PAPER • OPEN ACCESS

Electron–phonon coupling in semiconductors within the GW approximation

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
1 October 2018

REVISED
6 November 2018

ACCEPTED FOR PUBLICATION
30 November 2018

PUBLISHED
14 December 2018

PAPER

Electron–phonon coupling in semiconductors within the GW approximation

Ferenc Karsai[1], **Manuel Engel**[1], **Georg Kresse**[1] **and Espen Flage-Larsen**[2]

[1] University of Vienna, Department of Physics, Austria
[2] SINTEF Materials Physics, Oslo, Norway

E-mail: ferenc.karsai@univie.ac.at

Keywords: electron–phonon interaction, GW approximation, density-functional theory, band gap renormalization

Abstract

The magnitude of the renormalization of the band gaps due to zero-point motions of the lattice is calculated for 18 semiconductors, including diamond and silicon. This particular collection of semiconductors constitute a wide range of band gaps and atomic masses. The renormalized electronic structures are obtained using stochastic methods to sample the displacement related to the vibrations in the lattice. Specifically, a recently developed one-shot method is utilized where only a single calculation is needed to get similar results as the one obtained by standard Monte-Carlo sampling. In addition, a fast real-space GW method is employed and the effects of G_0W_0 corrections on the renormalization are also investigated. We find that the band-gap renormalizations inversely depend on the mass of the constituting ions, and that for the majority of investigated compounds the G_0W_0 corrections to the renormalization are very small and thus not significant.

1. Introduction

Studies of the interaction between electrons and phonons have recently gained momentum. For materials, this is mainly motivated by the lack of robust methods at the intersection between statistical physics and many-body perturbation theory, and the need for further technology and materials advancements. For instance, a proper description of the temperature dependence of materials properties is becoming important in fields like thermoelectrics [1–5], high-power electronics [6], organic electronics [7], batteries [8, 9] and photovoltaics [10, 11]. In addition to the temperature dependence, the interaction between electrons and phonons create scattering events that determine, for instance, the thermalization of carriers, limit the transport of heat and charge, and lay foundations for conventional superconductivity.

The interaction of electrons and phonons is already relevant at absolute zero temperature due to zero-point vibrations [12]. As a matter of fact, it is of utmost importance to obtain a satisfactory description of the zero-point vibrations before continuing to finite temperatures. The zero-point renormalizations (ZPR) are difficult to access from experiments. Experimental data for the ZPR exist for a few compounds [13], especially for diamond [14] and silicon [15], but the exact values of the existing data are highly debatable [14–17], since the results depend on the underlying model and interpretation of the experimental data. Theoretical calculations are therefore relevant and important, and with recent computational and methodological advances, the calculation of the band-gap renormalization from phonons is on the verge of becoming a routine task.

Historically, the first qualitative description of the electron–phonon interaction appeared in the 1950s [18]. In the late 1970s and early 1980s the interaction was quantitatively described at the perturbative level by Allen, Heine and Cardona [19–21] (AHC). Here, the electron–phonon interaction was treated adiabatically, an approach that is still frequently used and serves as a starting point or reference for more advanced methods. Another popular approach is to treat the electron–phonon interaction by vibrational averages. These statistical averages can be obtained for example by Monte-Carlo (MC) integration [22, 23], molecular dynamics [24, 25] or path-integral molecular dynamics [26, 27]. A comprehensive review of the historical developments of theories treating electron–phonon interactions was recently published by Giustino [28].

Parameter-free calculations of the effects of the electron–phonon interactions on the band gaps of materials have become quite popular in the last 15 years. To our knowledge the first studies utilized path-integral MC [29] and later the AHC theory [30] to calculate the ZPR of the direct gap in diamond. Since the reported values overestimated the experimental values [14, 15], the interest in the field was sparked and numerous reports were published on the ZPR [31–41] and the temperature dependence [39–41] of the band gap in diamond and silicon. With two exceptions [40, 41] the local density approximation [42, 43] (LDA) for the electron exchange and correlation was employed. Gonze and coworkers [40] (later repeated by Monserrat [41]) reported that the temperature dependence of the direct band gap in diamond is underestimated compared to experiment when using LDA. To remedy this, they employed G_0W_0 calculations on top of the LDA and obtained results that compare reasonably well to experiment.

Reviewing the recent literature suggests that the AHC theory has been the most widely used first-principles approach to calculate band-gap renormalizations. It has the advantage of utilizing small unit cells, and it allows a very accurate sampling of the Brillouin-zone. The disadvantages are, however, also substantial. (i) First, second-order perturbation theory must be used to calculate the electron–phonon coupling terms. This limits the applicability of AHC theory to mean-field methods, such as DFT, where perturbation theory is readily available. (ii) The first-order change of the one-electron eigenvalues due to the second-order change of the Hamiltonian upon displacing the ions needs to be approximated by what is called the rigid-ion approximation. (iii) At high-symmetry points, standard second-order perturbation theory, in principle, needs to be replaced by second-order perturbation theory for degenerate states, which to the best of our knowledge has not been done. All three points limit the applicability and accuracy of perturbational approaches per se. Sampling of the configuration space of the phonons using e.g. MC procedures is certainly much simpler. The advantage of simplicity and robustness is only offset by three disadvantages: (i) to converge the Brillouin-zone integrals large super cells need to be used, (ii) many structures need to be sampled, and (iii) direct access to state dependent properties of the perturbation is not possible. The latter point is less of a disadvantage if we are not interested in such properties, which is the case for this study. With recent advances in algorithms and high-performance computing, the first issue becomes progressively less relevant: modern codes allow to perform calculations for super cells with several hundred atoms in a few hours. With respect to (ii), another recent development seems to dip the balance towards super-cell approaches. Giustino and coworkers [39] proposed a method where the MC sampling of the quantum-harmonic oscillations is replaced by a calculation on a single structure. This selective sampling yields exact results in the limit of large super cells. Consequently, this means that one is now in the position to not only tackle the demanding physical problem of the electron–phonon interaction, but also doing so with very demanding methods such as GW to obtain results of unprecedented accuracy.

Instrumental for such GW calculations is another recent development. Traditionally, the GW method scales quartic with system size, so that large unit cells with and beyond 100 atoms are very demanding even on present-day high-performance computers. Already in the late 90s, Godby and coworkers suggested a cubic-scaling implementation of the GW method, where the polarizability and self-energy are evaluated in real-space and imaginary time [44, 45]. This method was recently implemented in the VASP code [46–48] and allows for routine calculations on low-symmetry systems with hundreds of atoms. In this work we conducted a large number of such G_0W_0 calculations, each containing typically 128 atoms to address the question whether the values for the band-gap renormalization differ between LDA and G_0W_0.

In sections 2 and 3 we present the theory and calculational methods. This is followed in section 4 by a detailed analysis of the methodology on the example of diamond, which is one of the most debated materials in literature. In that section we also compare the one-shot sampling of Giustino to the standard MC method and investigate the electron–phonon interactions in a large temperature range. In section 5 we investigate the ZPR on a representative number of semiconductors. We investigate the role of G_0W_0 and closely examine the dependence of the ZPR on the atomic mass before we finally conclude this paper in section 6.

