

Fully converged plane-wave-based self-consistent GW calculations of periodic solidsHuawei Cao,^{1,2} Zhongyuan Yu,¹ Pengfei Lu,¹ and Lin-Wang Wang^{2,*}¹*State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, P.O. Box 49, Beijing 100876, China*²*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

(Received 31 July 2016; revised manuscript received 18 November 2016; published 23 January 2017)

The GW approximation is a well-known method to obtain the quasiparticle and spectral properties of systems ranging from molecules to solids. In practice, GW calculations are often employed with many different approximations and truncations. In this work, we describe the implementation of a fully self-consistent GW approach based on the solution of the Dyson equation using a plane wave basis set. Algorithmic, numerical, and technical details of the self-consistent GW approach are presented. The fully self-consistent GW calculations are performed for GaAs, ZnO, and CdS including semicores in the pseudopotentials. No further approximations and truncations apart from the truncation on the plane wave basis set are made in our implementation of the GW calculation. After adopting a special potential technique, a ~ 100 Ry energy cutoff can be used without the loss of accuracy. We found that the self-consistent GW (sc- GW) significantly overestimates the bulk band gaps, and this overestimation is likely due to the underestimation of the macroscopic dielectric constants. On the other hand, the sc- GW accurately predicts the d -state positions, most likely because the d -state screening does not sensitively depend on the macroscopic dielectric constant. Our work indicates the need to include the high-order vertex term in order for the many-body perturbation theory to accurately predict the semiconductor band gaps. It also sheds some light on why, in some cases, the G_0W_0 bulk calculation is more accurate than the fully self-consistent GW calculation, because the initial density-functional theory has a better dielectric constant compared to experiments.

DOI: [10.1103/PhysRevB.95.035139](https://doi.org/10.1103/PhysRevB.95.035139)**I. INTRODUCTION**

Density-functional theory (DFT) [1,2] in the Kohn-Sham (KS) scheme has been widely used to study the electronic structure in solid-state physics. In spite of its success in describing the ground state properties, DFT suffers from a severe band gap problem [3]. Thus the KS eigenvalues cannot be used to interpret quasiparticle excitations as measured by photoemission spectroscopy or optical absorption. In the past two decades, the GW approximation [4] derived from many-body perturbation theory has been widely used to study the quasiparticle energy and excitation spectra for real materials. In practice, large numbers of procedures employing different approximations have been used [5–10]. The simplest GW approach is performed non-self-consistently for the evaluation of the quasiparticle self-energy [11,12]. The excitation energies are then obtained from the first-order perturbation theory as corrections to the DFT single-particle eigenenergies. For simple s - p bonded materials, the calculated band gaps with single-shot approximation (also called as G_0W_0) are considerably improved upon the DFT results and show good agreement with experiments [11–16]. However, there could be a strong dependence of G_0W_0 results on the initial single-particle Hamiltonian [e.g., DFT or Hartree-Fock (HF) or its hybrid method like HSE]. Different initial wave functions and eigenenergies can yield widely different band gaps, e.g., differ by 1 eV for some oxides [17]. Furthermore, traditional G_0W_0 approximation does not fulfill some microscopic conservation laws. Schindlmayr [18] found that there was a genuine violation of particle number conservation if the self-energy was

not calculated self-consistently. In addition, the total energy which can be regarded as an explicit function of the Green's function G varies a lot when it is implemented in different GW methods without self-consistency [19]. For all these reasons, it becomes interesting to try self-consistent GW (sc- GW) calculations with the hope that many of these problems will be rectified.

However, there could be drawbacks for the self-consistent procedure. Not only is it more expensive, Holm and von Barth [20] concluded that the self-consistency for homogeneous electron gas in the GW calculation tends to worsen the agreement of the band structure to the experimental results (when compared to G_0W_0). In addition, the weight of the plasmon satellite disappeared in the spectral function due to the self-consistent calculations. For real and nonmetallic systems, Schöne and Eguiluz [21] found that GW calculations under the shielded-interaction approximation and a full updated Green's function G and screened potential W can overestimate the band gap of silicon by as much as the DFT underestimates it. These conclusions seem to be in conflict with recent studies [22–24] on isolated molecule systems. They found that the accuracy of sc- GW ionization energies is comparable with that of non-self-consistent G_0W_0 with a DFT starting point. Thus, it is interesting to revisit this problem for bulk materials, especially using approaches where many of the approximations in the truncations are removed. We would like to know, what are the true effects of GW self-consistency for periodic systems?

There are many approaches to achieve self-consistency in the GW approximation. Many self-consistent GW calculations employ the noninteracting expression [25–28] to describe the Green's function G :

$$G(r_1, r_2, \omega) = \sum_i \psi_i^*(r_1) \psi_i(r_2) / (\epsilon_i - \mu - \omega \pm i\delta). \quad (1)$$

*lwwang@lbl.gov

Here $\psi_i(r)$ is the single-particle eigenwave functions, ε_i is its eigenenergy, and the $+/-$ sign depends on whether ε_i is above or below the Fermi energy μ . Although, as will be discussed below, in true self-consistent GW calculations, the Green's function can no longer be described by Eq. (1), in practice, the self-consistency in many studies is realized by only changing the eigenenergy ε_i and updating the wave function $\psi_i(r)$ while keeping the formulation of Eq. (1). There are also other options, e.g., only updating W in the G_0W method, or only updating G in the GW_0 approach. There could be an improved version of Eq. (1), replacing the energy-dependent denominator by a more general term $f_i(\omega)$, which is called the diagonal approximation. Unfortunately, there is no unique way to carry out the updating of Eq. (1) and one can propose different self-consistent schemes [17]. The true Green's function G should be described by solving the Dyson equation self-consistently:

$$G^{-1}(i\omega) = i\omega + \mu - H - \Sigma(i\omega). \quad (2)$$

Here the H is the single-particle Hamiltonian, μ is the Fermi energy, and Σ is the self-energy term. Under GW approximation, $\Sigma = iGW$, and the corresponding Dyson equation is a variational solution of the Klein total energy [29] expressed as a functional of the Green's function G . This is very much like the Kohn-Sham equation is the variational total energy minimum solution of the DFT energy. Furthermore, it can be proved that the quasiparticle eigenenergy of Eq. (2) is the difference of the Klein energies of $N + 1$ and N electron systems [30,31], much like the Kohn-Sham eigenenergy is the DFT total energy difference of the $N + 1$ and N electron systems. In addition, according to Baym and Kadanoff [32,33] many conservation laws (such as momentum, total energy, and particle numbers) are preserved following Dyson's equation. The conserving character is an important property in transport calculations [34].

In order to satisfy the variational Dyson equation (2), the Green's function G can no longer be described by the single-particle expression of Eq. (1). Instead, it is a full matrix for a given ω , either expanded by the plane wave basis set $\exp(iqr)$, or by the single-particle eigenstate basis set ψ_i . In contrast, Eq. (1) contains only the diagonal term under the basis set of ψ_i . Thus, using Eq. (1) is taking into account only the diagonal terms under the basis set ψ_i . There are some previous works for the importance of the off-diagonal terms. Fleszar and Hanke [35] concluded that the role of the off-diagonal elements is negligible. However, Sakuma *et al.* [36] reported that the off-diagonal elements of the self-energy are crucial and have a large influence on the quasiparticle band gap of correlated materials.

The choice of basis set is also an important issue in solving the GW problem. A majority of GW calculations are carried out using pseudopotentials. For some systems, this could be problematic. Ku and Eguluz [37] claimed that pseudopotential-based GW schemes carry a built-in error and the preferred procedure is to perform all-electron calculations based on the full-potential linearized augmented plane wave (FP-LAPW) or the linearized muffin-tin orbital, although this work was later questioned by Tiago *et al.* [38] for its conduction band convergence. Faleev *et al.* [28] also questioned the

validity of pseudopotential in the GW calculations and showed that G_0W_0 with pseudopotential can lead to systematic errors. This is because the pseudopotentials are generated for DFT calculation using a semilocal exchange-correlation functional. In GW calculation, the pseudo wave functions can yield an incorrect screened exchange integral. Thus a semicore will be needed, which will make the valence pseudo wave functions have the right shape, hence the correct exchange integral. In the work of Lilienfeld and Schultz [39], the effects of the semicore were discussed for DFT calculations. It shows that the inclusion of semicore can significantly change the band gap, making it closer to the all-electron results. It is now accepted by many groups that the inclusion of the semicore is necessary to make the pseudopotential-based GW result similar to that of the all-electron calculation. As we will show later, our pseudopotential results including semicores are indeed close to the all-electron results for G_0W_0 calculations. Another option of the basis set is the use of atomic orbitals or other localized basis sets (such as a Gaussian basis set) [23,24,34,40,41], which could be more efficient for molecular systems. In the current study, we will use plane wave pseudopotentials with semicores.

One common problem of the GW calculations is the lack of numerical convergence caused by a finite number of unoccupied states. Ideally, the complete sets of unoccupied states need to be included to expand the Green's function. In practice, this inclusion is often truncated. According to Delaney *et al.* [42], high energy (8–10 Ry above Fermi energy) eigenstates are required to provide accurate numerical results. Shih *et al.* [43] have calculated the quasiparticle band gap of ZnO and found that 3000 bulk conduction bands were needed to obtain a converged GW band gap. However, in many GW calculations, only a few hundred conduction bands are used, which can result in an unconverged band gap as shown for ZnO [8,10,44].

In this work, we employ a sc- GW calculation without resorting to diagonal- G approximation and conduction band truncations. We would like to know (1) whether the true Dyson equation results improve upon the G_0W_0 results; (2) how much error remains; and (3) what causes the remaining errors. The full solution of Eq. (2) is only made possible with the use of large supercomputers. In this work, the Green's function $G(r_1, r_2, \omega)$ is represented numerically in its full matrix form under the plane wave basis set at different ω points without the truncation for the conduction bands. Three prototype semiconductors—GaAs, ZnO, and CdS—are studied to elucidate the effects of self-consistency for periodic solids. The semicores are explicitly taken into account in the pseudopotential representation. We will introduce the numerical methods and techniques to deal with the Γ point divergence problem in the evaluation of the self-energy term and dielectric function. The computation is done with tens of thousands of processors on one of the largest supercomputers, Titan, at the Oak Ridge Leadership Computing Facility (OLCF). The rest of the paper is organized as follows. Our fully sc- GW approach is presented in Sec. II. In Sec. III, we elaborate the numerical methods and technical details in implementing our sc- GW approach. Results and discussions are then presented in Sec. IV, followed by the main conclusion in Sec. V.

II. THE BASIC FORMALISM

To avoid the singularity in the real axis, we follow the “space-time” method first used by Rojas *et al.* [16], where the Green’s function is solved along the imaginary axis $i\omega + \mu$ [to be denoted as $G(i\omega)$] in the ω complex plane. Here, μ is the electron Fermi energy (both ω and μ are real numbers). In our previous work for isolated systems [22], the Green’s function $G(i\omega)$ defined through the Dyson equation can be expressed in Eq. (2). In a periodic system, the Green’s function at one k point can be written down as (see the derivations from isolated molecule systems to periodic systems in Appendix A)

$$G^{-1}(k, i\omega) = i\omega + \mu - H(k) - \Sigma(k, i\omega), \quad (3)$$

where G , H , and Σ are all matrices for a given $(k, i\omega)$ and k is the wave vector in the first Brillouin zone (BZ). The G , H , and Σ are represented either in real-space r index or reciprocal space q index. The transformations between r and q space for matrix $X(k, z)$ (e.g., G and Σ) are

$$X(q_1, q_2, k, z) = \frac{1}{\Omega} \int X(r_1, r_2, k, z) e^{iq_1 r_1} e^{-iq_2 r_2} d^3 r_1 d^3 r_2; \quad (4)$$

$$X(r_1, r_2, k, z) = \frac{1}{\Omega} \sum_{q_1, q_2} X(q_1, q_2, k, z) e^{-iq_1 r_1} e^{iq_2 r_2}.$$

