19
Which Electronic Structure Method for The Study of Defects: A Commentary

Walter R. L. Lambrecht

19.1
Introduction: A Historic Perspective

The title of the CECAM workshop that gave rise to this compilation is "Which electronic structure method for the study of defects?" (see the Preface by Deák [1]). The first question, the reader might ask is why this question arises now. After all, the standard framework for dealing with defects in materials has been around for at least a few decades and has made significant contributions to our understanding of a wide variety of defects in semiconductors.

The "standard" toolkit includes density functional theory (DFT) combined with supercell band structure calculations. That is, the electronic structure problem is reduced to a standard band structure problem by using periodic boundary conditions. For surfaces, one uses a repeated slab geometry with vacuum regions to separate the surfaces; for interfaces, the artificial periodicity is introduced in only one dimension, while there still is a physical periodicity in the other two directions. For point defects, one needs to impose artificial periodicities in three dimensions, resulting in a faster increase of the size of the system that needs to be calculated. Dislocations, which are periodic in only one direction, are actually more complex because of their long-range strain fields and to avoid those, one either needs to use an opposing Burger's vector dipole pair of dislocations, or use a finite sample surrounded by vacuum regions for which one can then again restore periodicity in the two remaining dimensions. In any case, this compilation almost entirely focuses on point defects in the bulk of materials, not at surfaces or in nanostructures or not on extended defects, so we here restrict the discussion to point defects.

The main reason almost all calculations nowadays use periodic boundary conditions is a matter of convenience. One can use the general purpose computer programs that have been developed for band structure calculations, once these have been boosted to be able to deal with sufficiently large number of atoms. This is not the only choice, one could resort completely to finite models, or clusters, or one could embed the defect region into a perfect crystal by means of Green's function techniques.

Clusters are still popular with quantum chemists. Their drawback is that the surface of the clusters presents a more severe perturbation than the milder periodic boundary condition. One can avoid it to some extent by artificially satisfying the surfaces dangling bonds with pseudo hydrogens, but there is no obvious advantage to them compared to supercell techniques, except perhaps if one wants to use more advanced treatments of correlation than available in DFT, such as multiconfiguration interaction methods [2].

On the other hand, the most sophisticated approach to point defects, leading to an exact embedding of the local environment of the defect in the surrounding perfect crystal using correct boundary conditions is the Green's function method. Unlike in supercells, the band edges are defined precisely and one can calculate not only bound states in the gaps but also changes in the total and local densities of states within the bands, in the form of resonances and anti-resonances. One can make sure that the defect states, obey the exact symmetry of the defect site and are not influenced by the supercell geometry. One naturally treats an open system, into or out of which charge can flow as set by the chemical potential, without any uncertainties on how to define the chemical potential relative to the bands unlike the need in supercells to restore neutrality by artificial means such as a compensating homogeneous background charge density. In fact, there was a large amount of work done in the late 1970s–1980s to develop Green's function methods for point defects [3–13]. Strangely, these methods were abandoned in large part. The reason for this in my opinion is not because these methods were less intuitive or intrinsically less powerful, but rather their development could not keep up with the pace of the standard supercell approach. The original versions of these methods did not allow for relaxation of the structure of the defects and in many cases were even restricted to specific symmetries of the defect structure. On the other hand, practitioners using a standard all-purpose band structure method, were able to solve the problems that arose from experiment and required immediate attention. In particular, their ability to use the correct relaxed structure near the defects was a crucial advantage to get the essential physics correct. Meanwhile, practitioners of the Green's function methods, had to keep on generalizing their codes to handle such problems.

A second reason why the standard approach won the race is no doubt the almost universal use of plane wave basis sets as opposed to localized basis sets. For the latter it is notoriously more difficult to prove convergence and to calculate forces analytically. Their main advantages, a local "chemically intuitive" description and smaller basis sets were no match against the brute force plane wave approach as computer power increased and more importantly iterative minimization algorithms [14] for dealing efficiently with these intrinsically much larger basis sets were developed. In other words, it just became too easy to tackle the active research problems with the supercell technique. Why would one need to develop a special technique for point defects if the standard all-purpose method could do the job and provide answers the experimentalist needed?

One might thus have expected a lively discussion on which methods to use in the early 1980s, when the Green's function methods were still widely in use and cluster calculations were still competitive with the supercell approach. But nowadays, only a few practitioners of that approach remain. One at least is represented in this

compilation (chapter 17, [15]). Development of methodologies for dealing with ever larger systems, has of course continued, in particular, the development of order-N methods. The large driving force for developing those methods has been the advent of nanostructures, but by and large they are not being applied very much to point defects in solids. It appears that the sizes we need to handle point defects satisfactorily are compatible with the computing power available without these new large scale system methods. Nonetheless, at least one article in this compilation discusses progress in developing methods to solve standard Kohn–Sham equations more efficiently for truly large systems (chapter 16, [16]).

At the same time, while it was long known that the local density approximation (LDA) (or its slight modification, the generalized gradient approximation, GGA) underestimates band gaps in semiconductors, the attitude of many studying point defects, has been to just put up with this problem by focusing on quantities which were supposedly not affected by this shortcoming. Thus the emphasis of point defect studies shifted from calculating one-particle energy levels or changes in densities of states in the bands, to total energies, energies of formation of defects and transition energies. As long as one avoided explicitly calculating excited state properties, the thinking went, we were safe. After all, transition energies are defined as the position of the chemical potential relative to the band edge, where one charge state becomes lower energy than another, so we keep focusing on the ground states, which is after all the legitimate quantity to calculate in DFT. Other quantities, which seemed safe are charge densities, and spin densities and those define such things as the hyperfine parameters, so useful to electron paramagnetic resonance (EPR) experimentalists, or local vibrational modes as measured by infrared spectroscopy.

Nonetheless, it appears that in the last few years, increasingly the "festering" underlying problems of the supercell plus LDA (or GGA) paradigm have become more apparent and have been increasingly discussed in the literature. In part, this is probably because a lot of the recent applications have been on more challenging systems. For example, wide band gap semiconductors and oxides, including transition metal oxides, present a new challenge that brings out these underlying problems. Transition metal impurities with strongly correlated d-states have received increasing interest in the context of dilute magnetic semiconductors. The underestimate of the band gap by LDA (and GGA) in these systems is often larger. Defect levels that should be and are experimentally in the gap, appear as resonances in the bands in the LDA calculations leading to qualitatively wrong descriptions of the defect behavior. The more ionic nature means that screening is reduced and brings out the effects of the Coulomb interactions more vividly. In particular, the self-interaction error of LDA and the orbital dependent correlation and exchange effects are being put in the spotlight in these systems. Ionic systems also exhibit stronger polaronic effects, which as will be seen below are strongly suppressed by LDA because of the incomplete cancellation of the self-interaction. Due to the reduced screening, spurious interactions involving charged defect states are also exacerbated.

In any case, whatever may have been the reason why these problems resurfaced, whether they were always there and they were just temporarily ignored while the community was absorbed in the successes of the standard approach, until our

demands for accuracy and rigor overtook, or because the problems were becoming more apparent in the new systems to which the attention has shifted, this renewed debate on the methodologies is highly welcome.

At the same time, there has been a strong push in recent years to go beyond the limitations of DFT to ground state properties. The quasiparticle excitations can now be calculated in Hedin's GW approximation [17, 18] using pseudopotential plane wave [19–22], all-electron linearized augmented plane-wave (LAPW) [23, 24], linearized muffin-tin orbital (LMTO) [25–29], and projector augmented wave (PAW) implementations [30–32]. Even electron–hole interactions affecting optical properties can be treated by the Bethe–Salpeter equation approach [33–39]. Time dependent DFT provides an alternative way of dealing with excited states [36, 40]. Finally, the Quantum Monte Carlo method has continued to make strides and is represented in this compilation with chapter 2 by Hennig and coworkers [41]. While such methodologies are still computationally demanding and until recently only feasible for small systems like perfect crystal unit cells, parallelization of codes, and new algorithmic developments are now letting these methods make inroads in the defect world. At the same time, approximate methods to include some of the essential correlation or orbital dependent effects have continued to be developed, such as LDA $+$ U (or GGA $+$ U) and hybrid functionals.

