Editors' Suggestion

QS$G\hat{W}$: Quasiparticle self-consistent GW with ladder diagrams in W

Brian Cunningham,[1] Myrta Grüning,[1,2] Dimitar Pashov,[3] and Mark van Schilfgaarde[4]

[1]*School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, United Kingdom*

[2]*European Theoretical Spectroscopy Facilities (ETSF)*

[3]*King's College London, London WC2R 2LS, United Kingdom*

[4]*National Renewable Energy Laboratories, Golden, Colorado 80401, USA*

(Received 15 February 2023; revised 5 September 2023; accepted 13 September 2023; published 3 October 2023)

We present an extension of the quasiparticle self-consistent GW approximation (QSGW) [T. Kotani *et al.*, Phys. Rev. B **76**, 165106 (2007)] to include vertex corrections in the screened Coulomb interaction W. This is achieved by solving the Bethe-Salpeter equation for the polarization matrix at all k points in the Brillouin zone. We refer to this method as QS$G\hat{W}$. QSGW yields a reasonable and consistent description of the electronic structure and optical response, but systematic errors in several properties appear, notably a tendency to overestimate insulating band gaps, blueshift plasmon peaks in the imaginary part of the dielectric function, and underestimate the dielectric constant ϵ_∞. A primary objective of this paper is to assess to what extent including ladder diagrams in W ameliorates systematic errors for insulators in the QSGW approximation. For benchmarking we consider about 40 well-understood semiconductors, and also examine a variety of less well-characterized nonmagnetic systems, six antiferromagnetic oxides, and the ferrimagnet Fe_3O_4. We find ladders ameliorate shortcomings in QSGW to a remarkable degree in both the one-body Green's function and the dielectric function for a wide range of insulators. New discrepancies with experiment appear, and a key aim of this paper is to establish to what extent the errors are systematic and can be traced to diagrams missing from the theory. One key finding of this work is to establish a relation between the band gap and the dielectric constant ϵ_∞. Good description of both properties together provides a much more robust benchmark than either alone. We show how this information can be used to improve our understanding of the one-particle spectral properties in materials systems such as $SrTiO_3$ and FeO.

DOI: 10.1103/PhysRevB.108.165104

I. INTRODUCTION

The one-particle Green's function $G(\mathbf{r}, \mathbf{r}', \omega)$ provides essential information about material properties. Aside from having value in its own right, determining both ground-state properties (total energy, charge, and magnetic densities) and excitation energies, it is the starting point for transport and other two-particle properties, e.g., spin and charge response functions, and superconductivity.

As a consequence, knowledge of G is of the first importance, and a vast amount of effort has been dedicated to finding prescriptions to yield G both efficiently and with high-fidelity *ab initio* (without recourse to models or adjustable parameters). Density-functional theory [1] (DFT), where the electron density n replaces G as the fundamental variable, is an alternative, and indeed it is far more popular because of its efficiency and good scaling with system size. DFT is a ground-state theory, but it generates an auxiliary one-body H_0, with fictitious eigenvalues and eigenfunctions. H_0 often provides a reasonable approximation to excitations of the real system, but it is often unsatisfactory, e.g., its notorious tendency to underestimate splitting between occupied and unoccupied levels. Wave-function methods, widely used in quantum chemistry, but less so in materials physics, use the single-particle orbitals ψ_i as the fundamental variable. They can provide high-fidelity solutions to the many-body Schrödinger equation. As Kohn noted in his Nobel prize lecture [2], wave-function methods contain more information than is needed or useful, but nevertheless require concomitant effort needed to compute observables. For that reason they are expensive and scale poorly with system size. Also, spectral properties are not readily computed.

Green's function (GF) methods lie between the two: G has more information than n but less than the wave functions. As with DFT, G-based methods create an effective one-body potential $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$, but differ in that Σ is nonlocal and energy dependent. They are computationally more intensive than DFT; however, they can be made to scale reasonably well with system size, and because of their better fidelity it is likely they will ultimately outphase DFT methods for many functional materials, particularly when excitations are involved. Thus, GF theories might be called the "Goldilocks" approach. GF methods possess a key advantage in another respect: dynamical screening becomes the predominant many-body effect for systems involving many atoms. Hedin's equations [3] can be expanded diagrammatically in powers of the screened

Coulomb interaction W, and encapsulate this phenomenon in a natural way, even in the lowest order (GW). The traditional target applications are also different: quantum chemical methods focus mostly on ground-state properties while GF methods focus on spectral properties, especially two-particle spectra. GF methods do not yet possess the fidelity of wave-function methods, and to what extent their fidelity can eventually approach them remains a key open question.[1]

One key aim of this paper is to provide a partial answer to this question. The quasiparticle self-consistent GW approximation (QSGW) provides an effective way to implement GW theory without relying on a lower-level approximation as a starting point. This makes discrepancies with experimental data much more uniform, and it is essential to distinguish errors intrinsic to the theory itself, from accidents as a result of the starting point. We assess in some detail the extent to which discrepancies QSGW displays with experiment can be mitigated by adding the low-order diagram (ladder diagram) to the random-phase approximation (RPA)[2] for the bare polarizability. As noted, the GW approximation is the lowest-order diagram in the many-body perturbation theory (MBPT) of Hedin [3], and while GW shows significant improvement over DFT (including functionals designed to surmount the well-gap underestimate [7]), it has well-known problems. First, it is a perturbation theory, which typically starts from some reference noninteracting G_0 and generates a correction to it. The most common choice of G_0 is one based in DFT, but many kinds of choices have been made to improve on the final result. This situation is unsatisfactory in two respects:

(i) G_0 can be (and often is) tuned to improve agreement with experiments. G_0 plays the role of a free parameter, and in this sense the theory is not really *ab initio* any more.

(ii) The errors inherent in low-order MBPT, e.g., GW, can be masked by the arbitrariness in G_0. Sometimes qualitatively wrong conclusions can be drawn, or good agreement with experiment found, but for the wrong reason. This is a quite common, albeit not well-appreciated, difficulty with the theory (see, for example, Ref. [8]).

By employing the GW approximation in the QSGW form, we can circumvent these difficulties. QSGW is a procedure where G_0 is determined self-consistently. Self-consistency is used not to minimize the total energy, but instead some measure (norm) of the difference between G_0^{-1} and G^{-1} [9]. With a definition for optimal construction for G_0, it surmounts the ambiguities from arbitrariness in the starting point [10]. It provides a good and systematic G_0 so that discrepancies with experiment that appear tend to be similarly systematic, making it possible to associate these discrepancies with diagrams missing from the theory.

There is no unique definition of the norm, but one intuitively appealing definition leads to a static (quasiparticlized) self-energy Σ^0 generated from the dynamical one as

$$\Sigma^0(\mathbf{r},\mathbf{r}') = \sum_{ij} \psi_i(\mathbf{r})\, \Sigma^0_{ij}\, \psi^*_j(\mathbf{r}'),$$

$$\Sigma^0_{ij} = \frac{1}{2}\left\{\mathrm{Re}[\Sigma(\varepsilon_i)]_{ij} + \mathrm{Re}[\Sigma(\varepsilon_j)]_{ij}\right\}, \qquad (1)$$

where i and j are eigenstates of the one-particle Hamiltonian. Ismail-Beigi showed this construction satisfies a variational principle, not for the total energy but its gradient [11]. One other important consequence of Eq. (1) is that at self-consistency the poles of G and the poles of G_0 coincide: thus, in contrast to DFT, the energy bands of H_0 generated by QSGW correspond to true excitations of the system.

Results generated by QSGW may sometimes worsen agreement with experiment over other forms of GW. For example, ε_∞ generated from a Kohn-Sham band structure is often better than the QSGW one. We will argue that this stems from a fortuitous cancellation of errors (see Sec. III B 3). Fortuitous error cancellation in QSGW is much less pronounced and, as a result, discrepancies with experiment are better exposed, and moreover they are much more uniform. Several of the most salient discrepancies are connected to the inadequate description of the dielectric polarizability. This forms the primary motivation for this work: to make a detailed assessment of how the simplest extension to the RPA polarizability improves both G and the dielectric response. In this work, the excitonic contributions are taken into account by including ladder diagrams into the screened Coulomb interaction W through use of the Bethe-Salpeter equation (BSE) [12,13] for the polarization. A high fidelity G is essential for a good description of any response function, including the magnetic one, as shown for NiO [14] and for yttrium iron garnet [15], and the particle-particle correlation function that governs superconductivity (see, e.g., Ref. [16]).

Other works have considered the effect of vertex corrections to the dielectric screening on the band gap. For example, Refs. [17–19] included ladder diagrams through an effective nonlocal static kernel constructed within time-dependent density functional theory to mimic the BSE. More recently, Kutepov proposed several schemes for the self-consistent solution of Hedin's equations including vertex corrections [20]. In particular, for selected semiconductors and insulators [21], he included the vertex correction for the dielectric screening at the BSE level together with a so-called first-order approximation for the vertex in the self-energy, $\Sigma = iGW\Gamma$. In all the cases, an improvement over LQSGW was observed (see See. II D 2 for a comparison between LQSGW and QSGW). With respect to these previous works, we include vertex corrections to the dielectric screening only at the BSE level, but introducing the usual static approximation for the BSE kernel, which was lifted in Refs. [20,21].[3] Here we omit the first-order vertex for Σ, in keeping with our present objective: to find the best single-Slater determinant construction. The QSGW

[1] As regards total energy, not examined here, GF methods are immature [4] but they have advanced significantly in recent years. A particularly noteworthy example is the study of the "S66" test set, where the authors achieved quantum-chemical accuracy by adding only singles and second-order screened exchange to GW [5,6].

[2] May also be referred to as the independent-particle approximation.

[3] An analogy in the quantum-chemical literature is the fully self-consistent framework corresponding to "multireference" starting points while the quasiparticlized form corresponds to an optimized single reference.

philosophy incorporates this vertex in an approximate way, via a Ward identity in Γ that goes as $1/Z$ in the $q \to 0, \omega \to 0$ limit, canceling the Z factor that is the predominant difference between the quasiparticlized G_0 and the interacting G [22]. Adding this vertex explicitly jeopardizes this cancellation. It can be surmounted via a fully self-consistent G, as Kutepov did, but the cost is considerable. Ladder diagrams can be included in W while retaining $O(N^3)$ scaling [23], but there is no obvious analog to adding the vertex in Σ. We show here that including the vertex in W is more important: including it without the vertex in Σ usually yields quite satisfactory results. Kutepov noted the interacting G with both vertices performs better in CuCl: more generally it seems to matter when the highest occupied states are flat and nearly dispersionless (see point 2, Sec. II D 1 and discussion around Table IV). Finding a way to surmount this shortcoming in a quasiparticle framework is a work in progress.

Another readily identifiable source of error is the contribution lattice vibrations make to Σ. It can be a few tenths of an eV in diamond and in polar compounds with light elements, so we include that contribution here in an approximate way, by obtaining the reduction in the gap by an independent method (Sec. II G) and using a hybrid self-energy to reproduce this shift (Sec. II D 3).

II. THEORY AND NUMERICAL IMPLEMENTATION

Starting from the Hedin equations (Sec. II A), we outline how the original QSGW approximation (Sec. II C) is modified to include excitonic contributions (Sec. II B). The resulting approach, which is referred to as QS$G\hat{W}$ (where the substitution $W \to \hat{W}$ implies that vertex corrections are included in W), was numerically implemented within an all-electron framework using a linear muffin-tin orbital basis set (Sec. II E) in the QUESTAAL package [24]. Later sections show applications to a broad range of materials.

A. The GWA from the Hedin's equations

The approach described in this work, the standard GW approximation (GWA) and the QSGW, are all derived from the many-body perturbative approach developed by Hedin [3]. In this method, the following set of closed coupled equations [3,25,26] are to be solved iteratively:

$$\Sigma(1,2) = i \int d(34)\, G(1,3^+) W(4,1) \Gamma(3,2,4), \tag{2}$$

$$G(1,2) = G_0(1,2) + \int d(34)\, G_0(1,3) \Sigma(3,4) G(4,2), \tag{3}$$

$$W(1,2) = v(1,2) + \int d(34) v(1,3) P(3,4) W(4,2), \tag{4}$$

$$P(12) = -i \int d(34) G(1,3) G(4,1^+) \Gamma(3,4,2), \tag{5}$$

$$\Gamma(1,2,3) = \delta(1,2)\delta(1,3) + \int d(4567) \frac{\delta \Sigma(1,2)}{\delta G(4,5)} G(4,6) G(7,5) \Gamma(6,7,3), \tag{6}$$

where G is the Green's function, $v(\boldsymbol{r},\boldsymbol{r}') = 1/|\boldsymbol{r}-\boldsymbol{r}'|$ is the bare Coulomb interaction, W is the screened Coulomb interaction, Γ is the irreducible vertex function, P is the irreducible polarizability (the functional derivative of the induced density with respect to the total potential), and Σ is the self-energy operator. In Eqs. (2) and (5), the indices subsume position and time and the $+$ superscript implies $t' = t + \eta$, with $\eta \to 0^+$.

In the standard GWA, also known as one-shot GW or G_0W_0, Eq. (6) is approximated as

$$\Gamma(1,2,3) \approx \delta(1,2)\delta(1,3) \tag{7}$$

in both the expressions for the self-energy [Eq. (2)] and the irreducible polarization [Eq. (5)]. In addition, Σ and P are both evaluated for $G = G_0$, the independent-particle Green's function,[4] that in the frequency domain takes the form

$$G_0(\boldsymbol{r},\boldsymbol{r}',\omega) = \sum_n \frac{\psi_n(\boldsymbol{r})\psi_n^*(\boldsymbol{r}')}{\omega - \varepsilon_n \pm i\eta}. \tag{8}$$

In Eq. (8), ψ_n and ε_n are the single-particle wave functions and energies; the index n contains band, spin, and wave-vector indices and the $+$ ($-$) is for unoccupied (occupied) bands.

The approximation for the irreducible polarization obtained by neglecting the vertex is referred to variously as the independent-particle approximation, the time-dependent Hartree approximation, and the random-phase approximation (RPA) [27]. When using Eq. (8) in frequency space it takes the form

$$P^{\mathrm{RPA}}(\boldsymbol{r},\boldsymbol{r}';\omega) = \sum_{n_1 n_2} \big(f_{n_2} - f_{n_1}\big) \frac{\psi_{n_2}(\boldsymbol{r})\psi_{n_1}^*(\boldsymbol{r})\psi_{n_1}(\boldsymbol{r}')\psi_{n_2}^*(\boldsymbol{r}')}{\varepsilon_{n_2} - \varepsilon_{n_1} - \omega + i\big(f_{n_2} - f_{n_1}\big)\eta}, \tag{9}$$

where f_n are the single-particle occupations.

The Green's function in Eq. (8) can be constructed from the Kohn-Sham electronic structure, which is obtained from the self-consistent solution of Schrödinger-type equations with the single-particle Hamiltonian

$$H_0(\boldsymbol{r}) = -\tfrac{1}{2}\nabla^2 + V_{\mathrm{ext}}(\boldsymbol{r})[\rho] + V_H(\boldsymbol{r})[\rho] + V_{\mathrm{XC}}(\boldsymbol{r})[\rho]. \tag{10}$$

In Eq. (10), $V_{\mathrm{ext}}(\boldsymbol{r})$ is the external potential due to nuclei and external fields, $V_H(\boldsymbol{r})$ is the Hartree potential describing the classical mean-field electron-electron interaction, and $V_{\mathrm{XC}}(\boldsymbol{r})$ is the exchange-correlation potential describing correlation effects missing in $V_H(\boldsymbol{r})$. ρ is the electronic density calculated as $\sum_{\mathrm{occ}} |\psi_n(\boldsymbol{r})|^2$, that is constructed from the eigensolutions of H_0 (from which the self-consistent solution is obtained). Because of V_{XC}, the corresponding single particle G_0 effectively contains many-body effects. Then in Eq. (3) for the many-body Green's function, the self-energy is replaced by $\Delta\Sigma = \Sigma - V_{\mathrm{XC}}$ to avoid double counting. Equation (3) is rewritten as a nonlinear equation for the quasiparticle energies $E_{n\boldsymbol{k}}$:

$$E_{n\boldsymbol{k}} = \varepsilon_{n\boldsymbol{k}} + \langle \psi_{n\boldsymbol{k}} | \Sigma(E_{n\boldsymbol{k}}) - V_{\mathrm{XC}} | \psi_{n\boldsymbol{k}} \rangle, \tag{11}$$

and solved after it is linearized,

$$E_{n\boldsymbol{k}} = \varepsilon_{n\boldsymbol{k}} + Z_{n\boldsymbol{k}} \langle \psi_{n\boldsymbol{k}} | \Sigma(\varepsilon_{n\boldsymbol{k}}) - V_{\mathrm{XC}} | \psi_{n\boldsymbol{k}} \rangle, \tag{12}$$

[4]Equivalent to performing one iteration, starting with $\Sigma = 0$ in Eq. (2).

where the renormalization factor $Z_{n\boldsymbol{k}}$ is

$$Z_{n\boldsymbol{k}} = [1 - \langle\psi_{n\boldsymbol{k}}|\partial\Sigma(\omega = \varepsilon_{n\boldsymbol{k}})/\partial\omega|\psi_{n\boldsymbol{k}}\rangle]^{-1}. \quad (13)$$

B. Ladder diagrams in W

In previous works employing QSGW [9,22,28,29] the RPA was used to make W. This leads to errors noted in the Introduction, e.g., a significant band-gap overestimation. Here we go beyond the RPA and include ladder-diagram corrections in W through the BSE for the polarization [30]. To include the vertex in Hedin's equations we need to determine the interaction kernel $\delta\Sigma/\delta G$ in Eq. (6). Suppose we have adopted the GW approximation for Σ,[5] and assume that $\delta W/\delta G$ is negligible [27], then $\delta\Sigma(12)/\delta G(45) = iW(12)\delta(1,4)\delta(2,5)$, where W is determined in the GW approximation. This expression is then inserted in Eq. (6).

Before we present the BSE for the polarization we will introduce the expansion of the two-point polarization to its four-point counterpart: $P(12) = P(1122) = P(1324)\delta(1,3)\delta(2,4)$. We are now able to present an expression for the polarization that goes beyond the RPA using Eqs (5) and (6) and adopting the expression for the interaction kernel from above:

$$P(12) = P^{\mathrm{RPA}}(12) - \int d(34)P^{\mathrm{RPA}}(1134)W(34)P(3422), \quad (14)$$

where $P^{\mathrm{RPA}}(1234) = -iG(13)G(42)$. The W that appears in the interaction kernel, $\delta\Sigma/\delta G$, is calculated at the level of the RPA and this is usually assumed to be static,[6] i.e., $\delta\Sigma/\delta G = iW^{\mathrm{RPA}}(\omega = 0)$. To avoid confusion with W in Eq. (2) we will refer to the W in Eq. (14) from here on as K.

The Dyson-type equation for the polarizability [Eq. (14)] can be transformed to an eigenproblem for an effective two-particle Hamiltonian by introducing the basis of single-particle eigenfunctions that diagonalize the RPA polarization. Using the completeness of the eigenfunctions, any four-point quantity can be expanded as

$$S(\boldsymbol{r}_1,\boldsymbol{r}_2,\boldsymbol{r}_3,\boldsymbol{r}_4) = \sum_{n_1n_2n_3n_4} S_{n_1n_2n_3n_4} \times \psi^*_{n1}(\boldsymbol{r}_1)\psi_{n_2}(\boldsymbol{r}_2)\psi_{n3}(\boldsymbol{r}_3)\psi^*_{n4}(\boldsymbol{r}_4), \quad (15)$$

where we have again combined band, spin, and wave-vector indices, and $S_{n_1n_2n_3n_4} = \int d(\boldsymbol{r}_1\boldsymbol{r}_2\boldsymbol{r}_3\boldsymbol{r}_4)S(\boldsymbol{r}_1,\boldsymbol{r}_2,\boldsymbol{r}_3,\boldsymbol{r}_4) \times \psi_{n_1}(\boldsymbol{r}_1)\psi^*_{n_2}(\boldsymbol{r}_2)\psi^*_{n_3}(\boldsymbol{r}_3)\psi_{n_4}(\boldsymbol{r}_4)$.

Inserting the expression for the RPA polarization from Eq. (9) in Eq. (14), one arrives at the following expression for the polarization:

$$P_{\substack{n_1n_2\boldsymbol{k}\\n_3n_4\boldsymbol{k}'}}(\boldsymbol{q},\omega) = [H(\boldsymbol{q}) - \omega]^{-1}_{\substack{n_1n_2\boldsymbol{k}\\n_3n_4\boldsymbol{k}'}}\left(f_{n_4\boldsymbol{k}'+\boldsymbol{q}} - f_{n_3\boldsymbol{k}'}\right), \quad (16)$$

whereby the conservation of momentum we have $\boldsymbol{k}_{2(4)} = \boldsymbol{k}_{1(3)} + \boldsymbol{q}$ and[7]

$$H_{\substack{n_1n_2\boldsymbol{k}\\n_3n_4\boldsymbol{k}'}}(\boldsymbol{q}) = (\varepsilon_{n_2\boldsymbol{k}'+\boldsymbol{q}} - \varepsilon_{n_1\boldsymbol{k}'})\delta_{n_1n_3}\delta_{n_2n_4}\delta_{\boldsymbol{k}\boldsymbol{k}'} + (f_{n_2\boldsymbol{k}+\boldsymbol{q}} - f_{n_1\boldsymbol{k}})K_{\substack{n_1n_2\boldsymbol{k}\\n_3n_4\boldsymbol{k}'}}(\boldsymbol{q}),$$

$$K_{\substack{n_1n_2\boldsymbol{k}\\n_3n_4\boldsymbol{k}'}}(\boldsymbol{q}) = \int d\boldsymbol{r}_1d\boldsymbol{r}_2\psi^*_{n_1}(\boldsymbol{r}_1)\psi_{n_3}(\boldsymbol{r}_1) \times W^{\mathrm{RPA}}(\boldsymbol{r}_1,\boldsymbol{r}_2;\omega = 0)\psi_{n_2}(\boldsymbol{r}_2)\psi^*_{n_4}(\boldsymbol{r}_2). \quad (17)$$

The expression $(H - \omega)^{-1}$ can be expressed in the spectral representation as

$$[H(\boldsymbol{q}) - \omega]^{-1}_{ss'} = \sum_{\lambda\lambda'} \frac{A^{\lambda}_s(\boldsymbol{q})N^{-1}_{\lambda,\lambda'}(\boldsymbol{q})A^{*\lambda'}_{s'}(\boldsymbol{q})}{E_\lambda(\boldsymbol{q}) - \omega \pm i\eta}, \quad (18)$$

where $A^{\lambda}_s(\boldsymbol{q})$ is element $s = n_1n_2\boldsymbol{k}$ of the eigenvector of $H(\boldsymbol{q})$ with corresponding eigenvalue $E_\lambda(\boldsymbol{q})$ and $N(\boldsymbol{q})$ is the overlap matrix. When the Tamm-Dancoff approximation (TDA) is adopted [32], H is Hermitian and Eq. (18) reduces to $\sum_\lambda \frac{A^{\lambda}_s(\boldsymbol{q})A^{*\lambda}_{s'}(\boldsymbol{q})}{E_\lambda(\boldsymbol{q})-\omega\pm i\eta}$.

The polarization in Eq. (16)can then be expressed in real space according to Eq. (15) and contracted to its two-point form. This two-point polarization is then used in Eq. (4) to obtain W with ladder diagrams included. In what follows we will denote W^{RPA} with the symbol W, and refer to W with ladders included as $\hat{W}$. The updated W or $\hat{W}$ is then used in the expression for the self-energy with the vertex Γ in the exact self-energy ($iGW\Gamma$) omitted. The justification for omission, and the consequences of it, is taken up in Sec. II D.

In works that report optical absorption $\alpha(\omega)$, we construct it using the relation

$$\alpha(\omega) = \frac{2\omega}{c}\mathrm{Im}\sqrt{\varepsilon(\omega)}, \quad (19)$$

where ε is calculated from the macroscopic part of the dielectric matrix $\varepsilon = 1 - vP$.

C. Self-consistency: QSGW

In the DFT-based G_0W_0 approximation, the $E_{n\boldsymbol{k}}$ are obtained as a first-order correction of the Kohn-Sham single-particle energies. As mentioned in the Introduction, the GWA works when the Kohn-Sham system gives a qualitatively correct description of the physical system, i.e., when the Kohn-Sham single-particle energies are not "too far" from the quasiparticle energies. When this is not the case, the GWA does not give accurate results. It is often improved in practice by choosing some other G_0 constructed, e.g., from a hybrid functional. Another route is to replace the corrected energy E_n [Eq. (12)] either in the Green's function [33] [Eq. (8)], or in the RPA polarization [Eq. (9)] entering the screened potential W, or in both.

[5]The vertex in Eq. (2) can been shown to effectively cancel with the Z factor (see, for example, Appendix A in Ref. [9]).

[6]In few works this approximation has been relaxed (see, e.g., Ref. [31]).

[7]Note that if we are determining the reducible polarizability χ, such that $W = v + v\chi v$, then the kernel becomes $K = W - V$ [27].

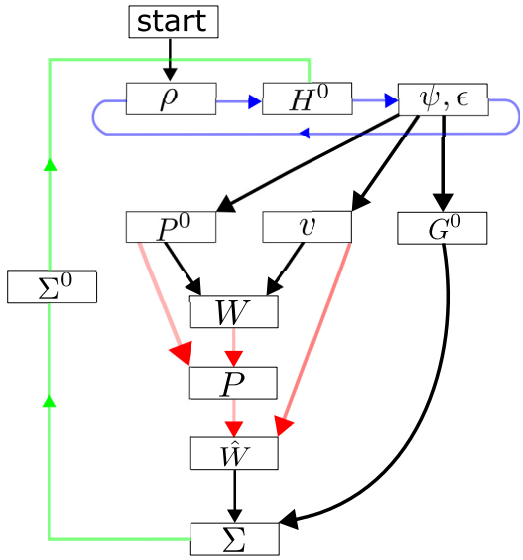


FIG. 1. Flow chart of the QSGW cycle. Noninteracting eigenfunctions and energies (ψ, ϵ) are calculated self-consistently (blue). These are used to construct the noninteracting Green's function G^0, Coulomb interaction v, RPA polarization $P^0 = -iG^0G^0$, and $W = (1 - vP^0)^{-1}v$. W is used to make a vertex and better P via Eq. (14), which gives the improved $\hat{W}$. One cycle makes the static self-energy Σ^0, that is passed to H^0 (green), and the cycle repeated to self-consistency.

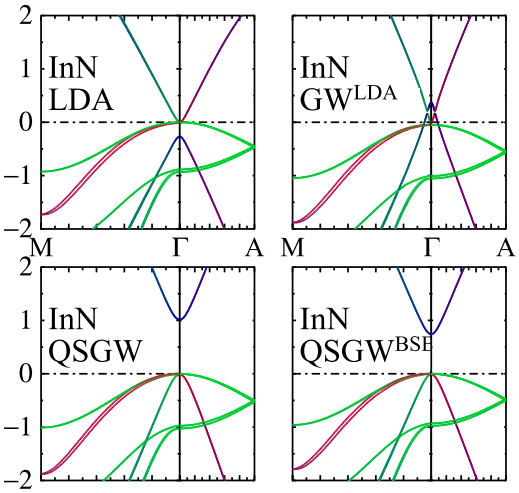


FIG. 2. Low-energy band structure of wurzite InN at four different levels of approximation. Colors depict orbital character of the bands: red for N p_{xy} character, green for N p_z character, blue for In s character. LDA bands are shown in top left panel: the state of In s character at Γ lies below the three states of N p character, reflecting an inverted gap. Upper right panel shows the effect of GW treated perturbatively from the LDA, i.e., Eq. (12), with $Z = 1$ for reasons explained in the text. GW rectifies the inverted gap at Γ, but without off-diagonal parts of Σ it cannot undo the wrong topology given by the starting point, and thus the bands cross near Γ. Bottom left panel is the classic QSGW result, Ref. [9]. It provides a good description of the InN energy bands; however, the gap is overestimated (1.01 eV) relative to experiment (0.6 eV). The QSGW dielectric constant ε_∞ is calculated to be 6.1, about $\frac{3}{4}$ of the measured value (8.4). Bottom right panel is the QSGW result with W augmented by ladder diagrams. The gap (0.74 eV) is slightly larger than experimental one, and differs by approximately the electron phonon interaction (estimated to be 0.07 eV [38]).

Here we use the QSGW approach in which the starting point is chosen to effectively minimize $\Delta\Sigma$: the difference between the dynamical self-energy and (static) quasiparticlized one. In practice, once the self-energy has been calculated within the GW approximation, a new effective single-particle static potential is determined by Eq. (1). Then, by substituting V_{XC} in Eq. (10) with Σ^0 [Eq. (1)], a new set of single-particle energies and wave functions can be determined. In turn, those can be used to recalculate the GW self-energy, and the whole procedure can be repeated until self-consistency in the energies and eigenvalues is achieved. In this procedure the resulting electronic structure does not depend on the quality of the Kohn-Sham DFT electronic structure for the system and, equally important, it removes the arbitrariness in starting point [10]. Figure 1 shows a flow chart of the process.

Why self-consistency is important

Self-consistency is not typically performed in weakly correlated materials. LDA-based GW can do very well (see description of Bi_2Te_3, Sec. III E 2) but self-consistency improves the theory and makes the discrepancies with experiment systematic. Recent work shows this to be the case even for simple sp metals such as Li, Na, and Mg [34] where errors in RPA for W are likely to be less important than in insulating systems. Similarly excellent agreement is found for the Fermi-liquid regime of Fe, in a detailed study examining several properties [35].

Perhaps the first study applying GW to a correlated material was the work of Aryasetiawan and Gunnarsson [33], in which case a starting point better than the LDA becomes essential. This issue arises for many kinds of narrow-band systems, and even in weakly or moderately correlated ones the starting point can be important. Narrow-gap semiconductors in which the LDA has a negative gap offer one notable illustration of this. Using GW in the usual manner [correcting the reference eigenvalues via Eq. (12)] cannot correct the wrong topology of the starting point [36]. InN is a classic example (Fig. 2). Even while the states at the k point Γ have the correct ordering, the improper initial ordering leads to unphysical dispersions in the band structure in the vicinity of Γ. Other systems which fall into this class are Ge, PbTe, InAs, and InSb. (In PbTe, a gap appears at L within the LDA, but with L_6^+ and L_6^- wrongly ordered; see Fig. 13 of Ref. [37]).

Another effect of self-consistency is to modify the one-body part of H or G. This is because not only the eigenvalues but the density is significantly renormalized relative to the LDA. GW induces a corresponding change in the effective potential through the inverse of the susceptibility, $\chi^{-1}(x_1, x_2) = \delta V(x_1)/\delta n(x_2)$. Starting from the perturbation $\delta V^0 = \Sigma^{QSGW}[G^0_{LDA}] - V^{xc}[G^0_{LDA}]$, and if we assume that $\chi^{-1}(x_1, x_2)$ is adequately approximated by the LDA, we can estimate the change in n from $\delta n = \chi^0 \delta V^0$, and from this obtain the attendant screening potential as

$$\delta V^{scr} \approx V^{HXC}[n_{LDA} + \delta n] - V^{HXC}[n_{LDA}]. \qquad (20)$$

TABLE I. Dependence of the band gap on various kinds of treatment in the off-diagonal parts of the self-energy. In all cases the starting Hamiltonian is the LDA. $E_G(\Sigma^{\rm diag})$ is the outcome from a treatment similar to the usual way *GW* is employed [Eq. (12) but with $Z = 1$]. $E_G(n_0)$ adds the full $\Sigma - V_{\rm XC}^{\rm LDA}$ to the LDA Hamiltonian, including the off-diagonal elements, but without updating the density. $E_G(n_0+\delta n)$ is similar to $E_G(n_0)$ but the density is updated in a "small" loop keeping Σ fixed, as described in the text around Eq. (20). QS*GW* is the quasiparticle self-consistent result (QS$G\hat{W}$ result in parentheses). Values reported for Ge, GaSb, and $TiSe_2$ are for the direct gap at Γ, with $TiSe_2$ in the high-temperature $P\bar{3}m1$ phase. In each of these three cases, the valence and conduction band edge states are inverted in the LDA, similar to Fig. 2.