Here Ω is the volume of the periodic unit cell and z can be either $i\omega$ or $i\tau$. $H(k) = -\frac{1}{2}\nabla_k^2 + V(r) + \sum_l |\phi_{l,k}\rangle\langle\phi_{l,k}|$ is the noninteracting one-electron Hamiltonian, including the kinetic energy operator $-\frac{1}{2}\nabla_k^2$, the nonlocal pseudopotential projector $\sum_l |\phi_{l,k}\rangle\langle\phi_{l,k}|$, and the single-particle potential $V(r)$. $V(r)$ is obtained as $V(r) = \sum_R v_{\text{at}}(r - R) + \int \frac{\rho(r')}{|r-r'|} d^3 r'$, where v_{at} is the local part of the atomic pseudopotential, R is atomic position, and ρ is the electron charge density calculated as $\rho(r) = -iG(r, r, i\tau)|_{\tau \rightarrow 0^+}$. During the self-consistent iterations, the potential $V(r)$ is recalculated through Pulay-Kerker potential mixing [45]. Σ is the electron self-energy that encompasses all exchange-correlation effects. Within Hedin’s GW approximation [4], the self-energy term is given by the product of Green’s function G and the dynamically screened interaction W . To avoid the time-consuming convolution in frequency domain, the self-energy Σ for each k is evaluated in real space and time domain as

$$\Sigma(r_1, r_2, k, i\tau) = i \sum_{k_2} G(r_1, r_2, k - k_2, i\tau) W(r_1, r_2, k_2, i\tau) w_{k_2}, \quad (5)$$

where w_{k_2} is used to represent a summation weight to represent the possible symmetry reduction of the k points. The W represents the dynamically screened Coulomb potential. The expression of W in reciprocal space with frequency dependence reads

$$W(q_1, q_2, k, i\omega) = \frac{4\pi}{|q_1 + k||q_2 + k|} \varepsilon^{-1}(q_1, q_2, k, i\omega), \quad (6)$$

where $4\pi/(q + k)^2$ is the Fourier transform of the bare Coulomb interaction and ε is the dielectric function expressed as

$$\varepsilon(q_1, q_2, k, i\omega) = \delta_{q_1, q_2} - \chi(q_1, q_2, k, i\omega) \frac{4\pi}{|q_1 + k||q_2 + k|}. \quad (7)$$

Finally, the irreducible polarizability χ is given by the product of two Green’s functions from different k vectors:

$$\chi(r_1, r_2, k, i\tau) = -i \sum_{k_2} G(r_1, r_2, k + k_2, i\tau) \times G(r_2, r_1, k_2, -i\tau) w_{k_2}. \quad (8)$$

All the matrices in the above equations are represented either in real space (r_1, r_2, z) or reciprocal space (q_1, q_2, z) (z can be either $i\omega$ or $i\tau$). The most time-consuming parts of the sc- GW calculations are the Fourier transformation between these two representations as well as the inversion of Green’s function G and dielectric function ε [Eqs. (3) and (6)]. The GW calculations using the “space-time” method on the imaginary $i\omega$ axis have been performed by Rojas *et al.* [16] a long time ago. However, only the one-shot G_0W_0 calculations were carried out, hence an analytical expression for $G_0(i\tau)$ was available, which avoided the need to do the ω space to τ space Fourier transform, which is particularly time consuming. Very often, the Matsubara time and frequency mesh with an artificial temperature [19] can also be used to facilitate the ω integration. Under such approximation, the final results are extrapolated from a series of artificial temperatures [19,41]. In this work, a special integration algorithm was carried out without the use of artificial temperature. Discrete exponential numerical grid points both in $i\omega$ and $i\tau$ are used, with the maximum ω being 3×10^6 hartree, while the minimum ω interval is 2×10^{-4} hartree. The details of the numerical Fourier transformation between the $i\tau$ and $i\omega$ space can be found in our previous work for isolated molecule systems [22]. For periodic bulk systems, particular attention is needed for Eqs. (5) and (8) to deal with the Γ point divergence problem. In the following, we will introduce the numerical methods and techniques to deal with these divergence problems. A plane wave energy cutoff E_{cut} is used to select the plane wave vectors q_1 and q_2 in $G(q_1, q_2, k, i\omega)$ and $\Sigma(q_1, q_2, k, i\omega)$. However, in the expressions of matrices of W , ε , and χ , the plane wave vectors should be defined by an energy cutoff $E_{\text{cut}2} = 4E_{\text{cut}}$. The reason for this is that these matrices are proportional to the squares of Green’s function [Eq. (8)]. In practice, a smaller $E_{\text{cut}2} = 2E_{\text{cut}}$ can be used to converge the final results, much like in traditional plane wave DFT calculations. After using these techniques, our sc- GW calculations are well converged with regard to k -point summation. Equations (3)–(8) constitute a close set of equations to find the self-consistent solution of the Green’s function G . Note, the first iteration of the calculation is equivalent to the conventional G_0W_0 calculations (in this work, the non-self-consistent G_0W_0 results are calculated using local-density approximation (LDA) Kohn-Sham eigenvalues and eigenfunctions as inputs).

III. IMPLEMENTATION OF THE GW METHOD

A. Evaluation of dielectric function

In the preceding section, we have mentioned that special care is required for the Γ point ($k = 0$) divergence problem in the periodic systems. For the calculation of dielectric function using Eq. (7), there is an obvious singularity if $q_1 = q_2 = 0$ (the “head”) or $q_1 = 0$ or $q_2 = 0$ (the “wings”) for very small k . One solution of this problem is to expand the polarizability

$\chi(q_1, q_2, k, i\omega)$ as a function of k for the “head” and “wings” of the polarizability matrices. Hybertsen and Louie [46] derived one such expression using the Adler-Wiser formulation [47,48] for single-particle expression more than 30 years ago. Here we use a similar technique to get the “head” and “wings” expansion of polarizability χ for a general matrix expression of G . For the “head” case, the polarizability $\chi(q_1 = 0, q_2 = 0, k, i\tau)$ at the limit of $k \rightarrow 0$ is expanded as

$$\begin{aligned} \chi(q_1 = 0, q_2 = 0, k, i\tau) &= \chi(q_1 = 0, q_2 = 0, k = 0, i\tau) \\ &+ \chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k, i\tau) k_\alpha k_\beta \\ &= \chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k = 0, i\tau) k_\alpha k_\beta, \end{aligned} \quad (9)$$

where k_α ($\alpha = x, y, \text{ or } z$) is the α th component of k approaching the Γ point ($k = 0$). Since the first term $\chi(q_1 = 0, q_2 = 0, k = 0, i\tau)$ equals zero, the “head” expression for $\chi(q_1 = 0, q_2 = 0, k, i\tau)$ goes to zero as $k_\alpha k_\beta$ when $k \rightarrow 0$. The second order $k_\alpha k_\beta$ in the “head” expansion as a function of k will cancel with the k_2 in the denominator of Eq. (7) for their magnitudes, although the result depends on the direction of the vector k (which gives rise to the well-known directional singularity of the dielectric constant for the low symmetry crystal). To get the term $\chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k = 0, i\tau) k_\alpha k_\beta$ in Eq. (9), a middle step term $\chi_{\alpha,\beta}^{(0)}(q_1 = 0, q_2 = 0, k = 0, i\tau) k_\alpha k_\beta$ based on Eq. (8) is defined as

$$\begin{aligned} \chi_{\alpha,\beta}^{(0)}(q_1 = 0, q_2 = 0, k = 0, i\tau) k_\alpha k_\beta &= -ik_\alpha k_\beta \sum_{k_2} \int [\nabla_{k_2}^\alpha H(r_1) G(r_1, r_2, k_2, i\tau) \nabla_{k_2}^\beta H(r_2) \\ &\times G(r_2, r_1, k_2, -i\tau) w_{k_2}] d^3 r_1 d^3 r_2. \end{aligned} \quad (10)$$

Here $\nabla_{k_2}^\alpha H(r)$ is the derivation of the single-particle Hamiltonian with respect to $k_{2\alpha}$ (k_2 belongs to the original k grid in the first BZ). In the calculation, $\nabla_{k_2} H(r)$ is written as a matrix to represent the nonlocal term.

In the noninteractive single-particle formalism, $\chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k = 0, i\tau)$ is related to $\chi_{\alpha,\beta}^{(0)}(q_1 = 0, q_2 = 0, k = 0, i\tau)$ with an extra eigenenergies square term in the denominator [46,47]. This extra term in the denominator can be obtained by a second-order integration of τ in the form

$$\begin{aligned} \chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k = 0, i\tau) &= - \int_{-\infty}^{\tau} \left[- \int_{-\infty}^{\tau} \chi_{\alpha,\beta}^{(0)}(q_1 = 0, q_2 = 0, k = 0, i\tau) d\tau \right] d\tau \\ &= \sum_{k_2} -i \left\{ - \int_{-\infty}^{\tau} \left[- \int_{-\infty}^{\tau} \left\{ \int [\nabla_{k_2}^\alpha H(r_1) G(r_1, r_2, k_2, i\tau) \right. \right. \right. \\ &\quad \left. \left. \left. \times \nabla_{k_2}^\beta H(r_2) G(r_2, r_1, k_2, -i\tau) w_{k_2} \right] d^3 r_1 d^3 r_2 \right\} d\tau \right] d\tau \right\}. \end{aligned} \quad (11)$$

Equation (11) is used to calculate $\chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k = 0, i\tau)$ from G when a full matrix form of G is represented. Note that Eq. (11) is only rigorous for the noninteractive Green’s function. Nevertheless, it should capture the main contribution

of the k expansion. Furthermore, in the final result (e.g., the self-energy), this only affects the dielectric constant at $k = 0$. As we use larger and larger k -point grids, the contribution of this $k = 0$ point becomes smaller and smaller. Thus, our final convergence in regard to the number of k points indicates that the approximation at $k = 0$ is fine, or at least the error in this approximation does not affect the final result.

After $\chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k = 0, i\tau)$ is in hand, we can get the polarizability $\chi(q_1 = 0, q_2 = 0, k, i\tau)$ for any k points near $k \rightarrow 0$ using Eq. (9). To get the frequency-dependent polarizability $\chi(q_1 = 0, q_2 = 0, k, i\omega)$ used in Eq. (7), a Fourier transform is carried out to change $\chi(q_1 = 0, q_2 = 0, k, i\tau)$ from $i\tau$ to $i\omega$ space.

$$\begin{aligned} \chi(q_1 = 0, q_2 = 0, k, i\omega) &= -ik_\alpha k_\beta \int \chi_{\alpha,\beta}^{(2)}(q_1 = 0, q_2 = 0, k = 0, i\tau) e^{-i\omega\tau} d\tau. \end{aligned} \quad (12)$$

The “wings” case can be dealt with similar fashion, where only one derivative is used. The detailed expression is given in Appendix B.

From Eqs. (10) to (12), we find that the polarizability $\chi(q_1 = 0, q_2 = 0, k, i\omega)$ at the limit of $k \rightarrow 0$ involves contribution from each k_2 point. For simplicity, we would like to write the frequency-dependent polarizability $\chi(q_1 = 0, q_2 = 0, k, i\omega)$ in the form below:

$$\chi(q_1 = 0, q_2 = 0, k, i\omega) = k_\alpha k_\beta \sum_{k_2} \bar{\chi}_{\alpha,\beta}(k_2) w_{k_2}. \quad (13)$$

Here $\bar{\chi}_{\alpha,\beta}(k_2)$ denotes the polarizability contribution from each k_2 point in Eq. (11). To obtain the dielectric function $\varepsilon(q_1, q_2, k, i\omega)$ in reciprocal space is straightforward after taking into account the nonanalyticities of the “head” and “wings” cases using the formula above. As discussed above, the $\varepsilon_{\alpha,\beta}(0, 0, k, i\omega)$ at the limit of $k \rightarrow 0$ depends on the direction $\hat{k} = k/|k|$. Taking into account the k_2 summation in Eq. (13), for small k , we can write

$$\begin{aligned} \varepsilon_{\alpha,\beta}(0, 0, k, i\omega) &= 1 - 4\pi \frac{k_\alpha k_\beta}{k^2} \sum_{k_2} \bar{\chi}_{\alpha,\beta}(k_2) w_{k_2} \\ &= 1 - 4\pi \cdot \hat{k}_\alpha \hat{k}_\beta \sum_{k_2} \bar{\chi}_{\alpha,\beta}(k_2) w_{k_2}. \end{aligned} \quad (14)$$

Here, \hat{k}_α ($\alpha = x, y, \text{ or } z$) is the α th Cartesian component of the unit vector: $k_\alpha/|k|$. For materials considered in this work, they all have cubic symmetry, thus the 3×3 tensor $\varepsilon_{\alpha,\beta}$ is an identity matrix multiplied by a constant. As a result, the orientation dependence disappears, and Eq. (14) can further be simplified as

$$\varepsilon(0, 0, k, i\omega) = 1 - 4\pi \sum_{k_2} \bar{\chi}_{\alpha,\beta}(k_2) w_{k_2}. \quad (15)$$

Thus, the dielectric constant ε can be directly approximated as $\varepsilon(q_1 = 0, q_2 = 0, k = 0, \omega)$ for small k . Note, when compared to experiments, the macroscopic dielectric constant ε is defined as $1/\varepsilon^{-1}(q_1 = 0, q_2 = 0, k = 0, \omega)$ including local field effects [ε^{-1} is the inversion of the $\varepsilon(q_1, q_2, k = 0, \omega)$]

TABLE I. The LDA macroscopic dielectric constant calculated for various k_2 grid sets. N_{k_2} is the number of k_2 points in the k_2 grid set. ε is the original calculated dielectric constant, $\varepsilon_{\text{inter}}$ is the interpolated dielectric constant using Eq. (18), and $\varepsilon_{\text{mask}}$ is the final fitted dielectric constant using Eq. (19).