Thus, the time is ripe not only to re-assess the accuracy of the standard approach now that we can push its limits by shear computational power but to incorporate these new methodologies beyond LDA into the world of point defects.

This compilation gives the reader a sampling of some of the problems under discussion. It will not provide definitive answers because there is as yet no consensus on many of these problems, but at least it will set the stage for further investigation and will allow a newcomer to the field to quickly get involved in the middle of the debate.

In the remainder of this article, I will comment on some of the highlights of the workshop and where the reader will find them in this issue.

19.2
Themes of the Workshop

Rather than commenting on the individual articles found in this book, the discussion is centered around a few themes that run throughout several articles and an attempt is made to place these in context and point out their connections. The themes are: (i) dealing with periodic boundary condition artifacts, (ii) dealing with the band gap underestimate by LDA, (iii) dealing with the self-interaction error of LDA, and (iv) developments of alternative methods to DFT and excited state methods.

19.2.1
Periodic Boundary Artifacts

Essentially the task is to extract the information on a single defect in the dilute limit from a calculation with periodically repeated defects at the smallest concentration one

can handle computationally. Periodicity imposes several artifacts: the defect levels broaden into bands because of their interaction, the structural distortion around the defect may result in long-range elastic forces, and for charged defects, there is a spurious Coulomb interaction between the image charges and between them and the compensating background one introduces to enable a meaningful definition of total energies. The total energy of an infinite periodic system is only well-defined if it is overall neutral.

The direct band broadening effects are presumably easiest to avoid if the defects wave functions are exponentially localized, which is the case for deep defects, while for shallow defects, an accurate description of the binding energies relative to the bands is probably better attempted in the framework of effective mass approximation methods [42–48].

Nonetheless, some care is needed to deal with defect band dispersion. Aradi *et al.* [49] for example use a tight-binding fit to the defect band dispersion to derive the center of gravity of the actual isolated defect level. This approach was used earlier by Louie *et al.* [50] and is more important the smaller the cells are. Wei and Yan [51] in chapter 13 of this compilation discuss an approach in which the one-electron defect levels calculated at the Γ-point, which reflect the correct symmetry of the isolated defect levels, are combined with the transition state approach. The point is that one generally uses a special k-point set to calculate total energies and the differences in defect level position at the special k-point from that at Γ needs to be taken into account.

The elastic effects are expected to behave like $1/L^3$ while the image charge electrostatic effects are longer range and expected to behave like $1/L$ with L the characteristic length scale of the supercell, say $1/V^{1/3}$ with V the volume. The fact that the image charges in the neutral background leads to a spurious Madelung contribution $\alpha_M q^2/\varepsilon L$ to the total energy of the system is known since the work of Leslie and Gillan [52]. Makov and Payne [53] identified a correction describing the interaction of the quadrupole moment of the defect density with the background going as $1/L^3$. Nonetheless, these proposed corrections were not universally adopted by practitioners in the field. It was for example argued by Segev and Wei [54] that if the defect density becomes delocalized the Madelung model, invoking point charges in jellium, overestimates the corrections. This point of view is discussed in Wei's contribution in this compilation (chapter 13, [51]). Instead these authors proposed to use an extrapolation scheme proportional to $1/V$ or proportional to the number of atoms in the cell. Gerstmann *et al.* [55] also criticized the approach for the case of delocalized defect wave functions. Other more sophisticated approaches were introduced but because they are more difficult to implement or focus on periodic boundary calculations of molecules, were rarely used [56]. One confusing point here is that even if the defect wave function is delocalized, the total electrostatic density perturbation contains the δ-function part of the nuclear charges and thus certainly the total charge density contains a point-like monopole contribution. A major reason why it has not always been that clear in practical studies to see the pure q^2/L behavior, is that other finite size effects may be dominating. To clarify the situation, it is best to consider the effect in unrelaxed

structures and one needs large enough cells of order several 100 atoms, before this behavior becomes apparent [57]. Attempts have been made by several authors to fit separately, $1/L$ and $1/L^3$ terms [58].

Recently, the problem was reformulated in a slightly different manner by Freysoldt *et al.* [59] and their method is described in additional detail in chapter 14 of this compilation [60]. Their analysis is based on plotting the defect minus perfect crystal potential and subtracting from it a long-range part, calculated with an assumed unscreened defect charge model, which in the simplest case is just a point charge. One important point arising from their analysis is that in doing so one must account for an arbitrary constant shift in the potential, a so-called alignment potential. The latter is chosen so that the remaining short-range part of the potential explicitly goes to zero far away from the defect. If necessary to make the potential flat in the region far away from the defect, a more sophisticated defect model charge density is introduced. This alignment term has the form $q\Delta_{q/b}$ and enters the defect formation energy together with the qE_F term which represents the chemical potential of the electron which one must add to describe correctly the change in Gibbs free energy of the charged defects, which is considered to be an open system in connection with an electron reservoir. Usually, this term is added separately but in Freysoldt *et al.*'s analysis, it emerges naturally from their consideration of the spurious interaction energies which one wishes to remove.

It is important to note that the alignment term goes as $1/V$ and was also emphasized by several other authors [57, 61–64]. It is being used by many other authors in some form or other, although the procedure used for determining it is not always explicitly mentioned in the literature. When a defect is created in charge state $+q$, the electron must be removed to the electron reservoir with energy its chemical potential $\mu_e = E_{vbm} + E_F$. Usually, one measures the Fermi level here relative to the valence band maximum (VBM). However, the question is how to calculate E_{vbm} of the perfect crystal in the defect containing cell. One might think this is just the highest occupied band (at the appropriate *k*-point), not counting the defect levels in the supercell itself, but the problem is that the defect may have perturbed the band edges in the supercell. It helps to plot the bands of the supercell, so one can recognize defect levels from host bands by their dispersion. The accepted alignment approach is to use a "local characteristic" of the potential, say the average of the electrostatic potential over an atomic sphere, or the potential at the muffin-tin radius in methods that use such spheres, or a core level. If one now knows the valence band energy in the perfect crystal relative to this "local potential marker," then all we need to determine is the same marker at atoms far away from the defect in the supercell containing the defect. Typically, one needs to average over a few atoms far away from the defect, and this is for instance illustrated in Lany and Zunger [65]. Freysoldt *et al.* [59, 60] essentially use a well defined separation approach of long and short range parts of the defect potential with a built-in check to make sure that after the long-range effects are subtracted the short-range potential indeed just becomes a constant. But essentially, it is just a different way of determining the alignment potential and the remaining correction is just the Leslie and Gillan [52] Madelung correction or the $1/L$ Makov– Payne term [53].

Now, there still remains the question of the Makov and Payne's quadrupole term. Lany and Zunger [57] recently pointed out that the net quadrupole of the defect charge density $Q \propto L^2$, i.e., it is not independent of the supercell size and the quadrupole correction $2\pi q Q/\varepsilon L^3$ then effectively behaves also as a $1/L$ term. The reason for this observed behavior is that Lany and Zunger's defect charge density from which the quadrupole moment is calculated includes the screening charge density which becomes almost constant at large distance from the defect in the supercell. The definition of Q involves an integral over $r^2 \varrho(r)$ and is thus dominated by large r. The same behavior in fact was pointed out earlier by Lento *et al.* [66]. Furthermore, Lany and Zunger [57] found that this reduces the monopole correction by a factor which is essentially independent of defect and amounts to about $-1/3$. Thus their prescription for the image charge correction becomes: take 2/3 of the point charge correction.