	$E_G(\Sigma^{\rm diag})$	$E_G(n_0)$	$E_G(n_0+\delta n)$	QS*GW*
Ge	1.11	1.10	1.07	1.18 (1.06)
GaSb	0.83	0.88	0.85	1.14 (1.01)
CdO	0.58	0.53	0.63	1.52 (1.18)
ZnO	3.13	3.04	3.15	4.12 (3.61)
CaO	6.92	6.81	6.69	7.61 (7.06)
MnO[a]	1.54	1.55	1.98	3.77 (3.05)
LiF[b]	14.5	14.6	14.7	15.9 (14.6)
MnTe	0.98	0.81	0.89	1.60 (1.36)
$SrTiO_3$[c]	2.54	2.19	1.89	4.56 (4.04)
$TiSe_2$[d]	0.23	0.30	–.37	–.25 (–.25)
CeO_2[e]	5.90	4.92	2.73	4.93 (4.24)
La_2CuO_4[f]	0.05	0.24	0.43	3.09 (1.67)

[a]Section III F 3.
[b]Section III E 1.
[c]Section III E 6.
[d]See Ref. [8].
[e]Section III E 4.
[f]Section III F 6.

Here $V^{\rm HXC}$ is the combined Hartree + (LDA) $V^{\rm XC}$. In practice the QUESTAAL codes execute an operation similar to this in the natural course of self-consistency: an internal loop is performed in the one-body code adding the fixed $\Sigma^{{\rm QS}GW}$ as an external potential and making the density self-consistent. This accelerates convergence to self-consistency, but for the present we use that process to estimate the effect of $\delta V^{\rm scr}$ on the band gap. In Table I we compare the band gaps for *GW* generated from the LDA, in various forms. It compares the usual Eq. (12) (with $Z = 1$), *GW* including the full matrix structure of Σ without updates to the one-body Hamiltonian [39]; an estimate for the change in one-body potential as just described; and finally QS*GW*. A key takeaway is that the off-diagonal parts of Σ are unimportant only in the simplest nearly homogeneous systems, such as Ge. Even in $SrTiO_3$, a simple d^0 transition metal compound, they are significant, modifying the eigenvalues both directly and indirectly through changes in the density. Another important finding is that both direct and indirect contributions vary widely in both magnitude and sign, and indeed the change is often larger than the well-recognized need to account for the electron-phonon contribution [40].

One solution is to perform partial eigenvalue-only self-consistency: i.e., use Eq. (12) in a self-consistent manner by updating the eigenvalues without changing eigenfunctions. There is a simpler way to approximate eigenvalue-only self-consistency by simply omitting the *Z* factor in Eq. (12). This was shown to be rigorously true for a two-level system in the Appendix of Ref. [36]. Eigenvalue-only self-consistency can significantly reduce the discrepancies with experiments, but it cannot resolve the topology problems or the modifications to the density noted above. Further, the off-diagonal parts of Σ can have nontrivial effects on the quasiparticle spectrum, as noted for example in the discussion around Fig. 2. Another solution is to choose a better starting point, e.g., based on an extension of LDA (as, for example, using a hybrid DFT approach [41] or LDA+*U* [42]) or the Coulomb-hole screened exchange approximation (COHSEX) [3]. Since the starting-point dependence can be chosen freely, the theory loses its *ab initio* flavor. This freedom is lost with QS*GW*, and errors that appear better reflect the nature of the approximations made.

D. Motivation for QS$G\hat{W}$

QS*GW* is already a good approximation in many systems, but it is well known that discrepancies with experiment appear. They tend to be very systematic, and mostly related to the RPA approximation to *W*. Band gaps being systematically overestimated, the high-frequency dielectric constant ε_∞ underestimated, and blueshifts in peaks in Im $\varepsilon(\omega)$, all fairly universal with QS*GW*, are connected to the RPA approximation to *W*. It has long been known, starting from independent work in the groups of Louie [43] and of Reining [44], that if the RPA is extended to include ladder diagrams, optical response is significantly improved in simple semiconductors.

Our primary focus here is to determine to what extent ladder diagrams in *W* ameliorate these discrepancies. As we will show here, when *W* is extended to $\hat{W}$ and the cycle carried through to self-consistency, many of the systematic errors in the QS$G W^{\rm RPA}$ self-energy are ameliorated to a remarkable degree for a wide range of weakly and strongly correlated insulators. (We restrict outselves to insulators since that is where ladders are most important [45].) While this is encouraging, some discrepancies remain, and these form a major focus of this paper. A very important feature of QS$G W^{\rm RPA}$ has been that when discrepancies with experiment appear, the origin can often be clearly associated with a particular missing diagram, enabling the possibility for a systematic, hierarchical extension of the theory. We will show that this remains mostly true with $\Sigma = iG\hat{W}$: QS$G\hat{W}$ improves on QS*GW* but systematic errors remain. The following omissions account for many of the shortcomings in results presented in this paper.

1. *Shortcomings in* QS$G\hat{W}$

(1) The electron-phonon interaction is a well-identified contribution to the self-energy and, if lattice vibrations are in the harmonic approximation, consists of two contributions (Fan and Debye Waller terms) [3]. The diagram usually reduces insulating band gaps; it also is needed to capture optical transitions between states of different wave numbers, e.g., in indirect gap semiconductors. For its effect on the index of refraction, see Sec. III D.

(2) Omission of Γ in the exact self-energy $iGW\Gamma$. As noted in the last paragraph of the Introduction, there is a partial renormalization *Z* factor connecting *G* to G^0 (see

Appendix A of Ref. [9]), which we rely on in the QS$G\hat{W}$ approximation. Typically this vertex pushes down all the states in an approximately uniform manner with a minimal effect on the band gap [46]. The effect is more pronounced for a nearly dispersionless d or f state, and when such a state comprises the valence band maximum, the gap is underestimated. Semicore d states in semiconductors such as CdTe and GaSb lie about 0.7 eV above photoemission experiments (Fig. 11). Also, band gaps in materials systems whose valence band consists of a $3d$ state, or a strong admixture of it, tend to be too small (Table IV). An extreme manifestation of this is EuO: the valence band maximum consists of a nearly dispersionless, atomiclike f state, and as a result the QSGW gap is underestimated [47]. From this calculation it was inferred that the Eu $4f$ state should be pushed down by $\sim$0.7 eV, a somewhat larger shift than for a flat $3d$ state (presumably it is even more atomiclike). A shift in corelike d levels of order 0.5 eV was first explicitly demonstrated by Grüneis *et al.*, who introduced a simple first-order vertex into $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ [46]. Very recently Kutepov added a first-order vertex in a somewhat more rigorous manner [48].

As regards this work the most important error seems to occur with systems with shallow corelike levels, particularly when they occur near the valence band maximum. See Sec. III D 2 and also the discussion around Table IV for instances where this neglect is important.

(3) Higher-order diagrams in the polarizability. The interaction kernel W [Eq. (14)] is taken from the RPA and moreover it is assumed to be static. Other diagrams have been considered in a few works, e.g., the second-order screened exchange [49]. This diagram when augmenting the RPA was quite successfully used to predict total energies in chemical systems [6]. We consider only one additional diagram, namely, to use W^{BSE} as the kernel in generating Eq. (14), and note its effects on a few systems (Sec. III C 1).

(4) Inadequate treatment of spin fluctuations. In the theory presented here, the only spin contribution to the self-energy comes from the Fock exchange. We present some spectral properties of correlated antiferromagnetic insulators (Sec. III F), and show that even in such correlated cases, the response in the charge channel seems to be reasonably described. This is likely because, in contrast to spin fluctuations, charge fluctuations sense the long-range Coulomb interaction. The situation may be different when the gap closes or becomes small on the scale of spin excitations ($\lesssim$0.1 eV). In such cases there may be cross coupling between spin and charge channels. Our solution to date has been to augment QSGW with dynamical mean field theory (DMFT). DMFT is a nonperturbative method and exact solutions are possible with, e.g., continuous time quantum Monte Carlo [50] that include all diagrams. However, the vertex is assumed to be local, which is reasonable for spin fluctuations as the vertex is thought to predominantly reside onsite among the correlated orbitals where the fluctuations occur [51]. Indeed, the QSGW^{++} framework ($^{++}$ referring generally to extensions of QSGW), augmented either by ladder diagrams or by DMFT, does seem to have unprecedented predictive power in a number of strongly correlated materials [16,35,52–56]. Yet, there are places where a nonlocal vertex may be important, e.g., to explain the nematic phase of FeSe. Some approaches have been formulated to improve on DMFT, e.g., the "DΓA" approximation, a nonperturbative, semilocal approach [57], but it is extremely demanding in practice.

In addition to the T matrix [51], somewhat more sophisticated low-order diagrams that treat spin fluctuations on the same footing as charge fluctuations have been proposed [58], but this has not been attempted yet in an *ab initio* context. As spin fluctuations tend to be low energy, many channels are possible, so whether a low-order theory is sufficient or not remains an open question. A low-order perturbation theory that could replace DMFT would be very advantageous since DMFT has its own unique set of challenges. We do not consider such cases in this work, but it should be noted that the claim that charge fluctuations are well described already in low order is not universal [59] and whether a low-order perturbative theory can be sufficient remains an open question in low-density and strongly correlated metals.

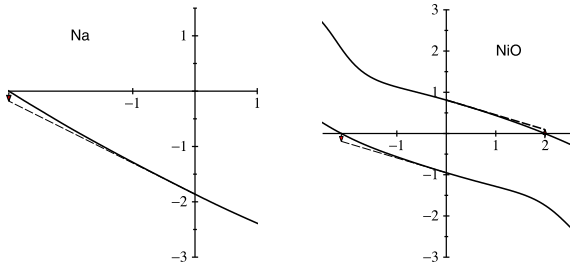
Perhaps surprisingly, this obvious deficiency does not seem to play a significant role in the systems we study here. This work considers only systems with band gaps, and the likely explanation is that spin-wave frequencies are typically small energy compared to the optical gaps, which suppress spin fluctuations.

Other discrepancies with experiment in zinc-blende semiconductors will be presented that do not appear to have a simple interpretation. Perhaps the most notable unexplained error are the errors in the k-dependent dispersion of the conduction band minimum in zinc-blende semiconductors (Sec. III C). Such systems are weakly correlated and the origin of the error is not readily explained. One significant possibility is that QUESTAAL's present implementation does not include a scheme to make the basis set truly complete [60]. This would not be a limitation of the theory itself, but in its implementation.

Finally, several outliers are noted often because the distinction between optical gaps and fundamental gaps is ignored, e.g., in ScN (Sec. III E 3), $SrTiO_3$ (Sec. III E 6), TiO_2 (Sec. III E 5), and $CuAlO_2$ (Sec. III E 7), or are likely artifacts of inaccurate experiments, e.g., hBN (see discussion around Table VII), and in correlated systems where the experiments are less reliable. FeO seems to be an extreme example of this (Sec. III F 4). The connection between discrepancies in one-particle properties and those in two-particle properties is discussed in Sec. III D 2.

The method is self-interaction free, however, so-called self-screening remains. As discussed in Ref. [61] the inclusion of exchange diagrams in the polarization (electron-hole interactions) remedies the self-polarization contribution.[8] The remaining self-screening contribution can be removed with the inclusion of the second-order exchange self-energy [61], incorporated with the inclusion of the vertex in the expression

[8]The inclusion of excitonic effects reduces self-polarization greatly with respect to methods that adopt the RPA, however, if we were to use the bare Coulomb interaction for the BSE kernel in Eq. (17), then we would have perfect cancellation between the direct and exchange terms in the polarization.



3.16[a]	2.90[b]	2.96[c]	2.78[d]

[a] LQSGW, Ref. [210]
[b] QSGW in LAPW basis Ref. [34]
[c] QSGW, this work and Ref. [22]
[d] Experimental result of Ref. [62]

FIG. 3. (Left) Dynamical self-energy $\Sigma(\omega)-\Sigma(\omega_{\rm QP})$ of the lowest valence band at the Γ point in Na, as function of ω (in eV). (Right) The same for the highest valence band and lowest conduction band at the X point in NiO. In all cases the QP level corresponds to the energy where $\Sigma(\omega)-\Sigma(\omega_{\rm QP})$ crosses zero (-2.96 eV in Na, -2.12 eV and $+1.99$ eV in NiO). The arrows indicate the potential difference between LQSGW and QSGW for a particular state, at the QSGW QP energy. Table: some values for the Na bandwidth calculated from different variants of QSGW, and the bandwidth from a recent experiment (in eV).

for the self-energy, $iGW\Gamma$. The self-screening effects that remain in this method will be more pronounced in systems with highly localized and correlated d and f orbitals, such as the systems investigated in Sec. III F.

2. Relation between LQSGW and QSGW

We noted earlier that Kutepov constructed both a quasiparticlized scheme and a fully self-consistent one. His quasiparticlized scheme (LQSGW) is somewhat different from QSGW. They are similar, but with the extensions to RPA presented here and in his work, the fidelity becomes high enough that the difference can be significant. To show this we present here a brief analysis of the relation between LQSGW and QSGW.

Kutepov quasiparticlizes the self-energy with $\Sigma(\omega=0)$, but folds in an effective energy dependence through $\Sigma'(\omega=0)$ (LQSGW) while preserving the ability to construct a static Hamiltonian. The Appendix derives a rough estimate for the expected difference in QP levels between LQSGW and QSGW, obtaining leading contribution from the omitted quadratic term [Eq. (A2)].

In the QSGW scheme, the diagonal element of static (quasiparticlized) Σ^0_{nn} is by construction equal to dynamical self-energy at the quasiparticle level $\omega_{\rm qp}$, so $\Sigma_{nn}(\omega)=\Sigma^{\rm QSGW}_{nn}$ at $\omega=\omega_{\rm qp}$. For LQSGW this is no longer true; thus, the quasiparticle levels of the static G_0 do not coincide with the poles of G. We can estimate the difference between the LQSGW and QSGW QP levels from the difference between $\Sigma_{nn}(\omega_{\rm qp})$ and the linear approximation to it, $\Sigma_{nn}(\omega=0)+\omega_{\rm qp}\Sigma'_{nn}(\omega=0)$. This is depicted in Fig. 3 for the first band in Na, and the highest valence band and lowest conduction band in NiO. In Na, it corresponds to the QSGW-LQSGW change in bandwidth; in NiO, the change is the QSGW-LQSGW difference in the direct gap at X.

For the Na case, according to the simple perturbative expression [Eq. (A2)], LQSGW and QSGW should differ by 0.11 eV. A better estimate is the difference noted in the previous paragraph. The graphs of Fig. 3 indicate that the LQSGW bandwidth in Na should be slightly larger than QSGW, and the NiO band gap also slightly larger. Numerically, the difference in self-energies in the Na case, at the QSGW QP energy, is 0.17 eV. According to first-order perturbation theory, this is the expected difference between LQSGW and QSGW QP energies. Indeed, the discrepancy between LQSGW and QSGW appears to be of this order: one LQSGW and two QSGW calculations have been reported for Na[9] (see table in Fig. 3). 0.17 eV is similar to the spread between QSGW and a recent photoemission measurement [62]; see the table in Fig. 3. As GW is known to break down at sufficiently low densities, an accurate determination of the bandwidth in Na is important since it is one of the best realizations of a nearly homogeneous low-density metal.

3. Hybrid QSGW self-energies

As QUESTAAL has no implementation of the electron-phonon vertex as yet, or the vertex modifying GW, we cannot evaluate its effect *ab initio*. However, by perturbing slightly the QS$G\hat{W}$ self-energy with an admixture of the QSGW Σ when the gap is underestimated or LDA $V^{\rm xc}$ when it is overestimated, we can modify to Σ^h to reach a target band gap without affecting the eigenfunctions too severely. That permits us to assess the effect of the error in E_G on ε_∞. Table VII presents cases where both E_G and ε_∞ are well known, and it establishes that discrepancies in the two are intimately connected for several systems. That provides an independent confirmation that the one-body Hamiltonian would be of high fidelity if this perturbation were properly included.

To make a reasonable proxy to the QS$G\hat{W}$ self-energy, e.g., for the electron-phonon self-energy and missing vertex noted in Sec. II D 1 (points 1 and 2), we will construct a hybrid one-body self-energy Σ^h defined as

$$\Sigma^h=\alpha\Sigma[{\rm QS}G\hat{W}]+\beta\Sigma[{\rm QS}GW]+\gamma V^{\rm LDA}_{\rm XC},$$
$$1=\alpha+\beta+\gamma \qquad (21)$$

This equation has often been used with $\alpha=0$, $\beta=0.8$, $\gamma=0.2$ because ε_∞ computed from QSGW has been found empirically to be very nearly 80% of the true value for a wide range of semiconductors [see Fig. 4(b)]. This formula has been empirically found to yield very good band gaps in many kinds of materials sytems [63,64]. In Sec. III D 2 we use it to show how the errors in Σ (whatever their origin) are closely connected to discrepancies in the dielectric function. A caveat should be noted here: even while the band gap can be rendered accurate with such a hybrid self-energy, ϵ_∞ computed from the RPA

[9] A precise comparison cannot be made because the three implementations presented in the table are all different, and yield slightly different results for ostensibly the same theory. Perhaps the most rigorous implementation of QSGW is the SPEX implementation used in Ref. [34].

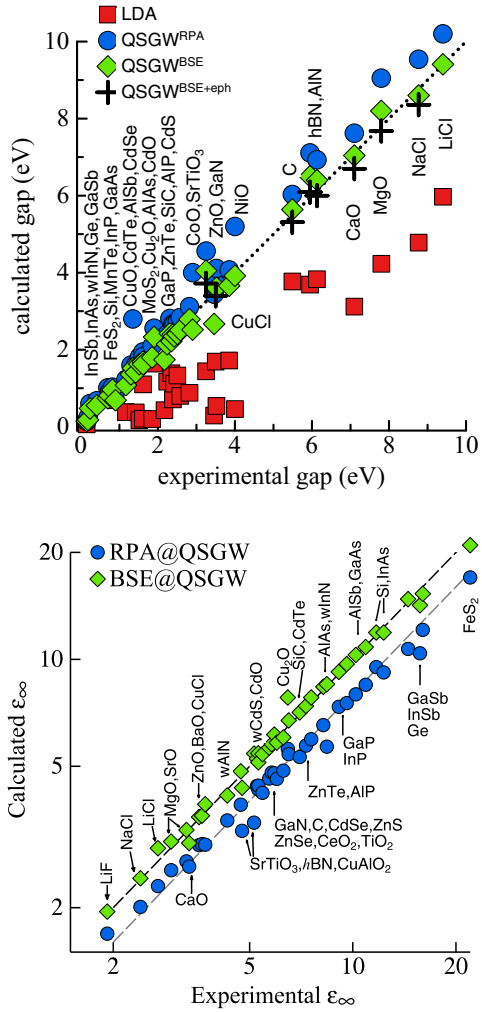


FIG. 4. (Top) Fundamental band gap for selected materials calculated within the LDA (red squares), QSGW (blue circles), and QS$G\hat{W}$ (green diamonds). The black crosses add to QS$G\hat{W}$ an estimate for the gap correction from the electron-phonon interaction when it exceeds 0.19 eV. Where available, this was taken from Ref. [40]; otherwise it was estimated from the Frölich expression, Eq. (24). (Bottom) $\varepsilon_\infty^{\mathrm{RPA}}$ calculated from G_0 generated QSGW (blue circles) and $\varepsilon_\infty^{\mathrm{BSE}}$ calculated from G_0 generated QS$G\hat{W}$ (green diamonds). The dark dashed line corresponds to perfect agreement with experimental data. The light dashed line corresponds to $\varepsilon_\infty^{\mathrm{th}}/\varepsilon_\infty^{\mathrm{expt}} = 0.8$. For hBN, we used ε in the basal plane; see Table VI. Benchmarking ε_∞ in the antiferromagnetic oxides CuO, MnO, FeO, CoO, NiO, and Cu is more complex. They are omitted here but discussed in Sec. III F. For a comparison between two inconsistent approximations, $\epsilon_\infty^{\mathrm{RPA}}$ calculated from G_0 generated QS$G\hat{W}$ and $\epsilon_\infty^{\mathrm{BSE}}$ calculated from G_0 generated QSGW, for a subset of the materials considered here, refer to Fig. 6.

is not similarly improved, so the theory cannot capture both quantities in a consistent manner. Section III B 3 discusses this at greater length.

We will also assume this connection to hold in cases where the fundamental gap is uncertain while ε_∞ is better known. By aligning ε_∞, or the frequency dependent $\varepsilon(\omega)$ with measured data, we can make a reasonable estimate for the fundamental gap. This is done for several systems, e.g., TiO_2 (Sec. III E 5) and FeO (Sec. III F 4).

E. Numerical evaluation of the kernel matrix elements

Our numerical implementation of the BSE relies on a generalization of the linear muffin-tin orbital basis [9,65,66]. The eigenfunctions are expanded in Bloch-summed muffin-tin orbitals in spheres around atom centers. The radial part of the eigenfunctions in these spheres is expanded by numerical solutions of the radial Schrödinger equation. In the region between the spheres, the eigenfunctions are then expanded in either smoothed Hankel functions [66] and/or plane waves. Expanding the interstitial in plane waves, the eigenfunctions are

$$\Psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{R}u} \alpha_{\boldsymbol{R}u}^{\boldsymbol{k}n} \varphi_{\boldsymbol{R}u}^{\boldsymbol{k}}(\boldsymbol{r}) + \sum_{\boldsymbol{G}} \beta_{\boldsymbol{G}}^{\boldsymbol{k}n} P_{\boldsymbol{G}}^{\boldsymbol{k}}(\boldsymbol{r}), \tag{22}$$

where $\boldsymbol{R}$ denotes the atomic site and u is a composite index that contains the angular momentum of the site along with an index that denotes either a numerical solution of the radial Schrödinger equation at some representative energy; its energy derivative (since the energy dependence has been linearized by expanding in a Taylor series about the representative energy [67]); or a local orbital which is a solution at an energy well above or below the representative energy. In GW and the BSE a basis is required that expands the product of eigenfunctions. Expanding the interstitial in plane waves, the product eigenfunctions will also be expanded in plane waves, and within the spheres the basis is expanded by $\varphi_{Ru}(\boldsymbol{r}) \times \varphi_{Ru'}(\boldsymbol{r})$. This mixed product basis (MPB) is denoted $M_I^{\boldsymbol{k}}(\boldsymbol{r})$.

Using the notation in Ref. [9], the kernel K in the MPB is read as

$$K_{\substack{n_1 n_2 \boldsymbol{k} \\ n_3 n_4 \boldsymbol{k}'}}(\boldsymbol{q}) = \sum_{I,J} \left\langle \psi_{n_3,\boldsymbol{k}'} \middle| \psi_{n_1,\boldsymbol{k}} \widetilde{M}_I^{\boldsymbol{k}'-\boldsymbol{k}} \right\rangle W_{IJ}^{\mathrm{RPA}}(\boldsymbol{k}'-\boldsymbol{k};\omega=0) \times \left\langle \widetilde{M}_J^{\boldsymbol{k}'-\boldsymbol{k}} \psi_{n_2,\boldsymbol{k}+\boldsymbol{q}} \middle| \psi_{n_4,\boldsymbol{k}'+\boldsymbol{q}} \right\rangle, \tag{23}$$

where the matrix elements and W_{IJ}^{RPA} are calculated as in Ref. [9].

Owing to the huge computational demands of the BSE only a subset of transitions that occur between bands within a selected energy range about the Fermi level are considered. Contributions from transitions not included in the BSE are, however, included at the level of the RPA. To include such contributions, we effectively have a matrix H that is diagonal except for a block corresponding to the coupled transitions discussed above. To calculate the polarization in this case, the RPA contribution from the subset of states that are treated at the level of the BSE are not included in the full P^{RPA} [calculated according to Eq. (32) in Ref. [9]] and the contribution from P^{BSE} is added to P^{RPA}. The corrected polarization is then transformed into the MPB and the dielectric matrix $\epsilon = 1 - vP$, and hence $W = \epsilon^{-1}v$, are thus obtained. The so obtained $W_{IJ}(\boldsymbol{q},\omega)$ is used as in Eq. (34) of Ref. [9] to calculate the correlation part of the self-energy.

F. Divergence of the macroscopic dielectric function at $q = 0$

The macroscopic ($\boldsymbol{G} = \boldsymbol{G}' = 0$) dielectric function (head of the dielectric matrix) is constructed from the divergent bare Coulomb interaction ($4\pi/|\boldsymbol{k}|^2$) and polarization function. Since the dielectric matrix contains a three-dimensional integral over $\boldsymbol{k}$, the dielectric matrix for $\boldsymbol{k} = 0$ itself remains finite but angular dependent, resulting in the dielectric tensor. In this work we employ the offset Γ method [9,68,69], to treat the divergent part of W, where an auxiliary mesh is introduced that is shifted from the original Γ-centered mesh. The averaged macroscopic dielectric function calculated in a small cell around Γ is then used to calculate the macroscopic part of the screened Coulomb interaction for $\boldsymbol{k} \to 0$, as in Ref. [68].

The $\boldsymbol{G} = \boldsymbol{G}' = 0$ component of the irreducible polarizability should vanish at $q = 0$. Owing to numerical errors this is not exactly the case, so to correct for this its value is subtracted from the irreducible polarizability for all q. This adjustment stabilizes the calculations and also improves on the **k** convergence of the polarizability and self-energy. We performed careful checks for the **k** convergence in ε_∞ in the RPA, and found for example in zinc-blende semiconductors an $8 \times 8 \times 8$ mesh was reasonably good, and a $12 \times 12 \times 12$ mesh converged ε_∞ to ∼1% in all cases but the smallest gap semiconductors.

G. Including the Frölich contribution to the band gap

To correct the value for the band gap in this method due to the neglection of electron-phonon interactions we can include an approximation for the contribution from the Frölich contribution to the Fan term, which, in polar insulators such as LiF, should be the dominant part. We include lattice polarization corrections (LPC) using the method outlined in Ref. [70]. The energy shift is determined from

$$\Delta E_{n\boldsymbol{k}} = \frac{e^2}{4a_p}\left(\epsilon_\infty^{-1} - \epsilon_0^{-1}\right), \tag{24}$$

$$a_P = \sqrt{\frac{\hbar}{2\omega_{\mathrm{LO}} m^*}} = a_0\left(\frac{m}{m^*}\frac{e^2}{2a_0\hbar\omega_{\mathrm{LO}}}\right)^{1/2}, \tag{25}$$

where a_P is the polaron length scale, which, in the effective mass approximation is computed from the optical mode phonon frequency ω_{LO} and the effective mass m^*. a_P is different for electrons and holes, and we take an average of the electron and hole contributions, following Ref. [70]. ϵ_∞ is the ion-clamped static (optical) dielectric constant and ϵ_0 contains effects accounting for nuclear relaxations. The values for ϵ_0, ϵ_∞, and ω_{LO} used in this work are taken from Refs. [70,71] for materials discussed there. For many of the systems studied here a more rigorous calculation of the gap shift has been published (Ref. [40]). Where available, we use those results.

For a given shift, we use a proxy Σ^h [Eq. (21)] to estimate the effect on the band structure and ϵ_∞.

H. Effective oscillator model for index of refraction

In Ref. [72] it was established that the frequency-dependent index of refraction of many compounds can be fit reasonably well by a single-oscillators model. The model has the form

$$n^2 - 1 = \frac{E_d E_0}{E_0^2 - (\hbar\omega)^2}, \tag{26}$$

where E_0 is the oscillator energy, and E_d is a measure of the strength of interband optical transitions. Empirically, $(n^2-1)^{-1}$ has been found to be a mostly linear function of $(\hbar\omega)^2$, for a wide range of ionic materials, which lends credence to the model. In some experiments where $n(\omega)$ is tabulated, we use Eq. (26) to extrapolate to $\omega = 0$.

III. RESULTS AND DISCUSSION

A. Computational details

All results have been obtained using QUESTAAL [24]. Table XII contains the relevant parameters used in the calculations. The ℓ cutoff for partial waves inside muffin-tin spheres was set to 4 and an *spdfg-spdf* basis was used in all calculations, except in some lighter systems where the g orbitals were omitted. Local orbitals were also used in some systems as indicated in Table XII. Empty sites were used as placements for additional site-centered Hankel functions (to $\ell_{\max} = 2$ or 3) without augmentation, to improve the basis in systems with large interstitial voids. When calculating the polarization within the RPA, the tetrahedron method is employed for integration over the Brillouin zone [9]. In the BSE implementation, a broadening was applied according to Eq. (18) and set to 0.01 Ha for vertex calculations.

The TDA was also adopted due the huge increase in com·pu·ta·tion required to store, calculate, and diagonalize the non-Hermitian matrix that has twice as many rows and columns as the Hermitian TDA one. We would, however, not expect going beyond the TDA to have too much of an impact on the systems investigated in this work [32]. We did remove the TDA in a few cases, e.g., InSb, ScN, and MgO, and found the effect to be minor, as anticipated (Sec. III C 1).

Treatment of the screened Coulomb interaction

For numerical reasons the QUESTAAL codes compute Fock matrix elements not of the bare Coulomb interaction $1/q^2$ but a slightly screened one, $1/(q^2 + \kappa^2)$. A small value for κ is taken, between 10^{-5} and 10^{-4}. The QS*GW* self-energy is not usually sensitive to the value of κ; however, the dielectric constant ε_∞ can vary by a few percent for κ ranging between 10^{-5} and 10^{-4}. For that reason, we compute ε_∞ for three values of κ between 10^{-5} and 3×10^{-5} and extrapolate to $\kappa = 0$.

Also, to avoid evaluating matrix elements at $q = 0$, we use the offset-Γ method [9], which requires generating the polarizability for small values of q near zero. For ε_∞ we evaluate $\varepsilon(\omega, \mathbf{q})$ for three small finite values of q and extrapolate to $q = 0$. The direction of approach to $q = 0$ gives us the orientation dependence of $\varepsilon(q = 0)$.

Both kinds of extrapolations are done in one process. The difference between extrapolated and finite-(q, κ) values can differ by a few percent.

B. Survey of results

We begin with a birds-eye view of some key results. Figure 4 shows the fundamental band gaps (E_G) and high-frequency dielectric constant (ε_∞) for a wide variety of materials, comparing classical QSGW results to QS$G\hat{W}$. This figure elucidates the general trends: QSGW tends to overestimate band gaps and underestimate ε_∞ by an almost universal constant factor 0.8. As anticipated, addition of ladder diagrams ameliorates both of these discrepancies. Apart from a few exceptions (see discussion in Sec. III D 2), QS$G\hat{W}$ greatly improves on QSGW. On the wide scale of the figure the ability of QS$G\hat{W}$ to predict optical properties (Fig. 4) looks stellar, but discrepancies appear on closer inspection. A main theme of this paper is to seek out these deviations, and associate them, where possible, with the missing diagrams noted in Sec. II D 1. Fortunately, most of the discrepancies with measured data are fairly systematic, which opens the possibility that the shortcomings can be rectified with relatively simple low-order diagrams.