Systems	N_{k_2}	ε	$\varepsilon_{\text{inter}}$	$\varepsilon_{\text{mask}}$
GaAs	$7 \times 7 \times 7$	22.33	15.69	11.81
	$9 \times 9 \times 9$	16.47	13.25	11.81
	$13 \times 13 \times 13$	13.18	12.05	
	$17 \times 17 \times 17$	12.67	11.83	
	$21 \times 21 \times 21$	11.94	11.80	
CdS	$23 \times 23 \times 23$	11.85		
	$25 \times 25 \times 25$	11.83		
	$7 \times 7 \times 7$	8.77	7.55	6.84
	$9 \times 9 \times 9$	7.78	7.21	6.84
	$13 \times 13 \times 13$	7.03	6.85	
ZnO	$15 \times 15 \times 15$	6.88	6.87	
	$7 \times 7 \times 7$	6.64	6.25	4.68
	$9 \times 9 \times 9$	6.63	5.66	4.68
	$13 \times 13 \times 13$	5.69	5.17	
	$15 \times 15 \times 15$	5.26	4.71	
	$19 \times 19 \times 19$	4.67	4.65	

matrix]. The local field effects are essential in predicting the correct quasiparticle spectrum [49].

The convergence of the dielectric constant ε is often related to k -point summation and the number of conduction bands used [46,50]. In our approach, we use the full matrices without conduction band truncation, so the only concern is the k -point BZ summation, especially for the ‘‘head’’ and ‘‘wing’’ using Eq. (15). The calculated LDA macroscopic dielectric constants for GaAs, CdS, and ZnO with respect to the number of k_2 points are illustrated in Table I. We can see that the convergence of the LDA dielectric constant ε is notoriously slow. It is also clear that the convergence is more difficult for small gap semiconductors. For GaAs, even a $21 \times 21 \times 21$ Γ -centered grid does not yield fully converged results. The main reason for this problem is that the term of $\bar{\chi}_{\alpha,\beta}(k_2 = 0)$ (at the Γ point) in Eq. (15) is very large, hence a large number of k_2

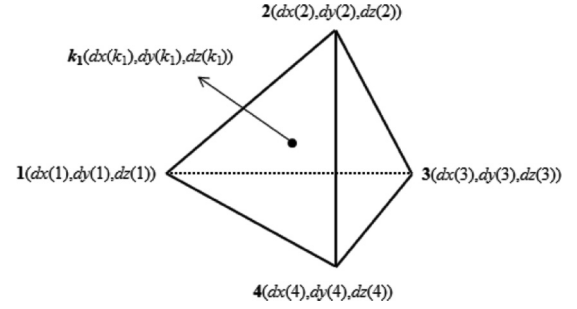


FIG. 1. Schematic diagram of the tetrahedron interpolation (also called the linear tetrahedron). The 1, 2, 3, and 4 points belong to the original k_2 grid and the k_1 point is from the dense grid.

points are needed to average out the influence of this single Γ point. The simplest solution is to use a shifted Monkhorst-Pack grid without the Γ point. We found that a shifted Monkhorst-Pack grid with $8 \times 8 \times 8$ k_2 points yields a converged LDA dielectric constant of 11.86 for GaAs. However, it is clear from Eq. (11), this will require one to define G at these shifted grid points, thus not being able to obtain the Γ -point band gap. Here we will introduce a numerical technique to yield converged $\varepsilon(q_1 = 0, q_2 = 0, k = 0, \omega)$ of Eq. (15) without an excessively large k_2 -point grid. We first extend the discrete k_2 sum in Eq. (15) to a continuous k_1 -point sum (in practice, with a much denser k_1 -point grid). The polarizability contribution $\bar{\chi}_{\alpha,\beta}(k_1)$ at each k_1 point is interpolated from $\bar{\chi}_{\alpha,\beta}(j)$ ($j = 1, m$) of the nearest m k_2 points of the original k_2 grid in the form

$$[\bar{\chi}_{\alpha,\beta}(k_1)]^{1/n} = \sum_{j=1}^m f_{k_1}(j) [\bar{\chi}_{\alpha,\beta}(j)]^{1/n}, \quad (16)$$

where $f_{k_1}(j)$ ($j = 1, m$) is defined as the j th-point interpolation weight such that they sum to 1 and the exponential factor $n = 3$ is chosen to make the resulting $\bar{\chi}_{\alpha,\beta}(k_1)^{1/n}$ as linear as possible. The linear interpolation coefficients $f_{k_1}(j)$ are determined from a tetrahedron interpolation scheme as illustrated from Fig. 1, and obtained by solving the linear equations:

$$\begin{aligned} f_{k_1}(1) + f_{k_1}(2) + f_{k_1}(3) + f_{k_1}(4) &= 1, & dx(1)f_{k_1}(1) + dx(2)f_{k_1}(2) + dx(3)f_{k_1}(3) + dx(4)f_{k_1}(4) &= dx(k_1), \\ dy(1)f_{k_1}(1) + dy(2)f_{k_1}(2) + dy(3)f_{k_1}(3) + dy(4)f_{k_1}(4) &= dy(k_1), \\ dz(1)f_{k_1}(1) + dz(2)f_{k_1}(2) + dz(3)f_{k_1}(3) + dz(4)f_{k_1}(4) &= dz(k_1). \end{aligned} \quad (17)$$

Here $dx(j)$, $dy(j)$, and $dz(j)$ ($j = 1, 2, 3, 4$ or k_2) are the coordinates of the j th corner (or k_2 point) of the tetrahedron. When the k_1 point falls to the edge or corner of the tetrahedron, multiple tetrahedrons are used, and the results are averaged from different tetrahedron interpolated results. With the interpolated polarizability contribution $\bar{\chi}_{\alpha,\beta}(k_1)$ in hand, the summation of Eq. (15) can be rewritten as a much denser grid sum over k_1 :

$$\begin{aligned} \varepsilon(0, 0, k, i\omega) &= 1 - 4\pi \sum_{k_1} \bar{\chi}_{\alpha,\beta}(k_1) w_{k_1} \\ &= 1 - 4\pi \sum_{k_1} \bar{\chi}_{\alpha,\beta}(k_1) \frac{1}{N_{k_1}}. \end{aligned} \quad (18)$$

Here N_{k_1} is the total number of k_1 points in the dense grid. The resulting dielectric constant $\varepsilon_{\text{inter}}$ obtained using Eq. (18) is also presented in Table I. We can see that this dielectric constant $\varepsilon_{\text{inter}}$ converges much faster than the original formula and the convergence for GaAs can be reached by a $17 \times 17 \times 17$ k_2 grid [the k_1 grid used is $(126)^3$]. Nevertheless, this is still computationally demanding. We note that much of this slow convergence is still due to the dramatic change of $\bar{\chi}_{\alpha,\beta}(k_1)$ near the Γ point. A possible good approximation is that the shape of $\bar{\chi}_{\alpha,\beta}(k_1)$ near the Γ point for a given system might not change much, from LDA to GW results, but the overall amplitude might change near that region. To capture this feature, we

have defined a mask prefactor $\lambda_{\text{LDA}}(k_1) = \bar{\chi}_{\alpha,\beta}^{\text{mask}}(k_1)/\bar{\chi}_{\alpha,\beta}(k_1)$ to describe the shape of $\bar{\chi}_{\alpha,\beta}(k_1)$. Here $\bar{\chi}_{\alpha,\beta}(k_1)$ is interpolated from a small k_2 grid (e.g., the $9 \times 9 \times 9$ grid to be used in our sc- GW calculations) using Eq. (16), while $\bar{\chi}_{\alpha,\beta}^{\text{mask}}(k_1)$ is interpolated from a dense grid (e.g., the $21 \times 21 \times 21$ grid) from Eq. (16). The k_1 grid is still a dense grid, e.g., $(126)^3$. To calculate $\lambda_{\text{LDA}}(k_1)$, both $\bar{\chi}_{\alpha,\beta}(k_1)$ and $\bar{\chi}_{\alpha,\beta}^{\text{mask}}(k_1)$ are calculated with LDA. Then this fixed LDA mask function $\lambda_{\text{LDA}}(k_1)$ will be used in GW calculations in the following formula:

$$\varepsilon(0,0,k,i\omega) = 1 - 4\pi \sum_{k_1} \bar{\chi}_{\alpha,\beta}(k_1) \lambda_{\text{LDA}}(k_1) \frac{1}{N_{k_1}}. \quad (19)$$

Here $\bar{\chi}_{\alpha,\beta}(k_1)$ are interpolated from the small (e.g., $9 \times 9 \times 9$) k_2 grid using Eq. (16) during the sc- GW iterations, and $\lambda_{\text{LDA}}(k_1)$ is fixed throughout the iterations. For the LDA calculation, almost by definition, different k_2 grids will get the same result (e.g., all equal to the $21 \times 21 \times 21$ grid result) under this procedure, as shown in Table I. To test the convergence of this procedure for sc- GW calculations, we have calculated the dielectric constant with a $7 \times 7 \times 7$ grid ($\varepsilon_{GW} = 4.96$) and a $9 \times 9 \times 9$ grid ($\varepsilon_{GW} = 4.98$) for GaAs; they only differ by 0.02. To be conservative, we will use a $9 \times 9 \times 9$ k grid in our following sc- GW calculations to guarantee a full convergence.

We would like to point out that all of the above discussions from Eq. (9) to Eq. (19) are concerning the $\varepsilon(q_1 = 0, q_2 = 0, k_2 = 0, \omega)$ value (also the ‘‘wing’’ values). The $\varepsilon(q_1, q_2, k_2, \omega)$ for all the other k_2 points, or nonzero q_1, q_2 for $k_2 = 0$ are well defined using Eq. (7). The small difference between the $7 \times 7 \times 7$ grid and $9 \times 9 \times 9$ k grid sc- GW results (including the small quasiparticle energy difference of 15 meV) indicates the adequacy of using Eq. (11). As the k_2 -point grid is getting larger, the relative roles of Eqs. (10) and (11) [which are only used to get the $k_2 = 0$ value of $\varepsilon(q_1 = 0, q_2 = 0, k_2 = 0, \omega)$] are getting smaller. Thus, even if there were some small errors in Eqs. (10) and (11), the final sc- GW quasiparticle energies would not be affected, as long as the results are converged regarding the k -point grid.