On the other hand, Makov and Payne [53] defined Q explicitly as "the second radial moment only of that part of the aperiodic density that does not arise from dielectric response or from the jellium, i.e., is asymptotically independent of L." This contradicts Lany and Zunger's statement that $Q \propto L^2$ but this is simply a question of whether or not one here includes the short-range dielectric screening. We here say short-range screening because the remaining long-range screening is included by dividing by the dielectric constant. It is not obvious how to determine Q according to Makov and Payne's strict definition and one might wonder why all (even short range) effects of the screening density should be excluded. Lany and Zunger define the defect charge density simply as the difference of the charge density in the cell with the defect minus the corresponding perfect crystal charge density calculated in the same cell.

In my opinion, this is closely related to the question, is it actually correct to include a compensating background density? In reality a charged defect is after all compensated by other defects far away and not by a homogeneous back ground. Near interfaces for example, it is well known that one has depletion layers which are actually charged. On the other hand, if we consider a finite small region around the defect, say one or two shells of neighbors, and consider this as an open system, then clearly charge can flow into this region. In principle it should be determined self-consistently in the presence of a given chemical potential of the electrons. The latter sets the energy up to which to integrate the local densities of states to determine the total charge inside the defect region. So, if a defect level is below it, it will be occupied and if above it, it will be empty. Even in a Green's function method, one would only treat a finite region of a few cells around the defect as the region where the potential is assumed to be different from bulk. But if the defect region retains a net charge, the potential outside still has a long-range Coulomb tail. Presumably that could be used as boundary condition for the potential in the defect region although I am not aware of calculations where this was explicitly done. Asymptotically a charged defect should behave as $q/(\varepsilon r)$ with ε the macroscopic dielectric constant, but this means that the net charge of the defect region is reduced to q/ε, meaning that $q[1 - (1/\varepsilon)]$ has indeed flown toward the defect. As long as ε is fairly large (perhaps a wrong assumption for very ionic oxides!) assuming local charge neutrality seems a reasonable approximation. In other words, the background density represents to some extent the physical screening charge density, more precisely, the short-range part of it.

Thus, the screening charge density if included in the definition of Q seems to arise in large part from the background density. The quadrupole correction of Makov and Payne is the interaction of this quadrupole with the background. So, it is somewhat puzzling whether this needs to be included since it appears to amount at least in part to an interaction of the background with itself. A detailed argument of Lany and Zunger's point of view can be found in Ref. [65].

A subtle point is to what extent the background density is really included in the codes that are commonly in use. In some codes, it appears the latter's presence is only assumed in order to give a well defined meaning to the reference electrostatic potential, which implicitly assumes charge neutrality but the interaction of the background density with the electrons and nuclei is not explicitly included in the total energy. This point is also discussed in [57, 65]. Other codes may explicitly add the uniform background density to the charge density in a systematic manner. This is for example the case in the FP-LMTO code [67].

Now, so far we have only considered corrections to the total energies of the system. In another recent paper, Lany and Zunger [68] showed that also one-electron energy levels are subject to finite size potential shifts.

On the other hand, it is worthwhile mentioning that some attempts have been made to go outside the standard practice of using a compensating background. Schultz [69–71] for example has advocated the use of a local moment counter charge approach, in which long-range effects of the net charge of the defect are treated with boundary conditions of a single isolated defect in calculating the electrostatic potential, while the remaining moment free (up to some moment order) density is treated with the usual periodic boundary conditions. The special role of an undetermined alignment potential also crops up in his theory. Its advantages and disadvantages versus the neutralizing back ground density have been studied by Wright and Modine [72] and by Lento *et al.* [66] One finds, in fact that the convergence with size of the system is slower than in the background charge approach and outside the defect region, still a classical continuum model polarization correction must be added. We cannot delve into the details of these other approaches here and conclude that likely, discussions of finite size effects will continue for some time.

As mentioned in Section 19.1, periodic boundary artifacts can in principle be entirely avoided by resorting to Green's function methods which provide an exact embedding of the defect region in the host crystal. A remarkable feat of such methods is that they are able to describe the correct defect wave functions well outside the region in which the potential (and possibly the structure) is perturbed. This is emphasized in the contribution by Gerstmann [15] in this monograph (chapter 17). It is evidenced by the degree of accuracy with which it can describe hyperfine and super-hyperfine interactions. While the Green's function methods were originally developed with deep defects in mind, they have now apparently found their most impressive performance on shallow defects. Even though in his approach, the long-range Coulomb tail of the charged defect is not included, and thus the shallow defect becomes a resonance in the conduction band, it is possible within this approach to accurately identify the resonance in the density of states and thus to reconstruct the charge density from the Green's function integrated over the energy range of the resonance.

Shallow defects have perhaps been overlooked for a while with the impression that the shallow defect problem was solved in terms of effective mass theory (EMT) long ago. Nonetheless Gerstmann's contribution here draws attention to the shortcomings of the EMT in terms of describing the so-called central cell correction. We also point out that some recent work is trying to revive the EMT with novel approaches to refine the central cell potential [46–48]. Perhaps a fruitful avenue will be to combine first-principles supercell or Green's function approaches to extract central cell potentials to be used in EMT.

19.2.2
Band Gap Corrections

As is well known, the Kohn–Sham eigenvalue gap is underestimated by the LDA. On the other hand, the prime question about point defects is where the defect levels, either one-electron levels or transition levels lie with respect to the band edges. Thus, a correction of the band gap is necessary before one can compare to experiment.

The most dramatic failure related to the band gap underestimate occurs when it leads to an erroneous occupation of host states rather than defect states because the defect level becomes a resonance in the bands. This situation for example occurs for the oxygen vacancy in ZnO [73], a defect that was discussed by several contributors at the workshop and has become a sort of benchmark [68, 74–78].

In the past, several approaches have been used: some are *a posteriori* corrections, some are addressing the problem at the level of the Hamiltonian by going beyond LDA in some way or other. *A posteriori* corrections come down to deciding what is the nature of the defect state. If the defect is essentially a shallow acceptor, the idea is that the defect level position relative to the VBM is correct, and one would then just shift up the conduction band minimum (CBM) without changing the defect level. If the defect is a shallow donor state, the defect level would be shifted up along with the CBM. If it is a deep defect, the intuitive idea is that the level would shift according to how much it is valence band or conduction band like. How to apply this intuitive idea in practice is another matter and different approaches give different results.

One approach goes by the name of a "scaled scissor correction." The approach consists in determining the projection of the defect state onto valence and conduction band edge states to determine its percentage valence and conduction band character [79]. However, its limitations were pointed out by Deák *et al.* [80] by comparing how much the one-electron levels shift in this approach compared to a GW quasiparticle calculation.

A related idea was used by Janotti and Van de Walle [77, 81] for the oxygen vacancy in ZnO. The LDA $+ U$ approach is designed originally to deal with localized orbitals such as semicore d states or open shell d and f systems [82, 83]. It produces a partial band gap correction because it reduces the p–d hybridization. Janotti and Van de Walle [81] reasoned that the extent to which this correction shifts the defect level, shows to what extent it is valence or conduction band like and thus used this as a basis for extrapolating to the full gap correction, even if the rest of the band gap correction is not arising from the p–d hybridization effect. However, it is not clear that one can

decompose the defect wave function in CBM and VBM like host states. Certainly for deep defect levels, the idea of pinning the defect level to the host state at one dominant k-point seems incorrect.

The scissor shift is easily incorporated in Green's function methods [4, 8, 15]. In that case, one could even relatively easily include different shifts at different k-points instead of a uniform shift.