1. Index of refraction

Typically, ε_∞ is obtained by extrapolating the frequency-dependent index of refraction $\varepsilon(\omega)$ to zero using, e.g., Eq. (26). Its value is known only to a resolution of a few percent even in the best cases, and the uncertainty is often larger. An extreme case is AlN, where several values have been reported ranging between 3.8 [73] and 4.8 [74], and hBN is another instance (Sec. III D). When several experimental values are available for the compounds in Fig. 4, we use an average value. Reported values for ε_∞ for antiferromagnetic transition metal oxides (not shown in Fig. 4) also show variations, and calculations show larger deviations from the average value. They are discussed in Sec. III F.

There is a small indeterminacy on the theory side also. Aside from the extrapolation noted in Sec. III A, care must also be taken to converge the uniform k mesh entering into numerical Brillouin zone integration: for narrow-gap semiconductors pushing this mesh beyond $12 \times 12 \times 12$ divisions was not feasible, leading to a slight tendency to underestimate ε_∞. These approximations lead to an uncertainty of a few percent.

The bulk of the remaining paper focuses on discrepancies where either E_G or ε_∞ are outside the experimental uncertainty, which appear in some systems. One primary aim of this work is to draw a connection between E_G and ε_∞. Generally speaking, in well-characterized systems the discrepancies in E_G and ε_∞ occur largely at the same time: when the gap is accurately described, ε_∞ is also. We believe this to be a significant finding, and it is taken up in Sec. III B 3. Sections III D and III D 2 present cases where E_G deviates the most strongly from experiment. We use hybrid self-energies [Eq. (21)] to correct the gap, to see how the change in ε_∞ tracks it.

2. NiO as archetype system

NiO is an archetype system that exhibits many of the phenomena that are the subject of this work. Figure 5 shows in greater detail how ladder diagrams renormalize the QSGW self-energy in NiO. This manifests as shifts in QS$G\hat{W}$ energy bands and peaks in the density of states (DOS). DOS are

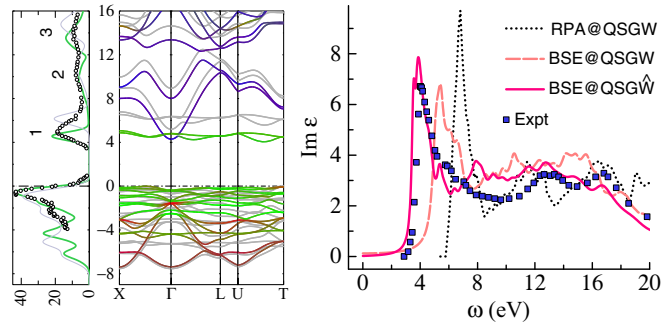


FIG. 5. Middle panel: Energy bands in NiO, within the QS$G\hat{W}$ approximation (colored bands), with valence band maximum at 0. Green projects onto Ni d character, blue onto Ni sp character, and red onto O p character. Shown for comparison are the QSGW results (light gray bands). Left panel: QS$G\hat{W}$ (green) and QSGW (gray) DOS, compared against PES data ($E < E_F$) and BIS data ($E > E_F$) from Ref. [75] (circles). Right: Experimental dielectric function Im $\epsilon(\omega)$ from Ref. [76], compared to results calculated at three levels of approximation: RPA@QSGW, BSE@QSGW, and BSE@QS$G\hat{W}$. As is typical, the shoulder of RPA@QSGW is blueshifted, by roughly 2 eV in this case. Adding ladders (BSE@QSGW) shifts the shoulder towards the experiment, but it is still ∼1 eV too high, as a consequence of overestimate of the QSGW fundamental gap (Table IX). BSE@QS$G\hat{W}$ describes Im $\epsilon(\omega)$ rather well, including peaks around 6, 13.5, and 17 eV. However, the shoulder around 3.5 eV is slightly redshifted compared to experiment, indicating that the fundamental gap is underestimated. ε_∞ is also overestimated. See Sec. III F 1 for more details.

compared to bremsstrahlung-isochromat spectroscopy (BIS) and x-ray photoemission (XPS) measurements in the left panel [77].

(i) BIS data exhibit three peaks between 0 and 9 eV, which the QS$G\hat{W}$ DOS captures quite well, except for a small underestimate of the fundamental gap seen in both BIS and optics (see Sec. III F 1). This shows that ladders do an excellent job of capturing the frequency dependence of the local (k-integrated) spectral function.

(ii) The corresponding QSGW peaks are blueshifted relative to experiment, but in varying amounts. Peak 1, which is composed almost entirely of flat Ni d states, is shifted about 1.5 eV while peak 2, derived essentially of dispersive Ni sp states, is shifted by about half of that. This reflects a universal tendency: flat bands are affected by ladders more than dispersive ones. Fe_3O_4 offers a particularly striking example (Sec. III F 7).

(iii) QS$G\hat{W}$ significantly narrows the occupied Ni d bands relative to QSGW. Red bands (depicting O character) are almost unaffected, while there is a significant narrowing of the green bands relative to QSGW. This is a potentially important finding. It is well known that the LDA severely overestimates d band widths in narrow-band transition metal compounds. Further, it has been shown in several works, e.g., Ref. [78], that QSGW narrows d bands relative to the LDA, but nevertheless continues to overestimate these bandwidths, especially in systems with strong spin fluctuations such as $BaFe_2As_2$ and FeSe. In cases we have studied where experimental information is also available, e.g., in Sr_2RuO_4 [56],

this overestimate is remedied very well by augmenting QSGW with DMFT, which includes vertices in both charge and spin channels. Whether the bandwidth can be captured entirely by a combination of low-order diagrams in both spin and charge channels remains an intriguing possibility. To the extent it is true, this greatly simplifies the complexity of the electronic structure problem in correlated systems. This will be explored in a future work.

Some more details for NiO are presented in Sec. III F 1. Also, there are some strong parallels with La_2CuO_4; see Sec. III F 6.

3. Consistency between one- and two-particle properties

The consistency between benchmarks for one- and two-particle quantities (E_G and ε_∞ in Fig. 4) is striking. Apart from some outliers to be discussed in Sec. III D, the calculated values ε_∞ agree with measured ones to within the available resolution. When this is not the case, usually there is a corresponding discrepancy in the fundamental gap: discrepancies in E_G and ε_∞ occur largely at the same time: overestimate of E_G yields underestimate of ε_∞, and underestimate of E_G yields overestimate of ε_∞.

The internal consistency between one- and two-particle properties is a signature of consistency of the theory since the same quantities (G and W) construct both $\varepsilon(\omega)$ and the potential $\Sigma(\omega)$ that makes G.

If we assume the fidelity of the theory is sufficient for this principle to be universally applicable, the extra information provides an ansatz to predict optical properties in materials with stronger correlations, where benchmarking is less simple. In such cases there is often a large uncertainty in the benchmark itself, not only owing to a wide variation in reported experimental data, but also the extraction from one-particle properties (e.g., fundamental gap) from two-particle response functions. This is reasonable for tetrahedral semiconductors where excitonic effects are small (see Fig. 9), but has less validity in general. For these more correlated cases our approach will be to compare optical experiments directly with calculated response functions. Combining such a comparison with the observed relation between calculated one- and two-particle properties, we can benchmark the theory, and sometimes provide values of quasiparticle levels where not well known, or new interpretation of accepted values.

Figure 4 was generated from two consistent approximations: $\varepsilon_\infty^{\mathrm{RPA}}$@QS$GW$ and $\varepsilon_\infty^{\mathrm{BSE}}$@QS$G\hat{W}$. Consider by contrast two inconsistent approximations $\varepsilon_\infty^{\mathrm{BSE}}$@QS$GW$ and $\varepsilon_\infty^{\mathrm{RPA}}$@QS$G\hat{W}$ (Fig. 6). Both of these approximations show more randomness than either $\varepsilon_\infty^{\mathrm{RPA}}$@QS$GW$ or $\varepsilon_\infty^{\mathrm{BSE}}$@QS$G\hat{W}$ of Fig. 4. Yet, there is a strong similarity between the green diamonds and the blue circles in the two figures. The green diamonds in Fig. 6 fall slightly below the ideal line, showing a modest but non-negligible effect of improving the reference G_0. The blue circles rise slightly above the 80% showing that the RPA continues to underestimate ε_∞, even with a nearly ideal reference G_0. This affirms that most, but not all, of the underestimate of ε_∞ originates from the RPA itself.

This sheds light on the commonly observed fact that $\varepsilon_\infty^{\mathrm{RPA}}$, when computed from the G_{LDA}, often provides a rather good

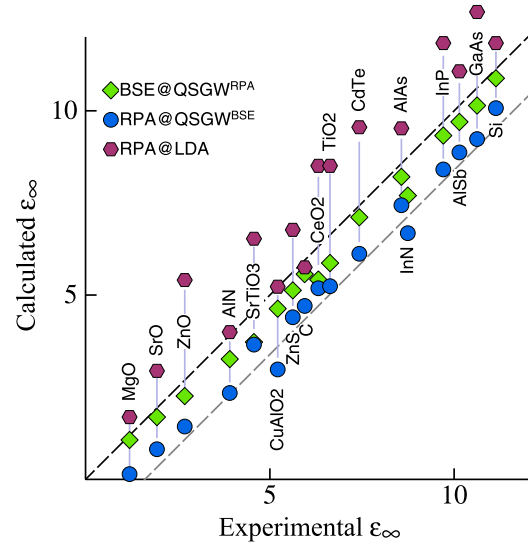


FIG. 6. ε_∞ generated by two inconsistent approximations: $\varepsilon_\infty^{\mathrm{RPA}}$@QS$G\hat{W}$ (blue circles) and $\varepsilon_\infty^{\mathrm{BSE}}$@QS$GW$ (green diamonds), to be compared against Fig. 4. Also shown are LDA results: $\varepsilon_\infty^{\mathrm{RPA}}$@LDA (red squares). Agreement is best for indirect gap semiconductors, where gaps for vertical transitions are relatively large.

estimate for $\varepsilon_\infty^{\mathrm{expt}}$, e.g., in sp semiconductors. The obvious, naive reason for this is a fortuitous error cancellation: LDA underestimates band gaps, which tend to overestimate ε_∞, while the RPA's neglect of electron-hole attraction tends to underestimate the screening, and thus tends to underestimate ε_∞. However, there has been some speculation that the good agreement is not accidental, but a consequence of characteristics inherent in the RPA and the LDA. In particular, a recent work [79] asserts that $\varepsilon^{\mathrm{RPA}}$@LDA should be a good approximation for insulators, based on two arguments. First, ladders involve tunneling processes, and are effective at short range but not long range; thus, the long-range screening that predominantly controls $\varepsilon(q=0)$ is well described by the RPA [80,81]. The first argument is rather appealing, and consistent with prior work establishing that the largest corrections to the RPA occur at short distances [80–82]. This argument can be rigorously checked by comparing blue circles in Fig. 6 to green diamonds in Fig. 4: both share the same eigenfunctions generating ε, the only difference being the presence or absence of ladder diagrams. Agreement is fairly good, differing by 15%–20%, which explains in part why QSGW is a good and consistent theory. The argument of van Loon *et al.* [79] is only partially true: even if the vertex part of P^0 is short range, v is long range so $\varepsilon = 1 - vP^0$ can have a long-range contribution from the short-ranged part of P^0. Notably, the difference does not get smaller as the gap becomes large, as Ref. [79] asserted based on the tunneling argument. This is apparently because as the gap closes the screening becomes large, so the long-range contribution from a short-range vertex becomes relatively less important. It is nevertheless striking that the BSE correction is so insensitive to the band gap.

The second argument of Ref. [79] is that the local vertex in insulators is approximately accounted for by using LDA eigenfunctions. The argument is based on a connection

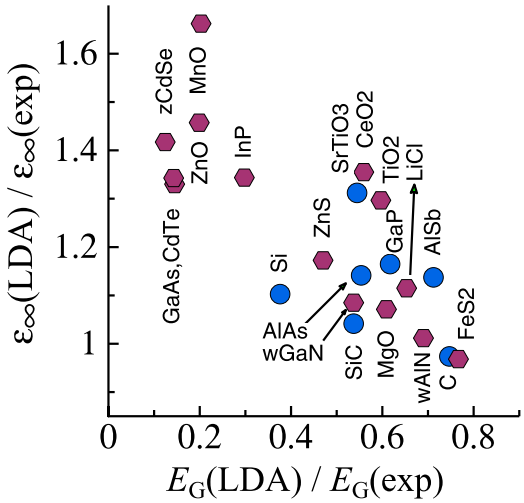


FIG. 7. Ratio $\varepsilon_\infty^{\mathrm{RPA}}$@LDA/$\varepsilon_\infty^{\mathrm{expt}}$, plotted as a function of the LDA relative gap error $E_G^{\mathrm{LDA}}/E_G^{\mathrm{expt}}$. Blue circles are indirect-gap semiconductors; red hexagons are direct gap.

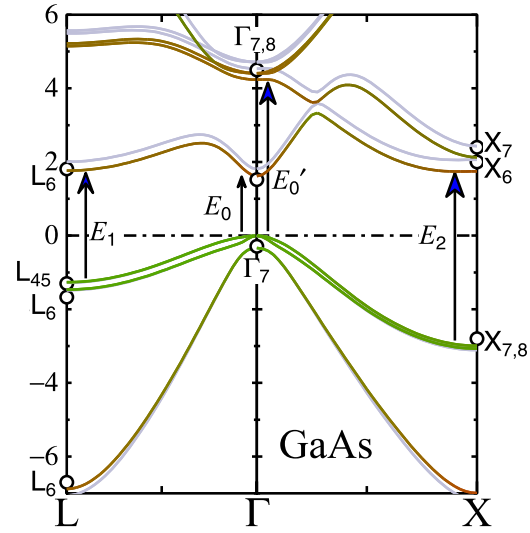


FIG. 8. Energy bands in GaAs, depicting vertical transitions E_0, E_0', E_1, and E_2 that can be measured by ellipsometry. Circles depict measurements of states at high-symmetry points. They have been determined by ARPES [84] to a resolution of about 0.1 eV. E_0, E_0', E_1, and E_2 are reported by Lautenschlager *et al.* in Ref. [85]. Combining these data provides one way to determine levels at X and L in the conduction band. Colored bands are taken from QS$G\hat{W}$ calculations, with red and green showing projections onto Ga and As, respectively. Gray lines show results of QSGW calculations. In the valence, QS$G\hat{W}$ and QSGW are nearly indistinguishable. QS$G\hat{W}$ and QSGW dispersions in the conduction band are very similar, with QSGW slightly higher in energy.

between the LDA derivative discontinuity and the missing vertex, which emerges in a model. To examine this proposition, $\varepsilon_\infty^{\mathrm{RPA}}$@LDA is also presented in Fig. 6 (red hexagons). The second argument is more difficult to assess quantitatively because screening modifies both G_0 and ε_∞, but roughly speaking the difference between blue circles and green diamonds should be similar to the difference between red hexagons and blue circles. One might attribute the poor agreement to inadequacy of the LDA functional (distinct from the derivative discontinuity in the exact functional), but at least in a few systems where it has been tested, the primary gap error has been shown to originate largely from the derivative discontinuity, and not inadequacy of the functional [83].

There is a distinct tendency for LDA to better predict ε_∞ for indirect gap tetrahedral semiconductors than for direct-gap ones: compare diamond, AlAs, AlSb, GaSb in Fig. 6 to ZnO, ZnS, CdTe, InP, and GaAs. Since the derivative discontinuity does not vary wildly between direct- and indirect-gap materials, this is a hint that some other parameter controls the errors in $\varepsilon_\infty^{\mathrm{RPA}}$@LDA. Note also that in the one-oscillator model [72], Sec. II H, the effective oscillator energy E_0 tends to better align with the smallest direct gap than the fundamental one. To disentangle the various effects, Fig. 7 plots the relative error in $\varepsilon_\infty^{\mathrm{RPA}}$@LDA against the relative error in the gap, which is a proxy for the derivative discontinuity. Excepting the d^0 and f^0 systems, the relationship between $E_G^{\mathrm{LDA}}/E_G^{\mathrm{expt}}$ and $\varepsilon_\infty^{\mathrm{RPA}}$@LDA/$\varepsilon_\infty^{\mathrm{expt}}$ is roughly linear. The sensitivity of $\varepsilon_\infty^{\mathrm{RPA}}$@LDA to the derivative discontinuity, together with the tendency of RPA to underestimate band gaps established earlier, provides strong indication that $\varepsilon_\infty^{\mathrm{RPA}}$@LDA yields reasonable $\varepsilon_\infty^{\mathrm{expt}}$ only sometimes, and since it generally produces values larger than experiment while the RPA underestimates the screening as we have shown, there is an additional hidden benefit from fortuitous error cancellation.

C. Benchmarks in weakly correlated semiconductors

The tetrahedrally coordinated sp^3 compounds form a good benchmark for weakly correlated systems in part because they are the best characterized of any family of materials, but also because weak correlations make it possible to well identify transitions between single-particle levels, especially associating peaks in ellipsometry measurements with them. The valence band maximum falls at or very near Γ for all tetrahedrally coordinated semiconductors, which simplifies the analysis. Aside from the lowest Γ-Γ transition E_0, the next Γ-Γ transition E_0' has been measured for some materials. Ellipsometry also measures E_1 and E_2 shown in Fig. 8. E_0 and E_1 are easier to measure accurately because there is larger volume in k space where the valence and conduction bands are parallel. E_2 has been measured for most semiconductors, but its determination is less certain (excepting compounds such as Si, C, and SiC where the global conduction band minimum lies near X). Some data for E_0' are available, but their values are also less well known.

The wider conclusions we draw from the detailed analysis to be described below are as follows.

(1) Band gaps in light (and especially polar) materials are overestimated (MgO, LiF, LiCl, NaCl, TiO_2, $SrTiO_3$, C). The primary cause is the electron-phonon interaction (Sec. III D). A diagrammatic electron-phonon contribution to Σ has long been known [3] though, historically speaking, reliably determining its magnitude has posed a challenge. A fairly high-fidelity calculation of it has recently appeared (Ref. [40]), and we use their results to estimate this term where available. In other cases we make a simple estimate using the Frölich approach of Ref. [71] (Sec. II G). The fundamental

TABLE II. Band gaps at X or L for some zinc-blende semiconductors. All the semiconductors listed above have a global conduction band minimum near X, except for Ge and GaSb. When the electron-phonon interaction is taken into account, gaps at X are systematically underestimated by ∼0.15 eV, while those at L are not.

	E_G (expt)		QS$G\hat{W}$	ZP (est)	QS$G\hat{W}$-ZP
C	5.40	X	5.64	−0.40	5.24
Si	1.17	X	1.08	−0.06	1.02
SiC	2.42	X	2.35	−0.15	2.20
GaP	2.35	X	2.22	−0.09	2.13
AlAs	2.24	X	2.14	−0.04	2.10
Ge	0.74	L	0.81	−0.05	0.76
GaSb	0.88	L	0.91	−0.03	0.88

gaps with these adjustments are shown as black crosses in Fig. 4. See also Sec. III D 2.

(2) The gap in compounds with shallow, nearly dispersionless, d levels are too small (Table IV), and semicore d levels are too shallow (Fig. 11). This is a consequence of the imperfect Z-factor cancellation noted in Sec. II D 1, point 2. To correct it would require the missing vertex Γ in the exact self-energy, $GW\Gamma$. Several instances of this are presented in Sec. III D 2.

(3) k dispersions in the conduction bands of zinc-blende semiconductors show systematic errors of the order ±0.1 eV (see discussion around Tables II and III). There is no obvious diagram that explains this discrepancy.

Figure 9 benchmarks E_0, E_0', E_1, and E_2 transitions in zinc-blende semiconductors where ellipsometry data are available. E_1 shows close agreement, but E_0 and E_2 exhibit discrepancies with distinct patterns: Tables II and III establish that there is a systematic, k-dependent error in the conduction band in zinc-blende semiconductors on the order of 100 meV. The consequences can be significant: note, for example, that QS$G\hat{W}$ predicts GaSb to have a global minimum at L, with $E_\Gamma - E_L = 0.05$ eV, while experimentally at 0 K it is a direct gap, with $E_\Gamma - E_L = -0.09$ eV [89]. Also, where gaps are overestimated, effective masses are too large (Fig. 10).

The **k**-dependent gap error is further discussed in Sec. III C 1, but we can find no obvious explanation for it.

TABLE III. E_0 in narrow-gap zinc-blende semiconductors is overestimated by 0.1 eV. The tendency does not hold for narrow-gap semiconductors that form in other structures, shown as the entries in the second half of the table. An estimate for low-temperature band gap for Ti_2Se_2 (24-atom $P\bar{3}c1$ CDW structure) is taken from Ref. [90].

	E_G (expt)	QS$G\hat{W}$	ZP (est)	QS$G\hat{W}$-ZP
Ge	0.90	1.06	−0.05	1.01
GaSb	0.81	0.96	−0.03	0.93
InAs	0.42	0.53	−0.02	0.51
InSb	0.24	0.42	−0.02	0.40
InN	0.70	0.74	−0.07	0.67
Bi_2Te_3	0.15	0.15		
PbTe	0.19	0.18		
$TiSe_2$	0.15	0.15		

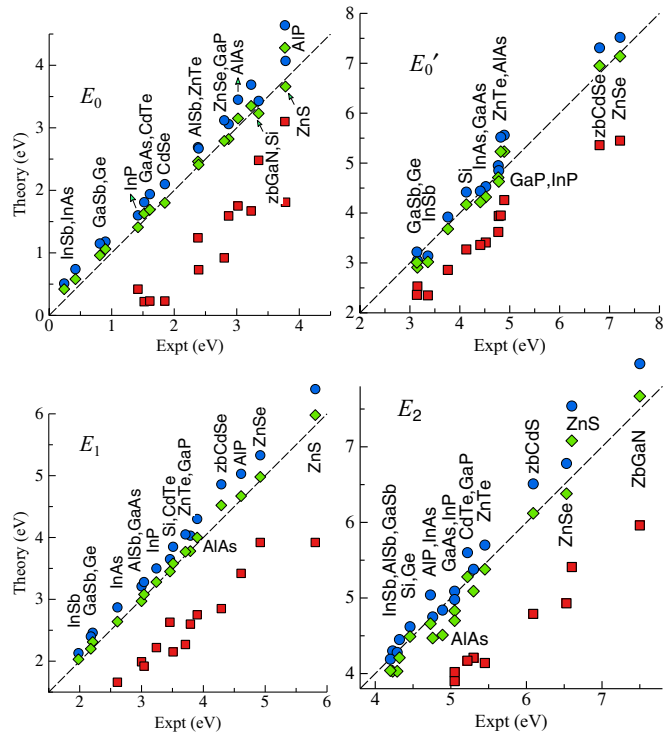


FIG. 9. E_0, E_0', E_1, and E_2 transitions in zinc-blende semiconductors, where experimental data are available. Red squares, blue circles, and green diamonds correspond to LDA, QSGW, and QS$G\hat{W}$ matching Fig. 4.

One possibility is that QUESTAAL's implementation of QSGW contains an error not inherent in QSGW itelf. In particular, the incomplete basis noted by Betzinger *et al.* in the $G^{\rm LDA}W^{\rm LDA}$ context [60] may be a factor. It cannot be ruled out that the nearly perfect agreement for so many systems is a fortuitous artifact of the implementation, or fortuitous cancellation of higher-order diagrams. At all events there is no simple explanation that reconciles these inconsistencies.

Figure 10 benchmarks for effective masses and band gaps in tetrahedral semiconductors. For the direct-gap systems (circles and hexagons), the discrepancy in m^* compared to the experimental value scales approximately in proportion to the discrepancy in E_G (compare to the light dashed gray line). $k\cdot p$ theory predicts a m^* to be proportional to E_G, assuming a fixed matrix element coupling valence and conduction bands, showing that errors in m^* have the same origin as whatever causes the gap to be too large.

	LDA		$(GW)^{\rm LDA}$		QSGW		QS$G\hat{W}$	
	Cation	Anion	Cation	Anion	Cation	Anion	Cation	Anion
Mean	3.1	5.2	1.2	2.0	0.7	1.1	0.9	1.3
RMS	1.5	2.5	0.4	0.6	0.4	0.5	0.4	0.4

Core levels. Figure 11 presents d-core levels at different levels of theory and compares to photoemission results. As is well known, their position is too shallow in the LDA. $(GW)^{\rm LDA}$ improves agreement with experiment, but levels remain too shallow. Self-consistency (QSGW) shows further

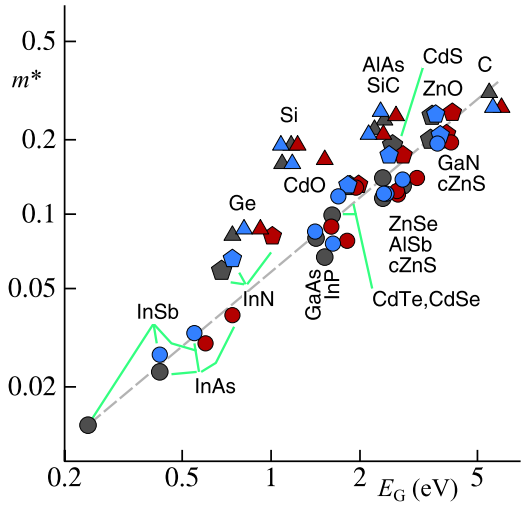


FIG. 10. Effective masses in weakly (and mostly tetrahedrally) coordinated semiconductors. Mass m^* is plotted as a function of band gap E_G. Black denotes experimental data, blue QS$G\hat{W}$ results, red QSGW results. Circles denote direct gaps in zinc-blende structures, pentagons denote direct gaps in wurtzite compounds, and triangles denote indirect gaps. The figure provides two independent kinds of information: the band-gap discrepancy can be seen comparing blue (or red) against black on the abscissa; the mass discrepancy compares the ordinate. The light dashed gray line shows a linear function $m^*(E_g)$. This is the dependence of mass on the gap in $k\cdot p$ theory, assuming the matrix element is fixed. Data for III-V semiconductors are taken from Ref. [86]; other data taken from Adachi's compilation [87]. CdTe, ZnTe, and GaAs are taken from a two-photon magnetoabsorption experiment [88] which is thought to be reliable. AlSb data are for the conduction band at Γ.

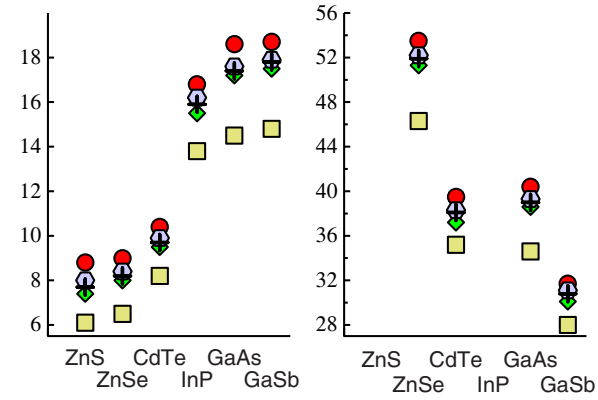


FIG. 11. (Left) Cation d-core levels relative to the valence band maximum, in eV. (Right) Anion d-core levels for systems where they are present. The center of gravity of the $d_{5/2}$ level was taken, except for ZnSe where it was not reported. Red circles are photoemission data taken from Refs. [91–93]. Yellow squares, green diamonds, blue hexagons, and crosses are results from LDA, $(GW)^{\mathrm{LDA}}$, QSGW, and QS$G\hat{W}$, respectively. Calculated anion $d_{5/2}$–$d_{3/2}$ splittings (1.5 eV, CdTe; 0.8 eV, GaAs; 1.3 eV, GaSb) are in close agreement with photoemission data. The table below the figure shows the mean difference with the photoemission data, and the RMS fluctations about the mean, for the different levels of approximation.

improvements. QS$G\hat{W}$ fares slightly worse than QSGW on average, but the consistency improves with the level of theory.

Table IV shows some materials system where the valence band maximum is nearly flat and dispersionless. The band gap is consistently underestimated in QS$G\hat{W}$. (QSGW fares better, but this is an artifact of fortuitous error cancellation).

Shortcomings shown in Fig. 11 and Table IV have a common origin, the missing vertex (Sec. II D 1, point 2). Note that the discrepancy with experiment increases with distance from the Fermi level: from ∼0.5 eV for the valence states near E_F, ∼0.9 eV for cation levels around −10 eV, and ∼1.3 eV for the deeper anion levels.

TABLE IV. Band gaps in systems with valence band maximum formed from a corelike d state. The band gap of FeS_2 is not well known, but the QS$G\hat{W}$ gap lies below the most likely value of 0.9 eV.

	E_G (expt)	QS$G\hat{W}$	QSGW
CuCl	3.46[a]	2.67	3.44
Cu_2O	2.17[b]	1.74	2.27
FeS_2	∼0.9[c]	0.69	0.81
VO_2	∼ 0.7[d]	0.43	0.76
FeO	∼1.1[e]	0.64	1.9

[a]Two-photon absorption [94].
[b]Inferred from interpretation of experiment with effective mass model and exciton observed at 2.0 eV [95].
[c]An estimate from many measurements, from Ref. [96].
[d]See Sec. III E 8.
[e]See Sec. III F 4.

Valence band parameters. The structure of the valence band around Γ provides less reliable benchmarks because of experimental uncertainty in the parameters. Key parameters are the effective masses, and in the wurtzite structure, the crystal-field splitting arising from inequivalence of the z and xy directions. As regards the masses, the matter is considerably complicated by intermixing three states at the valence band maximum near the Γ point and the nontrivial role of spin-orbit coupling that splits the threefold degeneracy at Γ and pushes the band maximum slightly off it.

To encapsulate the many different masses, a Luttinger model is typically used, which has only three independent parameters. The Luttinger parameters can be generated from the following effective masses:

$$\gamma_1 = \frac{1}{2m_{\mathrm{lh}}^{001}} + \frac{1}{2m_{\mathrm{hh}}^{001}},$$
$$\gamma_2 = \frac{1}{4m_{\mathrm{lh}}^{001}} - \frac{1}{4m_{\mathrm{hh}}^{001}},$$
$$\gamma_3 = \frac{1}{4m_{\mathrm{lh}}^{001}} + \frac{1}{4m_{\mathrm{hh}}^{001}} - \frac{1}{2m_{\mathrm{hh}}^{111}}, \qquad (27)$$

where m_{lh} and m_{hh} denote light-hole and heavy-hole masses. Table V shows Luttinger parameters for a few systems where they are best known. The range of values shown in the experimental columns correspond to the range collated from different measurements. In the two cases where the band gap is close to experiment (Si and InP) the calculated Luttinger parameters fall within the range of experimental data. In the

TABLE V. Top: Valence band Luttinger parameters in selected zinc-blende semiconductors. Bottom: Crystal-field splitting parameter in III-N compounds.

