

Lattice polarization effects on the screened Coulomb interaction W of the GW approximation

Walter R. L. Lambrecht

Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7079, USA

Churna Bhandari

Department of Physics, University of Missouri, Columbia, Missouri 65211, USA

Mark van Schilfgaarde

Department of Physics, King's College London, London WC2R 2LS, United Kingdom

(Received 29 June 2017; revised manuscript received 7 August 2017; published 19 September 2017)

In polar insulators where longitudinal and transverse optical phonon modes differ substantially, the electron-phonon coupling affects the energy-band structure primarily through the long-range Fröhlich contribution to the Fan term. This diagram has the same structure as the GW self-energy where W originates from the electron part of the screened Coulomb interaction. The two can be conveniently combined by combining electron and lattice contributions to the polarizability. Both contributions are nonanalytic at the origin, and diverge as $1/q^2$ so that the predominant contribution comes from a small region around $\mathbf{q} = 0$. Here we adopt a simple estimate for the Fröhlich contribution by assuming that the entire phonon part can be attributed to a small volume of \mathbf{q} near $\mathbf{q} = 0$. We estimate the magnitude for $\mathbf{q} \rightarrow 0$ from a generalized Lyddane-Sachs-Teller relation, and the radius from the inverse of the polaron length scale. The gap correction is shown to agree with Fröhlich's simple estimate $-\alpha_P \omega_{LO}/2$ of the polaron effect with α_P the polaron coupling factor.

DOI: [10.1103/PhysRevMaterials.1.043802](https://doi.org/10.1103/PhysRevMaterials.1.043802)

I. INTRODUCTION

The GW approximation [1,2] provides one of the most successful many-body-perturbation theoretical approaches to the electronic band structure of solids. It is based on an expansion of the self-energy in the screened Coulomb interaction W . In fact, the self-energy schematically is approximated as $\Sigma = iGW$ with G the one-electron Green's function. In its most recent quasiparticle self-consistent all-electron version [3,4], which we label QSGW, it has become applicable to a wide range of systems. Nonetheless, it significantly overestimates the band gap in strongly ionic systems. This effect has been attributed to the neglect of ladder diagrams, which describe electron-hole pair attraction, and thereby enhance the screening, which reduces W and the splitting between occupied and unoccupied states (see, e.g., Ref. [5]). However, the lattice polarization also contributes to W and enhances the screening. Previous estimates of both of these effects in literature [5–7] attribute most of the band gap overestimate by straight GW to just one of these but did not consider them together.

Both electronic and polaronic terms are bosonic in origin, have the same diagrammatic structure, and can be conveniently combined into a single GW diagram where W includes both electronic and lattice contributions to the polarizability. This fact was first used by Bechstedt [8] to modify W when solving both the GW band structure and the Bethe-Salpeter equation for the polarizability in strongly ionic materials. They apply the effect only in the static limit ($\omega = 0$) and within the Coulomb-hole static screened exchange (COHSEX) framework. Furthermore, using a model dielectric function, their correction amounts to replacing the macroscopic ϵ_∞ by ϵ_0 in the $\mathbf{q} \rightarrow 0$ limit. Vidal *et al.* [9] estimated the renormalization of band gaps in materials with a large Fröhlich coupling parameter. They adopted Bechstedt's approach and estimated a gap reduction of about 1 eV in CuAlO_2 . Subsequently Botti

and Marques (BM) made a refinement, taking into account the dynamics in W by using a generalized Lyddane-Sachs-Teller relation [7]. However, they did not properly take into account the volume confinement of W in \mathbf{q} space.

The main physics was already laid out by Hedin [2] in the framework of many-body perturbation theory. He partitioned out the Fan term as a separate contribution to the self-energy iGW^{ph} ; and indeed this has been the customary approach. The main effects on temperature dependent band structure and the zero-point motion corrections have been worked out in the Allen-Heine-Cardona (AHC) theory [10–12]. More recently both Fan and Debye-Waller contributions have been implemented in a density-functional framework [13,14]. A recent review by Giustino [15] describes the different approaches to the electron-phonon coupling problem and points out the relations between the adiabatic AHC theory and the more general Hedin [2] and Baym [16] field theoretical approaches and their modern implementation. The latter rely on interpolation of the electron-phonon coupling coefficients on a fine \mathbf{k} -point integration mesh by means of maximally localized Wannier functions [17]. This approach however becomes problematic for the long-range parts of the electron-phonon coupling in polar materials, the so-called Fröhlich part [18–21], because the latter decay as $1/q$ and lead to a divergent $1/q^2$ contribution (and which becomes $1/q^4$ near band extrema if applied straightforwardly to the AHC equations). While these problems can be overcome by removing the adiabatic approximation [15,22], it seems worthwhile pursuing simpler approaches to estimate the Fröhlich part of the Fan term, which dominates in compounds where ϵ_∞ is small compared to ϵ_{tot} .

Noting that the main problem with the Fröhlich term occurs near band edges, Nery and Allen [23] developed a $\mathbf{k} \cdot \mathbf{p}$ approach for the band near the band edge, dividing the Fan term into a nonanalytic Fröhlich part and a remainder. Using the resulting simple form of the Fröhlich term, the singular integral

can be done analytically and can then be combined with the numerical integration without the need for an excessively fine integration mesh. Here we take their approach a step further toward simplification. As they showed, the crucial length scale for the effect is the polaron length scale. Therefore, we use the inverse polaron length directly as an integration limit for the singular Fröhlich term. While this is a more approximate estimate than theirs, in principle, an exact approach, which subtracts the singularity from the numerical integration and replaces it by the analytical result for a simple model, we use directly the simple model and estimate the size of the region in \mathbf{q} space where it is applicable as the inverse polaron length.