2. MC and one-shot method in the harmonic approximation

The probability distribution of finding an atom within the coordinates $\kappa + \mathrm{d}\kappa$ (where κ denotes the Cartesian coordinates as well as the atom number) at temperature T in the harmonic approximation is given by the following expression [49, 50]

$$\mathrm{d}W_\nu(\kappa, T) = \frac{1}{2\pi\langle u^2_{\nu\kappa}\rangle} \mathrm{e}^{-\kappa^2/(2\langle u^2_{\nu\kappa}\rangle)} \mathrm{d}\kappa, \tag{1}$$

where the mean-square displacement of the harmonic oscillator is given as

$$\langle u_{\nu\kappa}^2 \rangle = \frac{\hbar}{2M_\kappa \omega_\nu} \coth \frac{\hbar\omega_\nu}{2k_B T}. \tag{2}$$

Here M_κ, ν and ω_ν denote the mass, phonon eigenmode and phonon eigenfrequency, respectively. Equation (1) is valid at any temperature and the high (Maxwell–Boltzmann distribution) and low temperature limits are easily regained. In order to obtain an observable $O(T)$ at a given temperature T, in this case the band gap, the average of the observable sampled at different coordinate sets $x_T^{\mathrm{MC},i}$ with sample size n is taken

$$\langle O(T) \rangle = \frac{1}{n} \sum_{i=1}^{n} O(x_T^{\mathrm{MC},i}). \tag{3}$$

Each set i is obtained from the equilibrium atomic positions x_{eq} as

$$x_T^{\mathrm{MC},i} = x_{\mathrm{eq}} + \Delta\tau^{\mathrm{MC},i} \tag{4}$$

with the displacement

$$\Delta\tau^{\mathrm{MC},i} = \sqrt{\frac{1}{M_\kappa}} \sum_{\nu}^{3(N-1)} \varepsilon_{\kappa,\nu} \mathcal{N}. \tag{5}$$

Here $\varepsilon_{\kappa,\nu}$ denotes the unit vector of eigenmode ν on atom κ. The magnitude of the displacement in each Cartesian direction is obtained from the normal-distributed random variable $\mathcal{N}$ with a probability distribution according to (1).

Motivated by the empirical observation that for increasing super-cell sizes the number of required structures in the MC method can be decreased, Giustino and coworkers [39] proposed a one-shot method where only a single set of displacements is used

$$\Delta\tau^{\mathrm{OS}} = \sqrt{\frac{1}{M_\kappa}} \sum_{\nu}^{3(N-1)} (-1)^{\nu-1} \varepsilon_{\kappa,\nu} \sigma_{\nu,T}, \tag{6}$$

where the summation over the eigenmodes runs in an ascending order with respect to the values of the eigenfrequencies, and the magnitude of each displacement is given by

$$\sigma_{\nu,T} = \sqrt{(2n_{\nu,T} + 1)\frac{\hbar}{2\omega_\nu}}. \tag{7}$$

Here $n_{\nu,T} = [\exp(\hbar\omega_\nu / k_B T) - 1]^{-1}$ denotes the Bose–Einstein occupation number. By using (6) the sum in (3) is reduced to a single calculation. In [39] it was shown that for super-cell sizes $N \to \infty$ the structural configuration obtained using (6) should lead to equivalent results as fully converged MC calculations. In practice, it was shown that already relatively small super-cell sizes are sufficient to achieve good accuracy, but the convergence with respect to the cell size can vary between different materials. We have also used a slightly modified approach, in which the signs of the displacements are chosen randomly instead of ± 1. This was necessary, when calculating volume-dependent ZPR, since the modes sometimes change the order as the volume changes. Using alternating signs for the displacement then causes small discontinuities in the ZPR volume curve of the order of 5 meV for carbon diamond. By averaging over many random phases this problem can be eliminated.

In perturbation theory, the electron–phonon interaction is described to a given order (second in standard AHC). The advantage of stochastic methods compared to perturbative methods is that these interactions are not in principle restricted to a given order, since full electronic-structure calculations are performed for the distorted structures including all electronic effects due to the changed geometry. However, the accuracy is limited by the harmonic sample distribution, in our case the density matrix of harmonic oscillators. To go beyond the harmonic approximation for the sampling of structures, one could possibly use finite-temperature path-integral methods, which is beyond the scope of the present work. Another effect of stochastic methods is that it captures the effects of the displacements leading to a lift of the degeneracy of eigenstates into sub levels. Each displacement along an eigenmode causes a change of the eigenenergy that is linear with respect to the magnitude of the displacement. For non-degenerate states, the linear shift cancels, since positive and negative displacements increase or decrease the eigenenergies by the same amount. For degenerate states this is, however, not the case. For positive displacements one of the degenerate states decreases its energy (while it increases for the other), whereas for negative displacements the energy change is flipped. This means, that the degenerate states, tend to split, on average into a lower and upper sub level. Figure 1 demonstrates this effect on the valence band maximum in diamond, which is three-fold degenerate in the equilibrium state (we only show the two orbitals that are involved in the splitting). The distortions are obtained by displacing the ions of the equilibrium structure along fractions of the eigenvectors of a single eigenmode (we arbitrarily chose the eigenmode with the highest eigenfrequency) in positive and negative directions. Such a splitting is not included in any of the AHC

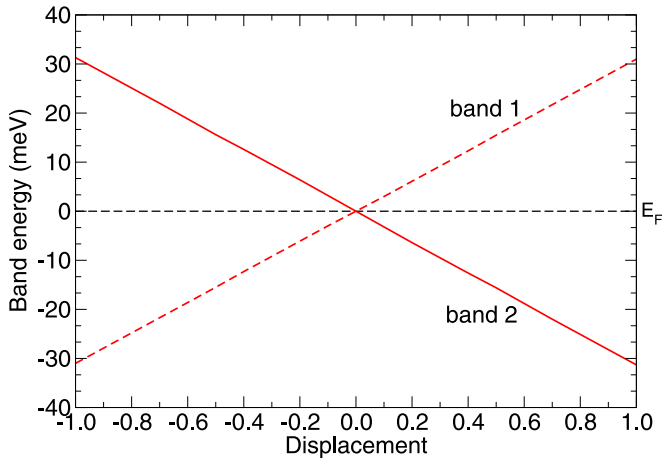

Figure 1. Band splitting of valence band maximum due to an arbitrarily chosen eigenmode within diamond using a $3 \times 3 \times 3$ cell. The displacements are given as fractions of $\langle u^2_{\nu\kappa} \rangle$ of the chosen eigenmode.