B. Evaluating the self-energy

In the GW approximation, the self-energy Σ is obtained from the product of the Green’s function G and the screened interaction W sum over many different k points as shown in Eq. (5). However, the screened interaction W at $k = 0$ is divergent as shown in Eq. (6). Thus, a discrete k -point sum including the $k = 0$ point in Eq. (5) will also get a divergent result. To solve this divergence problem, we use a technique similar to that proposed by Gygi and Baldereschi [51] for unscreened Fock exchange term calculation. A reference term which has the same singularities as the right side of Eq. (5) is added and subtracted in the formula as below:

$$\begin{aligned} \Sigma(r_1, r_2, k, i\tau) = & i \sum_{k_2} \{G(r_1, r_2, k - k_2, i\tau)W(r_1, r_2, k_2, i\tau) \\ & - G(r_1, r_2, k, i\tau)W(r_1, r_2, k_2, i\tau)\} w_{k_2} \\ & + iG(r_1, r_2, k, i\tau) \sum_{k_1} W(r_1, r_2, k_1, i\tau) w_{k_1}. \end{aligned} \quad (20)$$

Note the singularities will be canceled out when $k_2 = 0$ for the first two terms. The k_2 -point summation in the last term is replaced by a continuous k_1 point integration $\sum_{k_1} W(r_1, r_2, k_1, i\tau) w_{k_1}$. To avoid the divergence problem, k_1 is defined in a dense grid through the first BZ. In the reciprocal space, we have

$$\begin{aligned} & \sum_{k_1} W(q_1, q_2, k_1, i\omega) w_{k_1} \\ & = \sum_{k_1} \frac{4\pi}{|q_1 + k_1||q_2 + k_1|} w_{k_1} \varepsilon^{-1}(q_1, q_2, k_1, i\omega) \\ & = \sum_{k_1} \frac{4\pi}{|q_1 + k_1||q_2 + k_1|} w_{k_1} \left[\sum_{j=1}^m f_{k_1}(j) \varepsilon^{-1}(q_1, q_2, j, i\omega) \right]. \end{aligned} \quad (21)$$

Here, the same interpolation technique of Eq. (17) is used to get $\varepsilon^{-1}(q_1, q_2, k_1, i\omega)$ for an arbitrary k_1 from its nearest neighbor values on the k_2 grid. Note, since $\varepsilon^{-1}(q_1, q_2, k_1, i\omega)$ should be relatively smooth (compared to the $4\pi/|q_1 + k_1||q_2 + k_1|$ factor), such interpolation should work fine. From Eq. (21), it is clear that the singularities are caused by the bare Coulomb potential term $4\pi/|q_1 + k_1||q_2 + k_1|$ when $q_1 = q_2 = 0$ and $k_1 = 0$. The summation of this point should represent a k -space region of dimension $\Delta k_1 = 2\pi/(aN)$; here a is the lattice constant of the crystal, and N is the grid point number along each direction [e.g., for a $(400)^3$ k_1 grid, N is 400]. Since the volume is proportional to $(\Delta k_1)^3$, and the W is proportional to $\varepsilon^{-1}(0,0,0,i\omega)/(\Delta k_1)^2$, overall this single $k_1 = 0$ term in Eq. (21) should have a contribution of $\beta \varepsilon^{-1}(0,0,0,i\omega) \Delta k_1$. The β is a geometric factor depending on the crystal lattice. As a result, Eq. (21) for $q_1 = q_2 = 0$ can be rewritten as

$$\begin{aligned} & \sum_{k_1} \frac{4\pi}{(k_1)^2} w_{k_1} \varepsilon^{-1}(0,0,k_1,i\omega) \\ & = \sum_{k_1 \neq 0} \frac{4\pi}{(k_1)^2} w_{k_1} \varepsilon^{-1}(0,0,k_1,i\omega) + \beta \Delta k_1 \varepsilon^{-1}(0,0,0,i\omega). \end{aligned} \quad (22)$$

Note that β depends only on the crystal lattice, not on the dielectric constant. β can be obtained by fitting the unscreened Coulomb interaction result of Eq. (22) for a moderate N (e.g., 100) to the converged result using an extremely large N (e.g., 1000) (with an arbitrary preset β , e.g., $\beta = 1$). For the face-centered primary cell lattice used in this study, we found $\beta = 0.902$ can yield a very fast convergence of Eq. (22), as shown in Fig. 2. After this β is fixed in Eq. (22), in the following, we will use a 100^3 k_1 -point grid to carry out Eqs. (20)–(22). This will give us a converged self-energy Σ .

C. Calculating quasiparticle energy

The Green’s function G is updated by solving Eq. (3) after the self-energy Σ is obtained. Note, the Hamiltonian $H(k, i\omega) = H(k) + \Sigma(k, i\omega) = i\omega + \mu - G^{-1}(k, i\omega)$ is non-Hermitian. Although strictly speaking, the quasiparticle energy should be defined from the peaks in the spectral function, or say the poles of $G(k, i\omega)$, in practice, for a

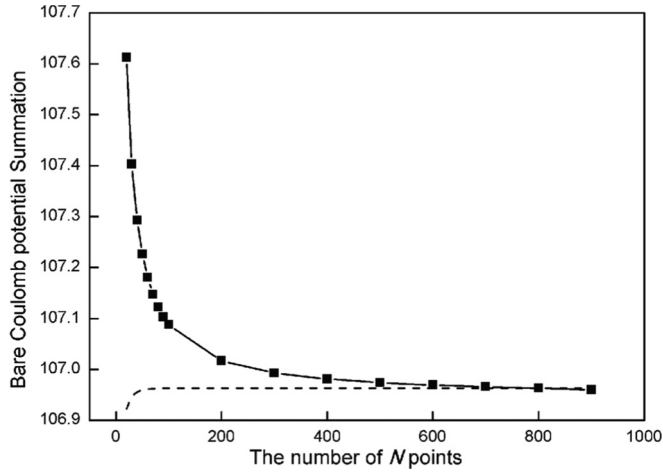


FIG. 2. The solid line is the convergence of the $\sum_{k_1 \neq 0} \frac{4\pi}{(k_1)^2} w_{k_1} + \beta \Delta k_1$ summation for $\beta = 1$ with respect to the number of N points. The continuous dashed line is the summation using prefactor $\beta = 0.902$.

post calculation analysis, it will be convenient to have a set of eigenstates of a Hermitian Hamiltonian, and use that to represent the quasiparticle states. To do so, we have first Hermitized $H(k, i\omega)$ as $H'(k, i\omega) = \frac{1}{2}[H(k, i\omega) + H^T(k, i\omega)]$, then diagonalized it to get the wave function $\psi_{jk}(i\omega)$, where j and k are the band and k -point indices, respectively. We found that $\psi_{jk}(i\omega)$ is almost independent of ω at least for the occupied states and a few conduction band states near the band gap [22]. As a result, the wave function $\psi_{jk}(i\omega)$ can be well approximated by $\psi_{jk}(0)$ [which is $\psi_{jk}(i\omega = 0)$]. This allows us to define an expectation value of the self-energy matrix $\bar{\Sigma}_{jk}(i\omega) = \langle \psi_{jk}(0) | \Sigma(k, i\omega) | \psi_{jk}(0) \rangle$, which is shown in Fig. 3 on the imaginary ω axis. It is clear that both the real and imaginary parts change significantly with $i\omega$. In addition, the expectation value of the non-Hermitian part (indicated by the imaginary part) can be as large as the Hermitian part. All of this

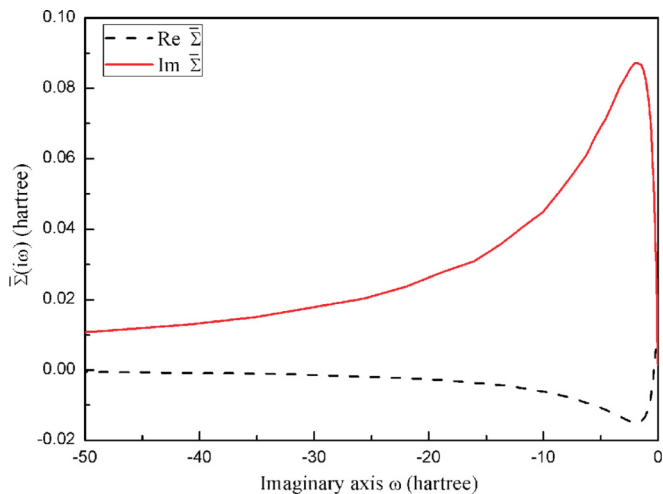


FIG. 3. The expectation value $\bar{\Sigma}_{jk}(i\omega) = \langle \psi_{jk}(0) | \Sigma(k, i\omega) | \psi_{jk}(0) \rangle$ of GaAs for the $j = 20$ state ($j = 22$ is the VBM) at the Γ point on the imaginary axis.

means the true Green's function is far from the noninteracting single-particle description of Eq. (1) [35,36].

To get the corresponding quasiparticle energies that can be measured in the experiment, knowledge of G and Σ in the real frequency domain is required. The corresponding poles can then be obtained with the ω solution of $\langle \psi_{jk}(0) | G^{-1}(k, \omega) | \psi_{jk}(0) \rangle = 0$. This requires us to get $\bar{\Sigma}_{jk}(\omega) = \langle \psi_{jk}(0) | \Sigma(k, \omega) | \psi_{jk}(0) \rangle$ on the real axis according to the Dyson equation. With the above calculated $\bar{\Sigma}_{jk}(i\omega)$ on the imaginary axis, $\bar{\Sigma}_{jk}(\omega)$ can be obtained by analytically extending $\bar{\Sigma}_{jk}(i\omega)$ to the real axis as proposed by Rojas *et al.* [16]. We have tested that this procedure is very reliable in obtaining $\bar{\Sigma}_{jk}(\omega)$ for ω within 1 or 2 hartrees from Fermi energy μ [22]. The quasiparticle energies $\varepsilon_{jk}(\omega)$ for each j state and k vector are then determined by solving the equation $\omega + \mu = \varepsilon_{jk}(0) + \text{Re}[\bar{\Sigma}_{jk}(\omega) - \bar{\Sigma}_{jk}(0)]$; $\varepsilon_{jk}(0)$ is the eigenenergy of $H'(k, i\omega)$ when ω equals 0.

The excitation spectrum is also an important quantity that is often measured in experiment. The spectral function which can be compared with experiment for the periodic solids is calculated as

$$A(k, \omega) = 1/\pi \text{Tr}[\text{Im}G(k, \omega)]. \quad (23)$$

Here “Tr” is the trace of the matrix G . Within the above approximation, the spectral functions can be directly obtained as

$$A(k, \omega) = \frac{1}{\pi} \sum_j \text{Im} \left[\frac{1}{\omega - \mu - \varepsilon_{jk}(0) - [\bar{\Sigma}_{jk}(\omega) - \bar{\Sigma}_{jk}(0)]} \right]. \quad (24)$$

In our test of the convergence of the sc-*GW* iterations, the eigenenergies $\varepsilon_{jk}(0)$ of $H'(k, i\omega)$ when ω equals 0 are used as marks for the convergence. The energy gaps measured by $\varepsilon_{jk}(0)$ as a function of self-consistent iterations are presented in Fig. 4(a). We can see that the energy gaps for GaAs, ZnO, and CdS converge within about four to five iterations with LDA input eigenenergies and eigenfunctions. In our previous work for isolated molecule systems [22], we have demonstrated that the final quasiparticle energies are independent of the initial input eigenfunctions and eigenenergies. We found this is also true for our bulk system calculations. The self-consistent loops are also measured with the change of dielectric constant ε in Fig. 4(b). On the one hand, it confirms the convergence of the self-consistent iterations. On the other hand, we found that there is a strong correlation between the energy gap and the dielectric constant; the energy gap of the $N + 1$ iteration increases with the decrease of the dielectric constant from the N iteration. An accurate prediction of the quasiparticle band gap requires an accurate prediction of the dielectric matrix.

D. The use of pseudopotentials

In this work, we use the plane wave basis set for the implementation of the fully self-consistent *GW* calculations. The norm-conserving pseudopotential with semicore electrons is used to give a good description of the valence pseudo wave functions. The outmost two shells of atomic orbitals are treated as valence states for Ga, As, Zn, and Cd atoms to avoid the errors in the quasiparticle energies as shown by Rohlfing

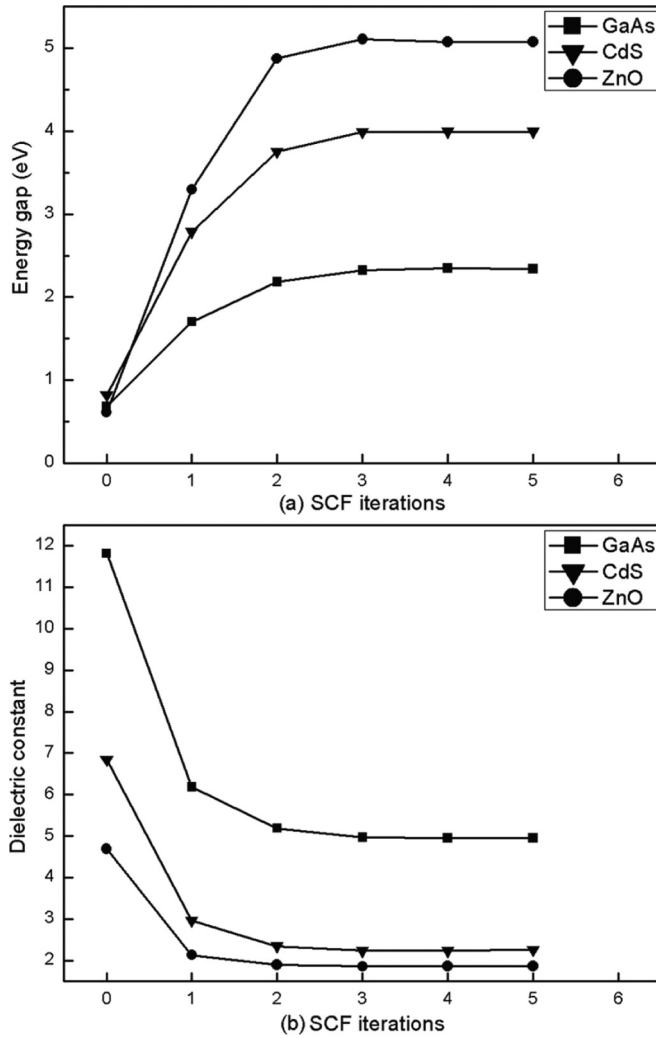


FIG. 4. (a) The minimum energy gaps measured by $\epsilon_{jk}(0)$ as a function of self-consistent iterations. (b) The change of the dielectric constant ϵ with respect to the iteration steps.

et al. [52]. More specifically, for Ga, As, Zn, and Cd atoms, their corresponding valence electrons in the pseudopotentials are $3s^23p^6 3d^{10}4s^24p^1$, $3s^23p^6 3d^{10}4s^24p^3$, $3s^23p^6 3d^{10}4s^2$, and $4s^24p^6 4d^{10}5s^2$, respectively, while the valence electrons for O and S are $2s^22p^4$ and $3s^23p^4$, respectively. Since the semicore states are highly localized, a large energy cutoff E_{cut} is often required to get well converged results, as indicated by LDA calculations. Figure 5 shows the dependence of LDA band gaps for CdS, GaAs, and ZnO on the energy cutoff E_{cut} . For CdS, the band gap can converge with a relatively small E_{cut} (around 80 Ry). However, a significantly larger E_{cut} (more than 300 Ry) is needed to get an accurate band gap with an accuracy better than 0.01 eV for GaAs and ZnO.