The BM approach is another option worth revisiting. This is the main goal of our paper. We identify the problem in BM's treatment of the $\mathbf{q} \rightarrow 0$ limit. Again, our solution is to base this on the polaron-length scale. We show that the BM approach, as is, depends crucially on the size of the integration mesh and the gap correction decreases proportional to $1/N_{\text{mesh}}$ and would thus go to zero at convergence. Instead, we assume that the correction applies in a finite \mathbf{q} region of the size of the inverse polaron length. The advantage compared to a full evaluation of the Fröhlich contribution to the Fan terms is that the computational effort is vastly simpler; moreover, the phonon contribution to the entire band structure is obtained in a single calculation.

We apply both approaches to a set of strongly ionic materials, MgO, NaCl, LiF, LiCl, and show that the modified BM approach leads to results in good agreement with the above simplified Nery-Allen polaronic estimate. We also consider zincblende GaN for a direct comparison to Nery and Allen's more complete approach, although the effect here is an order of magnitude smaller. Admittedly, our approach does not address the full electron-phonon coupling renormalization of the band gap, only the Fröhlich part. However, for strongly ionic materials, this is arguably the largest contribution. The other electron-phonon contributions to the zero-point motion correction are large only in systems with only light atoms. Finally, we consider the relative importance of this effect to the effects of missing electron-hole interactions based on literature data. The conclusion is that the latter are in fact a more important correction to the band-gap reduction.

II. THEORY

As is well known [24,25], optical phonon modes can strongly modify the screening in polar compounds. This is nicely encapsulated by the generalized Lyddane-Sachs-Teller (LST) relation in the $\mathbf{q} \rightarrow 0$ limit

$$\frac{\varepsilon_{\text{tot}}^{\alpha}(\mathbf{q} \rightarrow 0, \omega)}{\varepsilon_{\text{el}}^{\alpha}(\mathbf{q} \rightarrow 0, \omega)} = \prod_m \frac{(\omega_{\text{LO};m}^{\alpha})^2 - \omega^2}{(\omega_{\text{TO};m}^{\alpha})^2 - (\omega + i0^+)^2}. \quad (1)$$

The product runs over all optical modes m which are infrared active and have a longitudinal-transverse splitting ($\omega_{\text{LO};m} > \omega_{\text{TO};m}$) and belong to the irreducible representation corresponding to the polarization direction α . The superscript α indicates the direction along which $\mathbf{q} \rightarrow 0$ approaches zero and the LO modes depend on this direction. We next examine how this affects the screened Coulomb interaction W in the GW theory.

In practical calculations, $W(\mathbf{q}, \mathbf{r}, \mathbf{r}', \omega)$ is represented by an expansion in a basis set. In our all-electron implementation [4] this consists of a mixed product basis with Bloch sums of products of partial waves inside augmentation spheres and plane waves in the interstitial. Thus W becomes a matrix $W_{\mu\nu}(\mathbf{q}, \omega)$. More commonly plane waves are used for these bosonic degrees of freedom, in which case μ becomes \mathbf{G} . As noted already, the effect is dominant in the $\mathbf{q} \rightarrow 0$ limit. Treatment of $W_{\mu\nu}(\mathbf{q} \rightarrow 0, \omega)$ requires special care because of the divergence of the Coulomb interaction $v(\mathbf{q}) = 4\pi e^2/q^2$. (There is a similar divergence for Fröhlich contribution.) It is however integrable, because what is needed for the self-energy is a convolution integral over both energy and wave vector,

$$\begin{aligned} \Sigma_{nm}^c(\mathbf{k}, \omega) &= \frac{i}{2\pi} \int d\omega' \sum_{\mathbf{q}} \sum_{n'}^{\text{all}} G_{nn'}(\mathbf{k} - \mathbf{q}, \omega - \omega') \\ &\times \sum_{\mu\nu} W_{\mu\nu}^c(\mathbf{q}, \omega') e^{-i\delta\omega'} \\ &\times \langle \psi_{\mathbf{k}n} | \psi_{\mathbf{k}-\mathbf{q}n'} E_{\mu}^{\mathbf{q}} \rangle \langle E_{\nu}^{\mathbf{q}} \psi_{\mathbf{k}-\mathbf{q}n'} | \psi_{\mathbf{k}m} \rangle. \end{aligned} \quad (2)$$

The sum over \mathbf{q} becomes an integral and the contribution from the region near $\mathbf{q} = 0$ over a small sphere multiplies the divergence by $q^2 dq$. Here the Green's function $G_{nn'}(\mathbf{k} - \mathbf{q}, \omega - \omega') = [\omega - \omega' - \epsilon_{\mathbf{k}-\mathbf{q}} \pm i\delta]^{-1} \delta_{nn'}$ is a diagonal matrix in the basis of one-electron states $\psi_{\mathbf{k}n}$; the screened Coulomb interaction is expanded in an auxiliary mixed product basis set E_{μ} which diagonalizes the bare Coulomb matrix [26], and conversion factors from one basis to the other are included. The superscript c refers to taking the correlation part $W^c = W - v$.