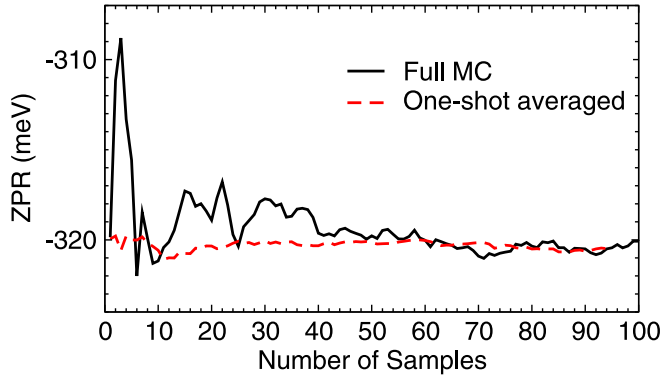

Figure 2. Mean value of the ZPR evaluated for a $6 \times 6 \times 6$ super cell of diamond over a certain number of sample configurations. The mean ZPR is shown for the Monte-Carlo sampling and for a random sampling of the signs in the one-shot method. Both methods converge to the same value, but the Monte-Carlo sampling requires more sample configurations.

calculations, since degenerate perturbation theory has not yet been executed, at least not to the authors knowledge. In our studies we found that the splitting is reduced as the super-cell size increases. However, we were not able to determine whether the splitting in fact converges exactly towards zero. As a pragmatic remedy, we determine the band-gap renormalization as the mean of the change of degenerate bands in the undistorted super cell. This has two advantages. (i) The mean converges fairly rapidly with super-cell size. (ii) The approach is compatible to published AHC calculations, which neglect linear changes of the band gaps from the outset. As already alluded to, such linear changes cancel out for non-degenerate states in the harmonic approximation. In passing, we note that Monserrat used the same approach [34, 51]. Whether the splitting indeed converges to zero for infinite super cells (experiments clearly indicate no splitting of degenerate states due to zero-point vibrations) needs to be investigated in future work. Another subtlety is how we determine the band-gap changes. Here, we have used a simple approach, in which only the average change of the one-electron levels is inspected. In the case of degenerate levels, the average change of all degenerate levels is taken. This does not necessarily correspond to the experimental measurements, where often Tauc fits are performed to the observed tails in spectroscopic experiments [39, 52]. Although these are important issues when comparison with experiment is the main priority, our choice is mandated by the need to remain compatible with the literature pertaining to computational studies.

3. Calculational parameters

All calculations were performed within the framework of the projector augmented-wave method [53] using the VASP code [54, 55]. Table 1 shows the calculated parameters for all chosen compounds. The lattice parameters in this table refer to the experimental lattice parameters at the lowest available temperatures. The phonon

Table 1. Calculational parameters of chosen compounds: E_{xc} is the method for treatment of electron exchange and correlation in the DFT calculation; E_{pw} is the plane-wave cut-off energy in eV, and a_{latt} is the experimental lattice parameter in Å. All lattice parameters with the exception of GaN and ZnO were taken from the supplemental material of [15] and references therein. The lattice parameters for GaN and ZnO were taken from [14, 30], respectively.

	E_{xc}	E_{pw}	a_{latt}	Crystal structure	Fundamental gap
C	PBE	414	3.567	Diamond	Indirect
BN	PBE	318	3.616	Zincblende	Indirect
SiC	PBE	414	4.358	Zincblende	Indirect
Si	PBE_0	245	5.431	Diamond	Indirect
AlP	PBE	255	5.463	Zincblende	Indirect
ZnO	PBE	402	4.584	Zincblende	Direct
GaN	PBE	405	4.535	Zincblende	Direct
ZnS	PBE	402	5.409	Zincblende	Direct
GaP	PBE	405	5.451	Zincblende	Indirect
AlAs	PBE	240	5.661	Zincblende	Indirect
ZnSe	PBE	402	5.669	Zincblende	Direct
CdS	PBE	259	5.818	Zincblende	Direct
GaAs	PBE_0	405	5.654	Zincblende	Direct
Ge	PBE_0	174	5.658	Diamond	Indirect
AlSb	PBE	263	6.136	Zincblende	Indirect
CdSe	PBE_0	254	6.077	Zincblende	Direct
ZnTe	PBE	402	6.481	Zincblende	Direct
CdTe	PBE	254	6.481	Zincblende	Direct

frequencies were obtained within density-functional-theory (DFT) using finite differences [56, 57] and super cells. In all cases the standard Perdew Burke Ernzerhof (PBE) functional [58, 59] was used for the electron exchange and correlation energy. For the majority of compounds PBE was also used in the determination of the band gap. Since some of the compounds yield no or only a marginal band gap for finite-temperature structures using PBE, hybrid functionals with $\alpha = 0.25$ exact exchange (PBE_0) [60–62] within the generalized Kohn–Sham [63, 64] theory were employed in the calculation of the band gap of Si, GaAs, Ge and CdSe. All calculations were carried out at a single Γ centered k-point.

4. Band-gap renormalization in diamond

Table 2 compares the ZPR for diamond using MC sampling and the one-shot method. From this table we can deduce that the difference is about 15 meV for the $3 \times 3 \times 3$ super cell but rapidly drops to about 5 meV for the $5 \times 5 \times 5$ super cell and even further if we increase the cell size. This we deem to be acceptable for this study. The error of the one-shot method can be further reduced by choosing random signs in the one-shot method and performing an averaging over several configurations. This approach converges faster than the full MC sampling, and reduces the error of the one-shot method somewhat. We have used this approach only in the calculation of volume-dependent ZPR, since the volume-dependent ZPR showed slight discontinuities otherwise.

The variance of the band-gap renormalization using the MC sampling method decays inversely proportional with respect to the number of atoms in the unit cell as shown in figure 3(a). From figure 3(b) we can deduce that for the MC sampling an error of the mean of 5 meV can be achieved using less than 10 structures for the $5 \times 5 \times 5$ super cell. One hundred fifty structures were used for the DFT in table 2, but for the G_0W_0 smaller sample sizes (60, 75 and 100 for the $5 \times 5 \times 5$, $4 \times 4 \times 4$ and $3 \times 3 \times 3$ super cells, respectively) were used. This leads to slightly larger errors for the G_0W_0 calculations, but the error of the mean is still below 5 meV for the $5 \times 5 \times 5$ super cell. The ZPR shows relatively large changes with the cell size. For cell sizes below $5 \times 5 \times 5$, convergence with respect to super-cell size is not achieved. The jump from the $4 \times 4 \times 4$ to the $5 \times 5 \times 5$ super cell is related to the $4 \times 4 \times 4$ cell being unable to sample the conduction band minimum located at a non high-symmetry point (0.727×). Unfortunately the $6 \times 6 \times 6$ unit cell with 432 atoms requires too much memory for G_0W_0 type calculations, but we observe already a very small change from the $5 \times 5 \times 5$ to the $6 \times 6 \times 6$ super cell for PBE. This indicates that $5 \times 5 \times 5$ super-cell calculations are reasonably well converged and we expect G_0W_0 calculations to also follow this trend.