The increased E_{cut} can significantly increase the dimension of matrices, computational cost, and the memory requirement. This is shown for GaAs in Table II. Here, N_q is the resulting number of plane waves within E_{cut} for $k=0$ in our calculations. As a result, the matrix for $G(q_1, q_2, k, i\omega)$ or $\Sigma(q_1, q_2, k, i\omega)$ at each k point and ω point is a $(N_q)^2$ matrix. It can be seen in Table II that the dimension of N_q has been increased by 20 times due to the inclusion of the

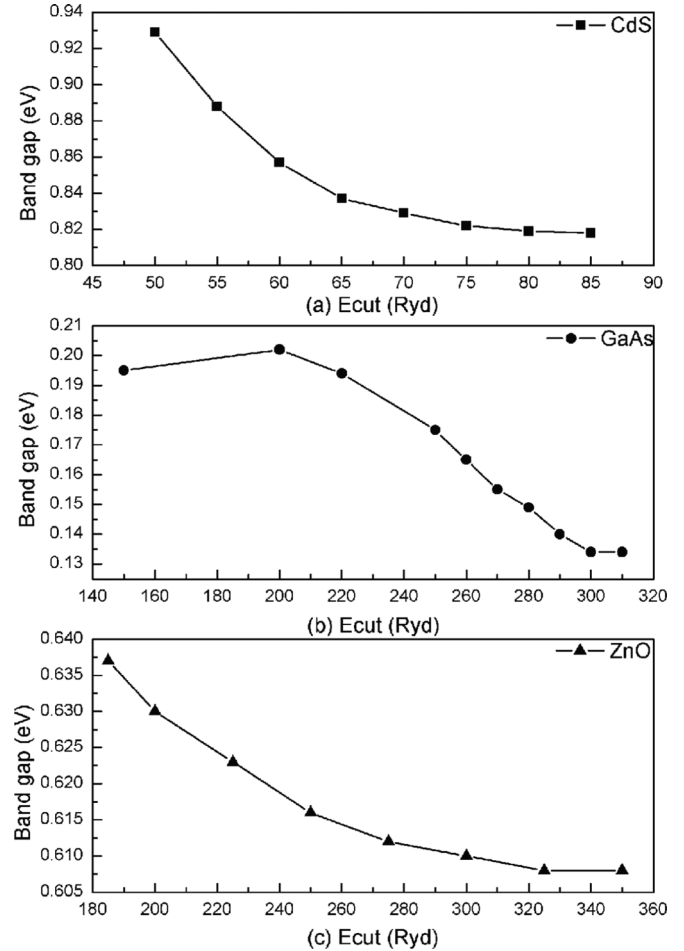


FIG. 5. The LDA band gaps (in eV) for (a) CdS, (b) GaAs, and (c) ZnO as a function of the energy cutoff E_{cut} .

semicore. For W , ϵ , and χ defined by the energy cutoff $E_{\text{cut}2}$, the dimension of these matrices for each k point and ω point is about three times bigger than those of G and Σ . All of this makes the computation extremely expensive with a large memory requirement.

We would like, however, to reduce this dimension for GaAs and ZnO to some degree. The purpose of including core level is twofold: one is to make the valence electron have the proper shape, so the screened exchange integral will be accurate; the second is to include the core level in the calculation in case they have any mixing with the valence states. We like to reduce the E_{cut} in some degree (e.g., reduce it to 100 Ry range), but still keep these two features intact. As shown in Fig. 5 and Table III, when $E_{\text{cut}} \sim 100$ Ry, the LDA band gap will have a small error (e.g., 0.1 eV), meanwhile the core

TABLE II. The energy cutoff E_{cut} ($E_{\text{cut}2}$) (in Ry) and resulting number of plane waves N_q (N_{qL}) in the initial converged LDA calculations for GaAs with and without semicores.

Systems	E_{cut}	N_q	$E_{\text{cut}2}$	N_{qL}
GaAs	38	1240	76	3480
GaAs with semicore	300	26700	600	75000

TABLE III. The comparison of the LDA results for GaAs with respect to the two different energy cutoffs. E_g is the minimum energy gap (eV) and E_d is the average cation d -states binding energy (eV) of Ga. Some selected eigenvalues (eV) relative to the valence band maximum (Γ_{22v}) at the Γ point are also presented.

Systems	$E_{\text{cut}} = 300 \text{ Ry}$	$E_{\text{cut}} = 85 \text{ Ry}$
E_g	0.132	0.135
$\Gamma_{1v}(\text{As } 3s)$	-183.43	-191.45
$\Gamma_{2v}(\text{Ga } 3s)$	-142.65	-146.79
$\Gamma_{3v}(\text{As } 3p)$	-127.67	-140.38
$\Gamma_{6v}(\text{Ga } 3p)$	-93.98	-100.69
$\Gamma_{12v}(\text{As } 3d)$	-35.18	-35.91
$\Gamma_{17v}(\text{Ga } 3d)$	-14.81	-14.77
E_d	-14.85	-14.82

level energy will have a slightly larger error (~ 10 eV out of 100–200 eV). The exact energy of the core level is a lesser concern due to the expected small mixing with the valence state. Overall, the plane wave truncation can be considered as a small perturbation to the Hamiltonian. We can thus add a counter term to balance the effects of this perturbation. More specifically, we have added a small Gaussian function $f(r)$ to the original s , or p or d potentials. The Gaussian function $f(r)$ reads as

$$f(r) = \beta e^{-[(r-r_{\text{peak}})/r_{\text{cut}}]^2}, \quad (25)$$

where r is the radius, r_{peak} is the position of the peak in radial direction, and r_{cut} is the width of the Gaussian in units of bohr. Using preselected r_{peak} and r_{cut} , by adjusting the factor of β , the modified pseudopotentials can recover the original converged LDA results by using $E_{\text{cut}} \sim 100 \text{ Ry}$. The comparison is given in Table III for GaAs with $E_{\text{cut}} = 85 \text{ Ry}$. Compared to the 300 Ry result, the band gap difference is less than 0.01 eV, the Ga $3d$ states energy difference is within 0.1 eV, and the deep semicore level difference is about 10 eV out of ~ 150 eV. Since the direct involvement of the semicore level is small, the 10 eV error in its energy is likely nonconsequential. The comparison for ZnO and the parameters of β , r_{peak} , and r_{cut} used for the modified pseudopotentials of GaAs and ZnO are given in Appendix C.

However, as we discussed above, the main purpose of including semicore is to correct the Fock exchange integral using pseudo wave functions. The LDA Hamiltonian does not test the exchange integral effects of reducing the 300 Ry cutoff to ~ 100 Ry. In order to test this, we have carried out one-shot (non-self-consistent, much like the G_0W_0 , with LDA wave function as the input) HF calculations for GaAs and ZnO, and compared the resulting band gaps of the 300 Ry cutoff and the 85/105 Ry cutoff with the Eq. (25) correction term. The results are shown in Table IV. The band gap difference between the two cutoff schemes is only about 20 meV, mostly coming from the exchange integral difference. Consider that, in the GW method, the exchange integral is screened (reduced), roughly by an order of 5; then the band gap error introduced by changing the 300 Ry to 85/105 Ry cutoff should be in the order of 4 meV. After using the above techniques, a numerically accurate self-consistent GW calculation can be achieved.

TABLE IV. The one-shot HF band gap E_g in (eV) for ZnO and GaAs with respect to the two different E_{cut} 's (in Ry) with Eq. (25) correction. In this test, we used a $3 \times 3 \times 3$ k -point mesh.

System	$E_{\text{cut}} \text{ (Ry)}$	$E_g \text{ (eV)}$
ZnO	105	11.09
	350	11.11
GaAs	85	6.68
	300	6.66

E. The computational details

The fully *sc-GW* calculation has been applied to study the quasiparticle energies of three prototype semiconductors: GaAs, CdS, and ZnO. We have adopted the experimental zinc-blende lattice constants for a meaningful comparison with experiment and other theoretical results. The lattice constants and cutoff energies are listed in Table V, along with the resulting number of plane waves N_q and the number of real-space grid points N_r . N_q shown in Table V is the number of plane waves for $k = 0$ within E_{cut} in our calculations. The $E_{\text{cut}2}$ is for the plane wave expansion of χ , W , and ε . We note that, in many of the previous works [38,53,54], although E_{cut} used could be large, the $E_{\text{cut}2}$ used were rather small, e.g., 30 Ry (much smaller than E_{cut}). The number of plane waves N_q is around 3000, while N_r could be about ten times bigger (e.g., ~ 30000). (A spherical E_{cut} is used to determine the plane wave vectors, while N_r is defined by the full fast Fourier transform grid points).

The matrix for $G(q_1, q_2, k, i\omega)$ at each k point and ω point is a 3000×3000 matrix, requiring about 0.2 GB of memory. Typically, we have used 400 ω points along the imaginary axis of $i\omega$ (from -3×10^6 to 3×10^6 hartree), and 40 τ points (from -200 to 200 hartree $^{-1}$). The exponential $i\omega$ and $i\tau$ grid points were shown in our previous publications [22]. The smallest intervals for ω and τ in our calculations are 2×10^{-4} hartree and 0.01 hartree $^{-1}$, respectively, while the maximum ω_{max} and τ_{max} are 3×10^6 hartree and 200 hartree $^{-1}$. The grid convergence has been tested to ensure that the resulting error in quasiparticle eigenenergy is less than 0.01 eV. The techniques to carry out the Fourier transformation between $G(i\omega)$ and $G(i\tau)$ were described in detail in our previous publication [22]; the accuracy of such numerical Fourier transform is shown to be 10^{-7} hartree $^{-1}$. In doing this transformation, massive parallelization is used to distribute the k points and q_2 vectors. The message passing interface communicator is first divided into the number of k subcommunicators. In each

TABLE V. The experimental equilibrium lattice parameter (in \AA) used in this work. The E_{cut} ($E_{\text{cut}2}$) is the plane wave cutoff energy (in Ry) used for the *sc-GW* calculation. N_q is the resulting number of plane waves, while N_r is the total number of grid points in real space.

Systems	Lattice constant	E_{cut}	$E_{\text{cut}2}$	N_q	N_r
GaAs	5.66	85	170	3735	32768
ZnO	4.62	105	210	2980	27000
CdS	5.83	80	160	3479	27000

TABLE VI. Results of LDA, G_0W_0 , and sc-GW band gaps (in eV) for bulk GaAs, ZnO, and CdS. Others' G_0W_0 calculations (using the LDA as initial input) and experimental values E_{exp} (in eV) are also provided. ϵ_{LDA} is the dielectric constant used for G_0W_0 calculations, ϵ_{GW} is the final converged dielectric constant, and ϵ_{exp} is the experimental high-frequency dielectric constant. E_{test} (in eV) is the converged GW band gap using the fixed experimental dielectric constant. (For ZnO, we use the ϵ_{LDA} as the fixed value.) Strictly speaking, the ϵ shown here are actually the $1/\epsilon^{-1}(q_1 = 0, q_2 = 0, k = 0, \omega = 0)$.

Systems	LDA	G_0W_0	Others' G_0W_0	sc-GW	E_{exp}	ϵ_{LDA}	ϵ_{GW}	ϵ_{exp}	E_{test}
GaAs	0.13	1.29	1.30, ^a 1.29 ^b	2.04	1.52 ^c	11.81	4.98	10.89 ^d	1.03
ZnO	0.61	2.51	2.31 ^e	4.43	3.44 ^f	4.68	1.86		2.37
CdS	0.82	2.10	2.06, ^g 2.03 ^e	3.35	2.42 ^g	6.84	2.44	5.20 ^h	2.28

^aReference [27].