		QS$G\hat{W}$			Expt		
	E_G	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
Si[a] |	1.22	4.24	0.32	1.42	4.26-4.29	0.34-0.38	1.45-1.56
Ge[a] |	0.81	10.7	3.00	4.41	13.2-13.4	4.20-4.24	5.56-5.69
AlAs[b] |	2.14	3.99	0.89	1.45	3.42–3.44	0.67–1.23	1.17–1.57
GaAs[c] |	1.63	6.75	1.83	2.74	6.79–7.20	1.90–2.88	2.68–3.05
InP[b] |	1.41	5.58	1.59	2.27	4.61–6.28	0.94–2.08	1.62–2.76

	AlN		GaN		InN	
	E_G	Δ_{cr}	E_G	Δ_{cr}	E_G	Δ_{cr}
QSGW	6.93	−0.224	3.93	0.023	1.02	0.010
QS$G\hat{W}$	6.40	−0.228	3.60	0.021	0.74	0.010
G_0W_0[d]	6.47	−0.295	3.24	0.034	0.69	0.066
Expt	6.13	−0.230	3.5	0.009-0.038	0.67	0.019-0.024

[a]Compilation in Ref. [99].
[b]Compilation in Ref. [86].
[c]Compilation in Ref. [86].
[d]Experimental data and G_0W_0 results taken from Ref. [98].

other two cases (Ge and GaAs) the parameters are underestimated for the same reason the conduction band effective masses are overestimated (see Fig. 10): the direct gap is somewhat overestimated (see, e.g., Table III). In Ge, for example, the conduction band mass at Γ was measured to be 0.037 [97], while our QS$G\hat{W}$ mass is 0.047.

Table V also show shows crystal-field splitting Δ_{cr} in the III-N semiconductors (splitting between states of p_z and p_{xy} character at Γ) in the absence of spin-orbit coupling. The QS$G\hat{W}$ result is within ∼0.01 eV of the measured values, which is quite satisfactory. This quantity is rather sensitive to find details of the potential. To obtain Δ_{cr} reliably, a fine k mesh of 9×9×6 divisions was needed: its value increased by 0.005 eV compared to the standard 6×6×4 mesh. Note that OEP-based GW reported in Ref. [98] yields quite different values for Δ_{cr}.

Two extensions to the theory and their effect on zinc-blende semiconductors

As two possible sources of error on the QP levels of zinc-blende semiconductors, we first considered eliminating the Tamm-Dancoff approximation (TDA). Here we focus on InSb as it has the largest relative gap error. Removing the TDA reduced the QS$G\hat{W}$ E_0 gap by 0.03 eV, considerably less than the discrepancy with experiment. We also considered whether eliminating the TDA improves the k dispersion, in particular, the wrong prediction of the global minimum in GaSb noted above. Removing the TDA reduces the gap in GaSb by 0.03 eV (similar to InSb), but the shift was essentially independent of k and did not rectify this shortcoming. Note that we also consider the effect of removing the TDA when examining the f-sum rules in Sec. III D 3.

We also considered the effect of using a better kernel in the BSE. In all the calculations presented here, we used W^{RPA} for the kernel [Eq. (23)]. It is possible that the dispersion errors in these compounds is a consequence of W^{RPA} being too removed from the exact vertex. We can assess this effect by using a better kernel, namely, BSE W as the kernel for the BSE. If we assume naively that the main effect of BSE is to reduce $W(q=0,\omega=0)$ (i.e., the change in ϵ_∞) then the substitution $W^{\mathrm{RPA}}\to W^{\mathrm{BSE}}$ in Eq. (23) would reduce the strength of the electron-hole attraction and shift the electronic structure to (e.g., the band gap) something intermediate between QSGW and QS$G\hat{W}$. This is roughly what happens in some cases, e.g., CrX_3 [100]. In sp semiconductors, however, using a better W in the vertex causes the gap to decrease still further by a small amount, e.g., by 0.03 eV in InSb. This is another manifestation of vertex corrections being short ranged, as noted earlier.

To conclude, the combined effect of eliminating the TDA and better W in the vertex are not sufficient to explain the tendency to overestimate the direct gap in small-gap zinc-blende semiconductors, or errors in the band dispersion.

D. Response functions in semiconductors

1. Birefringence

Birefringence occurs when the refractive index depends on the polarization and propagation direction of light. It is normally measured as a difference in the principal axes of the ellipsoid's index of refraction, sometimes called the "ordinary" and "extraordinary" indices when there are two inequivalent ones. We consider a few materials where n in the basal plane differs from n normal to it:

$$\bar{n} = (n_\parallel + n_\perp)/2, \quad \Delta n = n_\parallel - n_\perp, \tag{28}$$

$$\bar{\varepsilon} = (\varepsilon_\parallel + \varepsilon_\perp)/2, \quad \Delta\varepsilon = \varepsilon_\parallel - \varepsilon_\perp. \tag{29}$$

Birefringence is measured as the difference Δn. Table VI compares QSGW and QS$G\hat{W}$ predictions against experiment. With the possible exception of hBN, QS$G\hat{W}$ predicts n and

TABLE VI. Birefringence in selected insulators.

	QS$G\hat{W}$		QSGW		Expt	
	$\bar{n}$	Δn	$\bar{n}$	Δn	$\bar{n}$	Δn
ZnO	1.91	0.011	1.77	0.010	1.92	0.012[a]
					1.92	[b]
CdS	2.29	0.015	2.10	0.010	2.30	0.016[c]
TiO_2	2.45	0.22	2.37	0.21	2.50	0.26[d]
					2.55	0.24[e]
*h*BN	1.83	0.49	1.79	0.44	1.89	0.48[f]
					2.12	0.20[g]
AlN	2.04	0.038	1.82	0.048	2.06	0.046[h]
					2.16	0.040[i]

[a]Extrapolated from Ref. [101], using Eq. (26).
[b]Reference [72].
[c]Reference [102].
[d]Reference [103].
[e]Reference [104].
[f]Single crystal, Ref. [105].
[g]Polycrystalline, Ref. [106].
[h]Infrared frequencies, Ref. [107].
[i]Reference [108].

Δn approximately within the available resolution of the experiment.

2. Relation between gap and dielectric function

Figure 4(b) appears to predict ε_∞ very well, but there are discrepancies. Here we focus on systems for which $\varepsilon_\infty^{\rm BSE}$ falls outside the uncertainty of experimental values (estimated by the variation in reported values), and show that these errors directly correlate with errors in the fundamental gap.

Several known potential sources of error in Σ were enumerated in Sec. II D. Among them, the electron phonon interaction is significant for wide-gap, light-element compounds, especially polar ones where the narrow valence band enhances the Frölich interaction (24). The electron-phonon interaction usually reduces gaps, by as much as 0.5 eV in an extreme case such as MgO. Table VII selects some materials where this reduction exceeds 0.3 eV. In such cases ε_∞ is slightly underestimated. As we noted previously, at present QUESTAAL does not have the capability to incorporate the electron-phonon self-energy into the QSGW cycle; however, we can make a proxy by making a hybrid of the LDA and QS$G\hat{W}$ potentials to reduce the gap [Eq. (21)]. We choose $\beta = 0$ and pick the mixing parameter α to approximate the gap change from electron-phonon interaction calculated in Ref. [40]. This should be a reasonable proxy for $\Sigma^{\text{e-ph}}$ since for these systems the LDA and QS$G\hat{W}$ bands differ mostly in a simple rigid shift of the conduction band. Materials in Table VII above the dividing line show systems for which the electron-phonon interaction exceeds 0.3 eV, and where both E_G and ε_∞ are thought to be reliably known. Renormalization causes a modest increase in ε_∞, and the systematic tendency to underestimate it is reduced to approximately the experimental uncertainty.[10]

For compounds in the bottom half of the table, benchmarking becomes murkier because the electronic structure is not known or is poorly understood. *h*BN might have been put in the top half of the table, if so it would present a severe anomaly. Data for *h*BN in Tables VI and VII were computed from an average of Refs. [105,106]. To suggest the possible source of the anomaly, Table VII also shows an entry where experimental data are taken only from Ref. [105], and by such a comparison the agreement is in line with other materials. Further experiments are needed to determine the true values (both ordinary and extraordinary) for ε_∞ in *h*BN.

For less well-characterized systems, if we make the ansatz that the calculated ε_∞ should coincide with the experimental one when E_G also coincides, we can assess the effect of the error in the fundamental gap if ε_∞ is better known (this is a common situation). We can estimate what E_G should be by matching ε_∞ [more generally $\varepsilon(\omega)$] to experiment. In later sections we apply this technique to several materials systems, e.g., $CuAlO_2$ (Sec. III E 7) and FeO (Sec. III F 4).

3. Thomas-Reiche-Kuhn sum rule and the Tamm-Dancoff approximation

The Thomas-Reiche-Kuhn sum rule (f-sum rule)

$$\frac{2}{\pi\omega_p^2}\int_0^\infty \omega'\,\mathrm{Im}[\epsilon(\omega')]\,d\omega' = 1 \tag{30}$$

is readily derived within the RPA, assuming a local, static potential. $\omega_p = 4\pi\, n\, e^2/m$ depends only on the the number of free electrons/volume n, and fundamental constants.

This rule is strongly violated if $\epsilon(\omega)$ is calculated in the RPA from an interacting G [129]; indeed this fact was one of the original motivations for developing a quasiparticle form of self-consistency. The QSGW potential is static but nonlocal, so the position operator does not commute with the potential. Thus, the TRK sum rule is no longer satisfied within the RPA, and the left-hand side of Eq. (30) will typically exceed unity, when Imϵ is computed within the RPA. This can be seen from the special case when a potential is added to a local (e.g., LDA) potential, which is diagonal in the basis of its eigenfunctions. This is the most common way GW is applied (Sec. II C): it modifies eigenvalues but not eigenfunctions. Writing $E^{\rm LDA}$ and $E^{\rm QP}$ as the local (LDA) and QP-corrected eigenvalues, respectively, matrix elements of velocity and momentum operators are related by [130,131]

$$\langle i|\mathbf{v}|j\rangle = \langle i|\mathbf{p}/m|j\rangle\big[E_i^{\rm QP} - E_j^{\rm QP}\big]\big[E_i^{\rm LDA} - E_j^{\rm LDA}\big]^{-1}, \tag{31}$$

where i and j are unoccupied and occupied states, respectively. Since GW tends to widen gaps relative to the LDA, typically $\langle i|\mathbf{v}|j \gg \langle i|\mathbf{p}|j\rangle$ which makes $\mathrm{Im}\epsilon^{\rm QP} > \mathrm{Im}\epsilon^{\rm LDA}$.

[10]MgO is a mild anomaly: QS$G\hat{W}$ already overestimated ε_∞ and the gap reduction worsens the discrepancy to about 7%. Several independent experiments place ε_∞ at 2.95, so the experimental number is likely reliable.

TABLE VII. Estimated change to ε_∞ induced by adjusting the QS$G\hat{W}$ self-energy according to Eq. (21), using α given in the table. $\Delta E_G < 0$ shows the change in fundamental band gap, in eV: $\Delta E_G < 0$ indicates the gap is reduced by taking $\beta = 0$ and $\gamma = 1-\alpha$. $\Delta E_G > 0$ indicates the gap is increased by taking $\gamma = 0$ and $\beta = 1-\alpha$. $\bar{\varepsilon}$ and $\Delta\varepsilon$ are defined in Eq. (28). Column "org" indicates the probable predominant physical origin of ΔE_G: one of e-ph (electron-phonon); Γ (missing vertex in Σ), or * (unknown). Top box displays systems where both E_G and ϵ_∞ are fairly reliably known, or reliably known. Bottom box contains entries where E_G, and to some extent ϵ_∞, are not well known. For CoO and MnO no adjustment was made owing to uncertainty in ϵ_∞, and lack of information about the effect of the electron-phonon interaction.

	QS$G\hat{W}$		$[\alpha \cdot$ QS$G\hat{W} + \Delta\Sigma]$[a]					Expt	
	$\bar{\varepsilon}$	$\Delta\varepsilon$	$\bar{\varepsilon}$	$\Delta\varepsilon$	α	ΔE_G	org	$\bar{\varepsilon}$	$\Delta\varepsilon$
AlN	4.14	0.155	4.33	0.172	0.80[b]	−0.44	e-ph	4.47[c]	0.185
TiO_2	6.03	1.09	6.40	1.20	0.90[b]	−0.37	e-ph	6.39[d]	1.25
$SrTiO_3$	4.84		5.11		0.90[b]	−0.35	e-ph	5.17[e]	
C	5.64		5.82		0.80[b]	−0.31	e-ph	5.70[f]	
CaO	3.04		3.28		0.90[b]	−0.37	e-ph	3.28[g]	
MgO	3.07		3.16		0.85[b]	−0.53	e-ph	2.95[h]	
InSb	14.2		15.5		0.80[i]	−0.23	*	15.7[j]	
CuCl	3.91		3.69		0.0[k]	+0.77	Γ	3.71[l]	
NiO	6.14		5.97		0.9[m]	+0.23	Γ	5.73[n]	
*h*BN	3.42	1.78	3.56	1.97	0.80[b]	−0.48	e-ph	4.08[o]	1.33
*h*BN								3.63[p]	1.82
Cu_2O	7.81		6.80		0.0[q]	+0.53	Γ	6.46[r]	
$CuAlO_2$	5.43		5.22		0.8[s]	+0.18	*	5.13[t]	
CuO	7.86		7.12		0.8	+0.25	Γ	6.5%[u]	
FeO	17.6		12.8		0.7	+0.26	Γ	10.2[v]	
CoO	5.15							5.05[w]	
MnO	4.76							4.95[x]	

[a] α determines $\Delta\Sigma$ from Eq. (21), as described in the figure caption.
[b] To make ΔE_G correspond approximately to shift given in Ref. [40].
[c] Average of Refs. [107,108].
[d] Average of Refs. [103,109].
[e] Average of Refs. [103,104].
[f] Reference [110].
[g] Reference [72].
[h] Reference [111].
[i] So that E_G approximately match known gap, 0.24 eV at 0 K.
[j] Reference [112]. See discussion around Table III.
[k] Adjusted gap approximately matches 3.46-eV gap from two-photon absorption, Ref. [94].
[l] Average of Refs. [113,114].
[m] Adjust gap to average of Refs. [75,115].
[n] Average of Refs. [76,116,117], variation ±0.3. See Sec. III F 1.
[o] Average of Refs. [105,106].
[p] Data from Ref. [105] alone.
[q] QS*GW* gap approximately matches 2.17 eV reported in Ref. [95], and is slightly smaller than the 2.3-eV gap reported by Zimmermann [118].
[r] Reference [119].
[s] Adjust gap to approximately match measured ε_∞.
[t] Average of Refs. [120,121], with variation ±0.15. See Sec. III E 7.
[u] Average of Refs [122,123]. See Sec. III F 5.
[v] Average of Refs [124,125], variation ±0.9. See Sec. III F 4.
[w] Average of Refs. [76,126,127], variation ±0.3. See Sec. III F 2.
[x] Reference [128]. See Sec. III F 3.

It has been shown that the first-order vertex correction to the polarization reduces this violation [132,133]. In Appendix A of Ref. [132] it was shown that Eq. (30) should be satisfied for a particular form of nonlocal potential (self-consistent COHSEX approximation) because terms from the BSE kernel cancel the nonlocal part of the potential. Unfortunately, the self-consistent COHSEX approximation yields poor eigenvalues; but if the eigenvalues are bettered, e.g., with a scissor operator, the cancellation is no longer exact. QS*GW* is not COHSEX, but something akin to it, so we might expect Eq. (30) to be reasonably well satisfied if we generate $\epsilon^{\rm BSE}$ from the QS$G\hat{W}$ hamiltonian.

Figure 12 shows the TRK integral (30) but where the upper limit is taken to be a finite frequency ω, for Si, MnO, and NiO. This shows how the effective number of electrons within a frequency spectrum contribute to the sum. $n_{\rm eff}(\omega)$ is not a

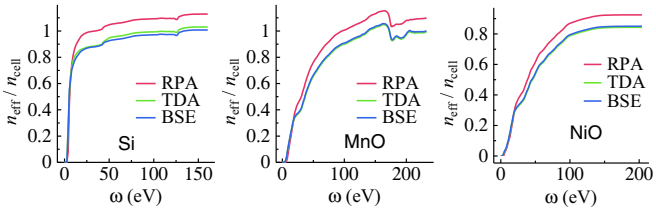


FIG. 12. Frequency-resolved sum rule in Si, MnO, and NiO. Dielectric functions $\epsilon^{\rm RPA}$, $\epsilon^{\rm BSE\text{-}TDA}$, and $\epsilon^{\rm BSE\text{-}full}$ are compared: all are generated from the same QS$G\hat{W}$ Hamiltonian.

strictly monotonic of ω in Si and MnO. This can be attributed to finite cutoff parameters which slightly affect $\varepsilon(\omega)$ at large ω.

We additionally assess the effect of the Tamm-Dancoff approximation (TDA) on the sum rule. We have compared QS$G\hat{W}$ with and without the TDA for a number of materials, and for all the materials we studied, the change in $\epsilon^{\rm BSE}$ was quite small, and this is reflected in small changes to the sum rule.

In all cases the sum rule derived from $\epsilon^{\rm RPA}$, $\epsilon^{\rm BSE\text{-}TDA}$, and $\epsilon^{\rm BSE}$ are compared. For Si and MnO the sum rule slightly exceeds unity for $\epsilon^{\rm RPA}$, and is close to unity for $\epsilon^{\rm BSE}$, entirely consistent with the preceding discussion. For NiO the behavior is similar, but the sum rule seems to be underestimated by about 20%.

We found that the TDA had little effect on the f-sum rules in the systems investigated. This finding is not surprising since the TDA has long been known to have a greater effect in finite systems than in extended systems [32,134]. In theory we believe the QS$G\hat{W}$ method with the TDA should satisfy the f-sum rules reasonably well within the numerical accuracy of the implementation in QUESTAAL, but the dependence of the result on interaction kernel and the use of the TDA both need further investigation.

E. Band structure and dielectric function in selected nonmagnetic materials

In this section we present a variety of selected materials. Where sufficient experimental information is available (e.g., LiF), those results are used to benchmark the theory. For most of the systems presented here, the available experimental information is partial, confused, or contradictory. For these systems we use a mix of theory and what experimental information seems sufficiently reliable, to arrive at a consistent picture where it seems reasonable to do so. In a few cases it is not fully possible (see $CuAlO_2$, Sec. III E 7).

The analysis relies on the ansatz stated in Sec. III D 2, namely, that if G_0 is good enough to well characterize one-particle properties, it also well characterizes two-particle properties provided an adequate theory for the vertex is used and, moreover, that ladder diagrams are sufficient for the vertex. This hypothesis was affirmed in nearly every case in this study where reliable information is available.

1. LiF

The macroscopic dielectric function of the polar insulator LiF was recently calculated in Ref. [29] within the BSE using QSGW as the starting G_0. Since the vertex corrections are omitted in QSGW the screening of the exchange was underscreened and the gap too large. Combined with the neglect of the electron-phonon self-energy, this results in a greatly overestimated band gap of ∼16.2 eV, i.e., about 2 eV larger than the experimental value. The underscreening also caused an overestimation of about 0.5–1 eV of the exciton binding energy. As a result of the partial cancellation of these errors, producing the optical absorption spectrum using the BSE with the QSGW electronic structure results in a blueshift of ∼0.9 eV with respect to experiment.

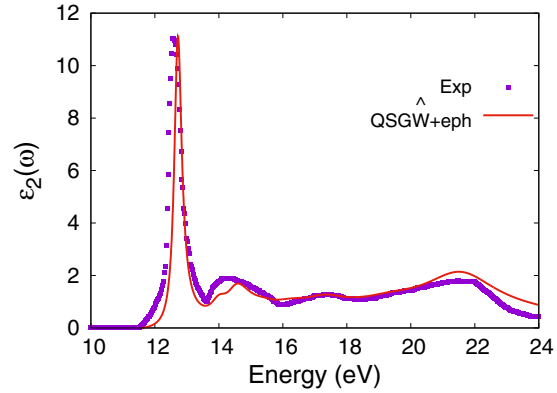


FIG. 13. Imaginary part of the macroscopic dielectric function for LiF. The experimental data (blue squares) [135] are compared with the results from the QS$G\hat{W}$ method. The spectrum is redshifted by 0.483 eV to account for lattice polarization effects (see text). The spectral broadening was increased linearly to match that of the experimental spectrum: the broadening at the first excitonic peak is 0.15 eV and at the X exciton peak (∼22 eV) is 1.25 eV.

Here, we repeat the BSE calculation of the optical spectrum but on top of the QS$G\hat{W}$ electronic structure and also consider the lattice polarization effects. Including ladder-diagram vertex corrections produces a fundamental band gap of 14.7 eV, a reduction of over 1.4 eV, in agreement with the vertex correction calculated in Ref. [18]. Including the 0.48-eV polaron shift correction from Ref. [70] gives then a band gap in excellent agreement with the experimental value of 14.2 ± 0.2 eV [136]. The exciton binding energy is around 2 eV, also in agreement with the experimental value [136] and, as a consequence, Fig. 13 shows an excellent overall agreement between the theoretical and experimental spectra. The BSE ϵ_∞ (1.95) is close to the experimental one (1.96, Ref. [137] and 1.92, Ref. [138]).

2. Bi_2Te_3

Bi_2Te_3 is widely studied because it has topological surface states protected by time-reversal symmetry [139]. It is a narrow-gap system with reported energy gaps between 130 and 170 meV [140–143]. The QSGW and QS$G\hat{W}$ bands are shown in Fig. 14, and are seen to be nearly identical. This system was studied previously [39] within the LDA and $G^{\rm LDA}W^{\rm LDA}$ approximations (albeit including the off-diagonal parts of Σ). Figure 14 is in close agreement with Fig. 1 of Ref. [39]. That the three many-body calculations ($G^{\rm LDA}W^{\rm LDA}$ of Ref. [39], QSGW and QS$G\hat{W}$) are so similar suggests

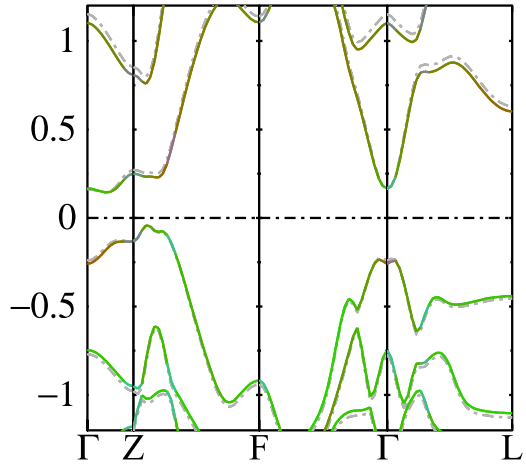


FIG. 14. Energy band structure for Bi_2Te_3 computed by QS$G\hat{W}$ (colored lines) and by QSGW (gray dashed lines).

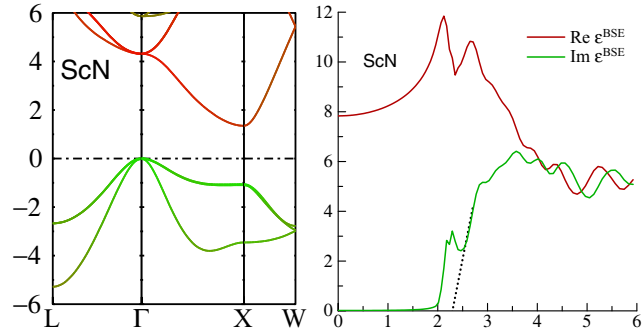


FIG. 15. (Left) Energy bands for ScN. Red and green depict projection onto Sc and N orbitals, respectively. (Right) Real and imaginary parts of the BSE dielectric function. The gray dashed line depicts the RPA dielectric function computed from the QS$G\hat{W}$ self-energy. The dotted line is a guide to the eye, extrapolating the second shoulder in Im ϵ to zero. Table below the figure: Experimental and QS$G\hat{W}$ gaps and values for ϵ_∞.

	E_G	E_G (dir)	ΔE (e-ph)	ε_∞
Expt[a]	0.92	2.07		7.7
QS$G\hat{W}$	1.27	2.33	−0.10[b]	7.8
QS$G\hat{W}$+e-ph	1.17	2.23		

[a]Reference [144].
[b]Based on analogy with InN, as described in the text.

that W is already well described by the LDA. Evidently, ladder diagrams have almost no effect. This is perhaps not surprising since the LDA and GW bands are also similar, with the LDA gap slightly smaller at 50 meV [39]. QSGW and QS$G\hat{W}$ energy gaps are both 145 meV, slightly larger than 120 meV reported in Ref. [39] (presumably because of self-consistency), and within the range of reported experiments [140–143].

3. ScN

ScN is a material of considerable interest in optoelectronics applications, especially as a buffer layer. It is an indirect gap material with the conduction band minimum at X. Its band gap has been controversial with many reported values ranging from 2.03–3.2 eV for the direct gap and 0.9–1.5 eV for the indirect one. Theoretical predictions similarly vary, with predictions ranging between 1.82–2.59 eV (direct) and 0.79–1.70 (indirect) (see Ref. [144] for a summary and detailed discussion).

The most recent and detailed experimental study taking into account prior work (Ref. [144]) yields an optical indirect gap of 0.92 ± 0.05 eV (table in Fig. 15). QS$G\hat{W}$ predicts a larger fundamental gap, 1.27 eV. The latter should be reduced by the electron-phonon interaction. Unfortunately no information is available in the literature, but it is likely to be similar to InN. The longitudinal and transverse mode phonon frequencies are similar, while the electron effective masses in ScN are heavier (0.34 eV for ScN, 0.07 eV for InN). Thus, according to the Frölich formula (24), the electron-phonon renormalization should be larger, by a factor between 1 and 2. A reasonable estimate is 0.1 eV, which is used in the table in Fig. 15.

A study of $\epsilon^{\rm BSE}(\omega)$ yields an exciton at 2.19 eV; thus, there is a spread of 0.14 eV between fundamental and optical gaps. This is apparent in the dielectric function (right panel of Fig. 15). The shoulder between 2 and 2.3 eV is the subgap excitonic transition. The dotted line is a guide to the eye, extrapolating the onset of the second shoulder to zero.

Combining this shift with an (admittedly crude) estimate of 0.1 eV for the electron-phonon interaction brings the QS$G\hat{W}$ and ellipsometry direct optical gaps to within 0.1 eV (approximately the uncertainty in both theory and experiment). ε_∞ also agrees well with Ref. [144] (table in Fig. 15). As $\varepsilon^{\rm BSE}(\omega)$ does not include indirect couplings via the electron-phonon interaction, and we did not consider contribution from $q \neq 0$ transitions, and cannot determine the exciton binding for the indirect transition, but it is reasonable to assume it is similar to the direct one.

To summarize, a consistent picture emerges in close agreement with the recent work of Ref. [144], with the proviso that the one- and two-particle gaps must be distinguished.

4. CeO_2

Electric conduction in CeO_2 takes place both by ionic and electronic conduction, and can be controlled by changing the O_2 pressure. Its unusual electrical properties make it a promising candidate anode in solid oxide fuel cells, or for intermediate-temperature electrolytes [145]. Its energy band gap has been measured optically by a number of groups, by absorption [146,147] or reflectance [145,148,149]. They vary in details, but all report optical band gaps ranging between 3.1 and 3.3 eV. We compare against absorption data in Ref. [146] because the reported energy range was wide enough to show two peaks (Fig. 16).

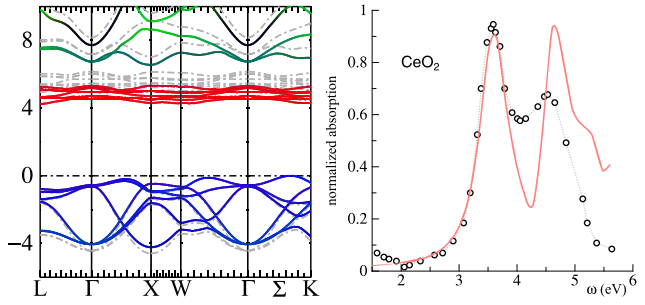


FIG. 16. Left: Energy band structure in CeO_2 computed by QS$G\hat{W}$. Red, green, and blue correspond to Ce-f, Ce-d, and O-p character, respectively. Light gray dashed lines show corresponding QSGW bands. Right: Corresponding dielectric function. Circles are absorbance data digitized from Ref. [146]. Red curve is the BSE absorption α generated from the QS$G\hat{W}$ Hamiltonian ($\alpha = 4\pi n/\lambda$, n^2 = dielectric function). The energy axis for the calculated function is redshifted by 0.5 eV as described in the text.

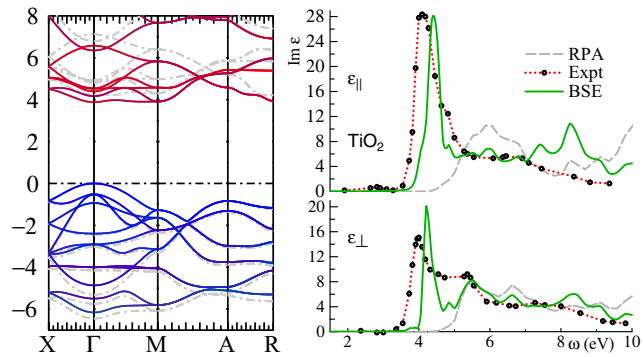


FIG. 17. (Left) Energy bands for TiO_2. Colored bands are taken from QS$G\hat{W}$ calculations, with red and blue showing projections onto Ti d and O, respectively. Dashed gray lines show corresponding QSGW results. (Right) Dielectric function for polarization parallel and perpendicular to the z axis. Circles connected by dotted red lines are data digitized from Ref. [103]. BSE Im $\epsilon_\perp(\omega)$ and Im $\epsilon_\parallel(\omega)$, computed from QS$G\hat{W}$, are shown in green. Light dashed lines compare the RPA results generated from the same QS$G\hat{W}$ Hamiltonian.

The QS$G\hat{W}$ band gap is computed to be 4.24 eV (4.93 eV in QSGW). The valence band is almost pure O-p character, the lowest conduction bands are nearly dispersionless Ce-4f bands (see left panel, Fig. 16). Above the narrow Ce-4f bands are states of mixed Ce-5d and Ce-6s character. LDA bands are not shown, but there is an orbital-selective shift: Ce-4f bands shift about 2.4 eV relative to the LDA, while the Ce-5d bands shift only 0.8 eV. (There is a smaller but orbital-selective shift of the opposite sign as ladders are added to QSGW, as happens for NiO, Fig. 5.) The LDA gap (1.8 eV) is not so far removed from the optical gap (∼3.2 eV), but this is largely fortuitous, for several reasons.

(i) The LDA badly underestimates the shift in empty Ce-4f states.

(ii) There is a large renormalization of the Hartree part of the Hamiltonian, which reduces the gap substantially (Table I), and this partially cancels the first error.

(iii) The optical gap and fundamental gap apparently differ by ∼0.6 eV. Note the absorption spectra in the right panel of Fig. 16. In that figure, the BSE optical spectra were redshifted by 0.5 eV to align them with the absorption data. Thus, the QS$G\hat{W}$ optical gap is ∼3.6 eV, about 0.6 eV less than the fundamental gap. QS$G\hat{W}$ still overestimates the optical gap by ∼0.5 eV (Fig. 16); however, some portion of this difference can be attributed to the electron-phonon interaction, as explained below.

The global valence band maximum falls on the Σ line, about $\frac{2}{3}$ between Γ and K (see Fig. 16), though other local maxima are nearly degenerate with it.