Direct comparison with literature values has to be done carefully since the calculational parameters such as i.e. lattice parameters and electron exchange-correlation functional can vary. All literature values in table 2 were calculated using LDA and either optimized lattice parameters, or the authors gave no information on the

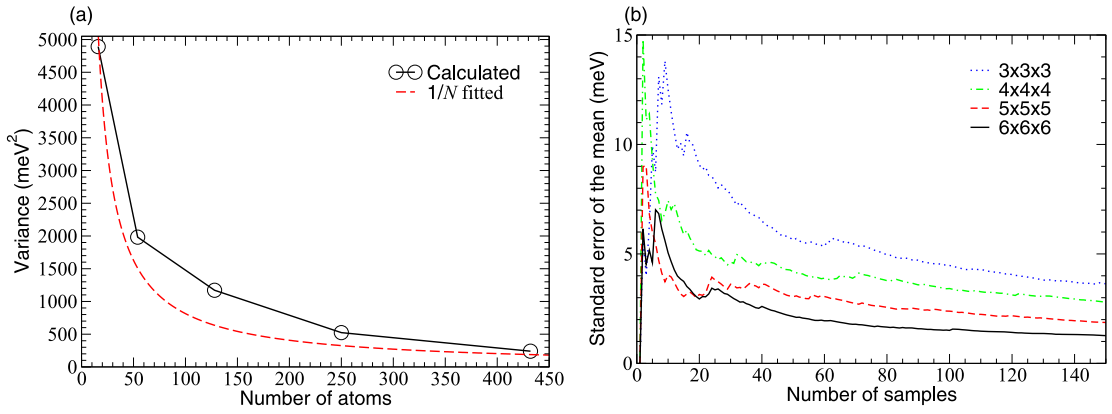

Figure 3. Statistical errors of the ZPR in diamond using the MC sampling method. (a) Variance with respect to number of atoms in the unit cell. A sample size of 150 was used for each point. (b) Standard error of the mean with respect to number of samples for different super-cell size.

Table 2. ZPR of the fundamental indirect gap in diamond. The standard error of the mean is given in brackets. Lattice parameters a_{latt} are in Å. All energies are in eV.

	One-shot		MC	
	PBE	G_0W_0	PBE	G_0W_0@PBE
$3 \times 3 \times 3$	−0.262	−0.298	−0.278 (0.0056)	−0.314(0.0064)
$4 \times 4 \times 4$	−0.363	−0.446	−0.365 (0.0035)	−0.439(0.0042)
$5 \times 5 \times 5$	−0.320	−0.337	−0.315 (0.0018)	−0.329(0.0020)
$6 \times 6 \times 6$	−0.318	—	−0.321 (0.0017)	—

Literature	a_{latt}	ZPR	References
MC (Tauc)	—	−0.345	[14]
MC	3.529	−0.344	[34]
AHC	3.529	−0.334	[32]
AHC	—	−0.330	[37]
Experiment	—	−0.340	[14]
	—	−0.370	[15]
	—	−0.410	[17]

employed lattice parameters. We, on the other hand, used PBE and the experimental lattice constant. PBE is known to give slightly larger band gaps than LDA, and we think that using the experimental lattice constant is a better choice for comparison with experiment. References [32, 37] were both utilizing perturbative methods (AHC or equivalent), while [14, 34] used similar super-cell methods as in this work. In [34] the sampling was done along so-called thermal lines using $6 \times 6 \times 6$ super cells and drawing sample distributions from the density matrix of a harmonic oscillator [14] introduced and used the one-shot approach; the main difference between our work and this is that in [14] the band gap was determined from the dielectric function [65] using Tauc fits [52]. Overall, there is no denying that our values tend to be somewhat smaller (10–15 meV) than most literature values. Using the LDA we obtain values of −0.321 and −0.315 eV using the $6 \times 6 \times 6$ cell at the experimental lattice parameter and the optimized lattice parameter (3.529 Å), respectively. In both cases the difference compared to PBE is 3 meV. Nevertheless, the agreement compared to previous literature calculations is satisfactory given the different lattice constants, exchange-correlation functionals, as well as pseudo potentials.

The comparison to experimental values is even more difficult, since several experimental values exist and none of them is unambiguously accepted in literature as the most accurate one. The experimental values strongly depend on how the experimental data is obtained and evaluated [14, 15, 17]. Our calculated ZPR value of −0.337 eV for the $5 \times 5 \times 5$ super cell using G_0W_0 agrees very well with the experimental result [14, 66] of −0.340 eV. In [14], Cardona claims that the experimental value of −0.410 eV is a more accurate estimate, since

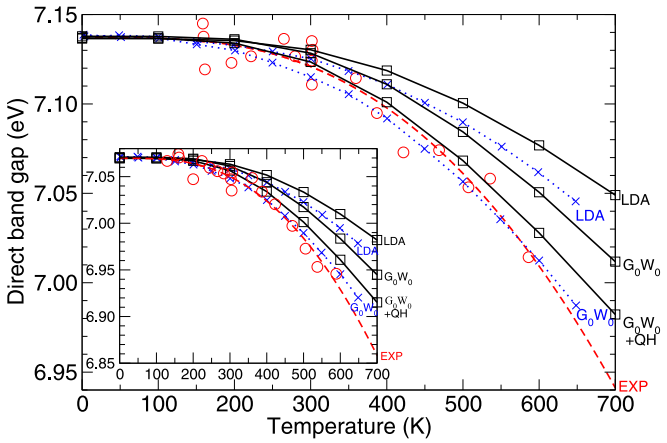

Figure 4. Temperature-dependent direct band gap in diamond: present results using a 5 × 5 × 5 super cell (black squares and continuous lines), literature results taken from [40] (blue crosses and dotted lines) and experimental data and Varshni fits (see (8)) for sample IIb in [67] (red circles and dotted lines). The inset shows the same calculational results but with respect to the experimental set IIa from [67].

it relies on a more recent evaluation of the experimental data. Furthermore, Cardona also suggested that G_0W_0 should yield a larger band-gap renormalization. Hence it seemed reasonable to believe that the measured band-gap renormalization to be larger than most calculations accessible in the literature at that time. Our calculations are the first to evaluate the renormalization of the fundamental indirect gap using G_0W_0, but the G_0W_0 renormalization only differs by 15 meV from the PBE renormalization. This means, that compared to the most recent experimental value (−0.410 eV) a sizable discrepancy between experiment and theory prevails.

We now move on to our results for the ZPR of the direct gap in diamond. G_0W_0 calculations of the ZPR have already been published for the direct band gap [40, 41], hence a comparison with literature data is possible. In [40] the authors reported a direct gap ZPR of −0.404 and −0.613 eV using LDA and G_0W_0, respectively, leading to GW corrections of $\Delta E_{\mathrm{GW}}^{\mathrm{bg}} = E_{\mathrm{GW}}^{\mathrm{bg}} - E_{\mathrm{LDA}}^{\mathrm{bg}} = -0.209$ eV. While the LDA values are already in reasonable agreement with the experimental results of −0.320 and −0.450 eV from [67], by adding the G_0W_0 corrections, the experimental values are largely overestimated. Using the 5 × 5 × 5 super cell, we find direct gap ZPR of −0.326 and −0.586 eV for PBE and G_0W_0, respectively, and $\Delta E_{\mathrm{GW}}^{\mathrm{bg}} = -0.260$ eV, indicating an even larger G_0W_0 correction. Our DFT results tend to agree with the lower experimental value of −0.320 from [67], while the LDA results from [40], as many other literature calculations, tend to agree more with the larger value from [67]. We note that besides different DFT functionals and lattice parameters, the authors of [40] used 4 × 4 × 4 super cells in the G_0W_0, whereas we used 5 × 5 × 5 cells. When using the 4 × 4 × 4 cell we obtain ZPR of −0.442 and −0.616 eV for PBE and G_0W_0, respectively, which are in much closer agreement with their results. Nevertheless, the behavior of a large G_0W_0 correction to the ZPR for the direct gap is consistent in all calculations. It is important to emphasize that the G_0W_0 corrections to the ZPR are very different for the fundamental indirect and direct gap.