A better approach clearly is to use an energy functional or Hamiltonian for the host that reproduces the correct gap, or at least gives a better approximation to it that results in a qualitatively correct starting point. We already mentioned that the LDA $+$ U approach at least partially opens the gap. An extension of this approach is applying it to the states that dominate the CBM, typically cation s-like states [73, 74]. Another is to add simply non-local external potentials that shift the appropriate states [57, 84], or to use modified pseudopotentials [77, 85]. The success of such approaches depends significantly on the details of the implementation [74]. While their main advantage is simplicity and computational efficiency, it is undesirable that one needs to adjust these potentials on a case by case basis. Much hope for a universally applicable approach is placed these days in hybrid functionals.

Hybrid functionals were discussed by several contributors to the conference [75, 77, 86–88]. Hybrid functionals essentially mix some Hartree–Fock with LDA or GGA exchange. Thereby they add an orbital dependence to the exchange correlation functional that is missing in the LDA. This is also what the LDA $+$ U methods are essentially trying to mimic. While hybrid functionals were first explored by chemists for small molecules, e.g., the Becke B3LYP functional [89, 90], more recent versions seem to be rather successful to reproduce band gaps of standard tetrahedrally bonded semiconductors. The main new development however, is that they are now being implemented in the popular plane wave programs and can thus more readily be applied to the systems of interest. They still are typically much more time consuming than semilocal functionals.

Among hybrid functionals, we can distinguish those that mix a fraction α of unscreened Hartree–Fock with GGA, and those that use screened Hartree–Fock. The former approach is called PBE0 or PBEh [91] if the fraction is $\alpha = 0.25$ and added to the original PBE–GGA functional [92, 93]. This fraction was argued to be optimal based on many-body perturbation theory [91]. The drawback of including unscreened Hartree–Fock is the $1/r$ singularity of Hartree–Fock, which manifests itself as a $1/q^2$ singularity in reciprocal space. How to treat this carefully is discussed by Alkauskas *et al.* [86, 94] and by Kotani *et al.* [28] Among the screened HF approaches, the HSE [95, 96] introduced by Scuseria and coworkers is most popular. In that approach, the Hartree–Fock is divided in a long-range and short range (or rather medium range, as explained by Scuseria in his contribution to this volume [88]) by means of an error-function cut-off similar to the well-known Ewald procedure, and a fraction of the medium range part is included in the final functional. Roughly speaking, the idea is that truly long-range behavior will be canceled by corresponding correlation and should be avoided because it would lead to unphysical behavior in metals for example. At the same time, this makes the computational approach more easy to implement. A detailed discussion of why it is believed to improve band gaps and which band gaps is

provided in chapter 6 [88]. This chapter also discusses to what extent one can expect that a universal materials independent range-separation of exchange and correlation can be expected to work.

An alternative implementation of mixing screened Hartree–Fock is the so-called screened exchange approach [97–99]. The latter is discussed here in the chapter by Clark and Robertson (chapter 5, [75]) and their recent papers [76, 100, 101] and is justifiable as a Generalized Kohn–Sham scheme [97]. In that case, the screening of the exchange is usually done by a Thomas–Fermi exponential screening and the screening length is not arbitrarily chosen but determined by the valence electron density (excluding d-electrons). The fraction of screened HF included is determined by the double counting correction in that case. A well defined LDA of the screened exchange exists and is subtracted from the usual LDA so that in practice again a mix of LDA and screened HF is effectively used.

The HSE approach has been implemented [102, 103] in Vienna *ab initio* simulation package (VASP) [104] and has begun to be tested for point defects by several contributors. Examples represented in this compilation are the chapters by Deák *et al.*, (chapter 8, [87, 105]) Van de Walle and Janotti (chapter 9, [77]), and Alkauskas *et al.* (chapter 7, [86]). Deák *et al.* (chapter 8, [87]) emphasize the importance of obtaining accurate defect energy levels as a prerequisite for obtaining accurate formation energies and other derived properties. They consider a variety of correction approaches of the one-electron levels and their impact on total energy properties. The SX approach was implemented by Clark and Robertson [75] in CASTEP [106] and has been much less used, so it is too early to compare with the HSE approach in terms of practical results. It has previously been implemented in FLAPW by Freeman and coworkers [99].

In spite of the successes of these hybrid functional approaches, there remain several issues under debate. One approach is to adjust the fraction α of mixing in (screened) HF so as to exactly adjust the band gap for a particular system. The other is to stick to the universal 0.25 mixing factor. The second freedom is what screening length or long-range cut-off parameter to use. Unfortunately, these choices still lead to significantly different results for defect levels relative to the band edges, notably again the oxygen vacancy in ZnO as can be seen by comparing results from the different groups [68, 74–77].

Another question is, if one adjusts the mixing parameter for each system, then what to do at an interface? This question is addressed by Alkauskas *et al.* [86, 107]. Another interesting point raised by these authors [108] is that at least for well localized defects, it appears that defect levels measured relative to the average electrostatic potentials are in much better agreement among different approaches than relative to the VBM. In other words, the problem appears to lie in determining the band edges rather than the defect levels! Thus, one needs to investigate not only how well these new functionals do on the band gaps but how they do on the individual band edges. This however is a tricky question because individual energy levels on an absolute scale cannot be defined in a periodic crystal [109]. Nonetheless, whether such a level has physical meaning or not, one can refer levels with respect to the average electrostatic potential as zero and ask how these differ in different approaches, such as GGA, HSE, and GW. Komsa *et al.* [110] recently discussed how the different mixing

and screening parameters in hybrid functionals affect the band edges and defect levels relative to the electrostatic potential reference.

The most accurate approach for band gaps in semiconductors at present is the GW method. On the other hand, it is also the most computationally challenging to apply to defects. Secondly, GW does not readily fit into the DFT framework for total energies. It is a many-body perturbation theory for quasiparticle excitations. As such there is now a refocus of interest in one-electron energies [80]. One inventive approach by Rinke *et al.* [111] and discussed here in Giantomassi *et al.* [112] is to use a mixed approach of GGA and or hybrid functionals with GW. The idea is that GW quasiparticle energies give correctly the vertical (unrelaxed) total energy difference occurring in a change of charge state of a defect. After all, this occurs by transferring an electron from the defect level to a band state or vice versa. So, in principle, GW gives this excitation energy correctly for fixed geometry. The hybrid functional or GGA approach is then used subsequently to study the relaxation energy in a given charge state. This is a promising approach, which was recently also applied to the oxygen vacancy in ZnO by Lany and Zunger [68]. The GW method was well represented with several talks at the work shop and is further discussed in the next section.

While for some time calculations have focused on energies of formation and transition levels, i.e., the Fermi energies where the formation energies for different charge states cross, the use of GW brings back the one electron levels into focus. Relatedly Deák *et al.* [87] remind us that the one electron levels are a significant part of the total energy expression and thus changing these one-electron levels by for example gap corrections indirectly also changes the total energy derived quantities.

19.2.3
Self-Interaction Errors

A second important error of LDA that was widely discussed at the workshop is the self-interaction error. In Hartree–Fock, the exchange term exactly compensates the Coulomb interaction of an electron in a specific one-electron eigenstate with itself. LDA and GGA make approximations to the exchange functional and hence the self-interaction error (SIE) is not exactly canceled. This among other leads to a tendency of the semilocal functionals (LDA and GGA) to favor states in which defect electrons delocalize over several atoms. Hartree–Fock on the other hand misses correlation effects entirely and too strongly favors localization. This problem has been known for some time now to affect the structural relaxation primarily for deep acceptors [113–120]. It prevents the formation of a localized hole on a specific atom and the accompanying symmetry breaking relaxations, so-called Jahn–Teller distortions. Such localized relaxations trapping an electron or hole, are also called self-trapped polarons.

One of the earliest systems in which this problem was noticed is the Al acceptor in SiO_2. While LDA spreads the defect wave function equally over the four nearest neighbors, Hartree–Fock finds a localized state on one of the oxygens and it was recognized that this results from the spurious self-interaction error of LDA [113]. Subsequently, it was shown that a calculation including explicitly the self-interaction correction for the defect state solves the problem [114]. However, applying an *ad hoc* self-interaction correction for a

specific defect state is cumbersome. Others found that LDA $+$ U could also solve the problem [115]. Other defects were soon identified that show similar problems, for example the Zn vacancy in ZnO, the Li_{Zn} in ZnO [116, 120, 121].