Comparing ϵ_∞ to experiment is not straightforward because of the wide dispersion in reported experimental data, as well as preparation conditions [150] and the crystallinity of the material. Reported values vary from 4.7 [151] to a range between 5.8 and 6.6 [152] on single crystals. A measurement of highly oriented crystalline films yields $\epsilon_\infty = 6.1$ [150]. ϵ_∞ from QS$G\hat{W}$+BSE is found to be 5.8, which fits comfortably within the range of reported experimental values.

An estimate of the zero-point motion can be made using Eq. (24). For this equation, a_P is needed separately for the conduction band and valence band; however, the effective mass approximation is not meaningful for the almost flat conduction band, and we consider only the valence band here. Various experiments put $\omega_{LO} = 30\pm5$ meV [153]; we compute the QS$G\hat{W}$ hole masses to be $(0.86,\ 1.3,\ 1.8)m_0$, for an average mass of 1.27. Using these values, we obtain $a_P = 18.9a_0$. The static dielectric constant is roughly ∼25 [154]. Using $\epsilon_\infty = 6$, Eq. (24) predicts the valence band contribution to gap reduction to be 46 meV. Effective mass theory cannot be applied to the nearly dispersionless conduction band, but it is reasonable to expect its contribution to the total gap reduction to be several times larger. A factor of 3 larger conduction band contribution would give a total gap correction of ∼0.2 eV. This accounts in part, but it would seem not entirely, for the apparent ∼0.3–0.5 eV overestimate of the optical gap predicted by QS$G\hat{W}$. It would seem a discrepancy of order 0.3 eV remains, but a better determination of the electron-phonon interaction is needed to know the discrepancy reliably. Assuming that, for whatever reason, the QS$G\hat{W}$ gap is too large by 0.5 eV, the fundamental gap should be about 3.75 eV.

The RPA dielectric function calculated from the LDA [153] also shows a peak around 3 eV, but this is an artifact of error cancellation: RPA omits strong excitonic effects and the LDA gap is too small.

5. TiO_2

TiO_2, also known as titania, is a widely used commercial compound as food coloring (E number E171) and as a pigment in paint, for example. We consider here only the rutile phase. TiO_2 is a d^0 system with valence band essentially O-2p character and conduction Ti-d character (Fig. 17). The optical absorption edge was measured to be ∼3.3 eV by Cardona and Harbeke from reflection measurements [103], and as 3.03 eV from optical transmission [155]. This is well below the calculated QS$G\hat{W}$ fundamental gap of 3.88 eV. Part of the difference may be attributed to the electron-phonon interaction, which was calculated in Ref. [40] to reduce the gap by 0.34 eV. Reducing the gap by this amount puts it in

line with a PES/BIS study, which reported a fundamental gap of 3.3 ± 0.5 eV [156].

The first peak in $\mathrm{Im}\,\epsilon(\omega)$ is blueshifted about 0.2 eV compared to experiment [103], and correspondingly ϵ_∞ is about 5% smaller than experiment (Table VII). The corresponding QS$G\hat{W}$ birefringence is also slightly underestimated (Table VI): $\Delta n^{\mathrm{BSE}} = 0.22$ compared to $\Delta n^{\mathrm{expt}} = 0.26$. The corresponding RPA values from QSGW are about 80% of experiment, as is typical, with $\Delta n^{\mathrm{RPA}} = 0.16$. Excitonic effects are strong in TiO_2: compare $\mathrm{Im}^{\mathrm{RPA}}\epsilon$ to $\mathrm{Im}^{\mathrm{BSE}}\epsilon$ in Fig. 17 (both were generated from the same Hamiltonian). The BSE redshifts the peak in $\mathrm{Im}\epsilon$ and significantly changes the shape.

Both $\mathrm{Im}_{\parallel}^{\mathrm{BSE}}\epsilon(\omega)$ and $\mathrm{Im}_{\perp}^{\mathrm{BSE}}\epsilon(\omega)$ show reasonable resemblance to the experiment below 6 eV: peak at 4 eV, shoulder at 5 eV in both $\mathrm{Im}\,\varepsilon_{\parallel}$ and $\mathrm{Im}\,\varepsilon_{\perp}$. Three bound, weakly active excitons are found in the region ($E_c-0.45$, $E_c-0.17$ eV) below the fundamental gap, and several bright ones between $E_c-0.13$ and E_c. Here E_c is the conduction band minimum. Thus, $\mathrm{Im}^{\mathrm{BSE}}\epsilon$ shows a strong peak close to the fundamental gap, with a tail extending below.

The electron-phonon interaction missing in QS$G\hat{W}$ well accounts for the blueshift in leading shoulders in $\mathrm{Im}_{\parallel}^{\mathrm{BSE}}\epsilon(\omega)$ and $\mathrm{Im}_{\perp}^{\mathrm{BSE}}\epsilon(\omega)$ relative to experiment. As a proxy to account for it, we repeat the calculation with a hybrid G_0, consisting of 90% QS$G\hat{W}$ and 10% LDA. This reduces the gap by the amount calculated in Ref. [40]. With this shift, ϵ_∞ and the birefringence both align closely to available experiments (Table VII).

Thus, discrepancies in QS$G\hat{W}$ fundamental gap and experimental optical gap are fully explained in terms of a combination of excitonic effects, and the electron-phonon interaction. Adding the latter to the QS$G\hat{W}$ fundamental gap, we conclude that its true value is 3.5 ± 0.1 eV, significantly larger than the widely accepted value of ~3 eV.

6. $SrTiO_3$

$SrTiO_3$ is a perovskite material that may exist in the usual different perovskite phases: cubic, tetragonal, and orthorhombic. Like TiO_2, it is a d^0 compound with valence band essentially O-2p character and conduction Ti-d character (Fig. 18). In this respect it is very similar to TiO_2, and care must be taken in interpreting the experiments to determine the fundamental band gap. The experimental band gap is reported to be 3.25 eV (indirect) and 3.75 eV (direct) [157] and according to Bhandari *et al.* [71] it is almost independent of the structure. According to QS$G\hat{W}$, the system has a indirect gap, of 4.06 eV, with the valence band maximum at R and conduction band minimum at Γ. The direct gap at Γ is 4.51 eV, larger than the indirect one by 0.45 eV; however, the QS$G\hat{W}$ fundamental gap and optical direct gap differ by about 0.75 eV (Fig. 18). Note the BSE code has no electron-phonon coupling and cannot detect indirect transitions.

Peak positions at 5, 9, and 12 eV correspond well to ellipsometry data, though peak amplitudes are different, especially at 6 and 7 eV. There are three reported experimental values for ϵ_∞: the low-frequency index of refraction [109] extrapolated to 0 (Sec. II H) yields $\epsilon_\infty = 4.71$. n between 2.2 and 2.3 (reported in Ref. [157]) yields ϵ_∞ ranging between 4.8 and 5.3; and finally a classic ellipsometry measurement by Cardona

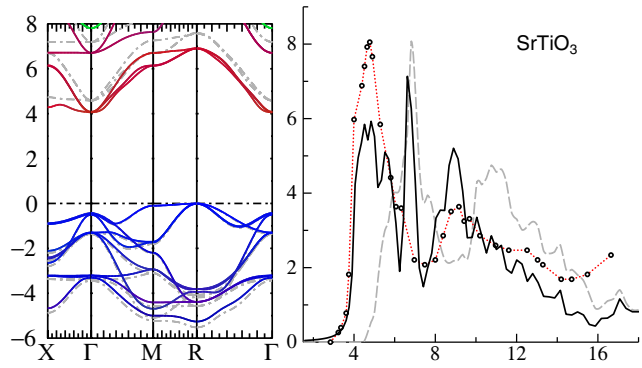


FIG. 18. (Left) Energy bands for $SrTiO_3$. Colored bands taken from QS$G\hat{W}$ calculations, with red and blue showing projections onto Ti 4d and O 2p, respectively. Gray lines show corresponding QSGW results. (Right) Dielectric function. Circles connected by dotted red lines are data digitized from Ref. [103]. The black line shows the $\mathrm{Im}\,\epsilon^{\mathrm{BSE}}(\omega)$ generated from a QS$G\hat{W}$ Hamiltonian on a $10\times10\times10k$ mesh (the waviness is an artifact of incomplete k convergence). The gray line shows $\mathrm{Im}\,\epsilon^{\mathrm{RPA}}(\omega)$ generated from the same Hamiltonian. $\mathrm{Im}\,\epsilon^{\mathrm{RPA}}(\omega)$ vanishes at the fundamental direct gap.

[158] reported $\epsilon_\infty = 5.3$. Thus, it is likely ϵ_∞=5.0±0.3. We find from QS$G\hat{W}$ $\epsilon_\infty = 4.84$ slightly lower than the average, consistent with the band gap being slightly overestimated. According to Ref. [40], the electron-phonon interaction should reduce the gap by 0.33 eV.

The shoulder in QS$G\hat{W}$ dielectric function in Fig. 18 lies about 0.1 eV above the ellipsometry measurement of Ref. [157]. This suggests that the band structure is close to the true one. ϵ_∞ is slightly lower than the average experimental value (Table VIII), which suggests that the uncorrected QS$G\hat{W}$ gap is slightly too large. Using a hybrid functional to reduce the gap, ϵ_∞ moves close to the average experimental value for ϵ_∞ (Table VIII), but the shoulder in $\varepsilon(\omega)$ is slightly redshifted relative to the Benthem data. Thus, there is a slight inconsistency. This excludes a precise determination of the fundamental gap, but we conclude it is 3.75 ± 0.1 eV, which is about 0.5 eV larger than the reported optical gap.

7. $CuAlO_2$

$CuAlO_2$ has received a great deal of attention because it is a p-type transparent conducting oxide (TCO); indeed, it seems to be the only known TCO that can be doped p type.

$CuAlO_2$ was not included in Fig. 4 because reports of its magnitude vary widely. Reports of the lowest (indirect) gap

TABLE VIII. Experimental and QS$G\hat{W}$ gaps and ϵ_∞ in $SrTiO_3$.

	E_G	E_G (dir)	ΔE (e-ph)	ε_∞
Expt[a]	3.25	3.75		5.0 ± 0.3[b]
QS$G\hat{W}$	4.05	4.42	−0.33[c]	4.84
0.9 QS$G\hat{W}$+0.1LDA	3.72			5.11

[a]Reference [157].
[b]Average of Refs. [109,157,158].
[c]Approximately the value in Ref. [40].

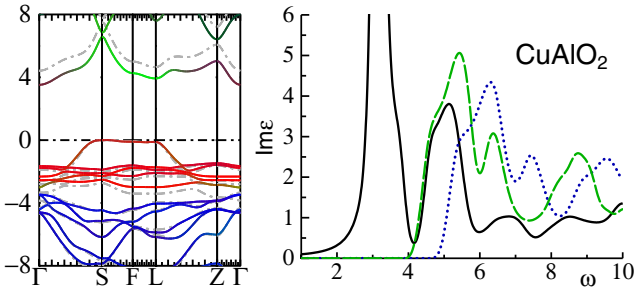


FIG. 19. (Left) Energy bands for $CuAlO_2$. Colored bands are taken from QS$G\hat{W}$ calculations, with red, blue, and green showing projections onto Cu d, O, and Cu sp, respectively. Dashed gray lines show corresponding QSGW results. Points S, F, and L correspond to $(\frac{2}{3}, \frac{1}{3}, 0)$, $(\frac{1}{2}, \frac{1}{2}, 0)$, and $(0, 0, -\frac{1}{2})$, respectively, as multiples of the reciprocal lattice vectors. (Right) Dielectric function (average of x and z directions) as function of frequency ω (eV). Solid line shows $\mathrm{Im}\,\epsilon^{\mathrm{BSE}}(\omega)$, generated from QS$G\hat{W}$. The green dashed line shows $\mathrm{Im}\,\epsilon^{\mathrm{RPA}}$ also generated from the QS$G\hat{W}$ Hamiltonian. Note $\mathrm{Im}\,\epsilon^{\mathrm{BSE}}$ and $\mathrm{Im}\,\epsilon^{\mathrm{RPA}}$ approach 0 near the fundamental direct gap at 4 eV; however, $\mathrm{Im}\,\epsilon^{\mathrm{BSE}}$ has additional subgap peaks from strongly bound excitons. The blue dotted line shows $\mathrm{Im}\,\epsilon^{\mathrm{RPA}}(\omega)$ generated from QSGW. The RPA functions look similar except for a 1-eV shift, because of the difference in band gaps.

range between 1.65 [159,160] and 2.99 eV [161,162], and a direct gap ranging between 3.3 and 4.2 eV [161,163,164]. QS$G\hat{W}$ finds that the valence band is almost exclusively Cu 3d character, with its maximum at a low-symmetry point near S in Fig. 19. The conduction band minimum at Γ is mixed Cu-sd and O-p character. The fundamental (indirect) gap is found to be 3.5 eV, and the lowest direct gap at L is 4.0 eV. There is a large region of k space where the highest valence band is nearly dispersionless. A prior calculation, using a hybrid functional, found similar gaps but not the nearly dispersionless highest valence band [165].

The principal axes for hole mass are along low-symmetry directions, with a moderate mass (1.3) on one principal axis and large masses (2.7, ∼10) on the other two. The conduction band at Γ, by contrast, has much smaller masses and they are along the Cartesian axes (1.3 in xy, 0.41 along z). It has been argued that there should be a large gap renormalization from the electron-phonon interaction [166]; however, the measured difference ϵ_∞ (5.1) and ϵ_0 (7.7) [120] is fairly small, and the eigenstates at the band edges are mostly Cu like instead of O (as it is in MgO), both of which reduce $\Sigma^{\text{e-ph}}$ in the Frölich model (Sec. II G).[11]

The fundamental gap is in line with the photoemission study of Ref. [159], which measures the DOS, a one-particle property. However, when two-particle properties are considered, QS$G\hat{W}$ predicts the situation to be more complicated. The right panel of Fig. 19 compares the BSE and RPA dielectric functions. Both approach 0 at the fundamental direct gap. However, the BSE shows strong peaks below the fundamental gap, around 3.2 eV; there are also several excitons between 3.7 eV and the fundamental gap at 4 eV. Such deep excitons are not typical in sp semiconductors, but it can be understood as an artifact of the nearly dispersionless Cu-like valence band, as well as a relatively small dielectric constant of 5.1. QS$G\hat{W}$ predicts ϵ_∞ rather well. If the strong correlation between the reliability of ϵ_∞ and the band gap (Sec. III B 3) applies equally to $CuAlO_2$, the gap should be close to the QS$G\hat{W}$ prediction of 3.5 eV.

$\epsilon(\omega)$ was computed without an electron-phonon contribution, so it can only measure direct transitions. Presumably there will be other excitons for bound electron-hole pairs coupling Γ and states in the valence band as well; thus, the optical response will show some intensity in a spread below the peak at 3.2 eV, larger than what is shown in Fig. 19, possibly as much as the difference between the direct and indirect gap. Since most determinations of the gap are performed with optical measurements, much of the confusion in the literature likely originates from these deep excitons. These excitons cannot explain a gap as low as 1.8 eV; however, such a gap likely originates from a defect band, which explains why it is not always seen. Indeed, recent work [167] shows that the optical absorption edge is strongly dependent on preparation and postannealing conditions. Defects apparently play an important role, which adds to the confusion about experimental reports on this materials system.

8. VO_2

In the low-temperature monoclinic (M_1) phase, VO_2 has a gap approximately 0.7 eV [168,169]. M_1 is a deformation of the high-symmetry rutile phase. The unit cell, consisting of four V atoms all equal in the rutile phase, dimerize into two pairs with short bond lengths. It is generally agreed that the V dimerization is what is responsible for the gap, splitting the V d manifold into a single occupied d bond per dimer, and a corresponding antibond (Peierls transition). Wentzcovitch *et al.* [170] calculated the energy band structure and suggested that despite the LDA yielding no gap, that the origin of the gap was more Peierls type than Mott type. This picture is further supported by the observation that LDA augmented by single-site DMFT is also metallic [171], which would not happen in a simple Mott description. A cluster form of DMFT added to LDA does yield a gap [172]. This indicates that the nonlocality of the self-energy is essential, and explains why it is too small in the LDA. Gatti *et al.* [173] employed GW to study this system, which captures the nonlocality quite well. While they found $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ failed to open a gap, a self-consistent GW scheme within the COHSEX approximation did so. Counterbalancing this view, a DMFT work [174] argued the M_1 phase should be characterized as the Mott transition in the presence of strong intersite exchange. In our view Gatti's work is the most definitive, as it does not rely on the LDA, partitioning, or adjustable parameters. It also confirms the original Wentzcovitch conjecture: VO_2 is a simple band insulator.

In further support of this conjecture, magnetism appears to play no role in this phase, as we show next. Here we computed the electronic structure of VO_2 in the QSGW and QS$G\hat{W}$ approximations. VO_2 has a nearly dispersionless core-like valence band maximum (Fig. 20). This makes it a prime candidate for the gap to be underestimated, owing to the

[11] It would seem that Ref. [166] did not properly take into account the volume confinement of W in $\mathbf{q}$ space [70].

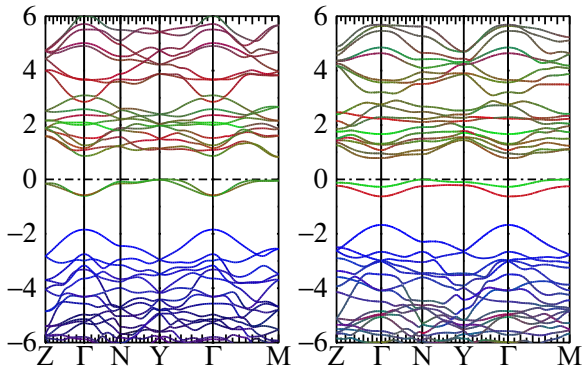


FIG. 20. Left: Energy band structure of VO_2 in the M_1 phase within the QSGW approximation. Red, green, and blue project onto V $d(m=-2,-1,0)$, V$d(m=1,2)$, and O orbitals, respectively. The QS$G\hat{W}$ bands (not shown) are essentially the same but with a 0.3-eV smaller band gap (see text). Right: Corresponding bands of the M_2 phase. In this figure red and green project onto the dimerized and undimerized V d orbitals, respectively.

missing vertex (see discussion around Table IV). Indeed, QSGW predicts a rather good gap ($E_G = 0.76$ eV) owing to a fortuitous error cancellation, while QS$G\hat{W}$ underestimates the gap (0.43 eV), reminiscent of CuCl.

If Mott physics were involved, magnetism should play a role. We find within QSGW that magnetism is totally suppressed in the M_1 phase: attempts to find a magnetic solution always reverted to a nonmagnetic one with self-consistency. The situation is very different in the metastable M_2 phase, where half of the V pairs dimerize and the other half do not. Nonmagnetic QSGW predicts a metallic phase. An insulating phase forms, however, if the system is allowed to be magnetic. To determine the magnetic structure, each of the four V atoms was assigned an arbitrary moment and the system driven self-consistent. We find that the magnetic moment on the dimerized pair vanishes, while spins on the undimerized pair becomes antiferromagnetically aligned with a local moment of $0.8\,\mu_B$, which opens a gap of 0.7 eV.[12] The band structure looks remarkably similar to the M_1 phase, even though the physical basis for the gap is very different. Strikingly, one of the two states forming the upper valence band consists almost purely of dimerized V, while the other is almost purely undimerized V. There is very little hybridization between them, or between V and O.

That the physical basis for gap formation differs in the M_1 and M_2 phases was already pointed out in a comment to the Wentzcovitch paper [175]. Their argument was based on NMR and EPR evidence for low-lying spin excitations in the M_2 phase, which is consistent with this work.

The picture from QS$G\hat{W}$ is similar to that of the DMFT calculation of Ref. [174] for the M_2 phase, but differs for the M_1 phase. It finds a simple Peierls distortion accounts for the known properties, and magnetism plays essentially no role for the latter. Reference [174] argues that the temperature dependence of the band gap is electronic in origin, and uses this as support for the Mott picture; however, QSGW calculations point to phonons playing an important role in controlling the band gap at high temperature, with strong support from experimental data [176]. Figure 3 of that work also presented the conductivity derived from the BSE dielectric function, with QSGW as a reference Hamiltonian. Agreement with ellipsometry data [177] is quite satisfactory.

[12]In the M_2 phase the magnetism is likely disordered. However, a paramagnetic state can maintain essentially the same gap as the antiferromagnetic one, as it does for many antiferromagnetic insulators such as NiO, CoO, and La_2CuO_4.

F. Antiferromagnetic insulating oxides

The monoxide crystal structures MnO, FeO, CoO, and NiO are all of rocksalt form. The magnetic structure consists of sheets of spins antiferromagnetically ordered, which doubles the size of the unit cell. According to the classic paper by Roth [178], the alternating sheets lie in the (111) plane, but the spin orientation depends on the monoxide. MnO and NiO are predicted to be band insulators even within the LDA. In these cases the spin orientation scarcely affects the electronic structure, and we assume the simpler [001] orientation. CoO and FeO are different: LDA predicts both to be metallic. Spins point in the $[\bar{1}\bar{1}7]$ direction in CoO, and perpendicular to the (111) plane in FeO [178]. For these systems we orient the spin quantization axis along these directions and also do not assume time-reversal symmetry.

All of the rocksalt structure oxides have sizable magnetic moments. By contrast, CuO is monoclinic with 8 formula units in the unit cell (Sec. III F 5) and a small local moment (Table IX).

Rödl and Bechstedt modeled the QP band structure of the rocksalt oxides with GW starting from a GGA+U functional [179], and later these authors used the BSE framework to examine the optical response, using a reference potential generated by a $G^{\mathrm{HSE03}}W^{\mathrm{HSE03}}$ functional [180].

In each of these systems (and probably Fe_3O_4, Sec. III F 7), QSGW significantly overestimates the band gap, but not the local moment (Table IX), the difference being more pronounced than in nonmagnetic counterparts. QS$G\hat{W}$ greatly amelioriates this overestimate, sometimes slightly overcorrecting QSGW because of the missing vertex. Precise benchmarking is difficult owing to the large uncertainty in experimental data, especially in the strongly correlated cases. One measure of correlation is the Z factor, Eq. (13). Table IX presents a band- and k-averaged Z factor, namely, the ratio of the interacting to noninteracting spectral functions $A(\omega)/A^0(\omega)$ at an energy just below the Fermi level. The degree of correlation differs in each case so each system is dealt with individually.

1. NiO

Figure 5 shows the (noninteracting) energy bands of NiO, and compares the DOS to BIS data. On the scale of the figure, agreement is excellent. However, QS$G\hat{W}$ apparently slightly underestimates the band gap, which is apparent in both the BIS data and the optics data of Fig. 5. Also, the BSE value for ε_∞, at 6.15, is outside the range of reported values (Table IX). Replacing Σ(QS$G\hat{W}$) with a hybrid $0.9\,\Sigma$(QS$G\hat{W}$)$+0.1\,\Sigma$(QSGW) [Eq. (21)], the gap increases by 0.17 eV, and ε_∞ decreases to 5.97. This is perhaps the

TABLE IX. Band gap; ε_∞; local magnetic moment; band- and k-averaged Z factor in selected antiferromagnetic insulators.

		MnO	NiO	CoO	FeO	CuO	LSCO
	Gap	0.71	0.55				0.01
LDA	ϵ_∞	8.81	32.0				
	m	4.48	1.21				0.27
QSGW	Gap	3.77	5.03	4.00	1.9	2.80	3.09
	ϵ_∞	3.72	4.27	3.87	4.08	4.86	4.04
	m	4.76	1.71	2.73	3.65	0.71	0.64
QS$G\hat{W}$	Gap	3.05	3.23	3.28	0.67	1.52	1.66
	ϵ_∞	4.76	6.15	5.15	17.6	7.85	5.5[a]
	m	4.73	1.67	2.70	3.66	0.66	0.53
	Z	∼0.75	∼0.7	∼0.75	∼0.4	∼0.5[b]	∼0.5[c]
Expt	Gap	3–3.9	4.0,4.3[d]	∼2.6[e]	<1[f]	1.3,1.4[g]	∼2[h]
	ϵ_∞	4.95[i]	5.43–6.0[j]	∼5[k]	9.24–11.1[l]	6.5[m]	∼5[n]
	m	4.79	1.64,1.77[o]	2.47[p]	3.32[q]	0.68[r]	0.64[s]

[a]Average of ϵ_{xx} [Eq. (6.5)] and ϵ_{zz} [Eq. (4.5)].
[b]Strongly orbital dependent: $Z \sim 0.65$ for the (mostly O) valence bands, and ∼0.45 for the (mostly Cu) conduction bands.
[c]Strongly state dependent: $Z \sim 0.65$ for the highest valence bands, and ∼0.4 for lowest conduction band.
[d]References [75,115,118].
[e]Optical absorption edge, Ref. [76]; BIS Refs [181]. See Sec. III F 2.
[f]Reference [118].
[g]Optical absorption edge [151]; PES/BIS and XPS [182,183].
[h]Reflectivity, Ref. [184], optical conductivity, Ref. [53].
[i]Reststrahlen spectrum, Ref. [128]. See Sec. III F 3.
[j]References [76,116,117].
[k]References [76,126,127]. See Sec. III F 2.
[l]Refserences [124,125].
[m]References [122,123].
[n]Reported in Ref. [184]. References [185,186] report anomalously large index of refraction (so that $\epsilon_\infty \sim 25$–50), which is likely connected to excess holes in nominal La_2CuO_4.
[o]Values cited in Ref. [187], taken from Refs. [188] and [189].
[p]Spin moment from Ref. [190]. Orbital moment estimated to be $\sim 1\,\mu_B$.
[q]Reference [178].
[r]References [191,192].
[s]A consensus value of $0.64\mu_B \pm 10\%$ from several sources, Ref. [193].

best-characterized correlated materials system, though even for NiO there is some spread in reported values for both the dielectric function and the fundamental gap. We can conclude that to within this experimental uncertainty, the close connection between gap and ε_∞ (Sec. III D 2) is affirmed.

Some ARPES experiments on this correlated antiferromagnet have been published [194]. Accordingly, we generated the fully dynamical self-energy to compute the k-resolved spectral function and compare to it (Fig. 21). The extent to which a particular band is broadened is strongly band dependent. Agreement with ARPES is satisfactory, in light of the fact that that the matrix element and final-state effects would have to be included for a direct comparison. Quite remarkably, the QS$G\hat{W}$ spectral function looks nearly identical to one generated by LDA+DMFT [195]. The close similarity between these two completely different approaches lends support to the

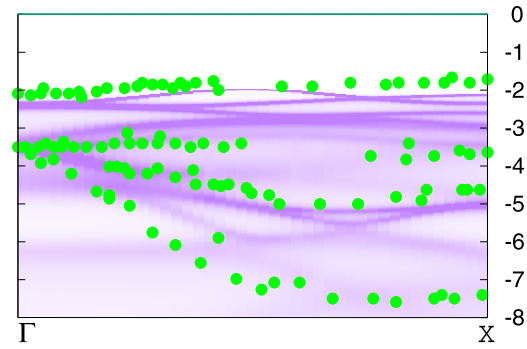


FIG. 21. Spectral function NiO along the Γ-X line, compared against ARPES measurements, Fig. 6 of Ref. [194] (green circles).

thesis that both are characterizing the actual spectral function of NiO.

2. CoO

Our QS$G\hat{W}$ gap is 3.28 eV (see the band structure in the left panel of Fig. 22), and larger than a gap of 2.5 eV measured by a combination of XPS and BIS [181]. Optical gap of similar size (∼2.6 eV) has been observed [76]. These two experimental findings are consistent only if there are no excitonic effects to reduce the gap. Our QS$G\hat{W}$ calculations show, however, that there are a multiplicity of excitons throughout the gap at $q = 0$. The deeper ones (ranging between 0.4 and 1.7 eV) are dark, but strongly active ones at 2.8, 2.96, 3.0, and 3.1 eV also appear. These are likely broadened somewhat, e.g., via some phonon-mediated transitions linking different q, which our calculation does not take into account. Finally, the QS$G\hat{W}$ dielectric constant (5.15) aligns well with the mean value of various experiments (5.43 [126], 5.29 [127], 4.75 [76]). If the consistency between gap and ε_∞ argued in Sec. III B 3 can be relied on, it provides another indication that the QS$G\hat{W}$ fundamental gap is close to the true one.

3. MnO

van Elp *et al.* measured the fundamental gap of MnO by x-ray photoelectron and BIS spectroscopies, and obtained a gap of 3.9 eV [197]. Three kinds of subgap transitions have been recorded by several groups, labeled as A, B, C transitions, and identified with the following symmetries ${}^6A_{1g} \rightarrow {}^4T_{1g}$

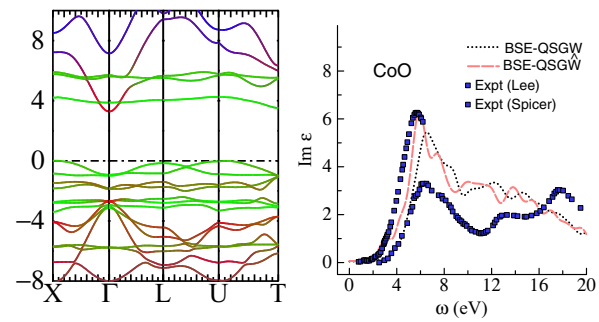


FIG. 22. Left: Energy band structure of CoO. Green and blue represents Co-centered orbitals (green: d, blue: spf), and red O-centered orbitals. Right: Im $\epsilon(\omega)$ measured by Messick *et al.* [196] compared to BSE@QS$G\hat{W}$.

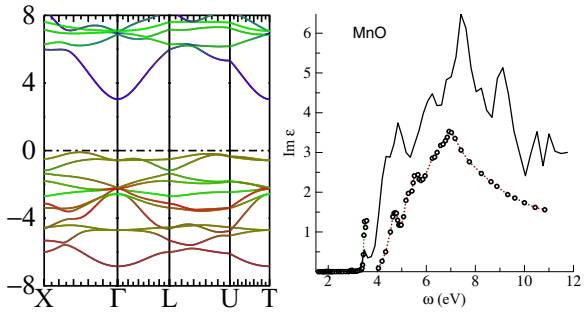


FIG. 23. Left: Energy band structure of MnO. Green and blue represents Mn-centered orbitals (green: d, blue: spf), and red O-centered orbitals. Right: Im $\epsilon(\omega)$ measured by Messick *et al.* [196] (circles connected by red dots) compared to BSE@QS$G\hat{W}$. Also shown at low energy (circles connected by green dots) is a scaled absorption, α^2/ω^2, taken from Ref. [200].

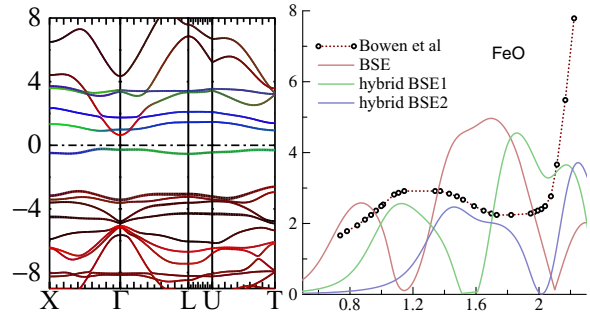


FIG. 24. Left: Energy band structure FeO. Green and blue represents Fe-centered d orbitals (green: m=−2, −1; blue: $m = 0, 1, 2$), and red O-centered orbitals. Right: Absorption $\alpha(\omega)$ measured by Bowen *et al.* [201] compared to BSE@QS$G\hat{W}$ (red) with absorption computed from Eq. (19). Also shown are BSE generated from a hybrid of QS$G\hat{W}$ and QSGW Σ [Eq. (21)], with $\beta = 0.3$ (BSE1) and $\beta = 0.6$ (BSE2) (see Table X).