The remainder of this section is devoted to the changes of the band gap as a function of the temperature. In many semiconductors, the temperature dependence of the band gap is well described by the modified Varshni relation [69, 70]

$$E_{\mathrm{GAP}}(T) = E_{\mathrm{GAP}}(0) - \frac{AT^4}{(B+T)^2}, \tag{8}$$

where A and B are fitting parameters for a given material. Using the one-shot approach, the band gap versus temperature can be calculated straightforwardly. Since theory yields a different band gap than experiment at $T = 0$ K, the theoretical results are shifted to match the experimental value at the lowest available temperature. As G_0W_0 corrected ZPR calculations have only been published for the direct gap in diamond, we start by comparing our results on the direct gap with literature and experiment. Figure 4 shows the dependence of the calculated and experimental direct band gaps on the temperature. Since two experimental data sets with different absolute values but similar slopes have been published in [67], we compare to the data set that was predominantly used for analysis in [67] in the main panel. For the sake of completion, we show the results for the second experimental set in the inset of figure 4. Although the deviations between LDA and PBE are small for the ZPR (below 10 meV) the discrepancies become larger at higher temperatures. To allow straightforward comparison with literature, we decided to employ the LDA (at the experimental lattice constant) for studying the temperature dependence of the direct gap. When comparing our calculations to those of Antonius *et al* in [40], we can clearly

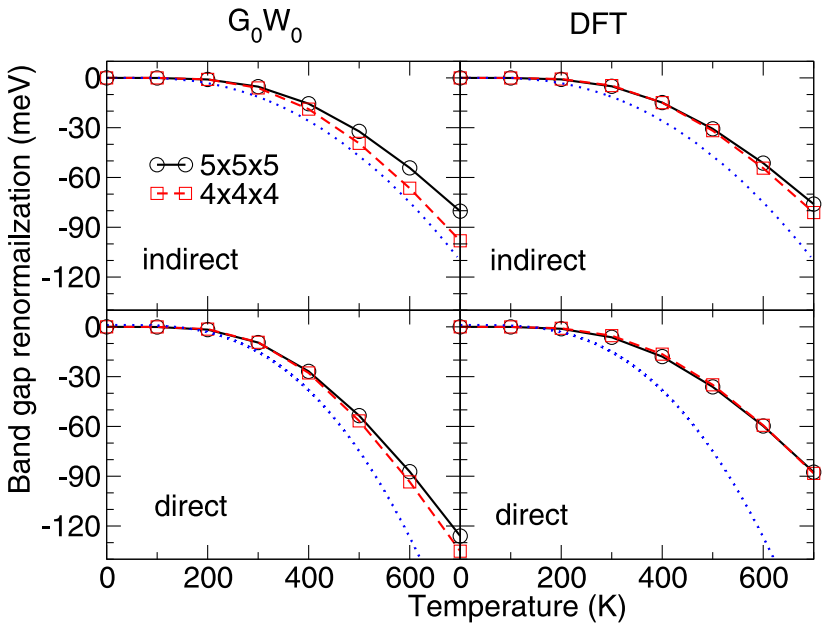

Figure 5. System size dependence of the relative temperature renormalizations of the band gaps in diamond. The blue dotted lines show the Varshni fit (see (8)) to the experimental data for the fundamental indirect and direct gap taken from [68] and [67], respectively. The offsets of all curves were shifted to zero.

see that our calculations show a less steep temperature dependence at the LDA level, with an even more pronounced difference at the G_0W_0 level. The most likely source of the difference is the use of a super cell in our approach, whereas Antonius *et al* used primitive cells with a very fine *k*-point sampling for the electronic as well as phononic dispersion. On the other hand, in [40] the G_0W_0 corrections were determined for a $4 \times 4 \times 4$ super cell. Figure 5 shows the temperature dependence of the band gap for two $4 \times 4 \times 4$ and $5 \times 5 \times 5$ super cells, clearly suggesting that the sampling errors are small even for the $4 \times 4 \times 4$ unit cell. So the differences between our calculations and the previous calculations remain largely unexplained, and more troublesome, our calculations do not agree very well with the experimental values.

An important effect that neither the authors of [40] nor we have yet explored is the temperature dependence of the lattice constant of diamond. We now show that this effect is sizable and brings our data back in agreement with experiment. To account for the lattice expansion of diamond we performed calculations in the quasi-harmonic (QH) [71] approximation to obtain the optimal volume at each temperature. This is achieved by first calculating the optimal lattice constant for each temperature as the minimum of the free energy over a range of volumes using the equation of state by Birch and Murnaghan [72, 73]. Then, we calculate a displaced representative finite-temperature structure at each volume including the electron–phonon interactions according to (6). The resulting curve employing this method (indicated by the label 'QH') in figure 4 shows a change of the high temperature slope towards the experimental values. Almost identical results are obtained by adding the volume induced change of the gap (calculated without ZPR) to the band gap renormalization calculated at fixed volumes. Lattice expansion effects and vibrational induced effects are therefore additive, and the former one are trivial to calculate even in the primitive cell. We would expect any theoretical calculation that neglects volume effects to give a noticeable discrepancy compared to experiment. This implies two things: (i) lattice expansions can certainly not be neglected when calculating the temperature dependence of the band gap. (ii) If the lattice expansion is disregarded and perfect agreement with experiment is obtained, this agreement is certainly fortuitous and will cease when the lattice expansion is taken into account.

For the fundamental indirect gap of diamond shown in figure 6 the inclusion of QH effects is even more dramatic. In this case, one can see that G_0W_0 and DFT yield almost identical slopes for the ZPR-temperature curve that, compared to experiment, are too small at high temperatures. The linear lattice expansion at larger temperatures shown in the inset of figure 6, increases the high temperature slope of the band gap quite dramatically above 300–400 K, bringing our curves in excellent agreement with experiment.

5. ZPR in semiconductors

In this section, we investigate the ZPR of the band gap of selected semiconductors. The compounds were selected to span a large set of band gaps, bonding types (ionic to covalent) and the atomic masses (compare [76]). The

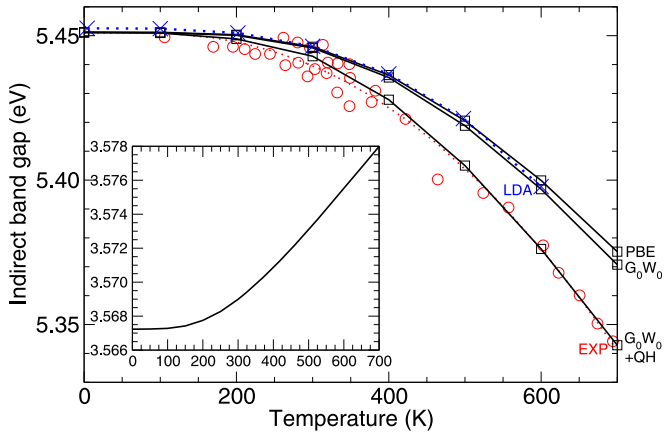

Figure 6. Quasi-harmonic temperature-dependent fundamental indirect band gap in diamond: present results using a 5 × 5 × 5 super cell (black squares and continuous lines), literature DFT calculations taken from [39] (blue crosses and dotted lines) and experimental data and Varshni fits (see (8)) taken from [74] (red circles and dotted lines). The inset shows the temperature dependence (in Å) of the lattice parameter obtained from quasi-harmonic calculations.