The contribution in this compilation by Lany (chapter 11, [122]) discusses the problem and his cure for it. While LDA $+$ U including Coulomb interactions U on the anion p states would help to localize the corresponding hole states, they have the disadvantage that applying a sufficiently strong U may perturb the host band structure. Lany and Zunger therefore construct a so-called hole state potential, which by construction is zero as long as the occupation of the anion p-orbitals is the same as that of the host. Only those orbitals on which a hole localizes, thus decreasing their occupation number feel a repulsive potential that further re-enforce the localization by pushing the level into the band gap. In order to answer the question, how strong should the localization potential be, they apply a so-called generalized Koopmans' theorem [123].

As shown by Perdew *et al.* [124] the total energy as function of occupation number in exact DFT for open systems with a continuously varying occupation number should be a piecewise linear function. The impacts of this on the delocalization problem were pointed out recently by Mori-Sánchez *et al.* [125]. Because of Janak's theorem [126] the linearity of the total energy implies constant Kohn–Sham eigenvalues as function of occupation numbers. LDA shows convex $d^2E/dn_i^2 > 0$ and Hartree–Fock concave $d^2E/dn_i^2 < 0$ behavior instead of linear behavior. A correct behavior leads to equality of the total energy difference (ΔSCF approach) with the Kohn–Sham eigenvalue for the defect levels, or satisfying a generalized Koopmans' theorem. Originally the theorem was derived within Hartree–Fock theory [123] but is only valid for delocalized states. Lany and Zunger thus adjust the strength of their hole-state potential so as to satisfy the generalized Koopmans' theorem and find that in many cases this leads to polaronic behavior where LDA fails to describe the correct relaxation of the system.

Others have used LDA $+$ U and adjusted the U-value to satisfy the same criterion, for example in a study of polaronic trapping in TiO_2 [127]. Interestingly, the same considerations about the required linear behavior of the total energy as function of occupation number are at the heart of a recently proposed method for determining U in LDA $+$ U methods [128]. The use of Janak's theorem to calculate transition energies in the context of LDA $+$ U theory was also advocated by Sanna *et al.* [129]. A point that should not be forgotten in this context is that Janak's or Koopmans' theorem refer to a specific eigenstate, while in the LDA $+$ U or hole-state potentials, one applies it to a basis set specific state or local atomic orbital. This is different from the explicit self-interaction correction by d'Avezac *et al.* [114].

While the correct physics, requiring an increased separation of empty and filled defect states is built into these methods and its strength is adjustable according to a well-described *ab initio* criterion, one might object against the *ad hoc* form of the hole state potential or worry whether applying such LDA $+$ U corrections would not spoil other aspects of the electronic structure for the host system. It is therefore of interest to see how hybrid functionals do for such polaronic systems. For example, Clark and Robertson [75, 76] note that their screened exchange method correctly describes the polaron hole trapping at the Zn-vacancy in ZnO with localization on a single atom. Deák *et al.* [105] discuss the satisfaction of Koopmans theorem by the

HSE functional. The application of the B3LYP hybrid functional to bound polarons in ZnO was discussed at the workshop by Du and Zhang [120]. A comparison between the hole-state potential and the HSE hybrid functional for N_0 can be found in Lany and Zunger [117] and Lyons *et al.* [130].

19.2.4
Beyond DFT

Besides the direct applications to defects, the workshop contained a good deal of discussion of the underlying methodologies that go beyond LDA. The main issue is how to make the more advanced methods, such as GW, Bethe–Salpeter equations, time-dependent DFT efficient enough to become applicable to large systems as required for point defect studies.

Two examples represented in this compilation are the chapters by Umari *et al.* (chapter 4, [131]) and Giantomassi *et al.* (chapter 3, [112]). Umari *et al.* [131, 132] discuss the use of a separate small orthogonal basis set for expanding the polarizability operator. It should be recognized here that the bottleneck of GW calculations is the calculation of the wave vector and frequency dependent polarizability that goes into the calculation of the screened Coulomb interaction *W*. In most GW calculations this quantity is expanded in plane waves. Umari *et al.* discuss the construction of a separate small basis set that spans the space of products of Wannier functions. This approach reminds me of the product basis set approach introduced by Aryasetiawan and Gunnarsson [133]. While they construct product basis functions of muffin-tin orbitals rather than Wannier functions, the underlying idea is similar. We note that the product basis supplemented with plane waves forms the basis for expanding polarizability, screened and bare Coulomb interactions in the FP-LMTO implementation of GW by van Schilfgaarde *et al.* [27–29] Umari *et al.*'s approach shows great promise to speed up GW as they show by applying it to supercells containing a few 100 atoms.

One of the remaining problems is the need to sum over a large number of empty states. This problem was also tackled by Umari *et al.* [134] by reformulating the calculation in such a way that no explicit summation of empty bands is required. Instead one uses the completeness and rewrites the sum over empty states as one minus the projection operator over all filled states. This is essentially similar to the Sternheimer approach that has been so successful in linear response theory [135]. A very similar approach was recently introduced by Giustino *et al.* [136].

The chapter by Giantomassi *et al.* [112] on the other hand discusses approaches beyond the GW approximation. For example, they discuss the inclusion of so-called vertex corrections as well as the so-called quasiparticle self-consistent GW approach. After all, GW as conceived by Hedin is only the first approximation in a perturbation series. The key here is that usually GW theory is applied as a one-shot correction to some underlying one-electron theory. The latter is usually LDA or GGA but could also be HSE or LDA + U. The idea of QS–GW is to construct the best one-electron (generalized) Kohn–Sham starting point with a non-local exchange potential extracted from the GW self-energy itself. The method was introduced by van Schilfgaarde *et al.* [27–29]. and implemented with a FP-LMTO basis set and product basis sets.

The way forward beyond GW is still a matter of debate. Early results indicated that self-consistent GW without vertex corrections gave worse results that single-shot GW. Other complications arise from the effects of semicore orbitals and differences between all-electron and pseudopotential implementations. The QSGW appears to be the most successful among those approaches and leads to small remaining and highly systematic errors on the band gaps of a wide variety of systems, not only standard semiconductors. One of the remaining errors is the use of the random phase approximation (RPA) in the dielectric screening. Including electron–hole interaction effects in the latter is expected to further improve the method.

Giantomassi *et al.* [112] also discuss various technical aspects of the GW method, such as the summation over empty states, the plasmon–pole approximation, and the PAW implementation. They show applications to the problem of band-offsets at interfaces and to point defects along the lines discussed earlier.

While at present GW is used to calculate one-electron excitations, there is also progress to turn GW into a total energy theory. In principle, it is closely related to the RPA total energy [28, 137]. However, it is not yet clear how stable this approach is to calculating total energies and how good they are compared to LDA or GGA and experiment. A recent evaluation of RPA was made by Harl *et al.* [138, 139].

At the workshop some results were also presented by M. Rohlfing applying the Bethe–Salpeter approach including excitonic effects to defect problems. While no contribution of his is included here, we refer the reader to [140–142].

Relatedly, a time-dependent density functional approach to point defects is presented in this compilation by Gali (chapter 18, [143]). His approach is part of a new direction that explores the calculation of excited states of defects. This is an extremely important new direction because much of the experimental information on point defects relates to optical excitations, within the defect. This first of all requires one to use non-equilibrium occupations of the defect levels (constrained DFT) but secondly, electron–hole excitonic effects can be expected to be important as they also are in low-dimensional systems and molecules. The contribution here describes a combination of such approaches, from the constrained DFT implemented with hybrid functionals in VASP to a time-dependent DFT approach implemented in a cluster calculation. He discusses two specific systems: the N–V centers in diamond and divacancies in SiC.