^bReference [7].

^cReference [59].

^dReference [60].

^eReference [55].

^fReference [61].

^gReference [62].

^hReference [63].

subcommunicator, q_2 is divided into different processors. Each processor group might only have a few q_2 points. To further decrease the computational cost, the crystal symmetry is used. Therefore, the first BZ can be represented with a reduced set of k vectors within the irreducible BZ. The calculations were carried out on the Titan supercomputer at OLCF using about 100 000 CPU processors. It takes about four hours for one fully converged sc-GW calculation.

IV. RESULTS AND DISCUSSIONS

We first study the quasiparticle energies. The calculated LDA, G_0W_0 (with LDA inputs) and sc-GW band gaps for bulk GaAs, ZnO, and CdS are tabulated in Table VI. When the non-self-consistent G_0W_0 calculations are performed, the band gaps for GaAs and CdS are 1.29 and 2.10 eV, respectively. These values are in relatively good agreement with other theoretical results [7,27,55], including the all-electron results [7,55]. For ZnO, the values of the G_0W_0 band gap are very scattered, ranging from 2.11 to 4.23 eV [27,44,55,56], due to different approximations, truncations, and initial input eigenenergies and wave functions. In one recent work [43], Shih *et al.* found that the conventional G_0W_0 method can yield a band gap that is very close to the experimental value for wurtzite ZnO, if one uses LDA + U as initial inputs, high cutoff energies, and enough conduction bands (about 3000 empty states). On the one hand, this example highlights the importance of high cutoff energies and the number of conduction bands to reach numerical convergence. On the other hand, it also shows that G_0W_0 quasiparticle energy could be highly sensitive to the input DFT eigenenergies and wave functions (e.g., using LDA instead of LDA + U will have a major difference). Some later studies [57,58] also discussed the issue of plasmon pole approximation used in the work of Shih *et al.* [43]. For zinc-blende ZnO as listed in Table VI, Chu *et al.* reported a 2.31 eV G_0W_0 band gap based on all-electron implementation [55] using LDA eigenenergies and wave functions as input. Our computed G_0W_0 band gap is 0.2 eV higher than their result. However, since only 150 conduction bands are included in their GW calculations, this

could lead to a numerically unconverged result. Figure 6 shows the dependence of the G_0W_0 band gap of ZnO on the number of conduction bands. The red dashed line is the band gap calculated using the full Green's function without truncation. It shows clearly that the quasiparticle gap of ZnO including 500 conduction bands does not converge completely, which is in agreement with the conclusion made by Shih *et al.* [43]. For systems with strongly localized states like ZnO, the convergence regarding the number of conduction band states can be very slow. As a result, it is crucial to use the nontruncated Green's function in the GW calculations. According to Fig. 6, our G_0W_0 using 150 conduction bands would yield a 2.2 eV band gap, which is quite close to the results in Ref. [55]. This close agreement of our semicore-pseudopotential G_0W_0 calculation and the all-electron calculation (when the same number of conduction bands is used) also confirms that the use of our semicore pseudopotential is accurate. For all three systems, it can also be noted that although the G_0W_0 band

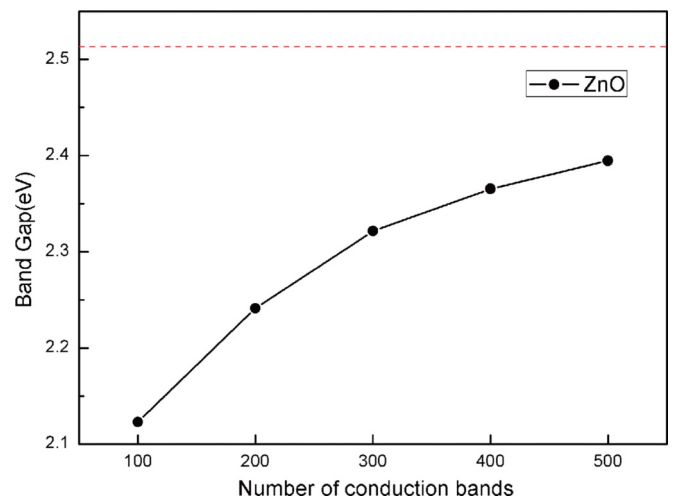


FIG. 6. G_0W_0 band gap of ZnO as a function of the number of conduction bands in evaluating the Green's function. The red dashed line is the full result without any conduction band truncation.

gaps considerably improved those at the LDA level, they systematically underestimate the experimental values.

The band gaps from sc-GW calculations are increased compared to the G_0W_0 band gaps and significantly overestimate the experimental values as can be seen in Table VI. Although similar phenomena have been discussed in previous literature [21,27,55], most of such works were under approximations more severe than the current work, and the previously reported overestimations were not as large as we reported here. We also noticed that our conclusion is different from a recent all-electron FP-LAPW sc-GW calculation by Chu *et al.* [55]. They emphasized the need of all-electron calculation and sc-GW calculation, and in many of their systems (including ZnO and CdS), their sc-GW results are much closer to the experiment than the G_0W_0 results. As we have shown above, our semicore-pseudopotential result is rather close to their all-electron G_0W_0 results (when the finite number of conduction band truncation in their calculation is taken into account), thus we believe the all-electron is not an issue here. There are several possible reasons for the differences between their results and ours. First is the diagonal approximation used in their approach, where the Green's function G has been approximated with a diagonal form similar to Eq. (1) [although the frequency-dependent denominator has been replaced by a general function of $f_i(\omega)$ during the self-consistent iteration], and the $\psi_i(r)$ basis is not updated. Instead, the G in our method is represented in the full matrices form that both the diagonal and off-diagonal elements are included in the self-consistent calculations. According to Shishkin and Kresse [27], the inclusion of off-diagonal elements has the tendency to open the band gaps. Thus, the relatively good agreement for the G_0W_0 result compared to Ref. [55], and the large difference for the sc-GW results, might indicate the importance of the off-diagonal term in the sc-GW calculation. Another possible reason is the finite number of conduction bands used in their diagonal representation of the G . Only 150 conduction bands are used and the energy of the highest eigenstate above the Fermi level is only 1.2 Ry. For strongly localized materials, these parameters might be far from converged as we discussed above [38].

From Table VI, we can see that, for GaAs, the sc-GW band gap is more than 30% larger than the experimental value, while G_0W_0 is about 14% smaller than the experimental one. It seems that the sc-GW results are worse than the G_0W_0 results. The same can be said for ZnO and CdS. This is in contrast to the conclusions made based on molecular sc-GW calculations [22,64], where the overall quality of the sc-GW results is similar to the G_0W_0 results. To analyze the origin for this overestimation, the converged macroscopic dielectric constants ϵ_{GW} of the sc-GW for GaAs, ZnO, and CdS are shown in Table VI as 4.98, 1.86, and 2.44, respectively. They are significantly smaller than the experimental values [60,63]. Due to the underestimated screening, it is natural the GW will give an overestimation of the band gap. Such sensitivity to the macroscopic dielectric constant does not exist in the molecular systems. We also noticed that the LDA dielectric constants ϵ_{LDA} are quite close to the experimental values. This partially explains why the G_0W_0 can yield a better band gap than the sc-GW results. To test this idea further, we have carried out the following tests. In these tests, we have used

the experimental macroscopic dielectric constant ϵ_{exp} . Then at every iteration step, according to the calculated macroscopic dielectric constant $1/\epsilon^{-1}(q_1=0, q_2=0, k=0, \omega=0)$, we defined a correction prefactor $\beta = \epsilon_{\text{exp}}/\epsilon^{-1}(0,0,0,0)$, and multiplied this prefactor β to all the $\epsilon^{-1}(q_1, q_2, k, i\omega)$ in the calculation of W in Eq. (6). The resulting band gaps E_{test} are also reported in Table VI. We can see that these test results significantly reduce the sc-GW band gaps, making them smaller than the experimental values. This indicates that the underestimation of the dielectric constant plays a determining role for the overestimation of the band gap. The fact that the resulting band gaps are smaller than experimental values is probably because the factor β overestimates the dielectric constant at other $(q_1, q_2, k, i\omega)$ points [since $\epsilon(q_1=0, q_2=0, k=0, \omega=0)$ is most sensitive to the band gap at $k=0$, while the other $(q_1, q_2, k, i\omega)$ is less sensitive, one might need a smaller prefactor]. To yield a better dielectric constant without fitting, one needs to include the higher order vertex terms in the Feynman diagram of the many-body perturbation theory.

We note that there are some recent works for the effects of lattice screening (electron-phonon coupling) to the semiconductor band gap [65–68]. Such lattice screening generally reduces the band gap, thus it could bring our sc-GW results into better agreement with the experiment. However, the reported lattice screening effects on GaAs [68] are rather small, only 0.06 eV, although there are reports of surprisingly large (0.7 eV) zero temperature lattice screening effects for diamond [65,66]. Nevertheless, in general, for semiconductors with a band gap smaller than 3 eV and for heavy elements, it looks like the lattice screening effect is less than 0.1 eV [66,68], and thus should not be enough to explain the difference between our sc-GW and experimental results.

Another important aspect is the accurate description of the d -state energies. We have computed the outmost cation d -state binding energy E_d at the Γ point, which is estimated as the average of all the corresponding d -state energies with respect to the valence band maximum (VBM), as presented in Table VII. It is clear that E_d predicted by LDA are underestimated by at least 2 eV compared to the experimental values. The too shallow LDA cation d -state energies are mainly caused by the well-known self-interaction error of d electrons within LDA. As expected, the G_0W_0 calculations perform better than LDA and place these d states at deeper binding energy for all the systems studied. However, the discrepancy can still be very large (around 1–2 eV). There are works [43,69] using LDA + U as initial inputs, in which the self-interaction is effectively removed, to give a good description of the ground states. In our self-consistent calculation, the energies of the

TABLE VII. The average semicore d -state binding energies (in eV) of GaAs, ZnO, and CdS at Γ calculated using LDA, G_0W_0 , and sc-GW methods. The experimental value Expt. (in eV) is given for comparison.

E_d	LDA	G_0W_0	sc-GW	Expt.
GaAs	-14.82	-16.73	-18.32	-(18.7–18.82) ^a
ZnO	-5.16	-5.97	-7.10	-(7.5–8.81) ^a
CdS	-7.51	-8.38	-9.67	-(9.2–10.0) ^a

^aReference [70].

cation d state are significantly improved when compared with experiment [70]. It seems like the d -state energy is relatively unaffected by the dielectric constant error, at least for the error at $1/\epsilon^{-1}(q_1 = 0, q_2 = 0, k = 0, \omega = 0)$; probably this is because the localized d state is mostly screened by finite k and q components of the dielectric constant, rather than by $1/\epsilon^{-1}(q_1 = 0, q_2 = 0, k = 0, \omega = 0)$. Since the ϵ at other k and

q points depends less sensitively on the band gap at the Γ point, they might have smaller errors. It is worth pointing out that the d -state energy is affected by the self-interaction energy, which has been corrected in the sc- GW calculation.

Finally, we reported the spectral functions using Eq. (24), and they are shown in Fig. 7 for $k = 0$. The GW spectral function shows sharp peaks at the quasiparticle energies. Note the peak positions are exactly the same as the ones shown in Tables VI and VII. As expected, the sharper peaks close to the Fermi energy are associated with the longer lifetimes of the corresponding quasiparticle states. In the work of Holm and von Barth for homogeneous electron gas [20], they observed a transfer of spectral weight from the plasmon satellite to the quasiparticle peak in the self-consistent GW calculations. This results in a weaker plasmon peak and a broader valence band. Similar to what they found, the valence band width of our sc- GW is slightly wider than that of G_0W_0 as shown in Fig. 7. In addition, we do not see any satellite peaks deep in the valence band. Such satellite peaks representing the plasmon excitations could be found by the cumulant method as a post process procedure [71].

V. CONCLUSIONS

In summary, we have implemented a fully self-consistent GW approach based on the solution of the Dyson equation using a plane wave basis set. We use this method for a detailed study of the quasiparticle energies and spectral properties for bulk GaAs, ZnO, and CdS using pseudopotentials with semicores. The Green's function is expressed as a full matrix without truncation. Algorithmic, numerical, and technical details of the self-consistent GW approach are presented to deal with the nonanalyticity and convergence issues in a bulk calculation. All these systems converge in four to five self-consistent iterations. We found that the sc- GW significantly overestimates the band gap due to the underestimation of the macroscopic dielectric constants during the self-consistent iterations. The results indicate that an accurate prediction of the quasiparticle band gap requires an accurate prediction of the dielectric function, which could be achieved by including the vertex correction beyond GW . Our work also sheds some light on why the G_0W_0 with LDA input could yield a better band gap compared to the sc- GW , since LDA often has relatively more accurate dielectric constants. We also demonstrated that the number of conduction bands and off-diagonal elements are very important in the GW calculation. After correcting the self-interaction error, sc- GW can yield accurate d -state energies, which less sensitively depend on the dielectric constant at $k = 0$.