(A band), ${}^6A_{1g}\rightarrow{}^4T_{2g}$ (B band), ${}^6A_{1g}\rightarrow{}^4A_{1g}+{}^4E_{1g}$ (C band) [198,199]. These transitions are forbidden due to spin and parity selection rules, though significant oscillator strengths have been observed. Huffman *et al.* reported two additional peaks [200], the highest at ∼3.5 eV.

The band structure depicted in the left panel of Fig. 23 is roughly similar to the one depicted by Rödl *et al.* [180]. Unlike NiO, the conduction band is essentially pure Mn s character, the Mn $d(t_{2g})$ appearing at 6–8 eV. The direct gap is at Γ, and is calculated to be 3.6 eV, slightly smaller than the XPS/BIS value (3.9 eV) reported in Ref. [197].

The BSE value for $\varepsilon_\infty^{\rm BSE}$ is slightly smaller than observed in a reststrahlen experiment [128] (Table IX). This suggests an inconsistency with the gap being underestimated; however, the maximum value of Im ϵ derived from n and k presented in that experiment is about an order of magnitude too large, so it is not clear how reliable the measurement is.

We find a dark exciton at 3.07 eV and several bright ones at ∼3.5 eV, which can be seen from the shoulder in Im $\epsilon(\omega)$ below the fundamental gap. These possibly correspond to the highest peaks observed by Huffman *et al.* [200]. We do not find the weak A and B excitons [198,199]; possibly these are associated with a phonon-assisted transition and an electronic part at finite q. The main shoulder in Im $\epsilon(\omega)$ starts to rise about 0.3 eV earlier than the reflectance data of Ref. [196] (Fig. 23). This is consistent with the discrepancy in the XPS/BIS measurement of Ref. [197]. Thus, we tentatively conclude that the QS$G\hat{W}$ gap is ∼0.3 eV too low, though reliable experimental evidence is too limited to draw strong conclusions.

4. FeO

FeO poses one of the most challenging benchmarks in this study. Its highest valence state consists of a single, almost dispersionless d orbital whose m character changes with wave number (Fig. 24). The small-Z factor (Table IX) provides a clear indication that FeO is strongly correlated.

Experimental information about FeO is sparse and somewhat inconsistent. Two values for ϵ_∞ have been reported: 9.24 [125] and 11.9 [124]. The former may be more reliable since the latter experiment was performed on Fe_xO, with x deviating several percent from unity. Bowen *et al.* investigated the infrared absorption [201], which we use here to benchmark against the BSE.

Regarding the fundamental gap, there is a general expectation in both experimental and theoretical literature that it is of order 2.5 eV [179,187,195,202]. Rödl *et al.* associated the sharp rise in $\alpha(\omega)$ observed around 2.4 eV in Ref. [201], with the fundamental gap. Hiraoka *et al.* [203] also assumed the fundamental gap was of this order but observed a peak in Im ϵ at ∼1 eV and tentatively assigned it to a defect band. Absorption data show peaks at both 1.2 and 2.4 eV [201] (see Fig. 24).

Turning to theory, at the QSGW level the fundamental gap is found to be 1.9 eV, with the smallest direct gap 2.4 eV. In all the other antiferromagnetic oxides of this study, QSGW overestimates the gap by ∼1 eV (Table IX), so these gaps are likely too high. At the QS$G\hat{W}$ level, the fundamental gap is much smaller, 0.64 eV. This leads to a puzzle: Why does QS$G\hat{W}$ yield such a gap so different from the accepted values in the literature?

Counterbalancing the experiments just mentioned, Zimmermann observed the one-particle spectral function XPS/BIS [118]. He did not attempt to extract a band gap, but based on his Fig. 15, it would be of order 1 eV. The XPS/BIS data and the optical measurements both point to the lowest excitation of order 1 eV, even though no deep excitons were found by QS$G\hat{W}$ to explain the absorption peak there. Thus, our QS$G\hat{W}$ analysis suggests a different interpretation, namely, that the observed peak in the absorption $\alpha(\omega)$ around 1.2 eV (Fig. 24) corresponds to the true fundamental gap.

The QS$G\hat{W}$ prediction for E_G is likely too small: FeO's practically dispersionless valence band strongly resembles that of VO_2 (Sec. III E 8), and we can expect the gap to be similarly underestimated in QS$G\hat{W}$. Using materials in Table IV as a guide, the gap can be expected to be underestimated by ∼0.5 eV. Two other pieces of evidence point to the gap being underestimated: $\varepsilon_\infty^{\rm BSE}$ is much larger than the two experiments noted earlier (Table X) and the peak in $\alpha(\omega)$ falls at ∼0.8 eV, well below the 1.2-eV peak reported in Ref. [201] (Fig. 24).

To adjust for the probable QS$G\hat{W}$ gap underestimate, we consider a hybrid of QS$G\hat{W}$ and QSGW, Eq. (21), and bench-

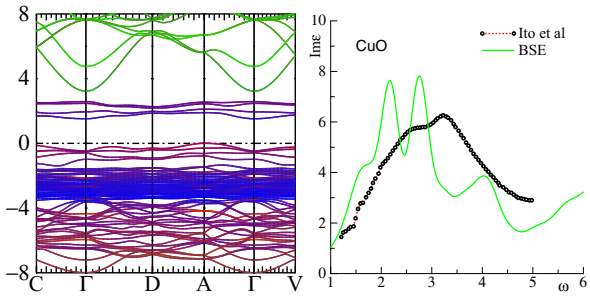


FIG. 25. Left: Energy band structure of CuO. Colors depict the following orbital characters: blue, Cu d; green, Cu sp, red, O sp. Right: Im $\epsilon(\omega)$ measured by Ito *et al.* [122] compared to BSE@QS$G\hat{W}$. Discrepancies with experiment are discussed in the text.

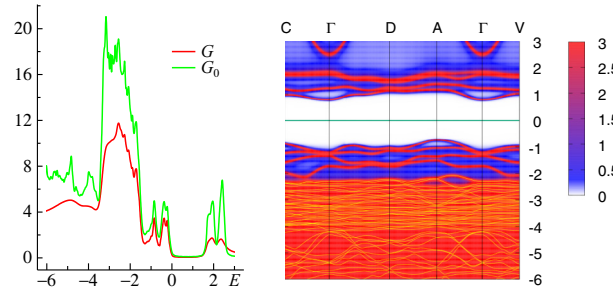


FIG. 26. (Left) Density of states generated from G compared to that generated from the QS$G\hat{W}$ G_0. Zero energy corresponds to the valence band maximum. (Right) Spectral function from interacting G generated by QS$G\hat{W}$. Yellow lines below −2 eV trace out the QS$G\hat{W}$ bands, and are equivalent to those in the left panel of Fig. 25.

mark both ϵ_∞ and the two peaks in absorption, for different admixtures β of QSGW into QS$G\hat{W}$ (Fig. 24). Table X shows the variation of E_G and ϵ_∞ with β. Perfect alignment with the best available experimental value for ϵ_∞ corresponds to $E_G = 1.05$ eV.

Thus, if FeO has a fundamental gap 1.05–1.10 eV, a consistent picture emerges. First, the two peaks in $\alpha(\omega)$ for BSE2 and BSE1 (Fig. 24 and Table X) bracket the two experimental peaks from above and below. Second, the one-particle DOS is consistent with XPS/BIS [118]. Finally, ϵ_∞ is consistent with the best available experimental data.

5. CuO

CuO has a monoclinic lattice structure of four formula units [204], while the magnetic structure is antiferromagnetic, and is a $\sqrt{2}\times1\times\sqrt{2}$ supercell of lattice with eight formula units [191,205]. The nominal configuration $Cu^{2+}O^{2-}$ would imply a single unpaired d electron; however, the magnetic moment is substantially smaller than $1\mu_B$/atom (Table IX).

The QS$G\hat{W}$ energy band structure is depicted in the left panel of Fig. 25. The valence band consists of approximately $\frac{2}{3}$ O-p character, and $\frac{1}{3}$ Cu-d character, and the conduction bands $\frac{1}{3}$ O-p character, and $\frac{2}{3}$ Cu-d character. Orbital weighting is quite different from cuprates such as La_2CuO_4, where both band edges are dominated by Cu. This finding is roughly in line with the DFT calculation of Filippetti and Fiorentini [205], who assigned the highest valence to O p (the LDA puts O p too high, so the O character will be overestimated).

A band gap has been measured optically from the absorption edge [151], and also by PES/BIS [182] and XPS [183]. All three measurements report band gaps in 1.3–1.4 eV range, slightly smaller than the 1.52-eV fundamental gap from QS$G\hat{W}$ (Table IX). However, QS$G\hat{W}$ overestimates ϵ_∞ (Table IX) by about 20%, which if the consistency between the gap and ε_∞ can be relied on (Sec. III B 3), the QS$G\hat{W}$ fundamental gap must be too small. Moreover, if we compare $\epsilon(\omega)$ against ellipsometry measurements of Ito *et al.* [122], the peaks of Im $\epsilon(\omega)$ are seen to fall ∼0.3 eV below the experimental data. From this we conclude it is likely that the fundamental gap is closer to 1.6 eV, assuming the dielectric data of Ito *et al.* [122] is reliable. This would mean the PES/BIS is underestimated. It is perhaps not unexpected since the BIS should be larger than the optical gap.

That experimentally Im $\epsilon(\omega)$ is smoother than the QS$G\hat{W}$ one can be attributed (at least in part) to $\epsilon(\omega)$ being generated from a noninteracting G_0 (QS$G\hat{W}$). The frequency dependence of Σ reduces quasiparticle weights (compare the DOS of the interacting G to that of G_0, left panel of Fig. 26), and the imaginary part smears out the quasiparticle (right panel of Fig. 26). CuO is very strongly correlated: note the sharp reduction in the DOS around 2 eV. These dynamical effects do not shift the average position of the bands (owing to the QSGW construction) but will smooth out transitions between occupied and unoccupied states and, correspondingly, the imaginary part of the longitudinal dielectric function in the basal plane, Im ϵ_{xx}.

TABLE X. Optical properties in FeO as a function of hybridization parameter β, Eq. (21). E_G is the fundamental gap in eV; E_G ($\Gamma\to\Gamma$) is the direct gap at Γ. Labels in the second column are used in Fig. 24.

β	label	E_G	E_G ($\Gamma\to\Gamma$)	ϵ_∞
0	BSE	0.64	0.94	17.6
0.3	BSE1	0.90	1.28	11.2
0.6	BSE2	1.19	1.63	8.04

6. La_2CuO_4

La_2CuO_4 is the parent compound for one of the most widely studied superconductors. A gap forms because Cu $d_{x^2-y^2}$ bands split into a bond-antibonding pair, owing to the formation of a local Cu moment. QSGW and QS$G\hat{W}$ energy band structures are shown in Fig. 27, with the Cu $d_{x^2-y^2}$ shown in red. There are several striking points of contrast:

(i) QS$G\hat{W}$ reduces the Cu-$d_{x^2-y^2}$ bond-antibond splitting relative to QSGW by about 1.5 eV. The relatively flat $d_{x^2-y^2}$ conduction band shifts more than the La d band (cyan), reminiscent of NiO. Thus, the the addition of ladders reduces the fundamental gap to 1.66 eV, from the QSGW gap of 3.1 eV.

(ii) The occupied Cu-$d_{x^2-y^2}$ band narrows relative to QSGW. QSGW itself narrows this band substantially relative to LDA or LDA+U (compare Cu $d_{x^2-y^2}$ in top right to bottom left panel), but the ladders narrow it still further, again reminiscent of NiO. Such mass renormalization plays a critical

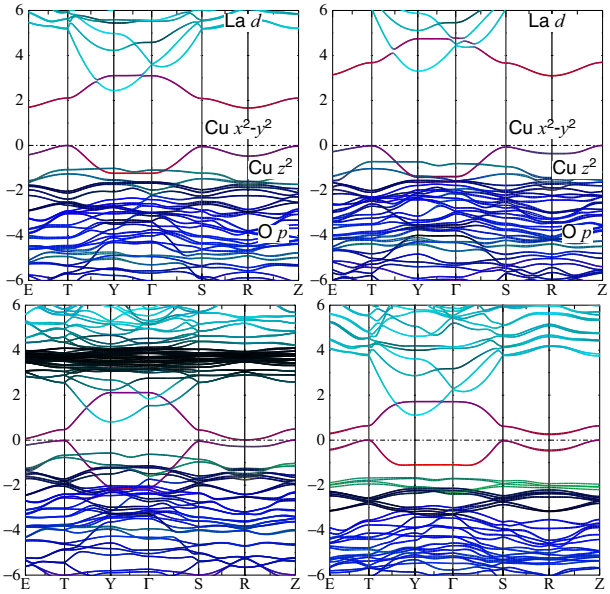


FIG. 27. Energy band structure of La_2CuO_4. Upper left and right panels apply to QS$G\hat{W}$ and QSGW approximations, respectively. Colors depict the following orbital character: red: Cu $d_{x^2-y^2}$; green: Cu d_{3z^2-1}; blue: O $2p$; cyan: La $5d$. Black valence bands are the remaining Cu-$3d$ orbitals. Lower left panel is the LDA band structure, with the same color scheme. The narrow window of black bands at 3–4 eV are of La $4f$ character. Bottom right is $G^{\rm LDA}W^{\rm LDA}$ result, with $Z = 1$ and including the off-diagonal parts of Σ.

role in the correlations of this orbital, which drives superconductivity. Spin fluctuations will narrow this band still further, but whether low-order perturbation theory will be sufficient to yield the true bandwidth remains an open question.

(iii) The O-p band is pushed down relative to QSGW, thus reducing the hybridization of O into the Cu-$d_{x^2-y^2}$ state. The LDA often misaligns orbitals of different character, but it is notable that ladder diagrams not only reduce the gap, but induce a shift to the QSGW valence bands.

(iv) For both QS$G\hat{W}$ and QSGW the La-$4f$ states are pushed well above the Fermi level, something the LDA fails to do.

(v) Self-consistency plays a very important role in this system (compare $G^{\rm LDA}W^{\rm LDA}$ to QSGW energy bands, and see Table I). As with NiO, the $G^{\rm LDA}W^{\rm LDA}$ band gap is severely underestimated [33]. The one-shot gap can be improved by using LDA+U or a hybrid functional instead of the LDA, but the resulting energy bands depend on the choice, as will other parts of the spectrum (e.g., the position of O-$2p$ states).

The left panel of Fig. 28 shows two measurements of the dielectric function, Im ϵ_{xx}, one inferred from reflectivity at 122 K [184], and the other from low-temperature optical conductivity [53]. Reference [184] also shows results from a photoconductivity measurement, which looks similar to the blue squares in the figure but slightly blueshifted. QS$G\hat{W}$ results are also shown: the peak in Im ϵ_{xx} appears at slightly higher energy (0.1–0.2 eV) than the experimental data. The QS$G\hat{W}$ result has sharper structure, in particular, there appears a pronounced subgap peak centered at ∼1.5 eV. A corresponding peak (albeit much weaker) is seen in the 122-K reflectivity data, though this peak is washed out as the temperature increases [184]. QS$G\hat{W}$ predicts a spectrum of 30 or so subgap excitons, ranging between 1.2 eV and the fundamental gap with widely varying oscillator strengths. A particularly bright exciton appears at 1.45 eV; it is is partly responsible for the peak in Im ϵ_{xx} there. As for the fundamental gap, QS$G\hat{W}$ predicts an indirect gap of 1.66 eV, but the lowest direct gap is ∼2.1 eV (Fig. 27). Reference [184] assigned a charge transfer gap of 2.1 eV, and Ref. [53] a similar gap (2.2 eV), which probably corresponds to the direct gap. The QS$G\hat{W}$ result for Im ϵ_{xx} shows sharper peaks than the experiment, as was shown for CuO (Sec. III F 5). For the same reason explained there, dynamical effects will smooth out Im ϵ_{xx}.

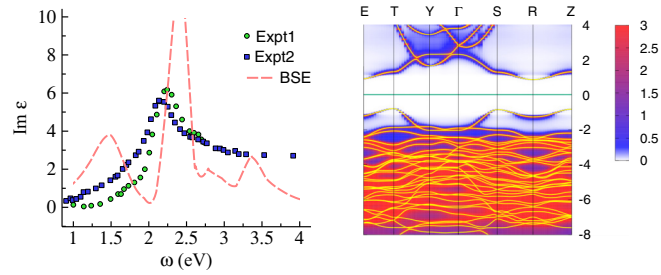


FIG. 28. (Left) BSE dielectric function Im ϵ_{xx} in La_2CuO_4, computed from QS$G\hat{W}$, compared to reflectivity data from Ref. [184] (Expt 1) and conductivity data from Ref. [53] (Expt 2). Data in Ref. [184] are anomalously small. They also report $\epsilon_\infty \sim 5$, which does not seem to be compatible with their scale for Im ϵ_{xx}, from the Kramers-Kronig relation. Expt 1 shown in the figure scales data taken from Ref. [184] by a factor of 5 to bring it approximately in line with Ref. [53]. (Right) Spectral function from the interacting G generated by QS$G\hat{W}$. Yellow lines are energy bands from the QS$G\hat{W}$ G_0.

7. Fe_3O_4

Magnetite, or Fe_3O_4, has a cubic inverted spinel structure above the Verwey transition at 123 K [206], with six Fe and eight O atoms in the unit cell. Two Fe are tetrahedrally bonded to O (O-Fe-O bond angles 109.5°) and four occupy octahedral sites with slightly larger bond lengths (bond angles 90° ± 2° and 180°). It is a ferrimagnet with the spins in the tetrahedral sites parallel, spins in the octahedral sites parallel, but the tetrahedral and octahedral sites are antiparallel.

The conventional picture, originating from Verwey, is that the tetrahedral sites are Fe^{3+} and octahedral sites consist of equal numbers of Fe^{2+} and Fe^{3+}. Below the Verwey temperature magnetite is a narrow-gap insulator with a band gap 0.14–0.3 eV. It was traditionally believed that above the Verwey temperature magnetite becomes a half-metal, in part because the conductivity increases by ∼100-fold across the transition, and evidence from interpretations of PES and STS experiments suggested a finite density of states at E_F. However, a more recent high-resolution PES experiment [207] found that the band gap persists above the Verwey temperature, reduced by ∼50 meV. That a gap of order 0.2 eV persists was confirmed by a subsequent STS measurement [208]. For a more detailed summary of the experimental status, see the work of Liu and Di Valentin [209]. These authors applied various one-body techniques (LDA+U, hybrid functionals) to study magnetite and concluded that the traditional picture of magnetic order yields a metallic ground state. They argued

TABLE XI. Spin moments and band gap of Fe_3O_4 calculated with the QSGW and QS$G\hat{W}$ approximations. Fe_t, Fe_{o1}, Fe_{o2} indicate Fe on tetrahedral sites, and the two inequivalent octahedral sites.

	E_G	μ (Fe_t)	μ (Fe_{o1})	μ (Fe_{o2})
QSGW	0.86 ± 0.1	-4.05	4.31	3.65
QS$G\hat{W}$	0.15 ± 0.15	-4.04	4.20	3.75

that a gap appears because the octahedral sites disproportionate into two kinds of atoms, one with a high spin (moment $\gtrsim 4\mu_B$) and one with a low spin (moment $\sim 3.5\mu_B$).

We performed QSGW and QS$G\hat{W}$ calculations with symmetry suppressed, so every spin could assume an independent value. Spins on the tetrahedral sites converged to a common value and those on the tetrahedral sites converged to two values, one high spin and one low spin (Table XI). The calculation was extremely difficult to stabilize and a fully converged solution was never found, even after 100 iterations. Local moments were stable as iterations proceeded, but the band gap fluctuated; for that reason error bars are given for the gaps shown in Table XI. Here we denote high- and low-spin sites as "o1" and "o2."

The Fe_{o2} $t_{2g}^{\downarrow}$ manifold splits off one band (actually two, because the unit cell consists of two Fe_3O_4 formula units which weakly couple), and this split-off band forms the valence band maximum (Fig. 29). On the other hand, the Fe_{o1} $t_{2g}^{\downarrow}$ manifold does not split in the same way, and it forms the conduction band minimum. Splitting of $Fe_{o2}t_{2g}^{\downarrow}$ is the main way in which cubic symmetry is broken and a gap is formed.

Ladder diagrams have two major effects: first, they sharply narrow all of the octrahedral Fe-d bandwidths, e.g., the $t_{2g}^{\downarrow}$ manifold forming the conduction band minimum is narrowed by $\sim$30%. Second, ladders cause band centers to shift in a highly orbital-dependent manner. Shifts on the occupied states are modest, but for the unoccupied states they can be quite large. Note, for example, the band center of the $t_{2g}^{\downarrow}$ level around 1.5 eV is pushed down by $\sim$1 eV, while the $t_{2g}^{\downarrow}$ and $e_g^{\downarrow}$ levels in the 5–6 eV range shift by $\sim$2 eV. Unfortunately, no experiments are available for benchmarking.

Ordering of the Fe levels is qualitatively in agreement with the picture of Ref. [209]; see their Fig. 5, and the QS$G\hat{W}$ spin moments are similar to their Table 1; we thus affirm their description of magnetite.

IV. CONCLUSIONS

We presented an extension of the QSGW approximation in which we include vertex corrections to W, calculated at the level of the BSE (QS$G\hat{W}$). The primary aim of this work was to establish to what extent QS$G\hat{W}$ rectifies the most severe errors in QSGW, with the ultimate aim to develop a high-fidelity, universally applicable theory. If low-order diagrams are sufficient to yield high-fidelity one- and two-particle properties, Green's function methods offer an enormous opportunity to be both high fidelity and relatively efficient. QS$G\hat{W}$ supplies excitonic effects which QSGW omits. They are known to be important, as already extensively discussed in the literature (see, e.g., Ref. [27]), so it is of interest to see to what extent it captures the quasiparticle spectrum in insulators, and whether systematic errors can be discerned.

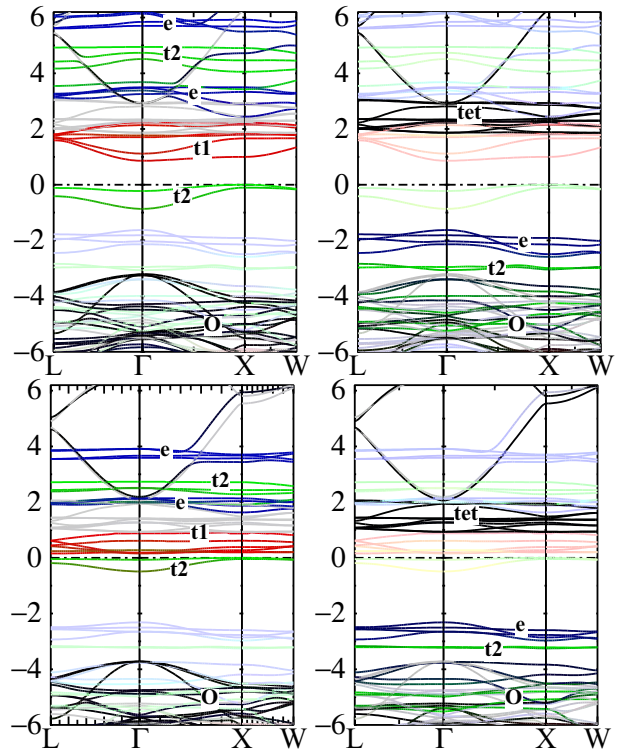


FIG. 29. Energy band structure Fe_3O_4, in the QSGW approximation (upper panels) and the QS$G\hat{W}$ approximation (lower panels). Left panels show minority spin with majority spin bleached out; right panels show the reverse. The character of relevant orbitals is labeled in the figure and the labels correspond to d orbitals on the octahedral sites with the following color scheme. t1: t_{2g}, Fe_{o1} (red); t2: t_{2g} Fe_{o2} (green); e: Fe_{o1}+Fe_{o2} e_g (blue). Labels tet and O signify bands centered mainly on Fe tetrahedral and O sites, respectively.

While the QSGW approximation has long been known to overestimate band gaps, the discrepancies with experiments are much more systematic than more commonly used G_0W_0 approaches with G and W constructed, e.g., from the LDA. Its systematic character is a consequence of self-consistency, in part because it does not rely on density functional theory. We presented a few contexts where a non-self-consistent approach is not important (e.g., Bi_2Te_3), and others (e.g., $TiSe_2$, La_2CuO_4) where it is fundamentally problematic. Even when such an approach yields a good band gap, it may occur for adventitious reasons. The relatively unsystematic nature of the errors in one-shot approaches makes it difficult to assess what diagrams are essential to realize the goal of a universally applicable, high-fidelity theory.[13] Thus, self-consistency is crucial for the aims of this work.

This work surveyed a wide range of insulators, including tetrahedrally coordinated semiconductors where experimental information is reliable and abundant, and also a variety

[13]The present method also employs an all-electron basis, which eliminates the dependence on the choice of the pseudopotential. This can, though it should not be the case, substantially influence the GW results.

of other *sp* systems, d^0 oxides, and polar compounds, and a family of 3*d* transition metal antiferromagnetic oxides. Each material system had a distinct set of characteristics, but apart from some important exceptions critically examined in this work, QS$G\hat{W}$ predicts with fairly high fidelity both one-particle and optical properties for all of the systems we studied. The exceptions are important and formed a major focus of this study. Two shortcomings clearly identified were the omission of electron-phonon interaction, which causes gaps to be too large in wide-gap systems, and the omission of the vertex in the exact self-energy. This vertex pushes down nearly dispersionless corelike states, and when they form the valence band maximum the band gap is consistently underestimated. By constructing hybrid self-energies, we could account for both of these shortcomings in an approximate way, and draw the following conclusions:

(1) At the QS$G\hat{W}$ level, there is a very close connection between the fidelity of the fundamental gap E_G and the dielectric constant ϵ_∞. When one is well described, so is the other, and vice versa. This provides a much more robust benchmark of a theory than benchmarking one-particle properties alone.

(2) If we take the first point as an ansatz for a general principle, it can be used in cases where experimental data are unavailable or inconsistent. We presented evidence for several systems (CeO_2, $SrTiO_3$, TiO_2, ScN, $CuAlO_2$, FeO) where the calculated results inform the experimental observations and indicate that accepted values of the one-particle properties need adjustment. For FeO, the revision is rather dramatic.

(3) A low-order diagrammatic theory appears to describe the dielectric response with high fidelity for all the systems in this study, to the extent we are able to reliably extract experimental data. Ladder diagrams appear to be sufficient to capture well the main part of the optical response functions and one-particle Green's functions in most insulators, even strongly correlated ones. While such an assertion is likely not universally true [45], it appears to be the case for broad classes of materials.

(4) Ladders not only shift the band gap but further narrow the *d* bandwidth in some systems (NiO, La_2CuO_4, Fe_3O_4). It may be that the addition of a low-order *GW*-like theory accounting for spin fluctuations, such as the dual-trilex formulation of Stepanov *et al.* [58], may adequately account for spin and charge response functions even in strongly correlated materials.

These last two observations suggest the tantalizing possibility that, with some modest extensions that may be added hierarchically, a broadly applicable, high-fidelity *ab initio* approach to solving one- and two-particle properties of the many-body problem is within reach.

ACKNOWLEDGMENTS

The authors would like to thank all those involved in the *CCP flagship project: Quasiparticle Self-Consistent GW for Next-Generation Electronic Structure*, especially S. Mckechnie for his help. We are grateful for support from the Engineering and Physical Sciences Research Council, under Grant No. EP/M011631/1. M.v.S. and D.P. were supported the Computational Chemical Sciences program within the Office of Basic Energy Sciences, U.S. Department of Energy under Contract No. DE-AC36-08GO28308. We are grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (Grant No. EP/P020194/1). The research was performed using computational resources sponsored by the Department of Energy's Office of Energy Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory. This research also used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC Award No. BES-ERCAP0021783.

APPENDIX A: THE LQS*GW* APPROXIMATION

Kutepov's LQS*GW* theory [210] is a linearized form of QS*GW*. He approximates the quasiparticlized self-energy as a Taylor series around zero frequency. Treating each band independently and suppressing band index for simplicity of presentation, Kutepov replaces the interacting *G*,

$$G^{-1}(k,\omega) = \omega + \mu - \epsilon - \Sigma(k,\omega),$$

by omitting the second-order and higher terms of an expansion of Σ in ω:

$$\Sigma(k,\omega) = \Sigma(k,0) + \omega\,\Sigma'(k,0) + \tfrac{1}{2}\omega^2\,\Sigma''(k,0) + \cdots. \tag{A1}$$

G^{-1} simplifies to a linear function of ω,

$$G^{-1}(k,\omega) = \bar{Z}^{-1}\,\omega + \mu - \epsilon - \Sigma(k,0),$$

and thus reduces to a linear algebraic eigenvalue problem. The bar over the *Z* factor indicates that is not equivalent to Eq. (13) since it is defined at zero frequency $\omega = 0$:

$$1 - 1/\bar{Z}^j = \Sigma'(k,0).$$

Evidently, $\epsilon - \mu + \Sigma(k,0)$ is the eigenvalue of a Hamiltonian defined as the one-body part of G^{-1}, but including the static part of Σ. The (linearized) energy dependence of Σ modifies this eigenvalue to read as

$$E - \mu = \bar{Z}[\epsilon - \mu + \Sigma(k,0)].$$

E is identical to the QS*GW* quasiparticle energy if Σ is a linear function of ω.

Now let us retain the quadratic term in Σ and determine the shift in *E* to estimate the difference between LQS*GW* and QS*GW*. Let us denote the LQS*GW* eigenvalue $E - \mu$ as E_0. Expanding G^{-1} to second order we obtain, to lowest order in $\Sigma''(k,0)$,

$$G^{-1} \approx \omega - \left(E_0 + \frac{\bar{Z}}{2}E_0^2\,\Sigma''(k,0)\right). \tag{A2}$$

The lowest-order difference between LQS*GW* and QS*GW* QP levels is the second term in parentheses.

APPENDIX B: SUPPLEMENTAL MATERIAL: COMPUTATIONAL DETAILS

Implementation of *GW* requires both a one-body framework and a two-body framework. Both are described in detail in QUESTAAL's methods paper [37], and the paper describing QUESTAAL's implementation of QS*GW* theory [9], which we denote here as papers I and II. I places heavier focus on the one-body part, while II focuses on the *GW* theory and its implementation.

QUESTAAL is an all-electron method, with an augmented wave basis consisting of partial waves inside augmentation spheres, constructed from numerical solutions of the radial Schrödinger equation on a logarithmic mesh (I, Sec. 2.2). The one-body basis consists of a linear combination of smooth, atom-centered Hankel functions as envelope functions, augmented by the partial waves. Two partial waves are calculated at some linearization energy ϕ_ℓ and energy derivative $\dot{\phi}_\ell$, which provides enough freedom to match value and slope to the envelope functions (I, Sec. 3).