Finally, as mentioned earlier Hennig and coworkers (chapter 2, [41]) discuss progress in Quantum Monte Carlo calculations. These calculations provide a benchmark for total energies for defects in relatively small supercells of order 16 atoms. The discussion is mostly on Diffusion Monte Carlo and the various types of controlled and uncontrolled error in this method and their application to self-interstitials in Si.

19.3
Conclusions

In this chapter, I have reviewed some of the main themes that were addressed during the CECAM workshop "Which electronic structure method for the study of defects?," held in Lausanne, June 8–10, 2009. It was not my intention to be complete in this

review but rather to give the reader a vivid impression of the issues that are currently under discussion in this field and the progress that is being made. The commentary given reflects my own point of view as a practitioner in this field and not necessarily that of the papers mentioned. If my understanding is incomplete, I apologize to the authors. I have tried to put the articles that can be found in this compilation in the context of the current open literature. The articles in this compilation provide only a limited glimpse at what was presented at the conference. First, not all presenters at the workshop chose to contribute to this compilation and secondly, some chose to focus on a particular part of their presentation. In any case, no set of separate articles can ever capture the gist of the many lively discussions that followed the presentations at the workshop.

I have focused this commentary on the original question in the title, focusing on methodology rather than on specific applications. At the same time, it should be said that the reader of this compilation will encounter a wide variety of defect problems in materials, from defects in Si to wide band gap semiconductors and oxides and interfaces, showing that the field is very much alive and not at all in an impasse over unsolved methodological questions. I have also tried to draw some attention to some of the "forgotten" problems and approaches, which were not very much represented during the workshop. In particular, the problem of EMT of shallow defects may see some revival and the chapter included here on the Green's function approach [15] will hopefully remind some people of the promise held by this method.

The newcomer or outsider to this field might at first get the impression that the field is in turmoil with many conflicting opinions. However, the fact that these methodological questions are now being discussed in the open literature is very healthy for the field and will hopefully help newcomers to avoid common pitfalls in how to apply these methods. It is pretty clear now that image charge corrections need to be made for charged defects and attention has to be paid to the proper alignment. Also, there is a consensus now that band gap corrections are important, in particular in cases where the low LDA gaps would lead to incorrect band filling for certain charge states. The different approaches to achieve it have their pros and cons: hybrid functionals and GW are more expensive but unbiased, not empirical. Non-local external potentials or LDA + U or hole-state potentials are less computationally expensive but require careful selection of the associated parameters. The broader availability of these new approaches will assist in their rapid deployment and will help in ascertaining their success. In any case, it does not mean that all prior work with LDA only is invalidated by the new approaches. It all depends on the system under study and on what questions the calculation is trying to answer.

We should also be reminded of the continuing successes of the field. In the end the task of computational physics in this field is to assist experimentalists in extracting the maximum information and understanding from their experiments. The standard approach for defect calculations has provided a lot of guidance on which defects are likely to form under what circumstances and what the basic characteristics of specific defects are. It has also been able to provide significant guidance on how to overcome doping problems [51], and enables one to estimate defect concentrations accurately taking into account rather complex defect chemistry and reactions. With the new

approaches, the accuracy with which defect level positions can be calculated with respect to the band edges is steadily improving. Theory still lags behind experiment: optics experiments can determine differences between sharp photoluminescence lines to better than a meV, and can purely spectroscopically distinguish different defects. However, determining the chemical indentity of defects is likely to remain a complex task requiring consistency between various experimental techniques, and computational results. The new focus in the theory on excited state properties which correlate more directly to optical studies is very promising. In fact, almost all experiments deal with excited states, in some way or another, whether optically activated EPR signals, or photoluminescence or optical absorption. The fact that we can now start addressing quasiparticle and optical (electron–hole pair) excitations including excitonic effects is an important advance in the theory. While the concepts of Franck–Condon diagrams have been around for a long time, we can now start actually calculating them just like for isolated molecules. Finally, defect wave functions, in particular the delocalized or localized character can now be addressed. In particular, great progress was made recently in realizing the importance of polaronic effects and such information can be directly tested by means of EPR fine structure. Even for delocalized shallow states, good agreement can be obtained between calculated wavefunctions and the hyperfine structure. Although not mentioned in this compilation, certain defects at or close to the surface can be visualized with scanning tunneling microscopy and this provides another testing ground for the theoretical capability to determine defect wave functions.

In short, the future of computational defect studies is bright. The theory can be increasingly applied to more complex solids, its accuracy is improving steadily and even defects in complex nanostructures are within reach.

Acknowledgements

References

1 Deák, P., *Phys. Status Solidi B*, in press.

2 Deák, P., Miró, J., Gali, A., Udvardi, L., and Overhof, H. (1999) *Appl. Phys. Lett.,* **75**(14), 2103–2105.

3 Baraff, G.A. and Schlüter, M. (1978) *Phys. Rev. Lett.,* **41**(13), 892–895.

4 Baraff, G.A. and Schlüter, M. (1979) *Phys. Rev. B,* **19**(10), 4965–4979.

5 Bernholc, J. and Pantelides, S.T. (1978) *Phys. Rev. B,* **18**(4), 1780–1789.

6 Bernholc, J., Lipari, N.O., and Pantelides, S.T. (1978) *Phys. Rev. Lett.,* **41** (13), 895–899.

7 Bernholc, J., Lipari, N.O., and Pantelides, S.T. (1980) *Phys. Rev. B,* **21**(8), 3545–3562.

8 Gunnarsson, O., Jepsen, O., and Andersen, O.K. (1983) *Phys. Rev. B,* **27**(12), 7144–7168.

9 Beeler, F., Andersen, O.K., and Scheffler, M. (1985) *Phys. Rev. Lett.,* **55**(14), 1498–1501.

10 Lindefelt, U. and Zunger, A. (1981) *Phys. Rev. B,* **24**(10), 5913–5931.

11 Lindefelt, U. and Zunger, A. (1982) *Phys. Rev. B,* **26**(2), 846–895.

12 Braspenning, P.J., Zeller, R., Lodder, A., and Dederichs, P.H. (1984) *Phys. Rev. B,* **29**(2), 703–718.

13 Zeller, R. and Dederichs, P.H. (1979) *Phys. Rev. Lett.,* **42**(25), 1713–1716.

14 Payne, M.C., Teter, M.P., Allan, D.C., Arias, T.A., and Joannopoulos, J.D. (1992) *Rev. Mod. Phys.,* **64**(4), 1045–1097.

15 Gerstmann, U. (2011) *Phys. Status Solidi B*, doi: 10.1002/pssb.201046237. (chapter 17 in this book).

16 Briddon, P.R. and Rayson, M.J. (2011) *Phys. Status Solidi B*, doi: 10.1002/pssb.201046147. (chapter 16 in this book).

17 Hedin, L. (1965) *Phys. Rev.* **139** (3A), A796–A823.

18 Hedin, L. and Lundqvist, S. (1969), Effects of electron–electron and electron–phonon interactions on the one-electron states of solids, in: *Solid State Physics, Advanced in Research and Applications*, edited by Seitz, D. Turnbull, and H. Ehrenreich (Academic Press, New York) pp. 1–181.

19 Hybertsen, M.S. and Louie, S.G. (1986) *Phys. Rev. B,* **34**(8), 5390–5413.

20 Godby, R.W., Schlüter, M., and Sham, L.J. (1988) *Phys. Rev. B,* **37**(17), 10159–10175.

21 Bruneval, F., Vast, N., and Reining, L. (2006) *Phys. Rev. B,* **74** (4), 045102.

22 Bruneval F. and Gonze X. (2008) *Phys. Rev. B,* **78**(8), 085125.

23 Aryasetiawan F. (1992) *Phys. Rev. B,* **46** (20), 13051–13064.

24 Friedrich C., Schindlmayr A., Blügel S., and Kotani T. (2006) *Phys. Rev. B,* **74** (4), 045104.

25 Aryasetiawan F. and Gunnarsson O. (1995) *Phys. Rev. Lett.,* **74**(16), 3221–3224.

26 Aryasetiawan F. and Gunnarsson O. (1998) *Rep. Prog. Phys.,* **61**(3), 237.

27 van Schilfgaarde, M., Kotani, T., and Faleev, S.V. (2006) *Phys. Rev. B,* **74** (24), 245125.

