.7). In practise, this pair of levels more or less coincides (see Fig.7), but we do get a clear view of how to treat the bandgap problem.

In the 64 atom cell the donor level (□ in Fig.7) lies roughly the experimental bandgap (1.3 eV) above the VBE, while the acceptor level (× in Fig.7) lies a little below the VBE. Finite size scaling places the $Zn_{In}^{0/-}$ acceptor level 0.03 eV above the VBE (× in Fig.7), in rather good agreement with experiment. The single donor level scales to the theoretical CBE. (This we calculate to be true both for the LDA and GGA approximation.) Hence, scheme 1, the extended gap scheme, is seen to be the most appropriate when only reporting uncorrected results from supercells of about 50-100 atoms. However, when the finite size errors are removed (by scaling or by some other technique) it becomes clear that the scissors scheme, scheme 2, is physically far more correct. Unscaled LDA or GGA results in supercells over a few 1000 atoms would also be best reported using the scissors scheme. For intermediate (100-1000 atom) supercells some kind of hybrid approach is required. The result also indicates why the controversy has lasted so long: ultimately the scissors scheme is correct, but this only shows up for very large supercells or with scaling.

6. A recipe for Practical Calculations

To draw the elements of this review together, we will outline our approach to making practical calculations of the formation energies and transfer levels of defects, using the examples of the indium vacancy (V_{In}) and the indium interstitial at both the octahedral sites ($In_{i(In)}$ and $In_{i(P)}$) in InP. Our approach is currently as follows:

1: Initial calculations and Kohn-Sham level analysis.
2: Calculating formation energies in several supercells.
3: Correction of non-supercell size dependent errors.
4: Correction of supercell size dependent errors.
5: Consideration of LDA bandgap errors.

Step 1) Initial calculations and Kohn-Sham level analysis. In a small-ish cell, (50-100 atoms, say), calculate relaxed structures for some probable charge states of each defect and then examine their Kohn-Sham (KS) defect levels. The spatial distribution of the partial charge densities of the calculated KS levels indicates which ones correspond to (perhaps perturbed) host band levels, and which correspond to localized defect levels. The multiplicities of these defect KS levels and their energies relative to the bandgap tell us which charge states are potentially stable or metastable.

For the three defects here, we previously reported[4] the localized KS levels in the vicinity of the gap for the neutral defects. For V_{In}^{+0} we have a triply degenerate level just below the valence band edge, which appears to enter the gap for sufficiently large supercells, allowing in principle any charge state from -3 to +3. For the two In interstitials a single non-degenerate level lies inside the bandgap. In the neutral charge state this level is filled, with an additional electron actually located in the conduction band. We therefore only consider the charge states +1 to +3
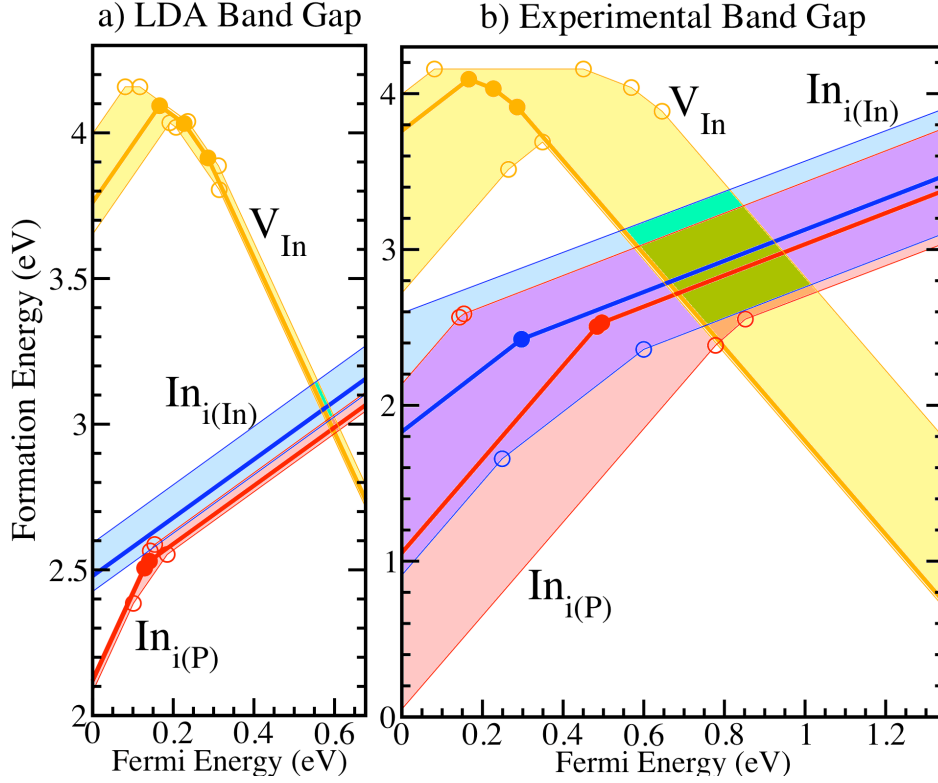
Figure 9. Plotting fully scaled and corrected formation energies (lines) and transfer levels (circles) versus Fermi level, for defects in InP. The energies are plotted over a) the LDA bandgap, and b) over the experimental bandgap using the specific scissors operation described in the text. Solid lines indicate the scaled and scissors shifted formation energies and shaded areas give a rough indication of the maximum combined error bars on those values.

