

## Corrections to density-functional theory band gaps

Kurt A. Johnson and N. W. Ashcroft

Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853-2501

(Received 12 January 1998)

An effective exchange-correlation potential for conduction-band states differing from that of the valence-band states is found within an extension of an extreme tight-binding model. Starting from Hedin's *GW* formulation of the self-energy, simplifications are made by including only near-neighbor interactions, and by an expansion of band energies carried out around their Brillouin-zone averages. The potential difference is applied as a perturbation to the conduction-band states from a Kohn-Sham calculation with the local-density approximation, and a scissors-type band-gap correction is then obtained in a simple and efficient manner. Although the model is valid for strictly insulating systems, it is found that the correction (when adjusted to reproduce the known silicon and carbon band gaps as in a Slater *X- $\alpha$*  method) leads to semiconductor band gaps within 0.1–0.3 eV of their experimental values. Both zinc-blende and wurtzite semiconductors of the IV, III-V, and II-VI groups are studied here. [S0163-1829(98)01244-2]

### I. INTRODUCTION

Significant theoretical effort has been devoted to the subject of correcting the excited-state spectrum of electronic systems resulting from density-functional theory (DFT) calculations<sup>1</sup> as normally carried out within the local-density approximation (LDA). It is a well-known problem that the LDA utilized within the Kohn-Sham (KS) formalism<sup>2</sup> leads to a substantial underestimate of calculated band gaps in semiconductors and insulators. Differences from experiment are typically about 50%, but deviations as large as 100% also occur. It is also well-known that the KS eigenvalues in general have no strict physical interpretation except for the highest occupied orbital;<sup>3,4</sup> therefore, any agreement between LDA band gaps and experiment might well be considered fortuitous. Valence-band states in semiconductors that contribute to the electronic ground-state density do tend to agree with experiment although the total valence-band width may be underestimated. On the other hand, conduction-band states of a semiconductor are often found to agree in *form*,<sup>5</sup> but generally not in location, with respect to the valence bands. The difficulty in achieving useful excitation energies with the DFT and the LDA has largely been attributed to the presence of a discontinuity in the true DFT exchange-correlation functional derivative, a property not manifested within the LDA.<sup>6,7</sup>

A major approach to correcting this state of affairs has been through the use of Hedin's *GW* approximation<sup>8,9</sup> in which an energy dependent, nonlocal self-energy replaces the LDA for the exchange-correlation energy, and true quasiparticle eigenvalues are then found by correcting the LDA eigenvalues through perturbation theory. The "full" *GW* method with direct calculation and inversion of the dielectric matrix in the random-phase approximation (RPA) has been shown to be capable of an accuracy near 0.1 eV for a range of materials such as Si, C, GaAs, SiC, and LiC,<sup>10–12</sup> and it is the standard by which any other technique is often judged. Although highly accurate, calculation of the necessary self-energy term requires the full frequency dependent, inverse dielectric matrix and it involves a considerable computa-

tional burden. Various simplifications for the dielectric matrix<sup>13–15</sup> have therefore been introduced to ease the computational effort and avoid the direct calculation by *ab initio* means. With the reduction in effort also comes a slight penalty in accuracy, perhaps at the level of 0.1–0.3 eV.<sup>16–19</sup>

There exist a few attempts at an analytic treatment and direct simplification of the electron self-energy operator leading to band-gap corrections showing some success.<sup>20–25</sup> Here, we approach the problem along the lines originally proposed by Sterne and Inkson<sup>20</sup> (SI) and later by Hanke and Sham.<sup>23</sup> The strategy is based on a self-energy correction within a simple analytic model containing the essential physics of excited electronic states in a semiconductor or insulator. In Sec. II, we introduce the goal of our method, discuss the approximations of the model, and finally propose a form for band-gap corrections. The ensuing algorithm is relatively simple, is quite efficient, and may be extended further. We emphasize now that the necessary inputs to the proposed corrections are obtained from prior LDA *ab initio* calculations. In Sec. III, we apply the correction to LDA band gaps of a range of common semiconducting materials. As will be seen, the results generally compare well with experiment [Figs. 2(a) and 2(b)]. In Sec. IV, we conclude with a discussion of possible extensions of this work.

### II. APPROACH TO THE MODEL *GW* CORRECTION

The general approach to *ab initio* DFT calculations proceeds from the solution of the KS equations

$$[T + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{\text{xc}}(\mathbf{r})]\Psi_i(\mathbf{r}) = \epsilon_i\Psi_i(\mathbf{r}), \quad (1)$$

where  $T$  is the kinetic-energy operator,  $V_{\text{ext}}$  the external potential,  $V_H$  the Hartree potential, and  $V_{\text{xc}}$  the exchange-correlation potential given by the standard functional derivative of the exchange-correlation energy with respect to the single-particle density  $\rho(\mathbf{r})$ , i.e.,

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})}, \quad (2)$$

and is usually given by the local-density approximation. In the ground state, the electron density  $\rho(\mathbf{r})$  is simply the sum of the single-particle densities of the lowest occupied KS states.