Table 3. Zero-point vibration band-gap renormalization energies. The experimental ZPR for C, Si, ZnS, GaAs, Ge, AlSb, CdSe, ZnTe and CdTe were taken from [13–16] respectively. The literature calculations for the ZPR for C, BN, SiC, Si, GaN and GaAs were taken from [32–34, 38, 75] and [39], respectively. All energies are in eV.

	3 × 3 × 3		4 × 4 × 4		5 × 5 × 5		Theory literature	Exp
	DFT	G_0W_0	DFT	G_0W_0	DFT	G_0W_0		
C	−0.262	−0.298	−0.363	−0.446	−0.320	−0.337	−0.343	−0.340
BN	−0.277	−0.283	−0.269	−0.276	−0.294	—	−0.262	—
SiC	−0.155	−0.175	−0.124	−0.145	−0.109	—	−0.109	—
Si	−0.056	−0.063	−0.064	−0.066	−0.065	—	−0.058	−0.064
AlP	−0.067	−0.075	−0.062	−0.072	−0.070	—	—	—
ZnO	−0.048	−0.085	−0.061	−0.069	−0.057	—	—	—
GaN	−0.117	−0.120	−0.102	−0.117	−0.094	—	−0.127	—
ZnS	−0.041	−0.060	−0.054	−0.057	−0.044	—	—	−0.064
GaP	−0.068	−0.044	−0.072	−0.104	−0.057	—	—	—
AlAs	−0.044	−0.061	−0.051	−0.058	−0.063	—	—	—
ZnSe	−0.032	−0.030	−0.027	−0.039	−0.028	—	—	—
CdS	−0.023	−0.030	−0.028	−0.037	−0.029	—	—	—
GaAs	−0.074	−0.079	−0.056	−0.054	−0.053	—	−0.032	−0.054
Ge	−0.064	−0.043	−0.049	−0.057	−0.050	—	—	−0.052
AlSb	−0.054	−0.055	−0.050	−0.052	−0.043	—	—	−0.039
CdSe	−0.028	−0.034	−0.020	−0.026	−0.021	—	—	−0.034
ZnTe	−0.026	−0.046	−0.027	−0.036	−0.024	—	—	−0.033
CdTe	−0.016	−0.016	−0.015	−0.018	−0.015	—	—	−0.017

calculated ZPR are presented in table 3. One observes that for the majority of compounds the differences between G_0W_0 and DFT are only a few meV. On first sight this result might be somewhat surprising, since the G_0W_0 method substantially increases the band gap. So why does it not increase the ZPR as well? Fundamentally, there is, however, no obvious reason why the band gap should be reduced by the inclusion of many-body effects in the ZPR. In the view of our results, this rather seems to be a misconception that emerged from one special case, namely, the direct band gap in diamond and the observation that the ZPR of that gap is about 200 meV larger in G_0W_0 than in DFT. Furthermore, even for diamond, we have shown that the effect is much smaller for the fundamental indirect gap than the direct gap.

In the previous section, we observed that the inclusion of QH effects in the calculations can have noticeable effects on the temperature dependence of the band gap, and we expect this to be also true for the materials studied here. For table 3 and figure 7, we have disregarded this effect, for simple reasons. The main purpose of table 3 is to determine how large the effect of zero-point vibrations are for a fixed volume, for instance at the experimental volume. Figure 6 inspects finite-temperature effects, but applying the QH approximation to all materials was computationally too demanding and beyond the scope of the present work.

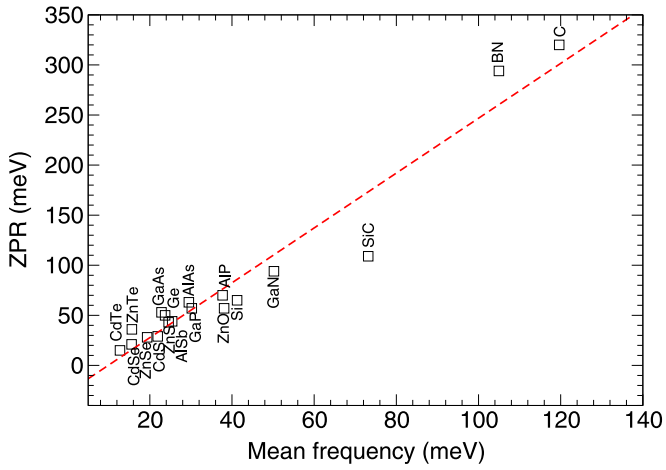

Figure 7. Magnitude of ZPR versus mean frequency.

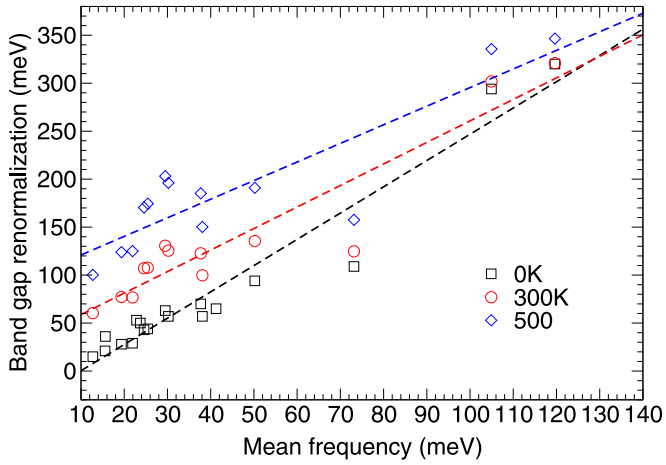

Figure 8. Band-gap renormalization with respect to mean frequency at elevated temperatures.

For most of the materials, our calculations agree reasonably well with the few available literature results, but differences are to be expected, since different parameters and methods were used in the calculations. The majority of the previous calculations used the LDA instead of the PBE, and optimized theoretical lattice constants were applied. Our calculations employ the experimental lattice constants. Among other reasons, our choice is motivated by the desire to know the magnitude of the zero-point corrections at the experimental volume, in order to allow comparison between high-level theory such as GW and beyond with experiment. For this purpose, a precision of about 10–20 meV is sufficient, due to the fact that current GW implementations still have a way to go to reach such precision.