to be valid for indium interstitials.

Such KS level checks are important, since if we naively proceed with finite size scaling more and more charge states, simply looking for the one with the lowest scaled formation energy, we can end up with some very unphysical results, owing to the presence of electrons at the CBE or holes at the VBE. For example, the unphysical neutral charge state of both In interstitials in InP has a lower scaled formation energy than the +1 charge state in n-type material, probably due to the omission of still higher order terms.

Step 2) Calculate formation energies for the plausible charge states in a series of cells of the same symmetry. Here we use the 8, 64, 216 and 512 atom simple cubic zinc-blende supercells. For the reasons discussed in section 4.4.3, we use fully converged k-point grids - with k-point convergence enhancement where needed.

Step 3) Correction of non-supercell size dependent errors (other than LDA bandgap errors). The significance of any errors which do NOT depend on supercell size are estimated numerically, using a medium sized supercell. The only significant such error here is due to treating In:4d electrons as core, so we re-do all defect calculations in the 64 atom supercell, treating the In:4d electrons as valence. For each charge state of each defect we then calculate the resulting difference in formation energy and add this on to the formation energy in all four supercells considered.

Step 4) Correction of supercell size-dependent errors by finite size scaling, assuming the form $1/L$ and $1/L^3$. Errors are estimating as described in section 4.4. Examples of such

Table 2. Formation energies of V_{In}^0, $In_{i(In)}$ and $In_{i(P)}$ in eV, with both supercell size dependent and independent errors corrected.

Defect	Energy	Defect	Energy
V_{In}^{-3}	4.77±0.05	$In_{i(In)}^{+1}$	2.48±0.11
V_{In}^{-2}	4.49±0.05	$In_{i(In)}^{+2}$	2.53±0.09
V_{In}^{-1}	4.26±0.03	$In_{i(In)}^{+3}$	2.96±0.05
V_{In}^{+0}	4.12±0.04		
V_{In}^{+1}	4.07±0.05	$In_{i(P)}^{+1}$	2.39±0.04
V_{In}^{+2}	3.76±0.24	$In_{i(P)}^{+2}$	2.25±0.07
V_{In}^{+3}	3.89±0.36	$In_{i(P)}^{+3}$	2.12±0.04

scalings were shown in Fig 4, but in practise we normally automate the processes to directly produce results like those in table 2. From these the formation energies and defect transfer levels as a function of Fermi level are obtained in the usual way, as shown in Figs 8 and 9a). Fig. 8 shows the energies plotted across the LDA bandgap, together with the error bars (shaded areas), alongside the *un*-scaled energies as calculated in the 64 (dotted lines) and 512 (dashed lines) atom supercells. It is apparent that even with the 512 atom supercell, significant errors can sometimes remain. Formation energies, transfer levels and even predicted charge states in $n-$ and $p-$type material can be wrong.