Given a set of eigenvalues, the gap  $E_g$  is the difference in energy between the lowest unoccupied and highest occupied states. As noted, with the exception of the highest occupied state, the KS eigenvalues from Eq. (1) cannot be directly identified with quasiparticle energies because the KS equations merely provide the formal machinery for finding the electron density and corresponding total energy within DFT.

The band gap is also given by the difference between the ionization potential  $I = E_{N-1} - E_N$  and the electron affinity  $A = E_N - E_{N+1}$ , where  $E_N$  represents the total energy of a system with  $N$  particles. Both are found within DFT by differences of well-defined ground-state total energies and are related to the highest occupied eigenvalues of the  $N$  and  $N+1$ -electron systems, respectively.<sup>3</sup> In principle, with two separate KS calculations, the band gap is given by

$$E_g = I - A = \varepsilon_{n+1}^{N+1} - \varepsilon_n^N, \quad (3)$$

where we have written the eigenvalue  $\varepsilon_n^N$  with  $N$  the total particle number and  $n$  the eigenstate of the  $N$  particle system. Any band-gap underestimate (which we will call  $\Delta$ ) resulting from the use of the  $n+1$  excited state eigenvalue from the KS calculation for an  $N$  particle system ( $\varepsilon_{n+1}^N$ ) is then formally given by

$$\Delta = (\varepsilon_{n+1}^{N+1} - \varepsilon_n^N) - (\varepsilon_{n+1}^N - \varepsilon_n^N) = \varepsilon_{n+1}^{N+1} - \varepsilon_{n+1}^N. \quad (4)$$

Given that a discontinuity exists in the exchange-correlation potential when one electron is added to an  $N$  particle system,  $\Delta$  will be finite. Accordingly, we need to determine a correction to the unoccupied orbital eigenvalues arising from this discontinuity.

We can derive the *difference* in the exchange-correlation energy of an electron in the valence-band and conduction-band states of a semiconducting system within the *GW* approximation, and this provides us with an approximation to the discontinuity  $\Delta$ . In Hedin's *GW* approximation (Ref. 8) the self-energy is given by

$$\Sigma(\omega) = i \int \frac{d\omega'}{2\pi} G(\omega - \omega') W(\omega') \exp(-i\delta\omega'), \quad (5)$$

where  $G$  is the Green's function

$$G(\mathbf{r}, \mathbf{r}'; \omega) = \sum_{\mathbf{k}} \frac{\psi_{\mathbf{k}}(\mathbf{r}) \psi_{\mathbf{k}}^*(\mathbf{r}')}{\omega - \varepsilon_{\mathbf{k}} + i\delta}, \quad (6)$$

with  $\varepsilon_{\mathbf{k}}$  the self-consistent single-particle energies,  $\psi_{\mathbf{k}}(\mathbf{r})$  the associated single-particle state, and  $W$  is the linearly screened Coulomb interaction

$$W(\mathbf{r}, \mathbf{r}'; \omega) = \int d\mathbf{r}'' \varepsilon^{-1}(\mathbf{r}, \mathbf{r}''; \omega) v(\mathbf{r}'' - \mathbf{r}'). \quad (7)$$

This screened interaction requires the frequency-dependent dielectric function

$$\varepsilon(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r}, \mathbf{r}') - \int d\mathbf{r}'' v(\mathbf{r} - \mathbf{r}'') P(\mathbf{r}'', \mathbf{r}'; \omega), \quad (8)$$

where  $P$  is just the random-phase-approximation polarization, namely,

$$P(\mathbf{r}, \mathbf{r}'; \omega) = 2 \sum_{\mathbf{k}, \mathbf{k}'} (f_{\mathbf{k}} - f_{\mathbf{k}'}) \frac{\psi_{\mathbf{k}}(\mathbf{r}) \psi_{\mathbf{k}'}^*(\mathbf{r}) \psi_{\mathbf{k}}^*(\mathbf{r}') \psi_{\mathbf{k}'}(\mathbf{r}')}{\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}'} - \omega - i\delta} \quad (9)$$

with all valence-band states assumed to be doubly occupied for an insulating system.