The approach dealing with this integrable divergence has been described by Freysoldt *et al.* [27] in the context of a plane wave basis set expansion of the bare and screened Coulomb interaction and by Friedrich *et al.* [26,28] in terms of a mixed product basis set expansion. The method consists of replacing the integral over the BZ by an exactly integrable function with the same type of divergence. The difference between the two is then a smooth function for which the integral can be replaced by a discrete sum. The approach originally was introduced by Massida *et al.* [29] in Hartree-Fock calculations because, in fact, the same problem already affects the bare exchange. In the context of GW theory, it requires a knowledge of the screened Coulomb interaction near $\mathbf{q} = 0$ and this can either be obtained by an analytical $\mathbf{k} \cdot \mathbf{p}$ approach [26] or by using the offset- Γ method. The actual approach used in the QSGW program [30] is described in Kotani *et al.* [31] and provides an improved version of the offset- Γ method used in Ref. [4]. To obtain the behavior $W(\mathbf{q})$ near $\mathbf{q} = 0$, one needs the macroscopic inverse dielectric constant, which is $1/\varepsilon_{00}^{-1}(\mathbf{q} \rightarrow 0, \omega)$. It is calculated by a block matrix inversion separating out the divergent term from the known behavior of the polarizability matrix as a function of \mathbf{q} . Here the subscripts 00 of the dielectric function matrix refer to the reciprocal lattice vector $\mathbf{G} = 0$ in a plane wave basis set, or equivalently the first mixed-product basis set function in the basis set that diagonalizes the bare Coulomb interaction [26]. Both of these in fact correspond to the average over the unit cell. The inverse of this quantity is then expanded in spherical harmonics and only the $L = 00$ spherical average is required for the integral of the "head" of ε_{00}^{-1} . This is if we neglect some higher order corrections, discussed by Betzinger *et al.* [32].

For a simple estimate of the phonon contribution, we use the fact that (1) the Fröhlich contribution to W originates predominately from the divergent, small- q region [23], and (2) we handle this region using the usual formulation of the GW self-energy calculation through a special treatment of the integrable divergence of W in the neighborhood of $\mathbf{q} = 0$ only. Equation (2) is integrated numerically on a discrete \mathbf{q} mesh, and the “central” cell term is treated specially to handle the divergence. Our approach simply modifies the central cell dielectric function $\varepsilon_{00}^{-1}(\mathbf{q} = 0)$ using the appropriate Lyddane-Sachs-Teller factor. The fact that its limit is nonanalytic, i.e., depends on the direction of \mathbf{q} means that it is a second rank tensor with nonzero components dictated by symmetry. For example for an orthorhombic crystal, it will have only diagonal components but the xx , yy , and zz diagonal elements are all different. In general in the anisotropic offset- Γ method [31] it is expanded in invariant tensors corresponding to the symmetry of the cell and requires at most six \mathbf{q} points close to $\mathbf{q} = 0$ where the macroscopic dielectric constant must be evaluated and for which we need to know the corresponding LO-TO splittings.

In a full approach to the electron-phonon coupling, one can arbitrarily cut out some small region near the singularity, subtract the standard mesh integration technique result for that region, and replace it by the properly integrated singularity. This is the approach followed by Nery and Allen [23]. Here we focus exclusively on the Fröhlich term, and thus we cannot rely on a cancellation of the two treatments to the self-energy integral. We thus need an accurate estimate of the range of the polaron effect. In the treatment of the $\mathbf{q} \rightarrow 0$ limit for the purely electronic screening, the relative weight of the specially treated Γ cell depends on the size of the \mathbf{q} -point mesh. The finer the mesh, the less the weight of the Γ cell. The electron-phonon contribution should not depend on the mesh spacing, but since we lump the entire contribution into the central cell and omit contributions from other microcells, the Fröhlich contribution to $W_{00}(\mathbf{q} = 0, \omega)$ should be rescaled as described below.

We may decouple the convergence in \mathbf{q} space of the electronic polarizability from that of the phonon contribution as follows. We define the LST factor f_{LST} to be the factor that corrects W , so that the additive correction is $\Delta W = (f_{\text{LST}} - 1)W$. At $\mathbf{q} = 0$, f_{LST} is the inverse of the factor in Eq. (1). Now, according to the above discussion, we want to correct the $\mathbf{q} = 0$ value of W but this represents an effective volume in \mathbf{q} space. So we need to estimate separately the size of this \mathbf{q} -space region over which the effect of the lattice polarization is to be taken into account. Let us call this q_{LP} and the corresponding volume $\Omega_{\text{LP}} = q_{\text{LP}}^3$. When we calculate the convolution integral $\Delta \Sigma(\mathbf{k}) = \sum_{\mathbf{q}} G(\mathbf{k} - \mathbf{q}) \Delta W(\mathbf{q})$ as a discrete sum, we assume only the $\mathbf{q} = 0$ microcell of volume $\Omega_{\text{GW}} = \Omega_{\text{BZ}}/N_{\text{mesh}}$ contributes, so $\Delta \Sigma(\mathbf{k}) = iG(\mathbf{k})\Delta W(0)$. If the GW mesh is coarser than q_{LP} , then we might overestimate the effect. On the other hand, if it is finer, then the phonon correction should be extended to GW -mesh points beyond $\mathbf{k} = 0$. Instead, we may simply rescale $\Delta W(0)$ to $W(0)\Omega_{\text{LP}}/\Omega_{\text{GW}}$. This means that for the pure electronic screening part, the usual compensation between the discrete sum (nondivergent part) and the special treatment of the Γ cell is still valid. But for the added $\Delta \Sigma = i(f_{\text{LST}} - 1)G(\mathbf{k})W(0)\Omega_{\text{LP}}/\Omega_{\text{GW}}$ we use a fixed volume of \mathbf{q} space corresponding to q_{LP}