Step 5) Addressing the LDA bandgap errors is done using a scissors operation for the reasons described earlier. In order to do this we need to decide which transfer levels we will move "up" with the CBE, which will remain tied to the VBE, and which will move part way. We do this by examining the structure and symmetry of the partial charge densities of the localized defect KS levels, and comparing them with the KS levels at the VBE and CBE. Various methods could be employed, but a simple one is to use the Wigner-Seitz projections of the KS levels onto individual host atoms and their approximate atomic orbital symmetries. These are summed separately over the host cations and anions (here In and P), assigning each KS level a vector in an abstract space with the basis (s^{In} p^{In} d^{In} s^P p^P d^P). We then calculate the "overlap" of the vectors for the localized KS defect levels with the band edges.

In this case, we find zero overlap between the band edges themselves (the VBE being purely $p^{In}+d^{In}+p^P$ and the CBE purely $s^{In}+s^P$). The triply degenerate V_{In}^{+0} level has a 99% overlap with the VBE, so we tie all the V_{In}^{+0} transfer levels there. For the interstitials the situation is a little more complex: for both we find a $\sim$59% overlap with the CBE, and a $\sim$52% overlap with the VBE. As a result we move the interstitial transfer levels up by 0.53 $\times$ the amount we move the CBE. Fig.9 b) shows the result of this. The shaded error bars combine the scaling error bars with a rough estimate of the possible range of errors resulting from the scissors operation. For the interstitials the latter error is estimated by taking the two extremes (moving the transfer levels purely with one band edge or the other). For V_{In}^{+0} this seems unnecessarily large, so we show simply the effect of moving the transfer levels half way up. Both alternatives are likely to give a large over-estimate of the error, but the task of developing a more appropriate error estimate is left for the future.

7. Summary
We have shown that it is possible to obtain defect formation energies in quantitative agreement with experiment using DFT calculations within the supercell approximation. We have systematically addressed the errors inherent in DFT calculations, categorizing them as non-

supercell size dependent errors, rational supercell size dependent errors and non-rational supercell size dependent errors. We have confirmed that these errors often lead to significant inaccuracies in the DFT results for defect energetics calculated in individual supercells, even for cells as large as 512 atoms.

The non-rational errors must be addressed supercell by supercell, for example using k-point convergence enhancements. The non-supercell size dependent errors can be found using higher level calculations in a single small or medium sized supercell, and the resulting corrections used in all other supercells. We have shown that the most reliable scheme to correct the rational supercell approximation errors is finite size scaling of otherwise *un*corrected energies, using terms inverse proportional to the linear dimension of the supercell and inverse proportional to the volume of the supercell. This produces accurate results, with systematic error estimates, but is, of course, costly in both human and computer time.

The best alternative is simply to use potential realignment in as large a supercell as time and resources permit. However, one should be aware that a) this does not help the elastic errors, b) potential realignment should *not* be combined with finite size scaling and c) there is no way to estimate the remaining errors or the reliability of the results. For our examples, the average errors using this method are ∼0.10 eV, but with some examples up to 0.21 eV, and nothing to say that much larger errors will never occur. If the conclusions being drawn from a calculation are not adversely affected by uncontrolled errors of 0.1-0.2 eV or so then this method is reasonably good. Otherwise, the only truly reliable method of controlling the errors in the supercell approximation, and defining the uncertainly in the results, is finite size scaling.

We have noted that finite size scaling using only the Γ-point for the k-point integration is a potentially very attractive scheme. However, before starting to use it, the functional form of the fitting equation has to be settled. Once this is done, the finite size scaling scheme using only the Γ-point for k-point integrations will be the most reliable error correction scheme because it also treats Jahn-Teller active defects more consistently.

Naturally, as computers and hence tractable supercell sizes improve, the errors in 1/L will come to dominate more and more, but we would argue that any analytical supercell size correction scheme relying on the dielectric constants or electrostatic expansions is likely to remain unreliable, and will certainly do so for the energies of relaxed defects. This is because such schemes cannot address the supercell size errors in the relaxation energies.

Finally, we have given a recipe for practical DFT calculations of defect properties which will hopefully be of help, especially to beginners trying to obtain reliable calculated defect formation energies.