Proceeding with a typical semiconductor, we may assume a two-atom basis associated with, for instance, the zincblende or diamond-crystal structures. The single-particle states  $\psi_{\mathbf{k}}$  are expanded in Wannier states  $\phi_{\nu}^n(\mathbf{r} - \mathbf{R})$ , localized in the unit cell designated by the lattice site  $\mathbf{R}$ . The index  $n$  represents the bond's character ( $v$  for a bonding orbital or  $c$  an antibonding orbital), and  $\nu$  is the bond number that identifies the location within the unit cell, i.e., one of the oriented molecular hybrids composed of two neighboring  $sp^3$  orbitals in a tetrahedrally bonded semiconductor. Therefore, the model includes  $\nu$  valence-band states derived from the bonding Wannier orbitals, and similarly  $\nu$  conduction bands derived from the antibonding orbitals. The resulting Bloch states are written as

$$\psi_{\mathbf{n}\mathbf{k}}(\mathbf{r}) = \sqrt{2/N} \sum_{\mathbf{R}} \exp(i\mathbf{k} \cdot \mathbf{R}) \phi_{\nu}^n(\mathbf{r} - \mathbf{R}), \quad (10)$$

where  $N$  is the number of atoms in the crystal ( $N/2$  being the number of unit cells).

Now we assume that each bond orbital indexed by  $\mathbf{n}$  (representing both  $\nu$  and  $n$ ) gives rise to a well-defined band, and we write the associated eigenvalue  $\varepsilon_{\mathbf{k}}$  as

$$\varepsilon_{\mathbf{k}} \equiv \varepsilon_{\mathbf{n}\mathbf{k}} = \varepsilon_{\mathbf{n}}^0 + \delta\varepsilon_{\mathbf{n}}(\mathbf{k}), \quad (11)$$

where  $\varepsilon_{\mathbf{n}}^0$  is some reference energy level for band  $\mathbf{n}$  with  $\delta\varepsilon_{\mathbf{n}}(\mathbf{k})$  describing the dispersion about that reference level. To establish such a reference level, we examine the denominator of Eq. (9) by expanding in terms of small  $\delta\varepsilon_{\mathbf{n}}(\mathbf{k})$ , writing the polarization with Eq. (10) as

$$\begin{aligned} P(\mathbf{r}, \mathbf{r}'; \omega) &= 2 \sum_{nn' \nu \mathbf{k}, \mathbf{k}'} \frac{f_{\mathbf{n}\mathbf{k}} - f_{\mathbf{n}'\mathbf{k}'}}{\varepsilon_{\mathbf{n}}^0 - \varepsilon_{\mathbf{n}'}^0 - \omega - i\delta} \\ &\times \left( 1 + \frac{\delta\varepsilon_{\mathbf{n}}(\mathbf{k}) - \delta\varepsilon_{\mathbf{n}'}(\mathbf{k}')}{\varepsilon_{\mathbf{n}}^0 - \varepsilon_{\mathbf{n}'}^0 - \omega} + \dots \right) \\ &\times \sum_{\nu \mathbf{R}, \mathbf{R}'} e^{i(\mathbf{k} - \mathbf{k}') \cdot (\mathbf{R} - \mathbf{R}')} \phi_{\nu}^n(\mathbf{r} - \mathbf{R}) \\ &\times \phi_{\nu}^{n'*}(\mathbf{r} - \mathbf{R}) \phi_{\nu}^{n'}(\mathbf{r}' - \mathbf{R}') \phi_{\nu}^{n'}(\mathbf{r}' - \mathbf{R}'). \end{aligned}$$

Here we ignore the overlap between all orbitals, i.e.,  $\phi^{\mathbf{n}}(\mathbf{r} - \mathbf{R}) \phi^{\mathbf{n}'}(\mathbf{r} - \mathbf{R}') \propto \delta_{\mathbf{R}, \mathbf{R}'} \delta_{\nu, \nu'}$ . Physically, this approximation misses contributions to the polarization for which  $\mathbf{r}$  and  $\mathbf{r}'$  are within the region of two overlapping neighbor-bond orbitals, but these terms are shorter range than the on-site terms because of the exponential decay of the Wannier functions. Although these contributions would be important if we were attempting to produce an accurate picture of the  $\mathbf{q}$  dependence of the dielectric function itself, here we only re-

quire a satisfactory approximation for use in obtaining the self-energy. Continuing, we now make the choice,

$$\sum_{\mathbf{k}} \delta\varepsilon_{\mathbf{n}}(\mathbf{k}) = 0$$

for each  $\mathbf{n}$  by selecting  $\varepsilon_{\mathbf{n}}^0$  as the zone *average* band energy, a choice that eliminates the terms linear in  $\delta\varepsilon_{\mathbf{n}}(\mathbf{k})$  and leaves the remaining terms necessarily of second and higher order in the ratio of the bandwidth to average band gap.