The essential problem to obtain meaningful results within this approach is thus to pick q_{LP} . We note that it cannot be obtained from considerations of the phonons or of the dependence of $\varepsilon_{\text{tot}}(\mathbf{q})/\varepsilon_{\text{el}}(\mathbf{q})$ alone because of the latter lack of information on the electron-phonon coupling to the bands, which must involve the effective masses of the bands. Following Nery and Allen’s idea [23], the relevant length scale here is the polaron length $a_P = \sqrt{\hbar/2\omega_{\text{LO}}m_*}$, where ω_{LO} is the longitudinal phonon for which we consider the electron-phonon contribution and m_* is the band effective mass. This means there is actually a different polaron length for electrons and holes, which we denote a_{Pe} and a_{Ph} . Using an electron (hole) effective mass of 0.35 (1.26) for MgO as an example, we obtain polaron length scales of 20.84 and 10.96 a_0 (a_0 is the Bohr radius). We use $(2m_{hh} + m_{lh})/3$ along the [100] direction for the holes. We use this type of average because the heavy hole band is doubly degenerate. The average $a_P = (a_{Pe} + a_{Ph})/2$ defines an inverse length scale of $0.06 a_0^{-1}$ which for MgO is about 1/12 of the BZ. Typically, for a two atom unit cell system like MgO, a $8 \times 8 \times 8$ mesh already gives both the GW and phonons very well converged. We also found that the gap correction in the BM approach varies as $1/N_{\text{mesh}}$. We can thus extrapolate to the appropriate $q = 1/a_P$ or use the approach for decoupling from the mesh size described in the previous paragraph. The precise way of averaging the effective masses here is not crucial because we are only trying to estimate the polaron length and the results are not very sensitive to this estimate.

The approach described here is similar to that of Botti and Marques [7], except that they did not take into account the range of the Fröhlich interaction. In their formulation the effect would have vanished in the limit of small microcell size.

They also appeared to confuse the relevant length scale: they said “It is easy to understand that the coupling of phonon waves and electromagnetic waves is effective only for $\mathbf{q} \rightarrow 0$, since the speed of sound is negligible if compared with the speed of light.” This is true but rather irrelevant to the problem considered here. In fact, the coupling of electromagnetic waves to phonon waves, i.e., polariton formation, following Pick [33], occurs only when $q = |\mathbf{q}| \gg \omega_0/c$ with ω_0 a typical phonon frequency and c the speed of light. Decoupling occurs for $q \ll \omega_0/c$. This is about $q = 10^{-4}$ of the Brillouin zone (BZ). In other words, this theory provides a cutoff *above* rather than below in which the effect comes into play. However, we are not concerned with the retardation effects of polariton formation here, we are interested in the applicability of the LST relation and the polaronic effect on the band gap.

We may also directly estimate the Fröhlich singularity integral. Following Nery and Allen [23], the singular contribution near the band edge to the zero-point motion self-energy is given by

$$\Delta E_{n\mathbf{k}} = -\alpha_P \hbar \omega_{\text{LO}} \tan^{-1} (q_F a_{\text{LO}}) \frac{2}{\pi}, \quad (3)$$

where

$$\alpha_P = \frac{e^2}{2a_P \hbar \omega_{\text{LO}}} \left[\frac{1}{\varepsilon_{\infty}} - \frac{1}{\varepsilon_0} \right] \quad (4)$$

is the dimensionless polaron coupling constant. The question now is what to use as an integration cut-off q_F for the

upper limit of the Fröhlich singularity integral. Clearly an overestimate will be obtained if we use $q_F = q_{\text{BZ}} = 2\pi/a$ because the expression is supposed to be valid only over the region where the band dispersion is parabolic. It is even customary to let $q_F \rightarrow \infty$, in which case we obtain $-\alpha\hbar\omega_{\text{LO}}$ as an upper limit [18]. In Nery and Allen's approach this choice is not crucial because they look at where this explicit approach becomes equivalent to the standard integration approach of the other electron-phonon coupling terms besides the long-range Fröhlich one. This occurs at about 1/6 of the BZ in their case of GaN. A better estimate would be $q_F = 1/a_P$; then the inverse tangent factor is simply $\pi/4$ and Eq. (3) simplifies to

$$\Delta E_{n\mathbf{k}} = -\alpha_P \omega_{\text{LO}}/2 = \frac{e^2}{4a_P} (\varepsilon_{\infty}^{-1} - \varepsilon_0^{-1}), \quad (5)$$

where if n corresponds to the conduction band minimum (CBM), we use a_{P_e} and if n corresponds to the valence band maximum (VBM), we would use a_{P_h} as polaron lengths. So, this is simply half the difference in Coulomb interaction at the polaron length calculated with purely electronic screening and electronic plus lattice screening. This differs from the upper limit by only a factor of 2, so the estimation of q_F is not very crucial if our goal is to obtain the right order of magnitude. For the example of GaN used by Nery and Allen, we find this already gives an excellent approximation to their full calculation. We will show that it also agrees well with the modified BM approach described above, in which q_{LP} is also set to $1/a_P$.

We emphasize again that if one applies the electron-phonon coupling fully at all \mathbf{q} points then one can subtract out the region of the singularity from the mesh sum and add it back in integrated analytically. However, in the BM approach we focus entirely on the singular contribution, and thus we are limited by how reliably we estimate the size of the singularity. This is true both if one thinks of it as a Fröhlich coupling strength dipole singularity or as the $1/q^2$ singularity in the W screened Coulomb potential. In fact, both are essentially the same and both give a correction proportional to $\varepsilon_{\infty}^{-1} - \varepsilon_0^{-1}$ and inversely proportional to a_P .

As far as the frequency dependence is concerned, it is clear from Eq. (1) that for $\omega > \omega_{\text{LO}}$ the factor quickly goes to 1. Therefore in the integral over frequency, it is sufficient to apply the effect only at $\omega = 0$ as long as the first nonzero ω -mesh point is already well above ω_{LO} . If one wishes to apply the effect including its frequency dependence, then one needs to use a sufficiently fine integration mesh near the ω_{TO} and ω_{LO} phonons to carry out the integrals over these poles correctly. We have done both and find that reliable results can be obtained using a coarse mesh, and scaling $W(\mathbf{q}=0, \omega=0)$ only. This is further discussed in the Appendix. Since a very fine frequency mesh is required if the pole is properly summed over, this greatly simplifies the computational effort.

III. COMPUTATIONAL DETAILS

All calculations before are carried out using the QSGW approach in the LMTO basis set implementation [4], which can be found online in Ref. [30]. The relevant phonons are taken from experiment or can be calculated using the ABINIT program [34].