One particle basis. In a conventional LMTO basis, envelope functions consist of ordinary Hankel functions, parametrized by energy E. QUESTAAL's smooth Hankel functions are composed of a convolution of Gaussian functions of smoothing radius r_s, and ordinary Hankel functions (I, Sec. 3.1); thus, two parameters are needed to define the envelope. In the periodic solid, Bloch sums of these functions are taken (I, Appendix C). In this work, E is constrained to a fixed value (-0.4 Ry for most systems), and r_s determined by optimizing the total energy of the free-atom wave function. These are kept fixed throughout the calculation, while the partial waves and linearization energy float as the potential evolves. By fixing E to a universal value, we are able to take advantage of the "screening transformation" to render the basis set short ranged (see I, Sec. 2.9). This can be useful for the interpolation of the self-energy to an arbitrary k mesh, as described below. A second envelope function of a deeper energy is needed to make the Hamiltonian reasonably complete. The latter energy is chosen to be 0.8 Ry deeper than the first. For most materials, the envelopes of orbitals $l = 0 \ldots 4$ the first energy, and $l = 0 \ldots 3$ for the second. At the *GW* level, a few other additions are made to make the basis closer to complete. Completeness of the envelope functions is sometimes improved by adding "floating orbitals," points in the interstitial regions where smooth Hankel functions are placed without an augmentation sphere (I, Sec. 3.11), usually for ℓ up to 2. N_{flt} in Table XII indicates how many points in the unit cell where floating orbitals are added. To expand the Hilbert space inside the augmentation spheres, a local orbital ϕ_z may be added (I, Sec. 3.7.3). ϕ_z is a solution of the radial Schrödinger equation at an energy either well below the linearization energy for deep corelike states, or well above it to better represent the unoccupied states. In the table, the ϕ_z used in the calculations here are listed, with a bar over the principal quantum number to indicate the high-lying states. For heavier elements, the p local orbital is sometimes replaced by the $p_{1/2}$ component of the Dirac equation. This has a modest effect but improves the accuracy of the spin-orbit coupling (I, Sec. 3.9). The total number of orbitals in the one-particle basis is listed in Table XII as N_{1p}. Another parameter is the sphere augmentation radius, r_{MT}.

k convergence. The *GW* mesh and the one-body mesh are generally different: the latter normally needs to be somewhat finer, as the self-energy is a relatively smooth function of k while the kinetic energy is less so. Since the cost is low, we use a finer mesh than necessary for the one-particle part, which obviates the need to test the mesh for k convergence. Careful tests of the *GW* mesh were made for each system. Most of the small unit cells used a mesh of six divisions along each axis: the number used in each materials system is listed as N_k in the Table XII. A finer mesh, e.g., $8\times8\times8$ divisions, changes the result only slightly (e.g., gap changes by $\sim$0.01 eV in *sp* semiconductors).

To enable inequivalent meshes, the self-energy must be interpolated. To render the interpolation everywhere smooth (I, Sec. 2G) eigenfunctions and self-energy are rotated to the LDA basis, and the full self-energy matrix is kept only up to a cutoff above the Fermi level in this basis, denoted Σ_{cut} in the table. Above this cutoff, only the diagonal part of Σ is kept. Σ_{cut} may be made arbitrarily high, but if it is too high the interpolation is no longer smooth. Fortunately, the result depends weakly on Σ_{cut}, and $\Sigma_{\text{cut}} \sim 2$ is typically sufficient to achieve a reasonably well-converged result.

A smooth Hankel function has a plane-wave representation; thus, any linear combination of them, e.g., an eigenfunction, does also. An eigenfunction represented in this form is equivalent to a representation in an LAPW basis: it is defined by the coefficients to the plane waves, the shape of the partial waves and their coefficients (which are constrained to match smoothly onto the envelope functions). The PW cutoff for the one-particle basis is listed as $G_{\text{cut}}(\psi)$ in Table XII.

Two-particle basis. The two-particle basis is needed to represent quantities such as the bare Coulomb interaction and the polarizability. As with the one-particle basis, it as a mixed construction with interstitial parts and augmentation parts (II, Sec. II A): envelope function products are represented as plane waves since product of plane waves is another plane wave. Thus, the interstitial parts of the mixed (product) basis are plane waves, and the PW cutoff is listed in Table XII as $G_{\text{cut}}(M)$. Inside augmentation spheres, all possible products of partial waves are called product functions B_ℓ, organized by ℓ with a form $B_I = B_\ell(r)Y_{\ell m}(\hat{\mathbf{r}})$. The set of all possible products of partial waves is somewhat overcomplete with a relatively large rank. It is reduced by diagonalizing the overlap matrix, and retaining the subset of functions above a cutoff eigenvalue of the overlap. It has been found from experience that eigenfunctions with eigenvalues below 3×10^{-4} for $\ell = 0,1$ and 10^{-3} for $\ell > 1$ have essentially negligible effect on the result, and are discarded. The product basis is truncated at a finite ℓ. For most systems, ℓ_{cut} was chosen to be 6 for elements with small to moderate r_{MT} and whose d orbitals are far from the Fermi level, such as O or P; 8 for elements of intermediate size; 10 for elements with large radii; and 12 for systems with f orbitals such as Ce.

Bare Coulomb interaction. To stabilize the calculation, the bare Coulomb interaction $v(q) = 1/q^2$ is approximated by a Thomas-Fermi form $v(q) = 1/(q^2 + V_{\text{TF}})$. This is because if

TABLE XII. Material parameters: a is lattice constant (quantity in parentheses is c/a where applicable). $G_{\rm cut}(\psi)$ and $G_{\rm cut}(M)$ are plane-wave cutoffs for the interstitial part of the one- and two-particle basis sets, in units of $2\pi/a$; $\Sigma_{\rm cut}$ the energy cutoff for above which Σ is restricted to a diagonal part, as described in the text. N_k is the number of divisions along each reciprocal lattice vector defining the k mesh. When two numbers appear, the c axis is assigned a mesh different than the basal plane. The latter number is selected to make the spacing between k points as similar as possible along the three directions. N_{1p} is the total number of basis functions in the unit cell. N_v and N_c are the number of occupied and unoccupied eigenstates included in the construction of the vertex. $N_{\rm flt}$ is the number of points in the interstitial where envelope functions were added to increase the basis completeness. φ_z lists the local orbitals (LO) for each element: $n-s$, $n-p$, $n-d$, where n is the principal quantum number of the LO. n without an overbar indicates the LO covers a corelike state, well below the linearization energy with a principal quantum number one less than that of the valence. $\bar{n}$ indicates the LO energy is far above the linearization energy, and is included to better treat unoccupied states well above the Fermi energy. Both kinds of LO are discussed in Ref. [37]. Partial waves marked as $\hat{p}$ replace the $l=1$ partial wave with the corresponding $p_{1/2}$ partial wave computed from the Dirac equation, as discussed in Ref. [37]. This is a small effect but it improves the matrix elements for spin-orbit coupling.

	a (a.u.)	$G_{\rm cut}(\psi, M)$	$\Sigma_{\rm cut}$(Ry)	N_k	N_{1p}	N_v	N_c	$N_{\rm flt}$	φ_z
C	6.740	4.2 3.5	3.0	6	104	4	8	2	$\bar{3}s\bar{3}p\bar{4}d$ C
Si	10.24	3.0 2.5	3.0	6	104	4	8	2	$\bar{4}s\bar{4}p\bar{4}d$ Si
Ge	10.68	3.0 2.5	3.0	6	104	4	8	2	$\bar{5}s\bar{5}p3d$ Ge
TiO_2	8.681 (0.6441)	3.5 2.8	3.0	4,6	330	12	10	8	$\bar{5}s3p\bar{4}d$ Ti; $\bar{3}s\bar{3}p\bar{4}d$ O
$SrTiO_3$	7.354	3.5 2.8	3.0	6	215	9	11	–	$4s4p\bar{5}d$ Sr; $\bar{5}s\bar{3}p\bar{4}d$ Ti; $\bar{3}s3p\bar{4}d$ O
$CuAlO_2$	3.121 (3.422)	4.1 3.3	3.5	6	177	8	8	4	$\bar{5}s\bar{4}d$ Cu; $\bar{4}s$ Al; $\bar{3}s\bar{3}p\bar{4}d$ O
LiF	7.597	4.2 3.5	3.0	6	99	3	8	2	$1s\bar{3}p\bar{4}d$ Li; $\bar{3}s\bar{3}p\bar{4}d$ F
LiCl	9.600	3.6 3.0	3.0	6	104	3	8	2	$1s\bar{3}p\bar{4}d$ Li; $\bar{4}s\bar{4}\hat{p}\bar{4}d$ Cl
NaCl	10.62	3.2 2.6	3.0	6	104	3	8	2	$2s2p$ Na
CuCl	10.23	3.2 2.6	3.0	6	104	8	8	2	$\bar{5}s\bar{5}\hat{p}\bar{4}d$ Cu; $\bar{4}s\bar{4}\hat{p}\bar{4}d$ Cl
Cu_2O	8.069	3.5 2.8	3.0	4	330	26	9	10	$\bar{5}s\bar{5}\hat{p}\bar{4}d$ Cu; $\bar{3}s\bar{3}p\bar{4}d$ O
MgO	7.933	4.0 3.3	3.0	6	104	3	8	2	$2s2p\bar{4}d$ Mg; $\bar{3}s\bar{3}p\bar{4}d$ O
CaO	9.077	3.5 2.9	3.0	6	104	3	8	2	$3s3p\bar{4}d$ Ca; $\bar{3}s\bar{3}p\bar{4}d$ O
SrO	9.751	4.0 3.3	3.0	6	104	3	8	2	$4s4p\bar{5}d$ Sr; $\bar{3}s\bar{3}p\bar{4}d$ O
BaO	10.43	3.5 2.8	3.0	6	85	3	4	2	$5s5p\bar{6}d$ Ba; $\bar{3}s\bar{3}p\bar{4}d$ O
CdO	8.874	3.9 3.2	3.0	6	104	4	4	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ Cd; $\bar{3}s\bar{3}p\bar{4}d$ O
ZnO	6.138 (1.602)	3.7 3.0	3.5	6,4	204	6	8	2	$\bar{5}s\bar{5}\hat{p}\bar{4}d$ Zn; $\bar{3}s\bar{3}p\bar{4}d$ O
ZnS	10.23	3.4 2.8	3.0	6	104	3	8	2	$\bar{5}s\bar{5}\hat{p}\bar{4}d$ Zn; $\bar{4}s\bar{4}\hat{p}\bar{4}d$ S
ZnSe	10.69	3.0 2.5	3.0	6	104	3	8	2	$\bar{5}s\bar{5}\hat{p}\bar{4}d$ Zn; $\bar{5}s\bar{5}\hat{p}3d$ Se
ZnTe	11.53	3.0 2.5	3.0	6	104	3	8	2	$\bar{5}s\bar{5}\hat{p}\bar{4}d$ Zn; $\bar{6}s\bar{6}\hat{p}4d$ Te
wCdS	7.861 (1.620)	2.7 2.2	3.5	6,4	204	6	8	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ Cd; $\bar{4}s\bar{4}\hat{p}\bar{4}d$ S
CdSe	11.43	2.7 2.2	3.0	6	104	3	8	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ Cd; $\bar{5}s\bar{5}\hat{p}3d$ Se
CdTe	12.24	2.8 2.3	3.0	6	104	3	8	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ Cd; $\bar{6}s\bar{6}\hat{p}4d$ Te
hBN	4.732	3.8 3.0	4.0	6,3	124	8	8	2	$\bar{3}s\bar{3}p$ B; $\bar{3}s\bar{3}p$ N
AlN	5.879 (1.601)	3.8 3.1	3.5	6,4	204	6	8	2	$\bar{4}s\bar{4}p\bar{4}d$ Al; $\bar{3}s\bar{3}p\bar{4}d$ N
AlP	10.32	3.0 2.5	3.0	6	104	4	8	2	$\bar{4}s\bar{4}p\bar{4}d$ Al; $\bar{4}s\bar{4}p\bar{4}d$ P
AlAs	10.70	3.3 2.7	3.0	6	104	4	8	2	$\bar{4}s\bar{4}p\bar{4}d$ Al; $\bar{5}s\bar{5}\hat{p}3d$ As
AlSb	11.59	3.0 2.5	3.0	6	104	4	8	2	$\bar{4}s\bar{4}p\bar{4}d$ Al; $\bar{6}s\bar{6}\hat{p}4d$ Sb
GaN	6.027 (1.626)	3.6 2.9	3.5	6,4	204	6	8	2	$\bar{5}s\bar{5}\hat{p}3d$ Ga; $\bar{3}s\bar{3}p\bar{4}d$ N
GaP	10.29	3.0 2.5	3.0	6	104	4	8	2	$\bar{5}s\bar{5}\hat{p}3d$ Ga; $\bar{4}s\bar{4}\hat{p}\bar{4}d$ P
GaAs	10.66	2.7 2.4	3.0	6	104	4	8	2	$\bar{5}s\bar{5}\hat{p}3d$ Ga; $\bar{5}s\bar{5}\hat{p}3d$ As
GaSb	11.50	2.7 2.3	3.0	6	104	3	8	2	$\bar{5}s\bar{5}\hat{p}3d$ Ga; $\bar{6}s\bar{6}\hat{p}4d$ Sb
InN	6.679 (1.624)	3.4 2.7	3.5	6,4	204	6	8	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ In; $\bar{3}s\bar{3}p\bar{4}d$ N
InP	11.09	2.9 2.4	3.0	6	104	4	8	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ In; $\bar{4}s\bar{4}\hat{p}\bar{4}d$ P
InAs	11.43	2.7 2.3	3.0	6	104	3	8	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ In; $\bar{5}s\bar{5}\hat{p}3d$ As
InSb	12.24	2.7 2.3	3.0	6	104	3	8	2	$\bar{6}s\bar{6}\hat{p}\bar{5}d$ In; $\bar{6}s\bar{6}\hat{p}4d$ Sb
ScN	8.504	3.5 2.9	3.0	6	104	3	8	2	$3s3p\bar{4}d$ Sc; $\bar{3}s\bar{3}p\bar{4}d$ N
PbTe	12.15	2.8 2.3	3.0	6	104	3	8	2	$\bar{7}s\bar{7}\hat{p}\bar{6}d$ Pb; $\bar{6}s\bar{6}\hat{p}4d$ Te
$TiSe_2$	6.689 (1.697)	2.8 2.2	3.0	3,2	664	16	16	16	$3p\bar{4}d$ Ti
FeS_2	10.22	2.7 2.2	3.0	4	720	32	12	24	$\bar{5}s\bar{4}d$ Fe; $\bar{4}s\bar{4}p\bar{4}d$ S
VO_2	8.536	3.4 2.8	2.0	4	332	20	8	–	$3p\bar{4}d$ V
CeO_2	10.23	3.5 3.0	2.5	6	124	6	16	1	$5s5p\bar{6}d\bar{5}f$ Ce; $\bar{3}s\bar{3}p$ O
Bi_2Te_3	4.783 (4.015)	3.0 2.5	3.0	6	260	18	12	5	$\bar{7}s\bar{7}\hat{p}5d$ Bi; $\bar{6}s\bar{6}\hat{p}4d$ Te
MnO	8.398	3.5 3.0	2.5	4	178	16	16	4	$\bar{5}s3p\bar{4}d$ Mn; $\bar{3}s\bar{3}p$ O
FeO	8.088	3.5 2.9	3.0	4	172	9	6	–	$\bar{5}s3p\bar{4}d$ Fe; $\bar{3}s\bar{3}p\bar{4}d$ O
CoO	8.050	3.5 2.9	3.0	4	172	12	12	–	$\bar{5}s3p\bar{4}d$ Co; $\bar{3}s\bar{3}p\bar{4}d$ O

TABLE XII. (*Continued.*)

	a (a.u.)	$G_{\rm cut}(\psi, M)$	$\Sigma_{\rm cut}$(Ry)	N_k	N_{1p}	N_v	N_c	$N_{\rm flt}$	φ_z
NiO	7.880	3.1 2.5	2.5	4	110	16	16	–	$\bar{4}d$ Ni
CuO	9.558[†]	3.5 2.8	3.0	3	616	48	12	–	$\bar{5}s\bar{5}\hat{p}\bar{4}d$ Cu; $\bar{3}s\bar{3}p\bar{4}d$ O
MnTe	7.823 (1.621)	2.7 2.3	2.2	6,4	170	8	8	2	$\bar{4}d$ Mn; $\bar{6}\hat{p}4d$ Te
Fe_3O_4	15.87	3.5 2.9	3.0	4	746	12	12	16	$\bar{5}s3p\bar{4}d$ Fe; $\bar{3}s\bar{3}p\bar{4}d$ O
La_2CuO_4	9.942 (1.245)	3.1 2.5	2.5	3	524	30	30	–	$5p\bar{6}d$ La; $3p\bar{4}d$ Cu; $\bar{3}s\bar{3}p\bar{4}d$ O

$V_{\rm TF}$ is set to zero, the result can become unstable. We use a small value $V_{\rm TF}$, typically 2×10^{-5} Ry, though sometimes somewhat larger values, up to 2×10^{-4} Ry were used. The dielectric constant ϵ_∞ can vary by a few percent over this range. For that reason ϵ_∞ was calculated for several values of $+V_{\rm TF}$, e.g., 1×10^{-5}, 1×10^{-5}, and 3×10^{-5} Ry, and the reported value is the result when extrapolated to zero.

Frequency mesh. To construct the self-energy, an energy integration on the real frequency axis is taken. A regular quadratic mesh of the form $\omega_i = dw\times i + dw^2 i^2/(2\omega_c)$ is used, with i spanning $\omega_i = 0$ and the largest eigenstate. Points are linearly spaced for $dw \ll \omega_c$, but the spacing increases for $dw \gtrsim \omega_c$. It has been found empirically that results are essentially independent of mesh for $dw < 0.08$ Ry and $\omega_c \gtrsim 0.1$ Ry. In practice we use $dw = 0.02$ Ry and $\omega_c = 0.2$ Ry to obviate the need for checking convergence. To pick up the poles of G and W to make Σ, the contour is deformed to include an integration on the imaginary axis of ω (I, Sec. 2F). In all the calculations used here, we used six points on a Legendre quadrature. A few checks showed that the result hardly depended on the number of points in the quadrature.

Manual vs autogenerated input. QUESTAAL has an automatic generator, `blm`, to construct input files from structural data. Most input parameters are automatically generated by `blm`, such as the MT radii $r_{\rm MT}$, the product basis cutoffs, and the plane-wave cutoffs, the Gaussian smoothing radius defining the envelope functions, and the placements for floating orbitals, when they are sought. Also, for the vast majority of parameters, the code uses default values if inputs are not explicitly specified. For a few parameters, manual intervention is needed to monitor convergence, especially the number of k points and the plane-wave cutoffs $G_{\rm cut}(\psi)$ and $G_{\rm cut}(M)$. Hankel function energies E must be manually set, but usually fixed values as noted above are sufficient. Occasionally, interpolation continues to be an issue and can be stabilized by making E deeper, e.g., $E=-0.6$ Ry was needed to stabilize $SrTiO_3$. Results are largely insensitive to the choice of E, provided it is not pushed too deep.

[1] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn, Nobel Lecture: Electronic structure of matter—wave functions and density functionals, Rev. Mod. Phys. **71**, 1253 (1999).

[3] L. Hedin, New method for calculating the one-particle green's function with application to the electron-gas problem, Phys. Rev. **139**, A796 (1965).

[4] T. Miyake, F. Aryasetiawan, T. Kotani, M. van Schilfgaarde, M. Usuda, and K. Terakura, Total energy of solids: An exchange- and random-phase approximation correlation study, Phys. Rev. B **66**, 245103 (2002).

[5] J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse, and M. Scheffler, Assessment of correlation energies based on the random-phase approximation, New J. Phys. **14**, 043002 (2012).

[6] X. Ren, P. Rinke, G. E. Scuseria, and M. Scheffler, Renormalized second-order perturbation theory for the electron correlation energy: Concept, implementation, and benchmarks, Phys. Rev. B **88**, 035120 (2013).

[7] P. Das and T. Bazhirov, Electronic properties of binary compounds with high fidelity and high throughput, J. Phys.: Conf. Ser. **1290**, 012011 (2019), a wide range of compounds computed with LDA and HSE. Useful test of range of validity of common functionals.

[8] S. Acharya, D. Pashov, A. N. Rudenko, M. Rösner, M. van Schilfgaarde, and M. I. Katsnelson, Importance of charge self-consistency in first-principles description of strongly correlated systems, npj Comput. Mater. **7**, 208 (2021).

[9] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Quasiparticle self-consistent *GW* method: A basis for the independent-particle approximation, Phys. Rev. B **76**, 165106 (2007).

[10] With the caveat that there not be a unique solution, there can be multiple solutions for G_0, similar to what is found in Hartree-Fock theory, or DFT. Nevertheless the fact that it satisfies a variational principle [11] implies that there are a limited, discrete set of solutions corresponding to metastable states as in Hartree-Fock or density functional theory. Such states, e.g., solutions found for both an antiferromagnetic and a ferromagnetic state that are connected to local minima.

[11] S. Ismail-Beigi, Justifying quasiparticle self-consistent schemes via gradient optimization in Baym–Kadanoff theory, J. Phys.: Condens. Matter **29**, 385501 (2017).

[12] F. Sottile, V. Olevano, and L. Reining, Parameter-free calculation of response functions in time-dependent density-functional theory, Phys. Rev. Lett. **91**, 056402 (2003).

[13] A. Marini, R. Del Sole, and A. Rubio, Bound excitons in time-dependent density-functional theory: Optical and energy-loss spectra, Phys. Rev. Lett. **91**, 256402 (2003).

[14] T. Kotani and M. van Schilfgaarde, Spin wave dispersion based on the quasiparticle self-consistent *GW* method: NiO, MnO and α-MnAs, J. Phys.: Condens. Matter **20**, 295214 (2008).

[15] J. Barker, D. Pashov, and J. Jackson, Electronic structure and finite temperature magnetism of yttrium iron garnet, Electron. Struct. **2**, 044002 (2020).

[16] S. Acharya, D. Pashov, F. Jamet, and M. van Schilfgaarde, Controlling T_c through band structure and correlation engineering in collapsed and uncollapsed phases of iron arsenides, Phys. Rev. Lett. **124**, 237001 (2020).

[17] C. Franchini, A. Sanna, M. Marsman, and G. Kresse, Structural, vibrational, and quasiparticle properties of the Peierls semiconductor $BaBiO_3$: A hybrid functional and self-consistent GW + vertex-corrections study, Phys. Rev. B **81**, 085213 (2010).

[18] M. Shishkin, M. Marsman, and G. Kresse, Accurate quasiparticle spectra from self-consistent GW calculations with vertex corrections, Phys. Rev. Lett. **99**, 246403 (2007).

[19] F. Bruneval, F. Sottile, V. Olevano, R. Del Sole, and L. Reining, Many-body perturbation theory using the density-functional concept: Beyond the *GW* approximation, Phys. Rev. Lett. **94**, 186402 (2005).

[20] A. L. Kutepov, Electronic structure of Na, K, Si, and LiF from self-consistent solution of Hedin's equations including vertex corrections, Phys. Rev. B **94**, 155101 (2016).

[21] A. L. Kutepov, Self-consistent solution of Hedin's equations: Semiconductors and insulators, Phys. Rev. B **95**, 195120 (2017).

[22] M. van Schilfgaarde, T. Kotani, and S. Faleev, Quasiparticle self-consistent *GW* theory, Phys. Rev. Lett. **96**, 226402 (2006).

[23] M. P. Ljungberg, P. Koval, F. Ferrari, D. Foerster, and D. Sánchez-Portal, Cubic-scaling iterative solution of the Bethe-Salpeter equation for finite systems, Phys. Rev. B **92**, 075422 (2015).

[24] QUESTAAL code website, https://www.questaal.org. Our *GW* implementation was adapted from the original ecalj package, now at https://github.com/tkotani/ecalj/.

[25] J. Schwinger, On the Greens functions of quantized fields. I, Proc. Natl. Acad. Sci. USA **37**, 452 (1951).

[26] P. C. Martin and J. Schwinger, Theory of many-particle systems. I, Phys. Rev. **115**, 1342 (1959).

[27] G. Onida, L. Reining, and A. Rubio, Electronic excitations: Density-functional versus many-body Green's-function approaches, Rev. Mod. Phys. **74**, 601 (2002).

[28] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, All-electron self-consistent *GW* approximation: Application to Si, MnO, and NiO, Phys. Rev. Lett. **93**, 126406 (2004).

[29] B. Cunningham, M. Grüning, P. Azarhoosh, D. Pashov, and M. van Schilfgaarde, Effect of ladder diagrams on optical absorption spectra in a quasiparticle self-consistent *GW* framework, Phys. Rev. Mater. **2**, 034603 (2018).

[30] E. E. Salpeter and H. A. Bethe, A Relativistic equation for bound-state problems, Phys. Rev. **84**, 1232 (1951).

[31] A. Marini and R. Del Sole, Dynamical excitonic effects in metals and semiconductors, Phys. Rev. Lett. **91**, 176402 (2003).

[32] M. Grüning, A. Marini, and X. Gonze, Exciton-plasmon states in nanoscale materials: Breakdown of the Tamm–Dancoff approximation, Nano Lett. **9**, 2820 (2009).

[33] F. Aryasetiawan and O. Gunnarsson, Electronic structure of NiO in the GW approximation, Phys. Rev. Lett. **74**, 3221 (1995).

[34] C. Friedrich, S. Blügel, and D. Nabok, Quasiparticle self-consistent GW study of simple metals, Nanomaterials **12**, 3660 (2022).

[35] L. Sponza, P. Pisanti, A. Vishina, D. Pashov, E. Weber, M. van Schilfgaarde, S. Acharya, J. Vidal, and G. Kotliar, Self-energies in itinerant magnets: A focus on Fe and Ni, Phys. Rev. B **95**, 041112(R) (2017).

[36] M. van Schilfgaarde, T. Kotani, and S. V. Faleev, Adequacy of approximations in *GW* theory, Phys. Rev. B **74**, 245125 (2006).

[37] D. Pashov, S. Acharya, W. R. L. Lambrecht, J. Jackson, K. D. Belashchenko, A. Chantis, F. Jamet, and M. van Schilfgaarde, Questaal: A package of electronic structure methods based on the linear muffin-tin orbital technique, Comput. Phys. Commun. **249**, 107065 (2020).

[38] M. Cardona and M. L. W. Thewalt, Isotope effects on the optical spectra of semiconductors, Rev. Mod. Phys. **77**, 1173 (2005).

[39] M. Michiardi, I. Aguilera, M. Bianchi, V. E. de Carvalho, L. O. Ladeira, N. G. Teixeira, E. A. Soares, C. Friedrich, S. Blügel, and P. Hofmann, Bulk band structure of Bi_2Te_3, Phys. Rev. B **90**, 075105 (2014).

[40] A. Miglio and V. Brousseau-Couture and E. Godbout and G. Antonius, Y.-H.Chan and S.G.Louie and M. Côté and M. Giantomassi and X. Gonze, Predominance of non-adiabatic effects in zero-point renormalization of the electronic band gap, npj Comput. Mater. **6**, 167 (2020).

[41] A. Schleife, F. Fuchs, C. Rdl, J. Furthmüller, and F. Bechstedt, Band-structure and optical-transition parameters of wurtzite MgO, ZnO, and CdO from quasiparticle calculations, Phys. Status Solidi B **246**, 2150 (2009).

[42] V. I. Anisimov, F. Aryasetiawan, and A. Lichtenstein, First-principles calculations of the electronic structure and spectra of strongly correlated systems: The LDA + U method, J. Phys.: Condens. Matter **9**, 767 (1997).

[43] M. Rohlfing and S. G. Louie, Electron-hole excitations and optical spectra from first principles, Phys. Rev. B **62**, 4927 (2000).

[44] L. Reining, V. Olevano, A. Rubio, and G. Onida, Excitonic effects in solids described by time-dependent density-functional theory, Phys. Rev. Lett. **88**, 066404 (2002).

[45] The screening is much larger in a metal than an insulator, so ladders in the charge channel are generally thought to be less important. A big exception to this rule is the low-density limit, e.g., when the average density is less than that of Na [59].

[46] A. Grüneis, G. Kresse, Y. Hinuma, and F. Oba, Ionization potentials of solids: The importance of vertex corrections, Phys. Rev. Lett. **112**, 096401 (2014).

[47] J. M. An, S. V. Barabash, V. Ozolins, M. van Schilfgaarde, and K. D. Belashchenko, First-principles study of phase stability of Gd-doped EuO and EuS, Phys. Rev. B **83**, 064105 (2011).

[48] A. L. Kutepov, Full versus quasiparticle self-consistency in vertex-corrected GW approaches, Phys. Rev. B **105**, 045124 (2022).

[49] F. Bechstedt, Correlation beyond the random phase approximation: A consistent many-body perturbation theory approach, Phys. Rev. B **97**, 241109(R) (2018).

[50] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte Carlo methods for quantum impurity models, Rev. Mod. Phys. **83**, 349 (2011).

[51] D. Nabok, S. Blügel, and C. Friedrich, Electron-magnon scattering in ferromagnets from first principles by combining GW and GT self-energies, npj Comput. Mater. **7**, 178 (2021).

[52] S. Acharya, D. Pashov, C. Weber, H. Park, L. Sponza, and M. van Schilfgaarde, Evening out the spin and charge parity to increase T_c in Sr_2RuO_4, Commun. Phys. **2**, 163 (2019).

[53] E. Baldini, M. A. Sentef, S. Acharya, T. Brumme, E. Sheveleva, F. Lyzwa, E. Pomjakushina, C. Bernhard, M. van Schilfgaarde, F. Carbone, A. Rubio, and C. Weber, Electron-phonon-driven three-dimensional metallicity in an insulating cuprate, Proc. Natl. Acad. Sci. USA **117**, 6409 (2020).

[54] S. Acharya, D. Pashov, F. Jamet, and M. van Schilfgaarde, Electronic origin of T_c in bulk and monolayer FeSe, Symmetry **13**, 169 (2021).

[55] S. Acharya, D. Pashov, and M. van Schilfgaarde, Role of nematicity in controlling spin fluctuations and superconducting T_c in bulk FeSe, Phys. Rev. B **105**, 144507 (2022).

[56] S. Acharya, D. Pashov, E. Chachkarova, M. van Schilfgaarde, and C. Weber, Electronic structure correspondence of singlet-triplet scale separation in strained Sr_2RuO_4, Appl. Sci. **11**, 508 (2021).

[57] A. Toschi, A. A. Katanin, and K. Held, Dynamical vertex approximation: A step beyond dynamical mean-field theory, Phys. Rev. B **75**, 045118 (2007).

[58] E. A. Stepanov, V. Harkov, and A. I. Lichtenstein, Consistent partial bosonization of the extended Hubbard model, Phys. Rev. B **100**, 205115 (2019).

[59] J. Koskelo, L. Reining, and M. Gatti, Short-range excitonic phenomena in low-density metals, arXiv:2301.00474.

[60] M. Betzinger, C. Friedrich, A. Görling, and S. Blügel, Precise all-electron dynamical response functions: Application to COHSEX and the RPA correlation energy, Phys. Rev. B **92**, 245101 (2015).

[61] F. Aryasetiawan, R. Sakuma, and K. Karlsson, *GW* approximation with self-screening correction, Phys. Rev. B **85**, 035106 (2012).

[62] D. V. Potorochin, R. Kurleto, O. J. Clark, E. D. L. Rienks, J. Sánchez-Barriga, F. Roth, V. Voroshnin, A. Fedorov, W. Eberhardt, B. Büchner, and J. Fink, Lifetime of quasiparticles in the nearly free electron metal sodium, Phys. Rev. B **106**, 125138 (2022).

[63] A. N. Chantis, M. van Schilfgaarde, and T. Kotani, *Ab initio* prediction of conduction band spin splitting in zinc blende semiconductors, Phys. Rev. Lett. **96**, 086405 (2006).