[1] J. Singh, *Physics of semiconductors and their heterostructures*, McGraw-Hill, 1993.
[2] W. Kohn and L. Sham Phys. Rev. **140**, A1133 (1965)
[3] M.C.Payne, M.P.Teter, D.C.Allan, T.A.Arias and J.D.Joannopoulos Rev. Mod.Phys. **64**, 1045 (1992)
[4] C.W.M. Castleton and S. Mirbt Physica B **340-342**, 407 (2003)
[5] C.W.M. Castleton and S. Mirbt Phys. Rev. B **70**, 195202 (2004)
[6] A. Höglund, C.W.M. Castleton and S. Mirbt Phys. Rev. B **72**, 195213 (2005)
[7] C.W.M. Castleton, A. Höglund and S. Mirbt Phys. Rev. B **73**, 035215 (2006)
[8] D. Vanderbilt Phys. Rev. B **41**, R7892 (1990); G. Kresse and J. Hafner J. Phys: Cond. Matt. **6**, 8245 (1994)
[9] G. Kresse and J. Furthmüller Comp. Mat. Sci. **6**, 15 (1996)
[10] C.W.M. Castleton and S. Mirbt Phys. Rev. B **68**, 085203 (2003)
[11] P.E. Blöchl, Phys Rev B **50** 17953 (1994); G. Kresse and J. Joubert Phys Rev B **59** 1758 (1999)
[12] H.A. Jahn Proc. R. Soc. London, Ser. A **161**, 220 (1937)
[13] A. Zywietz and J. Furthmüller Phys. Rev. B **59**, 15166 (1999)
[14] M.Leslie and M.J.Gillian J. Phys.C: Solid State Phys. **18**, 973 (1985)

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

[16] P.A. Schultz Phys Rev B **60** 1551 (1999): L.N. Kantorovich Phys. Rev. B **60** 15476 (1999); L.N. Kantorovich and I.I.Tupitsyn J. Phys.:Condens. Matter **11**, 6159 (1999); H. Nozaki and S. Itoh Phys. Rev. E **62** 1390 (2000); A.Castro, A.Rubio and M.J.Stott Can. J. Phys. **81**, 1151 (2003)

[17] C.G.Van de Walle and J.Neugebauer J. App. Phys. **95**, 3851 (2004)

[18] H. Monkhorst and P. Pack Phys. Rev. B **13**, 5188 (1976)

[19] A. Baldereschi Phys. Rev. B **7** 5212 (1973)

[20] D.J.Chadi and M.L.Cohen Phys. Rev. B **8**, 5747 (1973)

[21] D.K. Remler and P.A. Madden Mol. Phys. **70**, 921 (1990)

[22] M.J. Puska, S. Pöykkö, M. Pesola, and R.M. Nieminen Phys. Rev. B **58**, 1318 (1998)

[23] M.I.J. Probert and M.C. Payne Phys. Rev. B **67** 75204 (2003)

[24] P.A. Schultz Phys Rev Lett **84** 1942 (2000); U.Gerstmann, P.Deák, R.Rurali, B.Aradi, Th.Frauenheim and H.Overhof Physica B **340-342**, 190 (2003); B.Aradi, P.Deák, A.Gali, N.T.Son and E.Janzén Phys Rev B **69**, 233202 (2004)

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

[26] S.-H. Wei Comp. Mat. Sci. **30**, 337 (2004)

[27] M.Hedström, A.Schindlmayr and M.Scheffler Phys. Stat. Sol. B **234**, 346 (2002)

[28] K. Levasseur-Smith and N. Mousseau, Eur. Phys. J. B **64**, 165 (2008)

[29] K. Levasseur-Smith and N. Mousseau, J. Appl. Phys. **103**, 113502 (2008)

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

[31] A.F. Wright and R.R. Wixom, J. Appl. Phys. **103**, 083517 (2008)

[32] D.J. Wilson, A.A. Sokol, S.A. French, C.R.A. Catlow, Phys. Rev. B, **77**, 064115 (2008)

[33] N.D.M. Hine, K. Frensch, W.M.C. Foulkes and M.W. Finnis, Phys. Rev. B **79**, 024112 (2009)

[34] T.R. Paudel and W.R.L Lambrecht, Phys. Rev. B **77**, 205202 (2008)

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

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