TABLE I. Parameters used to calculate the polaron length and polaron coupling strength in various materials and the polaron shifts of the band edges and gap. Effective masses in units of free electron mass $m_h = (2m_{hh} + m_{lh})/3$. Phonon frequency ω_{LO} in cm^{-1} , and polaron lengths a_{P_e} and a_{P_h} in Bohr units a_0 . Polaron coupling constants α_{P_e} and α_{P_h} for electrons and holes are dimensionless. Energy shifts in meV.

| | MgO | NaCl | LiF | LiCl | GaN |
|---------------------------------------|------|------|------|------|------|
| m_e | 0.35 | 0.35 | 0.61 | 0.40 | 0.18 |
| $m_{hh}[100]$ | 1.70 | 2.10 | 2.83 | 1.06 | 1.70 |
| $m_{lh}[100]$ | 0.40 | 0.55 | 1.10 | 0.56 | 0.50 |
| m_h | 1.26 | 1.58 | 2.25 | 0.89 | 1.30 |
| $\omega_{\text{LO}} (\text{cm}^{-1})$ | 722 | 265 | 656 | 382 | 730 |
| $a_{P_e} (a_0)$ | 20.8 | 34.4 | 16.6 | 26.8 | 28.9 |
| $a_{P_h} (a_0)$ | 11.0 | 16.2 | 8.6 | 17.9 | 10.7 |
| $a_P (a_0)$ | 15.9 | 25.3 | 12.6 | 22.4 | 19.8 |
| ε_{∞} | 3.0 | 2.3 | 1.95 | 2.8 | 5.6 |
| ε_0 | 9.8 | 5.9 | 9.0 | 11.2 | 9.9 |
| α_{P_e} | 1.7 | 3.2 | 4.1 | 2.9 | 0.4 |
| α_{P_h} | 3.2 | 6.9 | 7.8 | 4.4 | 1.1 |
| $\Delta E_{\text{CBM}} (\text{meV})$ | 75 | 53 | 165 | 69 | 18 |
| $\Delta E_{\text{VBM}} (\text{meV})$ | 144 | 113 | 317 | 103 | 49 |
| $\Delta E_g (\text{meV})$ | 219 | 166 | 483 | 172 | 67 |

The experimental lattice constants were used, MgO (4.21 Å [35]), NaCl (5.64 Å), LiCl (5.14 Å [36]), LiF (4.02 Å [36]). In the QSGW calculation, for MgO we used semicore Mg $2p$ and high lying $3s$ states as local orbital for the completeness of the basis set. For NaCl we also used semicore $2p$ as local orbitals.

IV. RESULTS AND DISCUSSION

In Table I we show the polaron lengths and coupling strengths for electrons and holes as well as the parameters entering them. The effective masses are obtained from fitting the QSGW calculations before adding the lattice polarization correction. The dielectric constants are taken from experiment but could also be calculated in DFT using for example the ABINIT program or using the electronic band structure within QSGW for ε_{∞} and calculating ε_0 using the LST factor. Finally it shows the estimated band edge and gap shifts using Eq. (5). We note that for GaN, this amounts to 67 meV, close to Nery and Allen's own estimate of 50 meV, especially when considering that we expect an error bar of order 10 meV in view of the various approximations made. We may note that typically the shift is larger for the VBM because of the shorter polaron length because of the larger hole mass. For NaCl, we may compare our result with Fröhlich's own estimate of the conduction band shift [18]. He used an effective mass $m_e = 1$ and obtained 0.18 eV. Using $m_e = 1$ we would obtain 0.09 eV differing by a factor 2 because in Fröhlich's estimate the upper limit of the singularity integral is replaced by ∞ .

Next we compare the above polaron estimates of the band edge shift with the results of the modified BM approach, in which we use the average polaron length a_P to set the $q_{\text{LP}} = 1/a_P$. In Table II we show the LDA gaps, the QSGW gaps, and the QSGW gaps with the BM-type lattice

TABLE II. Band gaps in LDA, QSGW, QSGW+LPC in the BM approach with different k meshes and extrapolated to $q_{LP} = 1/a_p$, as well as the zero-point motion (ZPM) correction to the gap. All values in eV. The percent change in QSGW-LDA gap correction due to LPC is also given.

| | MgO | NaCl | LiF | LiCl | GaN |
|--------------------------|-------|-------|-------|-------|---------|
| LDA | 4.65 | 5.01 | 9.43 | 6.32 | 1.76 |
| QSGW | 8.69 | 9.44 | 16.19 | 10.19 | 3.54 |
| QSGW+LPC-BM-6 | 7.99 | 8.80 | 14.87 | 9.52 | 3.31 |
| QSGW+LPC-BM-8 | 8.20 | 8.99 | 15.32 | 9.71 | 3.50 |
| QSGW+LPC | 8.43 | 9.26 | 15.64 | 9.98 | 3.50 |
| ZPM-Fröhlich | -0.26 | -0.18 | -0.55 | -0.21 | -0.04 |
| ZPM-polaron | -0.22 | -0.17 | -0.48 | -0.17 | -0.07 |
| QSGW-LDA | 4.04 | 4.43 | 6.76 | 3.87 | 1.78 |
| % change LPC | -6 | -4 | -8 | -5 | -2 |
| QSGW-BM ^a | 8.94 | 9.52 | 15.81 | 10.28 | |
| QSGW+LPC-BM ^a | 7.71 | 8.37 | 13.69 | 9.05 | |
| ZPM-BM ^a | -1.23 | -1.15 | -2.12 | -1.23 | |
| Expt. gap | 7.8 | 8.5 | 14.2 | 9.4 | 3.5±0.1 |

^aFrom Botti and Marques [7].