[64] D. Deguchi, K. Sato, H. Kino, and T. Kotani, Accurate energy bands calculated by the hybrid quasiparticle self-consistent GW method implemented in the ecalj package, Jpn. J. Appl. Phys. **55**, 051201 (2016).

[65] M. Methfessel, M. van Schilfgaarde, and R. A. Casali, A full-potential lmto method based on smooth hankel functions, in *Electronic Structure and Physical Properies of Solids: The Uses of the LMTO Method Lectures of a Workshop Held at Mont Saint Odile, France, October 2–5,1998*, edited by H. Dreyssé (Springer, Berlin, 2000), pp. 114–147.

[66] T. Kotani and M. Van Schilfgaarde, Fusion of the LAPW and LMTO methods: The augmented plane wave plus muffin-tin orbital method, Phys. Rev. B **81**, 125117 (2010).

[67] O. K. Andersen, Linear methods in band theory, Phys. Rev. B **12**, 3060 (1975).

[68] T. Kotani, Quasiparticle self-consistent GW method based on the augmented plane-wave and muffin-tin orbital method, J. Phys. Soc. Jpn. **83**, 094711 (2014).

[69] C. Friedrich, S. Blügel, and A. Schindlmayr, Efficient implementation of the *GW* approximation within the all-electron FLAPW method, Phys. Rev. B **81**, 125102 (2010).

[70] W. R. L. Lambrecht, C. Bhandari, and M. van Schilfgaarde, Lattice polarization effects on the screened Coulomb interaction *W* of the *GW* approximation, Phys. Rev. Mater. **1**, 043802 (2017).

[71] C. Bhandari, M. van Schilfgaarde, T. Kotani, and W. R. L. Lambrecht, All-electron quasiparticle self-consistent *GW* band structures for $SrTiO_3$ including lattice polarization corrections in different phases, Phys. Rev. Mater. **2**, 013807 (2018).

[72] S. H. Wemple and M. DiDomenico, Behavior of the electronic dielectric constant in covalent and ionic materials, Phys. Rev. B **3**, 1338 (1971).

[73] D. Brunner, H. Angerer, E. Bustarret, F. Freudenberg, R. Hopler, R. Dimitrov, O. Ambacher, and M. Stutzmann, Optical constants of epitaxial AlGaN films and their temperature dependence, J. Appl. Phys. **82**, 5090 (1997).

[74] P. J. Dean, Lattice vibration spectra of aluminum nitride, Phys. Rev. **158**, 833 (1967).

[75] G. A. Sawatzky and J. W. Allen, Magnitude and origin of the band gap in NiO, Phys. Rev. Lett. **53**, 2339 (1984).

[76] R. J. Powell and W. E. Spicer, Optical properties of NiO and CoO, Phys. Rev. B **2**, 2182 (1970).

[77] Several photoemission studies can be found in the literature, e.g., for NiO [75,118,194]. Different studies yield somewhat different results: Ref. [118] in particular shows that the observed spectra depend significantly on the instrument used and the photon energy.

[78] J. M. Tomczak, M. van Schilfgaarde, and G. Kotliar, Many-body effects in iron pnictides and chalcogenides: Nonlocal versus dynamic origin of effective masses, Phys. Rev. Lett. **109**, 237010 (2012).

[79] E. G. C. P. van Loon, M. Rösner, M. I. Katsnelson, and T. O. Wehling, Random phase approximation for gapped systems: Role of vertex corrections and applicability of the constrained random phase approximation, Phys. Rev. B **104**, 045134 (2021).

[80] Z. Yan, J. P. Perdew, and S. Kurth, Density functional for short-range correlation: Accuracy of the random-phase approximation for isoelectronic energy changes, Phys. Rev. B **61**, 16430 (2000).

[81] A. Irmler, A. Gallo, F. Hummel, and A. Grüneis, Duality of ring and ladder diagrams and its importance for many-electron perturbation theories, Phys. Rev. Lett. **123**, 156401 (2019).

[82] A. Ruzsinszky, J. P. Perdew, and G. I. Csonka, A simple but fully nonlocal correction to the random phase approximation, J. Chem. Phys. **134**, 114110 (2011).

[83] M. Grüning, A. Marini, and A. Rubio, Density functionals from many-body perturbation theory: The band gap for semiconductors and insulators, J. Chem. Phys. **124**, 154108 (2006).

[84] T. C. Chiang, J. A. Knapp, M. Aono, and D. E. Eastman, Angle-resolved photoemission, valence-band dispersions $e(\vec{k})$, and electron and hole lifetimes for GaAs, Phys. Rev. B **21**, 3513 (1980).

[85] P. Lautenschlager, M. Garriga, S. Logothetidis, and M. Cardona, Interband critical points of GaAs and their temperature dependence, Phys. Rev. B **35**, 9174 (1987).

[86] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, Band parameters for III-V compound semiconductors and their alloys, J. Appl. Phys. **89**, 5815 (2001).

[87] S. Adachi, *Handbook on Physical Properties of Semiconductors* (Springer, Berlin, 2007).

[88] C. Neumann, A. Nöthe, and N. O. Lipari, Two-photon magnetoabsorption of ZnTe, CdTe, and GaAs, Phys. Rev. B **37**, 922 (1988).

[89] M. Wu and C. C. Chen, Photoluminescence of high-quality GaSb grown from Ga- and Sb-rich solutions by liquid-phase epitaxy, J. Appl. Phys. **72**, 4275 (1992).

[90] J. C. E. Rasch, T. Stemmler, B. Muller, L. Dudy, and R. Manzke, 1T-$TiSe_2$: Semimetal or semiconductor? Phys. Rev. Lett. **101**, 237602 (2008).

[91] C. J. Vesely, R. L. Hengehold, and D. W. Langer, uv photoemission measurements of the upper d levels in the IIB-VIA compounds, Phys. Rev. B **5**, 2296 (1972).

[92] L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, and D. A. Shirley, Total valence-band densities of states of III-V and II-VI compounds from x-ray photoemission spectroscopy, Phys. Rev. B **9**, 600 (1974).

[93] D. E. Eastman, T. C. Chiang, P. Heimann, and F. J. Himpsel, Surface core-level binding-energy shifts for GaAs(110) and GaSb(110), Phys. Rev. Lett. **45**, 656 (1980).

[94] K. Reimann and S. Rübenacke, Two-photon absorption in CuCl and CuBr under hydrostatic pressure, Phys. Rev. B **49**, 11021 (1994).

[95] C. Uihlein, D. Fröhlich, and R. Kenklies, Investigation of exciton fine structure in Cu_2O, Phys. Rev. B **23**, 2731 (1981).

[96] I. Ferrer, D. Nevskaia, C. de las Heras, and C. Sánchez, About the band gap nature of FeS_2 as determined from optical and photoelectrochemical measurements, Solid State Commun. **74**, 913 (1990).

[97] L. Roth, B. Lax, and S. Zwerdling, Theory of optical magnetoabsorption effects in semiconductors, Phys. Rev. **114**, 90 (1959).

[98] P. Rinke, M. Winkelnkemper, A. Qteish, D. Bimberg, J. Neugebauer, and M. Scheffler, Consistent set of band parameters for the group-III nitrides AlN, GaN, and InN, Phys. Rev. B **77**, 075202 (2008).

[99] *Semiconductors Group IV Elements, IV-IV and III-V Compounds. Part b: Electronic, Transport, Optical and Other Properties*, edited by O. Madelung, U. Rössler, and M. Schulz (Springer, Berlin, 1998).

[100] S. Acharya, D. Pashov, B. Cunningham, A. N. Rudenko, M. Rösner, M. Grüning, M. van Schilfgaarde, and M. I. Katsnelson, Electronic structure of chromium trihalides beyond density functional theory, Phys. Rev. B **104**, 155109 (2021).

[101] W. Bond, Measurement of the refractive indices of several crystals, J. Appl. Phys. **36**, 1674 (1965).

[102] M. P. Lisitsa, L. F. Gudymenko, V. N. Malinko, and S. F. Terekhova, Dispersion of the refractive indices and birefringence of CdS_xSe_{1-x} single crystals, Phys. Status Solidi **31**, 389 (1969).

[103] M. Cardona and G. Harbeke, Optical properties and band structure of wurtzite-type crystals and rutile, Phys. Rev. **137**, A1467 (1965).

[104] T. Toyoda, The temperature dependence of the optical dispersion parameters in $SrTiO_3$ and TiO_2, J. Phys. D: Appl. Phys. **18**, L129 (1985).

[105] T. Ishii and T. Sato, Growth of single crystals of hexagonal boron nitride, J. Cryst. Growth **61**, 689 (1983).

[106] R. Geick, C. H. Perry, and G. Rupprecht, Normal modes in hexagonal boron nitride, Phys. Rev. **146**, 543 (1966).

[107] W. J. Moore, J. A. Freitas, R. T. Holm, O. Kovalenkov, and V. Dmitriev, Infrared dielectric function of wurtzite aluminum nitride, Appl. Phys. Lett. **86**, 141912 (2005).

[108] I. Roskovcová, J. Pastrňáak, and R. Babuškováa, The dispersion of the refractive index and the birefringence of AlN, Phys. Status Solidi B **20**, K29 (1967).

[109] S. Ballard, J. Browder, and J. Ebersole, in *American Institute of Physics Handbook*, edited by D. Gray (McGraw-Hill, New York, 1972), Chap. 6, p. 12.

[110] T. Ruf, M. Cardona, C. S. J. Pickles, and R. Sussmann, Temperature dependence of the refractive index of diamond up to 925 K, Phys. Rev. B **62**, 16578 (2000).

[111] M. E. Lines, Bond-orbital theory of linear and nonlinear electronic response in ionic crystals. I. Linear response, Phys. Rev. B **41**, 3372 (1990).

[112] E. Burstein, H. Brodsky, and G. Lucousky, The dynamic ionic charge of zincblende type crystals. Int. J. Quantum Chem. **1**, 759 (1967).

[113] A. Feldman and D. Horowitz, Refractive index of cuprous chloride, J. Opt. Soc. Am. **59**, 1406 (1969).

[114] T. Sueta, T. Matsushima, T. Nishimoto, and T. Makimoto, Modulation of 10.6-micron laser radiation by CuCl, Proc. IEEE **58**, 1378 (1970).

[115] H. Hüfner, J. Osterwalder, T. Riesterer, and F. Hulliger, Photoemission and inverse photoemission spectroscopy of NiO, Solid State Commun. **52**, 793 (1984).

[116] G. Chern, S. D. Berry, H. Mathias, and L. R. Testardi, Observation of interfacial electrical polarization in Fe_3O_4/NiO superlattices, Phys. Rev. Lett. **68**, 114 (1992).

[117] C. Pecharroman and J. Iglesias, A method for the determination of infrared optical constants from reflectance measurements on powdered samples, J. Phys.: Condens. Matter **6**, 7125 (1994).

[118] R. Zimmermann, P. Steiner, R. Claessen, F. Reinert, S. Hüfner, P. Blaha, and P. Dufek, Electronic structure of 3d-transition-metal oxides: On-site Coulomb repulsion versus covalency, J. Phys.: Condens. Matter **11**, 1657 (1999).

[119] M. O'Keeffe, Infrared optical properties of cuprous oxide, J. Chem. Phys. **39**, 1789 (1963).

[120] J. Pellicer-Porres, A. Segura, and D. Kim, Refractive index of the $CuAlO_2$ delafossite, Semicond. Sci. Technol. **24**, 015002 (2009).

[121] V. H. Vu, B. Q. Tu, Q. X. Phung, V. T. Tran, N. N. Hoang, D. D. Pham, T. A. Mai, H. D. Tong, M. Van Nguyen, H. Q. Nguyen, H. M. Nguyen, H. Van Mai, D. C. Duong, Q. M. Doan, and T. Nguyen-Tran, Tailoring optical and resistance properties of the functional $CuAl_xO_y$ semiconductor for applications as thermal infrared imagers, J. Sci.: Adv. Mater. Devices **6**, 202 (2021).

[122] T. Ito, H. Yamaguchi, T. Masumi, and S. Adachi, Optical properties of CuO studied by spectroscopic ellipsometry, J. Phys. Soc. Jpn. **67**, 3304 (1998).

[123] D. Tahir and S. Tougaard, Electronic and optical properties of Cu, CuO and Cu_2O studied by electron spectroscopy, J. Phys.: Condens. Matter **24**, 175002 (2012).

[124] B. Prevot, J. Briellman, M. P. Meftah, and M. Sieskind, Infrared reflectivity of non-stoichiometric ferrous oxide, Phys. Status Solidi A **40**, 503 (1977).

[125] G. Kugel, C. Carabatos, B. Hennion, B. Prevot, A. Revcolevschi, and D. Tocchetti, Lattice dynamics of wustite (FeO), Phys. Rev. B **16**, 378 (1977).

[126] K. V. Rao and A. Smakula, Dielectric properties of cobalt oxide, nickel oxide, and their mixed crystals, J. Appl. Phys. **36**, 2031 (1965).

[127] P. J. Gielisse, J. N. Plendl, and L. C. Mansur, Infrared properties of NiO and CoO and their mixed crystals, J. Appl. Phys. **36**, 2446 (1965).

[128] J. Plendl, L. Mansur, S. Mitra, and I. Chan, Reststrahlen spectrum of MnO, Solid State Commun. **7**, 109 (1969).

[129] D. Tamme, R. Schepe, and K. Henneberger, Comment on "Self-consistent calculations of quasiparticle states in metals and semiconductors", Phys. Rev. Lett. **83**, 241 (1999).

[130] Z. H. Levine and D. C. Allan, Linear optical response in silicon and germanium including self-energy effects, Phys. Rev. Lett. **63**, 1719 (1989).

[131] R. DelSole and R. Girlanda, Optical properties of semiconductors within the independent-quasiparticle approximation, Phys. Rev. B **48**, 11789 (1993).

[132] D. Rocca, M. Vörös, A. Gali, and G. Galli, Ab initio optoelectronic properties of silicon nanoparticles: Excitation energies, sum rules, and Tamm–Dancoff approximation, J. Chem. Theory Comput. **10**, 3290 (2014).

[133] C. H. Patterson, Photoabsorption spectra of small Na clusters: TDHF and BSE versus CI and experiment, Phys. Rev. Mater. **3**, 043804 (2019).

[134] T. Sander, E. Maggio, and G. Kresse, Beyond the Tamm-Dancoff approximation for extended systems using exact diagonalization, Phys. Rev. B **92**, 045209 (2015).

[135] D. M. Roessler and W. C. Walker, Optical constants of magnesium oxide and lithium fluoride in the far ultraviolet, J. Opt. Soc. Am. **57**, 835 (1967).

[136] M. Piacentini, D. W. Lynch, and C. G. Olson, Thermoreflectance of LiF between 12 and 30 eV, Phys. Rev. B **13**, 5530 (1976).

[137] W. L. Smith, J. Bechtel, and N. Bloembergen, Picosecond laser-induced breakdown at 5321 and 3547 Å: Observation of frequency-dependent behavior, Phys. Rev. B **15**, 4039 (1977).

[138] R. Adair, L. L. Chase, and S. A. Payne, Nonlinear refactive index of optical crystals, Phys. Rev. B **39**, 3337 (1989).

[139] L. Fu and C. L. Kane, Topological insulators with inversion symmetry, Phys. Rev. B **76**, 045302 (2007).

[140] I. G. Austin, The optical properties of bismuth telluride, Proc. Phys. Soc. **72**, 545 (1958).

[141] C. Li, A. L. Ruoff, and C. W. Spencer, Effect of pressure on the energy gap of Bi_2Te_3, J. Appl. Phys. **32**, 1733 (1961).

[142] R. Sehr and L. Testardi, The optical properties of p-type Bi_2Te_3-Sb_2Te_3 alloys between 2-15 microns, J. Phys. Chem. Solids **23**, 1219 (1962).

[143] G. A. Thomas, D. H. Rapkine, R. B. van Dover, L. F. Mattheiss, W. A. Sunder, L. F. Schneemeyer, and J. V. Waszczak, Large electronic-density increase on cooling a layered metal: Doped Bi_2Te_3, Phys. Rev. B **46**, 1553 (1992).

[144] R. Deng, B. D. Ozsdolay, P. Y. Zheng, S. V. Khare, and D. Gall, Optical and transport measurement and first-principles determination of the ScN band gap, Phys. Rev. B **91**, 045104 (2015).

[145] Q. Li and V. Thangadurai, A comparative 2 and 4-probe DC and 2-probe AC electrical conductivity of novel co-doped $Ce_{0.9-x}RE_xMo_{0.1}O_{2.10.5x}$ (RE = Y, Sm, Gd; x = 0.2, 0.3), J. Mater. Chem. **20**, 7970 (2010).

[146] P. C. A. Brito, D. A. A. Santos, J. G. S. Duque, and M. A. Macédo, Structural and magnetic study of Fe-doped CeO_2, Phys. B (Amsterdam) **405**, 1821 (2010).

[147] L. Sandhya Kumari, P. Prabhakar Rao, and M. Lakshmipathy Reddy, Environment-friendly red pigments from CeO_2-Fe_2O_3-Pr_6O_{11} solid solutions, J. Alloys Compd. **461**, 509 (2008).

[148] G. Zhang, L. Li, G. Li, X. Qiu, and G. Yan, Hydrothermal processing and characterization of $Ce_{1-x}Pb_xO_{2-\delta}$ solid solutions, Solid State Sci. **11**, 671 (2009).

[149] F. Chevire, F. Munoz, C. F. Baker, F. Tessier, O. Larcher, S. Boujday, C. Colbeau-Justin, and R. Marchand, UV absorption properties of ceria-modified compositions within the fluorite-type solid solution CeO_2–Y_6WO_{12}, J. Solid State Chem. **179**, 3184 (2006).

[150] S. Guo, H. Arwin, S. N. Jacobsen, K. Järrendahl, and U. Helmersson, A spectroscopic ellipsometry study of cerium dioxide thin films grown on sapphire by rf magnetron sputtering, J. Appl. Phys. **77**, 5369 (1995).

[151] F. Marabelli and P. Wachter, Covalent insulator CeO_2: Optical reflectivity measurements, Phys. Rev. B **36**, 1238 (1987).

[152] T. Inoue, Y. Yamamoto, S. Koyama, S. Suzuki, and Y. Ueda, Epitaxial growth of CeO_2 layers on silicon, Appl. Phys. Lett. **56**, 1332 (1990).

[153] S. Shi, X. Ke, C. Ouyang, H. Zhang, H. Ding, Y. Tang, W. Zhou, P. Li, M. Lei, and W. Tang, First-principles investigation of the bonding, optical and lattice dynamical properties of CeO_2, J. Power Sources **194**, 830 (2009).

[154] N. Pradhani, P. K. Mahapatra, and R. N. P. Choudhary, Effect of cerium oxide addition on optical, electrical and dielectric characteristics of $(Bi_{0.5}Na_{0.5})TiO_3$ ceramics, J. Phys. Mater. **1**, 015007 (2018).

[155] H. Tang, F. Levy, H. Berger, and P. E. Schmid, Urbach tail of anatase TiO_2, Phys. Rev. B **52**, 7771 (1995).

[156] S. Shin, T. Ishii, T. Ejima, S. Suzuki, and S. Sato, Photoemission and bremsstrahlung isochromat spectroscopy studies of TiO_2 (rutile) and $SrTiO_3$, J. Phys. Soc. Jpn. **63**, 2612 (1994).

[157] K. van Benthem, C. Elsässer, and R. H. French, Bulk electronic structure of $SrTiO_3$: Experiment and theory, J. Appl. Phys. **90**, 6156 (2001).

[158] M. Cardona, Optical properties and band structure of $SrTiO_3$ and $BaTiO_3$, Phys. Rev. **140**, A651 (1965).

[159] H. Yanagi, S. Inoue, K. Ueda, H. Kawazoe, H. Hosono, and N. Hamada, Electronic structure and optoelectronic properties of transparent p-type conducting $CuAlO_2$, J. Appl. Phys. **88**, 4159 (2000).

[160] A. N. Banerjee, C. K. Ghosh, S. Das, and K. K. Chattopadhyay, Electro-optical characteristics and field-

emission properties of reactive DC-sputtered p-$CuAlO_{2+x}$ thin films, Phys. B (Amsterdam) **370**, 264 (2005).

[161] J. Pellicer-Porres, A. Segura, A. S. Gilliland, A. Muñoz, P. Rodríguez-Hernández, D. Kim, M. S. Lee, and T. Y. Kim, On the band gap of $CuAlO_2$ delafossite, Appl. Phys. Lett. **88**, 181904 (2006).

[162] J. Tate, H. L. Ju, J. C. Moon, A. Zakutayev, A. P. Richard, J. Russell, and D. H. McIntyre, Origin of *p*-type conduction in single-crystal $CuAlO_2$, Phys. Rev. B **80**, 165206 (2009).

[163] H. Kawazoe, H. Yasakuwa, H. Hyodo, M. Kurita, H. Yanagi, and H. Hosono, p-type electrical conduction in transparent thin films of $CuAlO_2$, Nature (London) **389**, 939 (1997).

[164] D. S. Kim, S. J. Park, E. K. Jeong, H. K. Lee, and S. Y. Choi, Optical and electrical properties of p-type transparent conducting $CuAlO_2$ thin film, Thin Solid Films **515**, 5103 (2007).

[165] D. O. Scanlon and G. W. Watson, Conductivity limits in $CuAlO_2$ from screened-hybrid density functional theory, J. Phys. Chem. Lett. **1**, 3195 (2010).

[166] J. Vidal, F. Trani, F. Bruneval, M. A. L. Marques, and S. Botti, Effects of electronic and lattice polarization on the band structure of delafossite transparent conductive oxides, Phys. Rev. Lett. **104**, 136401 (2010).

[167] S. Li, X. Zhang, and P. Z. *et al.*, Preparation and characterization of solution-processed nanocrystalline p-type $CuAlO_2$ thin-film transistors., Nanoscale Res. Lett. **13**, 259 (2018).

[168] S. Shin, S. Suga, M. Taniguchi, M. Fujisawa, H. Kanzaki, A. Fujimori, H. Daimon, Y. Ueda, K. Kosuge, and S. Kachi, Vacuum-ultraviolet reflectance and photoemission study of the metal-insulator phase transitions in VO_2, V_6O_{13}, and V_2O_3, Phys. Rev. B **41**, 4993 (1990).

[169] V. M. Bermudez, R. T. Williams, J. P. Long, R. K. Reed, and P. H. Klein, Photoemission study of hydrogen adsorption on vanadium dioxide near the semiconductor-metal phase transition, Phys. Rev. B **45**, 9266 (1992).

[170] R. M. Wentzcovitch, W. W. Schulz, and P. B. Allen, VO_2: Peierls or Mott-Hubbard? A view from band theory, Phys. Rev. Lett. **72**, 3389 (1994).

[171] A. Liebsch, H. Ishida, and G. Bihlmayer, Coulomb correlations and orbital polarization in the metal-insulator transition of VO_2, Phys. Rev. B **71**, 085109 (2005).

[172] S. Biermann, A. Poteryaev, A. I. Lichtenstein, and A. Georges, Dynamical singlets and correlation-assisted Peierls transition in VO_2, Phys. Rev. Lett. **94**, 026404 (2005).

[173] M. Gatti, F. Bruneval, V. Olevano, and L. Reining, Understanding correlations in vanadium dioxide from first principles, Phys. Rev. Lett. **99**, 266402 (2007).

[174] W. H. Brito, M. C. O. Aguiar, K. Haule, and G. Kotliar, Metal-insulator transition in VO_2: A DFT+DMFT perspective, Phys. Rev. Lett. **117**, 056402 (2016).

[175] T. M. Rice, H. Launois, and J. P. Pouget, Comment on "VO_2: Peierls or Mott-Hubbard? A view from band theory", Phys. Rev. Lett. **73**, 3042 (1994).

[176] C. Weber, S. Acharya, B. Cunningham, M. Grüning, L. Zhang, H. Zhao, Y. Tan, Y. Zhang, C. Zhang, K. Liu, M. Van Schilfgaarde, and M. Shalaby, Possible phonon-induced electronic bi-stability in VO_2 for ultrafast memory at room temperature, Phys. Rev. Res. **2**, 023076 (2020).

[177] K. Okazaki, S. Sugai, Y. Muraoka, and Z. Hiroi, Role of electron-electron and electron-phonon interaction effects in the optical conductivity of VO_2, Phys. Rev. B **73**, 165116 (2006).

[178] W. L. Roth, Magnetic structures of MnO, FeO, CoO, and NiO, Phys. Rev. **110**, 1333 (1958).

[179] C. Rödl, F. Fuchs, J. Furthmüller, and F. Bechstedt, Quasiparticle band structures of the antiferromagnetic transition-metal oxides MnO, FeO, CoO, and NiO, Phys. Rev. B **79**, 235114 (2009).

[180] C. Rödl and F. Bechstedt, Optical and energy-loss spectra of the antiferromagnetic transition metal oxides MnO, FeO, CoO, and NiO including quasiparticle and excitonic effects, Phys. Rev. B **86**, 235122 (2012).

[181] J. van Elp, J. L. Wieland, H. Eskes, P. Kuiper, G. A. Sawatzky, F. M. F. de Groot, and T. S. Turner, Electronic structure of CoO, Li-doped CoO, and LiCoO, Phys. Rev. B **44**, 6090 (1991).

[182] J. Ghijsen, L. H. Tjeng, J. van Elp, H. Eskes, J. Westerink, G. A. Sawatzky, and M. T. Czyzyk, Electronic structure of Cu_2O and CuO, Phys. Rev. B **38**, 11322 (1988).

[183] C. P. Massolo, M. Renteria, J. Desimoni, and A. G. Bibiloni, Electronic structure of Cu_2O and CuO, Phys. Rev. B **37**, 4743 (1988).

[184] J. P. Falck, A. Levy, M. A. Kastner, and R. J. Birgeneau, Charge-transfer spectrum and its temperature dependence in La_2CuO_4, Phys. Rev. Lett. **69**, 1109 (1992).

[185] D. Reagor, E. Ahrens, S. W. Cheong, A. Migliori, and Z. Fisk, Large dielectric constants and massive carriers in La_2CuO_2, Phys. Rev. Lett. **62**, 2048 (1989).

[186] C. Y. Chen, R. J. Birgeneau, M. A. Kastner, N. W. Preyer, and T. Thio, Frequency and magnetic-field dependence of the dielectric constant and conductivity of La_2CuO_{4+y}, Phys. Rev. B **43**, 392 (1991).

[187] V. I. Anisimov, Band-structure description of Mott insulators (NiO, MnO, FeO, CoO), J. Phys.: Condens. Matter **2**, 3973 (1990).

[188] H. A. Alperin, The magnetic form factor of nickel oxide, J. Phys. Soc. Japan Suppl. B **17**, 12 (1962).

[189] B. E. F. Fender, A. J. Jacobson, and J. F. A. Wedgwood, Covalency parameters in MnO, α-MnS, and NiO, Chem. Phys. **48**, 990 (1968).

[190] S. I. Csiszar, M. W. Haverkort, Z. Hu, A. Tanaka, H. H. Hsieh, H.-J. Lin, C. T. Chen, T. Hibma, and L. H. Tjeng, Controlling orbital moment and spin orientation in CoO layers by strain, Phys. Rev. Lett. **95**, 187205 (2005).

[191] B. X. Yang, J. M. Tranquada, and G. Shirane, Neutron scattering studies of the magnetic structure of cupric oxide, Phys. Rev. B **38**, 174 (1988).

[192] R. G. Graham, D. Fowler, J. S. Lord, P. C. Riedi, and B. M. R. Wanklyn, Pressure dependence of the nuclear resonance of antiferromagnetic CuO, Phys. Rev. B **44**, 7091 (1991).

[193] T. A. Kaplan and S. D. Mahanti, Spin fluctuations and covalence in neutron Bragg scattering from $YBa_2Cu_3O_6$ and La_2CuO_4, J. Appl. Phys. **69**, 5382 (1991).

[194] Z.-X. Shen, R. S. List, D. S. Dessau, B. O. Wells, O. Jepsen, A. J. Arko, R. Barttlet, C. K. Shih, F. Parmigiani, J. C. Huang, and P. A. P. Lindberg, Electronic structure of NiO: Correlation and band effects, Phys. Rev. B **44**, 3604 (1991).

[195] S. Mandal, K. Haule, K. Rabe, and D. Vanderbilt, Systematic beyond-DFT study of binary transition metal oxides, npj Comput. Mater. **5**, 115 (2019).

[196] L. Messick, W. C. Walker, and R. Glosser, Direct and temperature-modulated reflectance spectra of MnO, CoO, and NiO, Phys. Rev. B **6**, 3941 (1972).

[197] J. van Elp, R. H. Potze, H. Eskes, R. Berger, and G. A. Sawatzky, Electronic structure of MnO, Phys. Rev. B **44**, 1530 (1991).

[198] C. W. Pratt and R. Coelho, Optical absorption of CoO and MnO above and below the Néel temperature, Phys. Rev. **116**, 281 (1959).

[199] A. Treindl and K. H. Germann, Linear optical anisotropy investigations of the antiferromagnetic phase transition in MnO, Phys. Status Solidi B **80**, 159 (1977).

[200] D. R. Huffman, R. L. Wild, and M. Shinmei, Optical absorption spectra of crystal-field transitions in MnO, J. Chem. Phys. **50**, 4092 (1969).

[201] H. Bowen, D. Adler, and B. Auker, Electrical and optical properties of FeO, J. Solid State Chem. **12**, 355 (1975).

[202] P. Wei and Z. Q. Qi, Insulating gap in the transition-metal oxides: A calculation using the local-spin-density approximation with the on-site Coulomb U correlation correction, Phys. Rev. B **49**, 10864 (1994).

[203] N. Hiraoka, H. Okamura, H. Ishii, I. Jarrige, K. D. Tsuei, and Y. Q. Cai, Charge transfer and dd excitations in transition metal oxides, Eur. Phys. J. B **70**, 157 (2009).

[204] S. Åsbrink and L.-J. Norrby, A refinement of the crystal structure of copper(II) oxide with a discussion of some exceptional e.s.d.'s, Acta Crystallogr., Sec. B **26**, 8 (1970).

[205] A. Filippetti and V. Fiorentini, Magnetic ordering in CuO from first principles: A cuprate antiferromagnet with fully three-dimensional exchange interactions, Phys. Rev. Lett. **95**, 086405 (2005).

[206] E. Verwey, Electronic conduction of magnetite (Fe_3O_4) and its transition point at low temperatures, Nature (London) **144**, 327 (1939).

[207] J.-H. Park, L. H. Tjeng, J. W. Allen, P. Metcalf, and C. T. Chen, Single-particle gap above the Verwey transition in Fe_3O_4, Phys. Rev. B **55**, 12813 (1997).

[208] K. Jordan, A. Cazacu, G. Manai, S. F. Ceballos, S. Murphy, and I. V. Shvets, Scanning tunneling spectroscopy study of the electronic structure of Fe_3O_4 surfaces, Phys. Rev. B **74**, 085416 (2006).

[209] H. Liu and C. D. Valentin, Band gap in magnetite above Verwey temperature induced by symmetry breaking, J. Phys. Chem. C **121**, 25736 (2017).

[210] A. L. Kutepov, V. S. Oudovenko, and G. Kotliara, Linearized self-consistent quasiparticle GW method: Application to semiconductors and simple metals, Comput. Phys. Commun. **219**, 407 (2017).