ACKNOWLEDGMENTS

This work was supported by the Center for Computational Study of Excited-State Phenomena in Energy Materials at the Lawrence Berkeley National Laboratory, which is funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05CH11231, as part of the Computational Materials Sciences Program. H.C. is supported by the China Scholarship Council (Grant No. 201406470059),

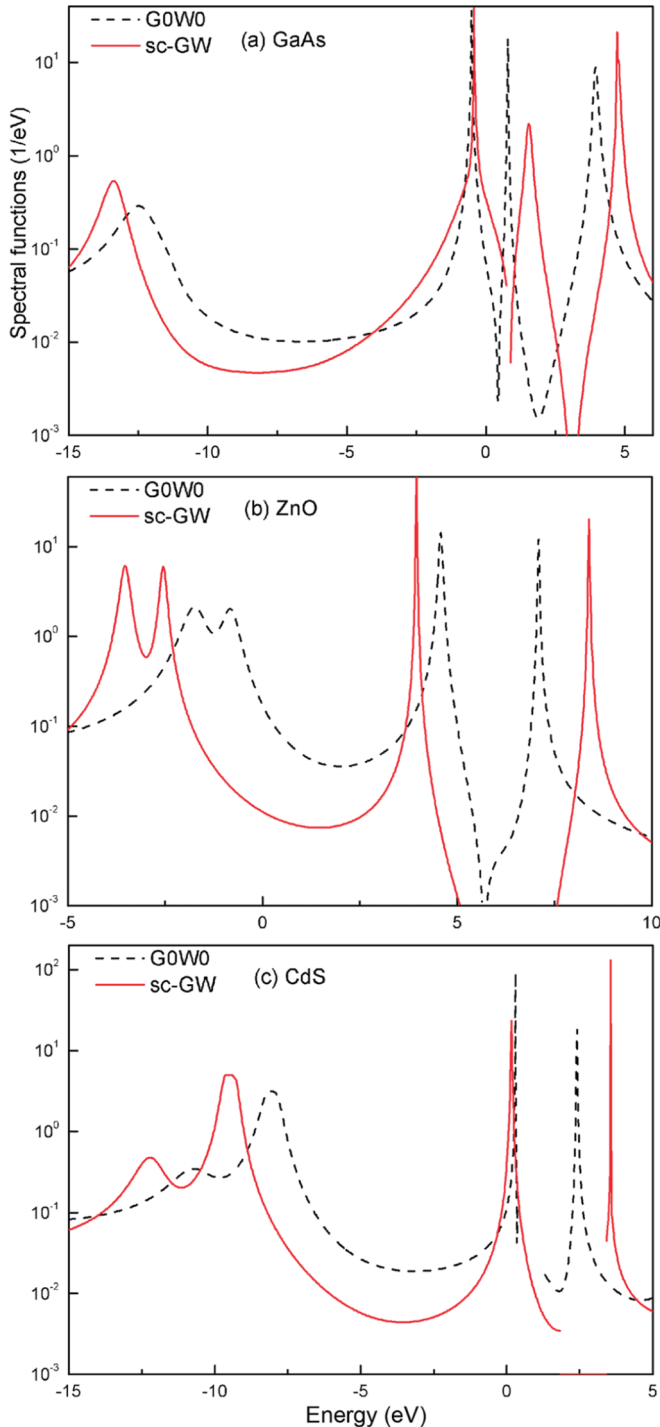


FIG. 7. The spectral functions for (a) GaAs, (b) ZnO, and (c) CdS. The dashed lines are for the G_0W_0 results and the solid lines are for the sc- GW results. The Fermi energies μ for GaAs, ZnO, and CdS are around 0.5, 6.1, and 1.5 eV, respectively.

the BUPT Excellent Ph.D. Students Foundation (Grant No.CX2015305), and the National Basic Research Program of China (973 Program Grant No. 2014CB643900). This work used the resources of the National Energy Research Scientific Computing Center (NERSC) and Oak Ridge Leadership Computing Facility (OLCF) with the computational time allocated by the ASCR Leadership Computing Challenge (ALCC) program.

APPENDIX A: DERIVATION OF GW FORMULA IN PERIODIC SYSTEMS

In this Appendix, the derivation of the GW formula for periodic systems is outlined. The noninteracting Green's function G_0 can be explicitly written in terms of the single-particle eigenfunctions $\psi_n(r)$ and eigenvalues ε_n as

$$G_0(r_1, r_2, \omega) = \sum_n \frac{\psi_n(r_1)\psi_n^*(r_2)}{\omega - \varepsilon_n \pm i\delta_n}, \quad (\text{A1})$$

where n is the orbital numbers. According to the Bloch's theorem, the noninteracting Green's function G_0 in the

periodic environment has the form

$$G_0(r_1, r_2, \omega) = \sum_k \sum_n \frac{u_{nk}(r_1)u_{nk}^*(r_2)e^{-ik(r_1-r_2)}}{\omega - \varepsilon_{nk} \pm i\delta_{nk}} w_k. \quad (\text{A2})$$

Here, $u_{nk}(r)$ is a periodic function with the same periodicity as the system and k is the wave vector and w_k is the weighting factor of each k point (w_k is omitted in the equations below for simplicity). For interacting systems, it can be proved that the Green's function G in the time space $i\tau$ has a similar form:

$$G(r_1, r_2, i\tau) = \sum_k G(r_1, r_2, k, i\tau)e^{-ik(r_1-r_2)}. \quad (\text{A3})$$

The basic formalism of our fully self-consistent GW approach for an isolated molecule system can be found in our previous paper [22]. Here we will present the derivation of these equations in periodic systems. The irreducible polarization χ within random-phase approximation is given by the product of two Green's functions as

$$\chi(r_1, r_2, i\tau) = -iG(r_1, r_2, i\tau)G(r_2, r_1, -i\tau). \quad (\text{A4})$$

Substituting Eq. (A3) into Eq. (A4), we can get

$$\begin{aligned} \chi(r_1, r_2, i\tau) &= -i \left\{ \sum_{k_1} G(r_1, r_2, k_1, i\tau)e^{-ik_1(r_1-r_2)} \right\} \left\{ \sum_{k_2} G(r_2, r_1, k_2, -i\tau)e^{-ik_2(r_2-r_1)} \right\} \\ &= \sum_k -i \left\{ \sum_{k_2} G(r_1, r_2, k+k_2, i\tau)G(r_2, r_1, k_2, -i\tau) \right\} e^{-ik(r_1-r_2)} = \sum_k \chi(r_1, r_2, k, i\tau)e^{-ik(r_1-r_2)} \end{aligned}$$

$$\text{So : } \chi(r_1, r_2, k, i\tau) = -i \sum_{k_2} G(r_1, r_2, k+k_2, i\tau)G(r_2, r_1, k_2, -i\tau). \quad (\text{A5})$$

Using Eq. (A5), the dielectric function ε is calculated as

$$\begin{aligned} \varepsilon(r_1, r_2, i\omega) &= \delta(r_1 - r_2) - \int \chi(r_1, r_3, i\omega)v(r_3 - r_2)dr_3 \\ &= \delta(r_1 - r_2) - \int \left\{ \sum_{k_1} \chi(r_1, r_3, k_1, i\omega)e^{-ik_1(r_1-r_3)} \right\} \left\{ \sum_{k_2} v(r_3 - r_2, k_2)e^{-ik_2(r_3-r_2)} \right\} dr_3. \end{aligned} \quad (\text{A6})$$

Note, $\int dr_3 = \int_{\Omega} dr_3 \sum_R$, where Ω is the unit cell and R is the lattice vector. So $\sum_R e^{-i(k_1-k_2)R} = \delta_{k_1, k_2}$. As a result, the dielectric function can be defined as

$$\varepsilon(r_1, r_2, i\omega) = \sum_k \delta(r_1 - r_2, k)e^{-ik(r_1-r_2)} - \int \left\{ \sum_k \chi(r_1, r_3, k, i\omega)v(r_3 - r_2, k)e^{-ik(r_1-r_2)} \right\} dr_3$$

$$\text{So : } \varepsilon(r_1, r_2, k, i\omega) = \delta(r_1 - r_2, k) - \int \chi(r_1, r_3, k, i\omega)v(r_3 - r_2, k)dr_3. \quad (\text{A7})$$

Unlike the polarization χ , there is no k mix in the formula of dielectric function ε , as well as in the inverse function ε^{-1} . Using a similar way, we can get the screened Coulomb potential W and the self-energy Σ for periodic systems in the below forms:

$$W(r_1, r_2, k, i\omega) = v(r_1 - r_3, k)\varepsilon^{-1}(r_3, r_2, k, i\omega), \quad (\text{A8})$$

$$\Sigma(r_1, r_2, k, i\tau) = i \sum_{k_2} G(r_1, r_2, k - k_2, i\tau)W(r_1, r_2, k_2, i\tau). \quad (\text{A9})$$

Finally, after Eqs. (A5)–(A9) are yielded, the Dyson equation for periodic solid is given as

$$\begin{aligned} G^{-1}(r_1, r_2, k, i\omega) &= (i\omega + \mu)\delta(r_1 - r_2, k) - H(r_1, r_2, k) \\ &\quad - \Sigma(r_1, r_2, k, i\omega). \end{aligned} \quad (\text{A10})$$

In the reciprocal space, it is expressed as

$$\begin{aligned} G^{-1}(q_1, q_2, k, i\omega) &= (i\omega + \mu)\delta_{q_1, q_2} - H(q_1, q_2, k) \\ &\quad - \Sigma(q_1, q_2, k, i\omega). \end{aligned} \quad (\text{A11})$$

TABLE VIII. The comparison of the LDA results for ZnO with respect to the two different cutoff energies. E_g is the minimum energy gap (eV) and E_d is the average d -state binding energy (eV) of Zn. Some selected eigenvalues (eV) relative to the valence band maximum (Γ_{13v}) at the Γ point are also presented.

Systems	$E_{\text{cut}} = 350 \text{ Ry}$	$E_{\text{cut}} = 105 \text{ Ry}$
E_g	0.613	0.617
$\Gamma_{1v}(\text{Zn } 3s)$	-122.65	-136.03
$\Gamma_{2v}(\text{Zn } 3p)$	-77.34	-83.27
$\Gamma_{5v}(\text{O } 2s)$	-17.34	-17.52
$\Gamma_{6v}(\text{Zn } 3d)$	-5.71	-5.70
E_d	-5.19	-5.16

APPENDIX B: CALCULATION OF POLARIZABILITY OF $k \rightarrow 0$

In this Appendix, details about the special case for $k \rightarrow 0$ are presented. To test the “head” expression of Eq. (11) in the main text, the noninteracting Green’s function G_0 in $i\tau$ space is used and its analytical expression can be written down as

$$\begin{aligned}
 G(r_1, r_2, k, i\tau) &= \sum_{n, \varepsilon_{nk} < \mu} \psi_{nk}(r_1) \psi_{nk}^*(r_2) e^{-(\varepsilon_{nk} - \mu)\tau}, \quad \text{for } \tau > 0 \\
 &= - \sum_{n, \varepsilon_{nk} < \mu} \psi_{nk}(r_1) \psi_{nk}^*(r_2) e^{-(\varepsilon_{nk} - \mu)\tau}, \quad \text{for } \tau < 0. \quad (\text{B1})
 \end{aligned}$$

Here $\psi_{nk}(r)$ is the single-particle eigenwave function, ε_{nk} is its eigenenergy, and n is the index for the band states. Substituting Eq. (B1) into Eq. (8), the single-particle polarizability $\chi(r_1, r_2, k, i\tau)$ in the limit for $k \rightarrow 0$ is obtained as (for $\tau > 0$)

$$\begin{aligned}
 \chi(r_1, r_2, k, i\tau) &= -i \sum_{k_2} \sum_{n, n_2} \psi_{n(k_2+k)}(r_1) \psi_{n_2 k_2}^*(r_1) \\
 &\quad \times \psi_{n_2 k_2}(r_2) \psi_{n(k_2+k)}^*(r_2) e^{-(\varepsilon_{n(k_2+k)} - \varepsilon_{n_2 k_2})\tau}. \quad (\text{B2})
 \end{aligned}$$

Transforming to reciprocal space, the resulting expression for $\chi(q_1, q_2, k, i\tau)$ is

$$\begin{aligned}
 \chi(q_1, q_2, k, i\tau) &= \frac{-i}{\Omega} \sum_{k_2, n, n_2} \langle n, k_2 + k | e^{i(q_1+k)r_1} | n_2, k_2 \rangle \\
 &\quad \times \langle n_2, k_2 | e^{-i(q_2+k)r_2} | n, k_2 + k \rangle e^{-(\varepsilon_{n(k_2+k)} - \varepsilon_{n_2 k_2})\tau}. \quad (\text{B3})
 \end{aligned}$$

Using first-order perturbation theory, the wave function at $(k_2 + k)$ can be obtained in terms of those at the k_2 point. The result is

$$\langle n_2, k_2 | e^{-ikr} | n, k_2 + k \rangle = \frac{\langle n_2, k_2 | k_\alpha \nabla_{k_2}^\alpha H(r) | n, k_2 \rangle}{\varepsilon_{nk_2} - \varepsilon_{n_2 k_2}}. \quad (\text{B4})$$

TABLE IX. The parameters of β , r_{peak} , and r_{cut} used for the pseudopotentials of GaAs and ZnO. Each row of the table stands for the parameters used in the Gaussian function $f(r)$ added to the original s , or p or d potentials of Zn or Ga atoms.