28 Kotani, T., van Schilfgaarde, M., and Faleev S.V. (2007) *Phys. Rev. B,* **76** (16), 165106.

29 Faleev, S.V., van Schilfgaarde, M., and Kotani, T. (2004) *Phys. Rev. Lett.,* **93** (12), 126406.

30 Lebègue, S., Arnaud, B., Alouani, M., and Bloechl, P.E. (2003) *Phys. Rev. B,* **67**(15), 155208.

31 Shishkin, M. and Kresse, G. (2006) *Phys. Rev. B,* **74**(3), 035101.

32 Shishkin, M. and Kresse, G. (2007) *Phys. Rev. B,* **75**, 235102.

33 Rohlfing, M. and Louie, S.G., (1998), *Phys. Rev. Lett.,* **80**(15), 3320–3323.

34 Rohlfing, M. and Louie, S.G. (1998) *Phys. Rev. Lett.,* **81**(11), 2312–2315.

35 Albrecht, S., Reining, L., Del Sole, R., and Onida, G. (1998) *Phys. Rev. Lett.,* **80**(20), 4510–4513.

36 Onida, G., Reining, L., and Rubio, A. (2002) *Rev. Mod. Phys.,* **74**(2), 601–659.

37 Schmidt, W.G., Glutsch, S., Hahn, P.H., and Bechstedt, F. (2003) *Phys. Rev. B,* **67**(8), 085307.

38 Fuchs, F., Rödl, C., Schleife, A., and Bechstedt, F. (2008) *Phys. Rev. B,* **78**(8), 085103.

39 Benedict, L.X., Shirley, E.L., and Bohn, R.B. (1998) *Phys. Rev. Lett.,* **80**(20), 4514–4517.

40 Reining, L., Olevano, V., Rubio, A., and Onida, G. (2002) *Phys. Rev. Lett.,* **88** (6), 066404.

41 Parker, D.W., Wilkins, J.W., and Hennig, R.G. (2010) *Phys. Status Solidi B*, DOI: 10.1002/pssb.201046149. (chapter 2 in this book).

42 Luttinger, J.M. and Kohn, W.(1955)*Phys. Rev.,* **97**(4), 869–883.

43 Pantelides, S.T. (1978) *Rev. Mod. Phys.,* **50** (4), 797–858.

44 Baldereschi, A. and Lipari, N.O. (1973) *Phys. Rev. B,* **8**(6), 2697–2709.

45 Baldereschi, A. and Lipari, N.O. (1974) *Phys. Rev. B,* **9**(4), 1525–1539.

46 Wang, H. and Chen, A.B. (2001) *Phys. Rev. B,* **63**(12), 125212.

47 Chen, A.B. and Srichaikul, P.(1997) *Phys. Status Solidi B*, **202**, 81.

48 Mireles, F. and Ulloa, S.E. (1998) *Phys. Rev. B,* **58**(7), 3879–3887.

49 Aradi, B., Gali, A., Deák, P., Lowther, J.E., Son, N.T., Janzén, E., and Choyke, W.J. (2001) *Phys. Rev. B,* **63**(24), 245202.

50 Louie, S.G., Schlüter, M., Chelikowsky, J.R., and Cohen, M.L. (1976) *Phys. Rev. B,* **13**(4), 1654–1663.

51 Wei, S.H. and Yan, Y. (2011) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046292. (chapter 13 in this book).

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

53 Makov, G. and Payne, M.C. (1995) *Phys. Rev. B,* **51**(7), 4014–4022.

54 Segev, D. and Wei, S.H. (2003) *Phys. Rev. Lett.,* **91**(12), 126406.

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

56 Blöchl, P.E. (1995) *J. Chem. Phys.,* **103**, 7482.

57 Lany, S. and Zunger, A. (2008) *Phys. Rev. B,* **78**(23), 235104.

58 Castleton, C.W.M., Hoglund, A., and Mirbt, S. (2006) *Phys. Rev. B,* **73**(3), 035215.

59 Freysoldt, C., Neugebauer, J., and Van de Walle, C.G. (2009) *Phys. Rev. Lett.,* **102**(1), 016402.

60 Freysoldt, C., Neugebauer, J., and Van de Walle, C.G. (2011) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046289. (chapter 14 in this book).

61 Laks, D.B., Van de Walle, C.G., Neumark, G.F., Blöchl, P.E., and Pantelides, S.T. (1992) *Phys. Rev. B,* **45**(19), 10965–10978.

62 de Walle, C. G. V. and Neugebauer, J. (2004) *J. Appl. Phys.,* **95**(8), 3851–3879.

63 Mattila, T. and Zunger, A. (1998) *Phys. Rev. B,* **58**(3), 1367–1373.

64 Persson, C., Zhao, Y.J., Lany, S., and Zunger, A.(2005) *Phys. Rev. B,* **72**(3), 035211.

65 Lany, S. and Zunger, A.(2009) *Model. Simul. Mater. Sci. Eng.* **17**,.

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

67 Kotani, T. and van Schilfgaarde, M. (2010) *Phys. Rev. B,* **81**(12), 125117.

68 Lany, S. and Zunger, A.(2010) *Phys. Rev. B,* **81**(11), 113201.

69 Schultz, P.A. (1999) *Phys. Rev. B,* **60**(3), 1551–1554.

70 Schultz, P.A. (2000) *Phys. Rev. Lett.,* **84**(9), 1942–1945.

71 Schultz, P.A. (2006) *Phys. Rev. Lett.,* **96**(24), 246401.

72 Wright, A.F. and Modine, N.A. (2006) *Phys. Rev. B,* **74**(23), 235209.

73 Paudel, T.R. and Lambrecht, W.R.L. (2008) *Phys. Rev. B,* **77**(20), 205202.

74 Boonchun, A. and Lambrecht, W.R.L. (2011) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046328. (chapter 10 in this book).

75 Clark, S.J. and Robertson, J. (2010) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046110. (chapter 5 in this book).

76 Clark, S.J., Robertson, J., Lany, S., and Zunger, A. (2010) *Phys. Rev. B,* **81**(11), 115311.

77 Van de Walle, C.G. and Janotti, A. (2011) *Phys. Status Solidi B,,* doi: 10.1002/pssb.201046384. (chapter 9 in this book).

78 Oba, F., Togo, A., Tanaka, I., Paier, J., and Kresse, G. (2008) *Phys. Rev. B,* **77**(24), 245202.

79 Castleton, C. W. M., Höglund, A., and Mirbt, S. (2006) *Phys. Rev. B,* **73**(3), 035215.

80 Deák, P., Frauenheim, T., and Gali, A. (2007) *Phys. Rev. B,* **75**(15), 153204.

81 Janotti, A. and Van de Walle, C.G. (2005) *Appl. Phys. Lett.,* **87**(12), 122102.

82 Anisimov, V.I., Zaanen, J., and Andersen, O.K. (1991) *Phys. Rev. B,* **44**(3), 943–954.

83 Anisimov, V.I., Solovyev, I.V., Korotin, M.A., Czyżyk, M.T., and Sawatzky, G.A. (1993) *Phys. Rev. B,* **48**(23), 16929–16934.

84 Lany, S., Raebiger, H., and Zunger, A. (2008) *Phys. Rev. B,* **77**(24), 241201.

85 Segev, D., Janotti, A., and Van de Walle, C.G. (2007) *Phys. Rev. B,* **75**(3), 035201.

86 Alkauskas, A., Broqvist, P., and Pasquarello, A. (2010) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046195. (chapter 7 in this book).