The dependence of the magnitude of the ZPR of the band gap on the mean vibrational frequency is shown in figure 7. The mean frequency was calculated as the mean of all non-translational modes obtained in the unit cell. The renormalizations linearly depend on the mean frequency, which can be seen from the linear fit (dashed line) in figure 7. The dependence of the band-gap renormalizations on the mean frequency is only loosely retained at elevated temperatures (plotted in figure 8) and the deviations from linearity become larger with increasing temperature. This is expected since the temperature dependence of the band gap for each element is nonlinear with a different slope, hence each point in figure 8 is shifted by a different amount when the temperature is increased. According to the equations for the dynamical matrix, the eigenfrequencies of the phonons are inversely proportional to the masses of the ions. The linear relation of the masses of the constituting formula units with respect to the mean calculated frequency is shown in figure 9(a). In this figure the inverse mass was calculated as the sum of the experimental inverse masses of the constituting ions. Since the ZPR depends linearly on the mean frequency and there is a linear relationship between frequencies and masses, the ZPR also depends linearly on the inverse mass. This is shown in figure 9(b).

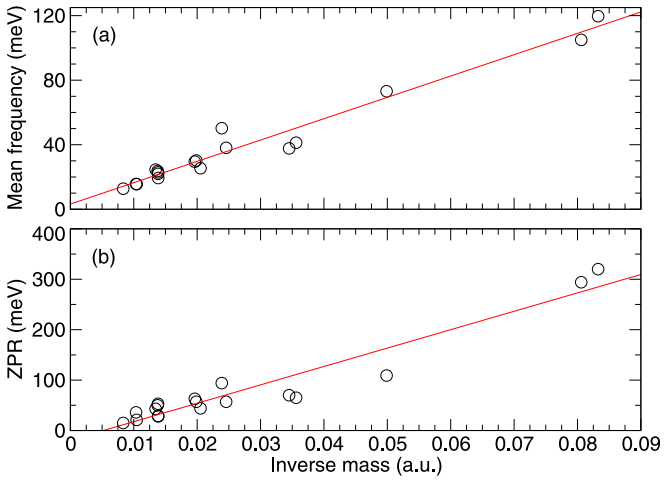

Figure 9. (a) Mean frequency versus inverse mass. (b) Absolute value of ZPR versus inverse mass.

6. Conclusion

In summary, we have used super-cell methods to obtain the band-gap renormalization due to electron–phonon interactions for a set of selected semiconductors. To test our implementation, we have investigated the accuracy of our approach for diamond. The calculated ZPR for different super-cell sizes indicates that for sufficiently-large super cells, one can achieve a similar accuracy using the single structure suggested in [39] as with a statistical MC sampling. For the rest of this work, we have hence used this one-shot method, since it obviously reduces the computational requirements significantly.

For the case of diamond our results can be summarized as follows. Our calculated ZPR values of -0.320 and -0.326 eV using PBE and the $5 \times 5 \times 5$ super cell agree reasonably well with the experimental values of -0.340 and -0.320 eV for the indirect and direct gap, respectively. The agreement to literature values for the indirect gap is satisfactory, while the values for the direct gap ZPR are significantly smaller than most literature values, which are commonly above 0.4 eV. Although all literature calculations employed the LDA and optimized lattice parameters, our test calculations show that deviations between LDA and PBE at optimized and experimental lattice parameters are below 10 meV and the reason for the discrepancies remains unexplained. Adding G_0W_0 on top of PBE, we obtain values of -0.337 and -0.586 eV for the ZPR of the indirect and direct gap, respectively. In the case of the indirect gap, G_0W_0 gives only very small corrections, below 20 meV, yielding a ZPR in close agreement with experiment. In contrast to this, G_0W_0 gives a too large correction and hence results in a strong overestimation compared to the experimental ZPR. This is in agreement with previous literature calculations employing the G_0W_0 method for the direct band gap in diamond.

Moving on to the temperature dependence of the band gap, our standard DFT calculations clearly yield a too weak slope of the temperature-dependent band-gap renormalization at high temperatures compared to experiment. This result is in accordance with computational literature. The description of the temperature dependence is improved when G_0W_0 calculations are performed. However, using a fixed theoretical volume for each temperature, a noticeable discrepancy between experiment and calculations prevails. To incorporate the temperature induced lattice expansions, the lattice parameters were optimized for each temperature using the QH approximation. For each temperature and volume, we then recalculate the band-gap renormalization, now finding significantly improved agreement with experiment. This clearly indicates that changes in the lattice parameters must be accounted for when temperature-dependent band-gap renormalizations are calculated.

Furthermore, we have calculated the ZPR of a set of representative semiconductors using the one-shot method from [39]. For the majority of the investigated compounds the inclusion of G_0W_0 corrections only marginally changes the band gap renormalizations. Considering the large computational cost for G_0W_0 calculations, it seems advantageous that this step can be usually avoided, and instead we can rely on standard DFT. However, even in the present work we found one exception to the rule, the direct gap in diamond. Here, the ZPR is 200 meV larger in G_0W_0 approximation than in DFT. Finally, by inspecting the relationship between the ZPR of the band gap and calculated phonon frequencies, we find that the renormalizations scale linearly with respect to the mean vibrational frequency. Since the phonon frequencies are also inverse proportional to the atomic masses, one can even roughly estimate the effect of zero-point vibrations from the atomic masses alone.

Acknowledgments

We want to acknowledge B Monserrat for correspondence and discussions on the topic of evaluation of degenerate band gaps. F Karsai acknowledges A Tröser for discussions on statistical approaches. M Engel gratefully acknowledges support by the Austrian Science Fund (FWF) No. I2460-N36. Furthermore, E Flage-Larsen would like to thank the Norwegian Research Council of Norway (project number 262339) and SINTEF for providing financial support for this work.