Atom	B (hartree)	r_{peak} (bohr)	r_{cut} (bohr)
Zn vs	-0.90	0.49	0.97
Zn vp	0	0.06	1.15
Zn vd	0	0.06	0.97
Ga vs	0	0.47	0.60
Ga vp	0.60	0.05	1.18
Ga vd	-0.36	0.05	1.18

$k_\alpha \nabla_{k_2}^\alpha H(r)$ is the same as that defined in the main text. The “head” expression of $\chi(q_1 = 0, q_2 = 0, k, i\tau)$ at the $k = 0$ point is given by

$$\begin{aligned}
 \chi(q_1 = 0, q_2 = 0, k = 0, i\tau) &= \frac{-i}{\Omega} \sum_{k_2, n, n_2} \langle n, k_2 | k_\alpha \nabla_{k_2}^\alpha H(r_1) | n_2, k_2 \rangle \langle n_2, k_2 | k_\beta \nabla_{k_2}^\beta H(r_2) | n, k_2 \rangle \\
 &\quad \times \frac{1}{(\varepsilon_{nk_2} - \varepsilon_{n_2 k_2})^2} \\
 &\quad \times e^{-(\varepsilon_{nk_2} - \varepsilon_{n_2 k_2})\tau}. \quad (\text{B5})
 \end{aligned}$$

After twice derivation of τ , Eq. (B5) is just the same as the intermediate term $\chi_{\alpha, \beta}^{(0)}(q_1 = 0, q_2 = 0, k = 0, i\tau) k_\alpha k_\beta$ defined in Eq. (10). As a result, the “head” expression of $\chi(q_1 = 0, q_2 = 0, k, i\tau)$ for the special case of $k \rightarrow 0$ expanded using Eq. (9) is correct. Similarly, the “wings” of $\chi(q_1 = 0, q_2, k, i\tau)$ used in our approach is

$$\begin{aligned}
 \chi_\alpha^{(0)}(q_1 = 0, q_2, k, i\tau) k_\alpha &= -ik_\alpha \sum_{k_2} \int [\nabla_{k_2}^\alpha H(r_1) G(r_1, r_2, k_2, i\tau) \\
 &\quad \times G(r_2, r_1, k_2, -i\tau) w_{k_2} e^{-q_2 r_2}] d^3 r_1 d^3 r_2 \\
 \text{So : } \chi(q_1 = 0, q_2, k, i\tau) &= - \int_\infty^\tau [\chi_\alpha^{(0)}(q_1 = 0, q_2, k, i\tau) k_\alpha] d\tau. \quad (\text{B6})
 \end{aligned}$$

The first-order k_α in the “wings” expansion as a function of k will cancel with the k in the denominator of Eq. (7).

APPENDIX C: POTENTIAL DETAILS

In this Appendix, the comparison for ZnO with respect to the above two different energy cutoffs is given in Table VIII and the parameters of β , r_{peak} , and r_{cut} used for the modified pseudopotentials of GaAs and ZnO are presented in Table IX.

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

[2] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).

[3] L. J. Sham and M. Schlüter, *Phys. Rev. Lett.* **51**, 1888 (1983).

[4] L. Hedin, *Phys. Rev.* **139**, A796 (1965).

[5] E. Luppi, H.-C. Weissker, S. Bottaro, F. Sottile, V. Veniard, L. Reining, and G. Onida, *Phys. Rev. B* **78**, 245124 (2008).

[6] A. Schleife, C. Rödl, F. Fuchs, J. Furthmüller, and F. Bechstedt, *Phys. Rev. B* **80**, 035112 (2009).

- [7] R. Gómez-Abal, X. Li, M. Scheffler, and C. Ambrosch-Draxl, *Phys. Rev. Lett.* **101**, 106404 (2008).
- [8] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, *Phys. Rev. B* **76**, 115109 (2007).
- [9] M. Shishkin, M. Marsman, and G. Kresse, *Phys. Rev. Lett.* **99**, 246403 (2007).
- [10] M. van Schilfhaarde, T. Kotani, and S. V. Faleev, *Phys. Rev. B* **74**, 245125 (2006).
- [11] M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* **34**, 5390 (1986).
- [12] R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. B* **37**, 10159 (1988).
- [13] E. L. Shirley, X. Zhu, and S. G. Louie, *Phys. Rev. Lett.* **69**, 2955 (1992).
- [14] E. L. Shirley, X. Zhu, and S. G. Louie, *Phys. Rev. B* **56**, 6648 (1997).
- [15] M. Rohlfing, P. Krüger, and J. Pollmann, *Phys. Rev. B* **57**, 6485 (1998).
- [16] H. N. Rojas, R. W. Godby, and R. J. Needs, *Phys. Rev. Lett.* **74**, 1827 (1995).
- [17] L. I. Bendauid and E. A. Carter, in *First Principles Approaches to Spectroscopic Properties of Complex Materials*, edited by C. Di Valentin, S. Botti, and M. Cococcioni (Springer, Berlin/Heidelberg, 2014), p. 47.
- [18] A. Schindlmayr, *Phys. Rev. B* **56**, 3528 (1997).
- [19] A. Kutepov, S. Y. Savrasov, and G. Kotliar, *Phys. Rev. B* **80**, 041103 (2009).
- [20] B. Holm and U. von Barth, *Phys. Rev. B* **57**, 2108 (1998).
- [21] W.-D. Schöne and A. G. Eguiluz, *Phys. Rev. Lett.* **81**, 1662 (1998).
- [22] L.-W. Wang, *Phys. Rev. B* **91**, 125135 (2015).
- [23] P. Koval, D. Foerster, and D. Sánchez-Portal, *Phys. Rev. B* **89**, 155417 (2014).
- [24] F. Caruso, P. Rinke, X. Ren, M. Scheffler, and A. Rubio, *Phys. Rev. B* **86**, 081102 (2012).
- [25] S. Lany, *Phys. Rev. B* **87**, 085112 (2013).
- [26] H. Jiang, R. I. Gomez-Abal, P. Rinke, and M. Scheffler, *Phys. Rev. Lett.* **102**, 126403 (2009).
- [27] M. Shishkin and G. Kresse, *Phys. Rev. B* **75**, 235102 (2007).
- [28] S. V. Faleev, M. van Schilfhaarde, and T. Kotani, *Phys. Rev. Lett.* **93**, 126406 (2004).
- [29] A. Klein, *Phys. Rev.* **121**, 950 (1961).
- [30] L.-W. Wang, *Phys. Rev. B* **82**, 115111 (2010).
- [31] Y. M. Niquet, M. Fuchs, and X. Gonze, *Phys. Rev. A* **68**, 032507 (2003).
- [32] G. Baym, *Phys. Rev.* **127**, 1391 (1962).
- [33] G. Baym and L. P. Kadanoff, *Phys. Rev.* **124**, 287 (1961).
- [34] M. Strange, C. Rostgaard, H. Häkkinen, and K. S. Thygesen, *Phys. Rev. B* **83**, 115108 (2011).
- [35] A. Fleszar and W. Hanke, *Phys. Rev. B* **71**, 045207 (2005).
- [36] R. Sakuma, T. Miyake, and F. Aryasetiawan, *Phys. Rev. B* **78**, 075106 (2008).
- [37] W. Ku and A. G. Eguiluz, *Phys. Rev. Lett.* **89**, 126401 (2002).
- [38] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, *Phys. Rev. B* **69**, 125212 (2004).
- [39] O. A. von Lilienfeld and P. A. Schultz, *Phys. Rev. B* **77**, 115202 (2008).
- [40] A. Kutepov, K. Haule, S. Y. Savrasov, and G. Kotliar, *Phys. Rev. B* **85**, 155129 (2012).
- [41] A. Stan, N. E. Dahlen, and R. van Leeuwen, *Europhys. Lett.* **76**, 298 (2006).
- [42] K. Delaney, P. Garcia-Gonzalez, A. Rubio, P. Rinke, and R. W. Godby, *Phys. Rev. Lett.* **93**, 249701 (2004); Wei Ku and A. G. Eguiluz, *ibid.* **93**, 249702 (2004).
- [43] B. C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, *Phys. Rev. Lett.* **105**, 146401 (2010).
- [44] M. Usuda, N. Hamada, T. Kotani, and M. van Schilfhaarde, *Phys. Rev. B* **66**, 125101 (2002).
- [45] G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
- [46] M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* **35**, 5585 (1987).
- [47] S. L. Adler, *Phys. Rev.* **126**, 413 (1962).
- [48] N. Wiser, *Phys. Rev.* **129**, 62 (1963).
- [49] M. S. Hybertsen and S. G. Louie, *Phys. Rev. Lett.* **55**, 1418 (1985).
- [50] V. Olevano, M. Palummo, G. Onida, and R. Del Sole, *Phys. Rev. B* **60**, 14224 (1999).
- [51] F. Gygi and A. Baldereschi, *Phys. Rev. B* **34**, 4405 (1986).
- [52] M. Rohlfing, P. Kruger, and J. Pollmann, *Phys. Rev. Lett.* **75**, 3489 (1995).
- [53] W. Luo, S. Ismail-Beigi, M. L. Cohen, and S. G. Louie, *Phys. Rev. B* **66**, 195215 (2002).
- [54] M. R. Filip, C. E. Patrick, and F. Giustino, *Phys. Rev. B* **87**, 205125 (2013).
- [55] I.-H. Chu, J. P. Trinastic, Y.-P. Wang, A. G. Eguiluz, A. Kozhevnikov, T. C. Schulthess, and H.-P. Cheng, *Phys. Rev. B* **93**, 125210 (2016).
- [56] S. Massidda, R. Resta, M. Posternak, and A. Baldereschi, *Phys. Rev. B* **52**, R16977 (1995).
- [57] C. Friedrich, M. C. Müller, and S. Blügel, *Phys. Rev. B* **83**, 081101(R) (2011).
- [58] M. Stankovski, G. Antonius, D. Waroquiers, A. Miglio, H. Dixit, K. Sankaran, M. Giantomassi, X. Gonze, M. Cote, and G. M. Rignanese, *Phys. Rev. B* **84**, 241201(R) (2011).
- [59] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 2005).
- [60] J. B. McKitterick, *Phys. Rev. B* **28**, 7384 (1983).
- [61] A. Mang, K. Reimann, and S. Rübenacke, *Solid State Commun.* **94**, 251 (1995).
- [62] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices* (Wiley, New York, 2006).
- [63] L. Guo, S. Zhang, W. Feng, G. Hu, and W. Li, *J. Alloys Compd.* **579**, 583 (2013).
- [64] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, *Phys. Rev. B* **88**, 075105 (2013).
- [65] R. Ramírez, C. P. Herrero, and E. R. Hernández, *Phys. Rev. B* **73**, 245202 (2006).
- [66] S. Zollner, M. Cardona, and S. Gopalan, *Phys. Rev. B* **45**, 3376 (1992).
- [67] F. Giustino, S. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* **105**, 265501 (2010).
- [68] S. Botti and M. A. L. Marques, *Phys. Rev. Lett.* **110**, 226404 (2013).
- [69] T. Miyake, P. Zhang, M. L. Cohen, and S. G. Louie, *Phys. Rev. B* **74**, 245213 (2006).
- [70] L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, and D. A. Shirley, *Phys. Rev. B* **9**, 600 (1974).
- [71] B. Holm and F. Aryasetiawan, *Phys. Rev. B* **56**, 12825 (1997).