87 Deák, P., Gali, A., Aradi, B., and Frauenheim, T. (2010) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046210. (chapter 8 in this book).

88 Henderson, T.M., Paier, J., and Scuseria, G.E. (2011) *Phys. Status Solidi B,*

doi: 10.1002/pssb.201046303. (chapter 6 in this book).

89 Becke, A.D. (1993) *J. Chem. Phys.*, **98**(2), 1372–1377.

90 Becke, A.D. (1996) *J. Chem. Phys.*, **104**(3), 1040–1046.

91 Perdew, J.P., Ernzerhof, M., and Burke, K. (1996) *J. Chem. Phys.*, **105**(22), 9982–9985.

92 Perdew, J.P., Burke, K., and Ernzerhof, M. (1996) *Phys. Rev. Lett.*, **77**(18), 3865–3868.

93 Perdew, J.P., Burke, K., and Ernzerhof, M. (1997) *Phys. Rev. Lett.*, **78**(7), 1396.

94 Broqvist, P., Alkauskas, A., and Pasquarello, A. (2009) *Phys. Rev. B*, **80**(8), 085114.

95 Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2003) *J. Chem. Phys.*, **118**(18), 8207–8215.

96 Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2006) *J. Chem. Phys.*, **124**(21), 219906.

97 Seidl, A., Görling, A., Vogl, P., Majewski, J.A., and Levy, M. (1996) *Phys. Rev. B*, **53**(7), 3764–3774.

98 Bylander, D.M. and Kleinman, L. (1990) *Phys. Rev. B*, **41**(11), 7868–7871.

99 Asahi, R., Mannstadt, W., and Freeman, A.J. (1999) *Phys. Rev. B*, **59**(11), 7486–7492.

100 Robertson, J., Xiong, K., and Clark, S.J. (2006) *Phys. Status Solidi B*, **243**(9), 2071–2080.

101 Xiong, K., Robertson, J., Gibson, M.C., and Clark, S.J. (2005) *Appl. Phys. Lett.*, **87**(18), 183505.

102 Paier, J., Hirschl, R., Marsman, M., and Kresse, G. (2005) *J. Chem. Phys.*, **122**(23), 234102.

103 Paier, J., Marsman, M., Hummer, K., Kresse, G., Gerber, I.C., and Ángyán, J.G. (2006) *J. Chem. Phys.*, **124**(15), 154709.

104 http://cms.mpi.univie.ac.at/vasp/.

105 Deák, P., Aradi, B., Frauenheim, T., Janzén, E., and Gali, A. (2010) *Phys. Rev. B*, **81**(15), 153203.

106 http://www.castep.org/.

107 Alkauskas, A., Broqvist, P., Devynck, F., and Pasquarello, A. (2008) *Phys. Rev. Lett.*, **101**(10), 106802.

108 Alkauskas, A., Broqvist, P., and Pasquarello, A. (2008) *Phys. Rev. Lett.*, **101**(4), 046405.

109 Kleinman, L. (1981) *Phys. Rev. B*, **24**(12), 7412–7414.

110 Komsa, H.P., Broqvist, P., and Pasquarello, A. (2010) *Phys. Rev. B*, **81**(20), 205118.

111 Rinke, P., Janotti, A., Scheffler, M., and Van de Walle, C.G. (2009) *Phys. Rev. Lett.*, **102**(2), 026402.

112 Giantomassi, M., Stankovski, M., Shaltaf, R, M. Grüning, F., Bruneval, R.P., and Rignanese, G.M. (2010) *Phys. Status Solidi B,*, DOI: 10.1002/pssb.201046094. (chapter 3 in this book).

113 Lægsgaard, J. and Stokbro, K. (2001) *Phys. Rev. Lett.*, **86**(13), 2834–2837.

114 d'Avezac, M., Calandra, M., and Mauri, F. (2005) *Phys. Rev. B*, **71**(20), 205210.

115 Nolan, M. and Watson, G.W. (2006) *J. Chem. Phys.*, **125**(14), 144701.

116 Lany, S. and Zunger, A. (2009) *Phys. Rev. B*, **80**(8), 085202.

117 Lany, S. and Zunger, A. (2010) *Phys. Rev. B*, **81**(20), 205209.

118 Lany, S. and Zunger, A. (2010) *Appl. Phys. Lett.*, **96**(14), 142114.

119 Raebiger, H., Lany, S., and Zunger, A. (2009) *Phys. Rev. B*, **79**(16), 165202.

120 Du, M.H. and Zhang, S.B. (2009) *Phys. Rev. B*, **80**(11), 115217.

121 Carvalho, A., Alkauskas, A., Pasquarello, A., Tagantsev, A.K., and Setter, N. (2009) *Phys. Rev. B*, **80**(19), 195205.

122 Lany, S. (2011) *Phys. Status Solidi B,*, DOI: 10.1002/pssb.201046274. (chapter 11 in this book).

123 Koopmans, T.C. (1934) *Physica (Utrecht)*, **1**(1–6), 104–113.

124 Perdew, J.P., Parr, R.G., Levy, M., and Balduz, J.L. (1982) *Phys. Rev. Lett.*, **49**(23), 1691–1694.

125 Mori-Sánchez, P., Cohen, A.J., and Yang, W. (2008) *Phys. Rev. Lett.*, **100**(14), 146401.

126 Janak, J.F. (1978) *Phys. Rev. B*, **18**(12), 7165–7168.

127 Morgan, B.J. and Watson, G.W. (2009) *Phys. Rev. B*, **80**(23), 233102.

128 Cococcioni, M. and de Gironcoli, S. (2005) *Phys. Rev. B*, **71**(3), 035105.

129 Sanna, S., Frauenheim, T., and Gerstmann, U. (2008) *Phys. Rev. B,* **78**(8), 085201.

130 Lyons, J.L., Janotti, A., and de Walle, C. G. V. (2009) *Appl. Phys. Lett.,* **95**(25), 252105.

131 Umari, P., Qian, W., Marzari, N., Stenuit, G., Giacomazzi, L., and Baroni, S. (2011) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046264. (chapter 4 in this book).

132 Umari, P., Stenuit, G., and Baroni, S. (2009) *Phys. Rev. B,* **79**(20), 201104.

133 Aryasetiawan, F. and Gunnarsson, O. (1994) *Phys. Rev. B,* **49**(23), 16214–16222.

134 Umari, P., Stenuit, G., and Baroni, S. (2010) *Phys. Rev. B,* **81**(11), 115104.

135 Baroni, S., de Gironcoli, S., Dal Corso, A., and Giannozzi, P. (2001) *Rev. Mod. Phys.,* **73**(2), 515–562.

136 Giustino, F., Cohen, M.L., and Louie, S.G. (2010) *Phys. Rev. B,* **81** (11), 115105.

137 Miyake, T., Aryasetiawan, F., Kotani, T., van Schilfgaarde, M., Usuda, M., and Terakura, K. (2002) *Phys. Rev. B,* **66** (24), 245103.

138 Harl, J. and Kresse, G. (2009) *Phys. Rev. Lett.,* **103**(5), 056401.

139 Harl, J., Schimka, L., and Kresse, G. (2010) *Phys. Rev. B,* **81**(11), 115126.

140 Ma, Y. and Rohlfing, M. (2008) *Phys. Rev. B,* **77**(11), 115118.

141 Ma, Y., Rohlfing, M., and Gali, A. (2010) *Phys. Rev. B,* **81**(4), 041204.

142 Kaczmarski, M.S., Ma, Y., and Rohlfing, M. (2010) *Phys. Rev. B,* **81**(11), 115433.

143 Gali, A. (2011) *Phys. Status Solidi B,* doi: 10.1002/pssb.201046254. (chapter 18 in this book).