References

[1] Lu W, Baek J B and Dai L 2015 *Carbon Nanomaterials for Advanced Energy Systems: Advances in Materials Synthesis and Device Applications* (New York: Wiley) p 1
[2] Gibbs Z M, Kim H, Wang H, White R L, Drymiotis F, Kaviany M and Snyder G J 2013 *Appl. Phys. Lett.* **103** 262109
[3] Tan G, Shi F, Hao S, Zhao L D, Chi H, Zhang X, Uher C, Wolverton C, Dravid V and Kanatzidis M 2016 *Nat. Commun.* **7** 12167
[4] Pei Y, Shi X, Lalonde A, Wang H, Chen L and Snyder G 2011 *Nature* **473** 66
[5] Tang Y, Gibbs Z, Agapito L, Li G, Kim H S, Nardelli M, Curtarolo S and Snyder G 2015 *Nat. Mater.* **14** 1223–8
[6] Prakash A, Xu P, Faghaninia A, Shukla S, Ager J, Lo C and Jalan B 2017 *Nat. Commun.* **8** 15167
[7] Urban J 2017 *Nat. Mater.* **16** 157
[8] Zheng F, Jiang J, Sun B, Zhang W and Pecht M 2016 *Energy* **113** 64
[9] Park M, Zhang X, Chung M, Less G B and Sastry A M 2010 *J. Power Sources* **195** 7904
[10] Khan F, Baek S H and Kim J H 2016 *Appl. Energy* **183** 715
[11] Moser J E 2016 *Nat. Mater.* **16** 4
[12] Mahan G 1990 *Many Particle Physics* 2nd edn (New York: Springer)
[13] Cardona M and Thewalt M L W 2005 *Rev. Mod. Phys.* **77** 1173–224
[14] Cardona M 2001 *Phys. Status Solidi* a **188** 1209–32
[15] Cardona M 2005 *Solid State Commun.* **133** 3–18
[16] Meyer T, Karaiskaj D, Thewalt M and Cardona M 2003 *Solid State Commun.* **126** 119–23
[17] Monserrat B, Conduit G J and Needs R J 2014 *Phys. Rev.* B **90** 184302
[18] Shockley W and Bardeen J 1950 *Phys. Rev.* **77** 407
[19] Chakraborty B and Allen P B 1978 *J. Phys. C: Solid State Phys.* **11** L9
[20] Allen P B and Cardona M 1981 *Phys. Rev.* B **23** 1495–505
[21] Allen P B and Cardona M 1983 *Phys. Rev.* B **27** 4760–9
[22] Monserrat B, Drummond N D, Pickard C J and Needs R J 2014 *Phys. Rev. Lett.* **112** 055504
[23] Patrick C and Giustino F 2013 *Nat. Commun.* **4** 2006
[24] Lahnsteiner J, Kresse G, Kumar A, Sarma D D, Franchini C and Bokdam M 2016 *Phys. Rev.* B **94** 214114
[25] Pan D, Wan Q and Galli G 2014 *Nat. Commun.* **5** 3919
[26] Ramírez R, Herrero C P and Hernández E R 2006 *Phys. Rev.* B **73** 245202
[27] Morales M A, McMahon J M, Pierleoni C and Ceperley D M 2013 *Phys. Rev.* B **87** 184107
[28] Giustino F 2017 *Rev. Mod. Phys.* **89** 015003
[29] Ramírez R, Herrero C P and Hernández E R 2006 *Phys. Rev.* B **73** 245202
[30] Giustino F, Louie S G and Cohen M L 2010 *Phys. Rev. Lett.* **105** 265501
[31] Monserrat B, Drummond N D and Needs R J 2013 *Phys. Rev.* B **87** 144302
[32] Monserrat B and Needs R J 2014 *Phys. Rev.* B **89** 214304
[33] Lloyd-Williams J H and Monserrat B 2015 *Phys. Rev.* B **92** 184301
[34] Monserrat B 2016 *Phys. Rev.* B **93** 014302
[35] Poncé S, Antonius G, Gillet Y, Boulanger P, Laflamme Janssen J, Marini A, Côté M and Gonze X 2014 *Phys. Rev.* B **90** 214304
[36] Poncé S, Antonius G, Boulanger P, Cannuccia E, Marini A, Côté M and Gonze X 2014 *Comput. Mater. Sci.* **83** 341–8
[37] Poncé S, Gillet Y, Janssen J L, Marini A, Verstraete M and Gonze X 2015 *J. Chem. Phys.* **143** 102813
[38] Antonius G, Poncé S, Lantagne-Hurtubise E, Auclair G, Gonze X and Côté M 2015 *Phys. Rev.* B **92** 085137
[39] Zacharias M and Giustino F 2016 *Phys. Rev.* B **94** 075125
[40] Antonius G, Poncé S, Boulanger P, Côté M and Gonze X 2014 *Phys. Rev. Lett.* **112** 215501
[41] Monserrat B 2016 *Phys. Rev.* B **93** 100301
[42] Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566–9
[43] Perdew J P and Zunger A 1981 *Phys. Rev.* B **23** 5048–79
[44] Rojas H N, Godby R W and Needs R J 1995 *Phys. Rev. Lett.* **74** 1827–30
[45] Steinbeck L, Rubio A, Reining L, Torrent M, White I and Godby R 2000 *Comput. Phys. Commun.* **125** 105–18
[46] Liu P, Kaltak M, Klimeš J c v and Kresse G 2016 *Phys. Rev.* B **94** 165109
[47] Kaltak M, Klimeš J C V and Kresse G 2014 *Phys. Rev.* B **90** 054115
[48] Kaltak M, Klimeš J and Kresse G 2014 *J. Chem. Theory Comput.* **10** 2498–507
[49] Bloch F 1932 *Z. Phys.* **74** 295–335
[50] Landau L D and Lifshitz E M 1959 *Course of Theoretical Physics: Statistical Physics* vol 5 (London: Pergamon)
[51] Monserrat B private communication
[52] Tauc J 1968 *Mater. Res. Bull.* **3** 37–46
[53] Blöchl P E 1994 *Phys. Rev.* B **50** 17953–79
[54] Kresse G and Hafner J 1993 *Phys. Rev.* B **47** 558–61
[55] Kresse G and Furthmüller J 1996 *Phys. Rev.* B **54** 11169–86
[56] Kresse G, Furthmüller J and Hafner J 1995 *Europhys. Lett.* **32** 729–34
[57] Parlinski K, Li Z Q and Kawazoe Y 1997 *Phys. Rev. Lett.* **78** 4063–6
[58] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865–8
[59] Perdew J P, Burke K and Ernzerhof M 1997 *Phys. Rev. Lett.* **78** 1396–1396
[60] Heyd J, Scuseria G E and Ernzerhof M 2003 *J. Chem. Phys.* **118** 8207–15

[61] Heyd J and Scuseria G E 2004 *J. Chem. Phys.* **120** 7274–80
[62] Krukau A V, Vydrov O A, Izmaylov A F and Scuseria G E 2006 *J. Chem. Phys.* **125** 224106
[63] Gilbert T L 1975 *Phys. Rev.* B **12** 2111–20
[64] Seidl A, Görling A, Vogl P, Majewski J A and Levy M 1996 *Phys. Rev.* B **53** 3764–74
[65] Zacharias M, Patrick C E and Giustino F 2015 *Phys. Rev. Lett.* **115** 177401
[66] Zollner S, Cardona M and Gopalan S 1992 *Phys. Rev.* B **45** 3376
[67] Logothetidis S, Petalas J, Polatoglou H M and Fuchs D 1992 *Phys. Rev.* B **46** 4483–94
[68] O'Donnell K P and Chen X 1991 *Appl. Phys. Lett.* **58** 2924–6
[69] Varshni Y 1967 *Physica* **34** 149–54
[70] Li C C, Gong M, Chen X D, Li S, Zhao B W, Dong Y, Guo G C and Sun F W 2017 *Diam. Relat. Mater.* **74** 119–24
[71] Baroni S, Giannozzi P and Isaev E 2010 *Rev. Mineral. Geochem.* **71** 39
[72] Birch F 1947 *Phys. Rev.* **71** 809–24
[73] Murnaghan F D 1944 *Proc. Natl Acad. Sci.* **30** 244
[74] O'Donnell K P and Chen X 1991 *Appl. Phys. Lett.* **58** 2924–6
[75] Kawai H, Yamashita K, Cannuccia E and Marini A 2014 *Phys. Rev.* B **89** 085202
[76] Hinuma Y, Grüneis A, Kresse G and Oba F 2014 *Phys. Rev.* B **90** 155405