polarization correction using $N_{\text{mesh}} = 6$ or 8. We note that the QSGW results are well converged already with a $6 \times 6 \times 6$ mesh, which differ from the $8 \times 8 \times 8$ mesh by only 0.1 eV. However, the LPC correction is then effectively applied only to a smaller region and has less weight. Therefore the corresponding QSGW+LPC-BM-8 gap is less reduced than the QSGW+LPC-BM-6 one. We then extrapolate from the $q_{LP} = (2\pi/a)(1/N_{\text{mesh}})$ to $q_{LP} = 1/a_p$ assuming linear dependence. This is the result labeled QSGW+LPC. Finally, the zero-point motion correction, due to the lattice-polarization correction, labeled ZPM-Fröhlich in Table II is the difference between the QSGW+LPC and QSGW gaps and should be compared with the polaron effect given in Table I. Rounding the values of ΔE_g of Table I to 0.01 eV, a more realistic estimate of the uncertainty, we obtain the results in the row labeled ZPM-polaron. We can see that the two estimates agree with each other to within a few 0.01 eV. Comparing with the gaps and gap reduction values given by Botti and Marques in the next few rows, we see that their calculation significantly overestimated the effect. This is primarily because they used a $6 \times 6 \times 6$ mesh but there are also differences in the QSGW results themselves which result from their use of a pseudopotential approximation and a plane wave basis set compared to our all-electron and LMTO basis set. Finally, we also give the experimental gaps.

We note that our QSGW gap for MgO is lower than that of BM (8.94 eV) or Shishkin *et al.* [5] (9.16 eV) or Chen and Pasquarello [6] (9.29 eV). We note that if we use a less converged basis set (leaving out the higher energy 3s local orbitals on O for example, we also find a higher gap). We therefore caution that comparisons of the lattice-polarization effect in GW between different methods should also keep in mind that differences between all-electron and pseudopotential methods as well as various convergence issues may play a role.

Our adjusted gaps still overestimate the experimental gaps. This indicates that the electron-hole effects on the gap reduction may be more important than the lattice polarization

correction. For MgO, the results of Shishkin *et al.* [5] and Chen *et al.* [6] indicate effects of the order of 20% of the QSGW-LDA gap correction. The underscreening in the QSGW due to the lack of electron-hole interactions or random phase approximation was noted before and for most semiconductors amounts to about 20%. This has led to a commonly used correction factor of 0.8Σ [37,38]. A universal factor of 0.8 can be approximately justified because ϵ_∞ is uniformly underestimated by a factor of 0.8 in a wide range of semiconductors. The results of Chen *et al.* [6] and Shishkin [5] MgO, NaCl, and LiCl further support this. Our results indicate a further reduction of this gap correction by the lattice polarization by about 5%. The percentage reduction of the QSGW-LDA gap correction due to the lattice polarization effect is given in Table II and varies from 2% to 8%. Taken together this would reduce the QSGW-LDA gap correction by 25%. The zero-point motion correction of the gap in MgO was previously estimated to be 0.15–0.19 eV for MgO [13], but it is not clear whether this properly includes the long-range Fröhlich contribution. Assuming it is not and adding the present Fröhlich lattice polarization correction, the total zero-point motion correction would then amount to 0.4 eV.

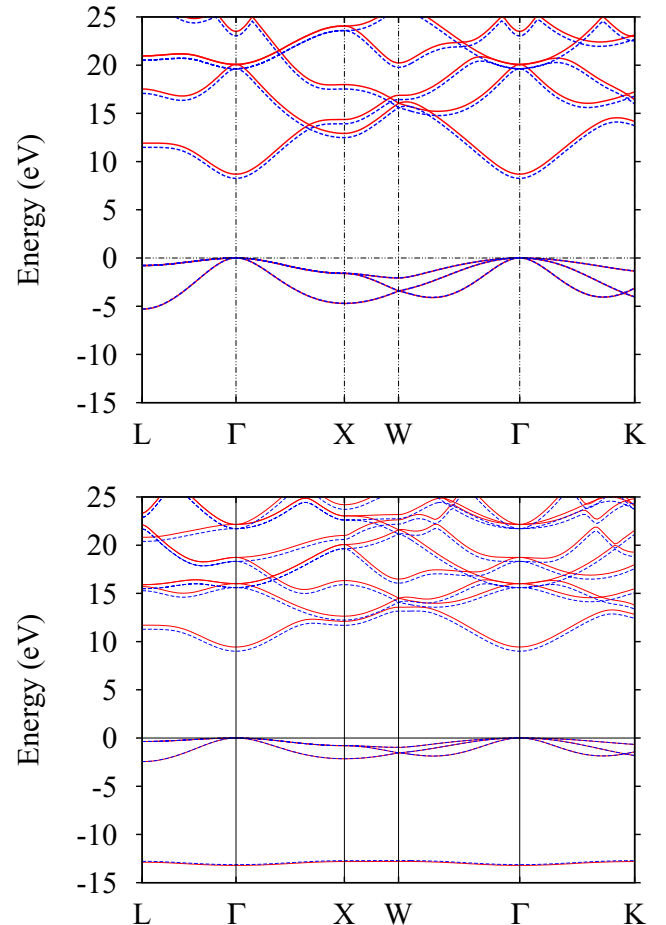


FIG. 1. Band structure of MgO (top) and NaCl (bottom), the red color for QSGW and blue dashed for QSGW+LPC with $N_{\text{mesh}} = 8$.

One advantage of our modified BM approach is that in principle it not only corrects the gap but the entire band structure. Plotting the full band structures, as shown in Fig. 1 for two examples MgO and NaCl, and for $N_{\text{mesh}} = 8$ we find that the effect amounts pretty much to a constant shift of the conduction band once the valence bands are aligned. Using the fixed mesh size may still somewhat overestimate the effect on the gap as seen from Table II but nonetheless allows us to estimate how it affects the rest of the band structure. This constant shift is by no means guaranteed because it is not clear immediately whether the same polaron length-scale estimates of the required integration region of the $\mathbf{q} \rightarrow 0$ singularity q_{LP} applies also to other bands. Our calculation assumes that this can be taken the same for all bands. At present we are not aware of experimental evidence to test this result of a constant shift.

V. CONCLUSIONS

In this paper we revisited the approach proposed by Botti and Marques [7] to estimate the lattice polarization effect on W in the GW method and hence on the band structure in ionic materials. As pointed out by Giustino [15] the BM approach is equivalent to the Fröhlich contribution to the Fan self-energy, which has thus far only received limited attention in spite of the large amount of work on electron-phonon coupling renormalization of the band gaps of materials. This is primarily due to the technical difficulties in calculating it, which require a very fine \mathbf{k} -space integration mesh.

We develop here a simplified approach which takes advantage of the fact that the Fröhlich interaction is dominant in a small region around $\mathbf{q} = 0$. The effective volume of \mathbf{q} is fixed by the polaron length scale $a_P = \sqrt{\hbar/2\omega_{LO}m^*}$. With this length scale for the Fröhlich interaction, and the LST relations at $W(\mathbf{q} = 0)$, we can construct a simple description that modifies W directly, which can be used both in GW calculations and for higher order diagrams involving W , e.g., incorporation of ladder diagrams via the Bethe-Salpeter equation [39], with minimal cost.

We compared our results of the BM method with a simple estimate based on directly integrating the Fröhlich electron-phonon coupling singularity near the band gap and find that the latter can also be estimated simply by setting the cutoff of the singularity integral to the inverse polaron length. We note that the present method allows us to estimate the gap correction to at best a few 0.01 eV only because of the remaining uncertainty in the polaron length a_P . For a more refined treatment to meV precision a full electron-phonon calculation of the band shifts will be required and would allow adjusting a_P to fit it.

ACKNOWLEDGMENTS

This work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences under Grant No. DE-SC0008933. Calculations made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University. M.v.S. was supported by EPSRC CCP9 Flagship Project No. EP/M011631/1.

APPENDIX: DISCUSSION OF THE $\omega \rightarrow 0$ LIMIT

Here we address the issue of whether we need the ω dependence of the LST factor. As far as the frequency dependence is concerned, we can in principle include the frequency dependent LST factor as given in Eq. (1) in the main text. However, in evaluating Σ we then need to make sure the extra pole due to the phonon is properly integrated over. The method for calculating the energy integral in Eq. (2) in the main text is described in detail in Sec. II F of Ref. [4]. It is done with a contour integral mostly along the imaginary axis but including pole contributions from the energy bands along the real axis. The inclusion of the lattice polarization effect through the LST factor introduces poles in $W(\omega)$ close to $\omega = 0$ in the energy range of the phonons, more precisely at the ω_{LO} frequencies. We thus need to ensure that these do not lead to spurious effects and are adequately integrated over. The behavior of W near $\omega \rightarrow 0$ along the imaginary axis is already taken care of especially in Ref. [4]. The band structure poles, however, lead to a contribution $W^c(\omega - \epsilon_{\mathbf{k}-\mathbf{q}n})$. These are tabulated on an ω mesh along the real axis for later interpolation of $\Sigma(\omega)$ to the values required. For example, in the QSGW, method one needs $[\Sigma_{nm}(\epsilon_{\mathbf{k}m}) + \Sigma_{nm}(\epsilon_{\mathbf{k}n})]/2$. Thus this mesh must be chosen fine enough so that adequate interpolation is possible if $\omega - \epsilon_{\mathbf{k}n} \approx 0$ for some energy band. Unphysical values would result if this energy band is close to the pole. One may avoid a divergence by adding a small imaginary part to the ω and using a fine mesh in the region of the phonons. However, for a reasonable ω -mesh spacing for the electronic part of W^c , all except the first point $\omega = 0$ are usually well above the phonon frequencies where the LST factor goes to 1. Thus, we may also only correct the $\omega = 0$ mesh point where the correction factor is simply $\prod_m \omega_{LOm}^2/\omega_{T0m}^2$. We have tested both approaches and found that they give the same result for the final band gap. In fact, intuitively, one does not expect the detailed behavior near each phonon to have a specific effect. Such an effect would occur whenever some energy band difference $\epsilon_{\mathbf{k}n} - \epsilon_{\mathbf{k}-\mathbf{q}n'}$ is close to an LO phonon energy and leads to an almost divergent contribution $W^c(\omega_{LO})$. This means that the ω dependence of the LST factor, which is one of the distinguishing features of BM compared to the previous work of Bechstedt [8] who only applied the correction to the static part of Σ , is not as important as one might guess at first sight and Bechstedt's approach is adequate, especially in view of the other approximations we are already making and the overall goal to keep this approach as simple as possible.

We compare results with different real $d\omega$ mesh sizes for MgO. This $d\omega$ corresponds to the mesh spacing used in the interpolation of $\Sigma(\omega)$ as mentioned earlier. From Table III we see that for $d\omega = 0.0002$ Ry the band gap is 8.06 eV which is very close to the value when we use a large mesh spacing $d\omega = 0.02$ Ry. So, either we pick the mesh so fine the

TABLE III. The band gap (in eV) in MgO as function of $d\omega$ (in Rydberg), the mesh size along the real frequency axis used in the calculation of Σ .

| $d\omega$ | 0.0002 | 0.0005 | 0.004 | 0.01 | 0.02 |
|-----------|--------|--------|-------|------|------|
| E_g | 8.06 | 7.95 | 7.89 | 8.12 | 8.08 |

phonon region is integrated and interpolated over correctly, or we pick it big enough so we just skip over it and effectively only include the correction at the first $\omega = 0$ point. However, for intermediate values of $d\omega$, the gap seems to vary in a somewhat uncontrolled way. This results from the difficulty of interpolating the rapidly varying LST correction factor of Eq. (1) in the region near the phonon frequency. It is clear that the latter goes through an asymptote at ω_T , or its inverse goes through an asymptote at ω_L . Thus the value can rapidly change from positive to negative and unreliable results are

obtained if the integration mesh samples just a few random points on this curve. In fact, note that the $\omega_T \approx 0.003$ Ry. In this case of MgO and hence $d\omega = 0.0002$ is sufficiently fine compared to ω_T and $d\omega = 0.02$ is sufficiently large to just skip over the whole range, while $d\omega \approx 0.004$ is troublesome. We conclude from this that correcting only the $\omega = 0$ value is more efficient and sufficiently accurate. Similar results are obtained for NaCl. However in that case, the relevant phonon frequencies are significantly smaller and hence an even finer and ultimately, unpractical mesh would be required.

-
- [1] L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- [2] L. Hedin and S. Lundqvist, in *Solid State Physics, Advanced in Research and Applications*, Vol. 23, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), pp. 1–181.
- [3] M. van Schilfgaarde, T. Kotani, and S. Faleev, *Phys. Rev. Lett.* **96**, 226402 (2006).
- [4] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, *Phys. Rev. B* **76**, 165106 (2007).
- [5] M. Shishkin, M. Marsman, and G. Kresse, *Phys. Rev. Lett.* **99**, 246403 (2007).
- [6] W. Chen and A. Pasquarello, *Phys. Rev. B* **92**, 041115 (2015).
- [7] S. Botti and M. A. L. Marques, *Phys. Rev. Lett.* **110**, 226404 (2013).
- [8] F. Bechstedt, K. Seino, P. H. Hahn, and W. G. Schmidt, *Phys. Rev. B* **72**, 245114 (2005).
- [9] J. Vidal, F. Trani, F. Bruneval, M. A. L. Marques, and S. Botti, *Phys. Rev. Lett.* **104**, 136401 (2010).
- [10] P. B. Allen and V. Heine, *J. Phys. C* **9**, 2305 (1976).
- [11] P. B. Allen and M. Cardona, *Phys. Rev. B* **23**, 1495 (1981).
- [12] P. B. Allen and M. Cardona, *Phys. Rev. B* **27**, 4760 (1983).
- [13] G. Antonius, S. Poncé, E. Lantagne-Hurtubise, G. Auclair, X. Gonze, and M. Côté, *Phys. Rev. B* **92**, 085137 (2015).
- [14] A. Marini, S. Poncé, and X. Gonze, *Phys. Rev. B* **91**, 224310 (2015).
- [15] F. Giustino, *Rev. Mod. Phys.* **89**, 015003 (2017).
- [16] G. Baym, *Ann. Phys.* **14**, 1 (1961).
- [17] F. Giustino, M. L. Cohen, and S. G. Louie, *Phys. Rev. B* **76**, 165108 (2007).
- [18] H. Fröhlich, *Adv. Phys.* **3**, 325 (1954).
- [19] P. Vogl, *Phys. Rev. B* **13**, 694 (1976).
- [20] C. Verdi and F. Giustino, *Phys. Rev. Lett.* **115**, 176401 (2015).
- [21] J. Sjakste, N. Vast, M. Calandra, and F. Mauri, *Phys. Rev. B* **92**, 054307 (2015).
- [22] S. Poncé, Y. Gillet, J. L. Janssen, A. Marini, M. Verstraete, and X. Gonze, *J. Chem. Phys.* **143**, 102813 (2015).
- [23] J. P. Nery and P. B. Allen, *Phys. Rev. B* **94**, 115135 (2016).
- [24] A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation*, 2nd ed. (Academic, New York, 1971).
- [25] X. Gonze and C. Lee, *Phys. Rev. B* **55**, 10355 (1997).
- [26] C. Friedrich, S. Blügel, and A. Schindlmayr, *Phys. Rev. B* **81**, 125102 (2010).
- [27] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R. W. Godby, and M. Scheffler, *Comput. Phys. Commun.* **176**, 1 (2007).
- [28] C. Friedrich, A. Schindlmayr, and S. Blügel, *Comput. Phys. Commun.* **180**, 347 (2009).
- [29] S. Massidda, M. Posternak, and A. Baldereschi, *Phys. Rev. B* **48**, 5058 (1993).
- [30] Our *GW* implementation was adapted from the original *ecalj* package, now at <https://github.com/tkotani/ecalj/>. It is available at <http://www.questaal.org>, together with the one-body part.
- [31] T. Kotani, *J. Phys. Soc. Jpn.* **83**, 094711 (2014).
- [32] M. Betzinger, C. Friedrich, and S. Blügel, *Phys. Rev. B* **81**, 195117 (2010).
- [33] R. M. Pick, *Adv. Phys.* **19**, 269 (1970).
- [34] X. Gonze, J. M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. M. Rignanese, M. Sindic, L. Verstraete, G. Zerah, and F. Jollet, *Comput. Mater. Sci.* **25**, 478 (2002).
- [35] Y. Fei, *Am. Mineral.* **84**, 272 (1999).
- [36] P. P. Ewald and C. Hermann, *Stukturbericht, 1913–1928* (Akademische, Berlin, 1931), Zeitschrift für Kristallographie, Kristallgeometrie, Kristallphysik, Kristallchemie: Ergänzungsband [Strukturbericht, bd. I] .
- [37] A. N. Chantis, M. van Schilfgaarde, and T. Kotani, *Phys. Rev. Lett.* **96**, 086405 (2006).
- [38] D. Deguchi, K. Sato, H. Kino, and T. Kotani, *Jpn. J. Appl. Phys.* **55**, 051201 (2016).
- [39] M. Rohlfing and S. G. Louie, *Phys. Rev. B* **62**, 4927 (2000).