Hence, the polarization reduces to

$$P(\mathbf{r}, \mathbf{r}'; \omega) = 2 \sum_{n, n'} \sum_{\mathbf{v}\mathbf{R}} \frac{f_n - f_{n'}}{\varepsilon_n^0 - \varepsilon_{n'}^0 - \omega - i\delta} \phi_{\mathbf{v}}^n(\mathbf{r} - \mathbf{R}) \times \phi_{\mathbf{v}'}^{n'*}(\mathbf{r} - \mathbf{R}) \phi_{\mathbf{v}'}^{n'*}(\mathbf{r}' - \mathbf{R}) \phi_{\mathbf{v}}^{n'}(\mathbf{r}' - \mathbf{R}), \quad (12)$$

which has the same form as in the extreme tight-binding model (ETBM) examined by Ortuno and Inkson<sup>26</sup> but now with the energy levels  $\varepsilon_n^0$  defined as the average energy of band  $n$ . With Eq. (12) the dielectric matrix is next found [see Eq. (A13) in the Appendix]. The expansion is retained to all orders as we wish to make a resummation of the self-energy resulting from this screened interaction; this allows for a more general form for the exchange-correlation potential within the ETBM.

Equation (A13) can now be inserted into Eq. (5) to develop the self-energy analytically as

$$\begin{aligned} \Sigma(\mathbf{r}, \mathbf{r}'; \omega) = & -\rho^v(\mathbf{r}, \mathbf{r}') v(\mathbf{r} - \mathbf{r}') \\ & -\rho^v(\mathbf{r}, \mathbf{r}') w(\mathbf{r}, \mathbf{r}') s(\omega, \varepsilon_v) \\ & + \sum_{n \in \{v, c\}} \rho^n(\mathbf{r}, \mathbf{r}') w(\mathbf{r}, \mathbf{r}') \frac{s(\omega, \varepsilon_n)}{2} \\ & -\rho^v(\mathbf{r}, \mathbf{r}') w'(\mathbf{r}, \mathbf{r}') s(\omega, \varepsilon_v)^2 \\ & + \sum_{n \in \{v, c\}} \rho^n(\mathbf{r}, \mathbf{r}') w'(\mathbf{r}, \mathbf{r}') \frac{s(\omega, \varepsilon_n)^2}{2} + \dots, \end{aligned} \quad (13)$$

where we introduce the definitions

$$\rho^n(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{v}, \mathbf{R}} \phi_{\mathbf{v}}^n(\mathbf{r} - \mathbf{R}) \phi_{\mathbf{v}'}^{n'*}(\mathbf{r}' - \mathbf{R}),$$

$$w(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{R}, \mathbf{v}} D_{\mathbf{v}}(\mathbf{r} - \mathbf{R}) D_{\mathbf{v}'}^*(\mathbf{r}' - \mathbf{R}),$$

$$w'(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{R}, \mathbf{v}\mathbf{v}'} D_{\mathbf{v}}(\mathbf{r} - \mathbf{R}) V_{\mathbf{v}\mathbf{v}'} D_{\mathbf{v}'}^*(\mathbf{r}' - \mathbf{R}),$$

and

$$s(\omega, \varepsilon_n) = 4E_g / [(\omega - \varepsilon_n)^2 - (E_g^2 + \omega_p^2)].$$

Terms in Eq. (13) involving only the bonding orbital density  $\rho^v$  are associated with the screened exchange and those corresponding to the Coulomb hole are symmetric in the band index  $n$ . The Coulomb hole component of the self-energy is

equivalent for both the valence- and conduction-band states, and the main difference arises solely from the screened exchange.<sup>27</sup>

With Eq. (13), we are now in position to calculate the self-energy of a valence-band state  $\Sigma_v \equiv \langle \Psi_v | \Sigma(\varepsilon_v) | \Psi_v \rangle$ . Performing the integration (within the constraints of the ETBM) and simplifying the remaining electron-hole field terms  $D_{\mathbf{v}}(\mathbf{r} - \mathbf{R})$  with a dipole approximation for the Coulomb interaction, i.e.,

$$D_{\mathbf{v}}(\mathbf{r} - \mathbf{R}) \cong e \left( \frac{\mathbf{d}_{\mathbf{v}} \cdot \mathbf{R}_{\mathbf{v}}}{|\mathbf{R}_{\mathbf{v}}|^3} \right), \quad (14)$$

with  $\mathbf{d}_{\mathbf{v}} = e f \phi_{\mathbf{v}}^v(\mathbf{r}) \mathbf{r} \phi_{\mathbf{v}}^{v'*}(\mathbf{r})$ , leads to the result

$$\begin{aligned} \Sigma_v = & \sum_{\text{HF}}^{vv} + \frac{e^2}{2} \left( \frac{N_0}{\epsilon_0} \right) \sum_{\mathbf{R}_{\mathbf{v}} \neq 0} \left( \frac{\mathbf{d}_{\mathbf{v}} \cdot \mathbf{R}_{\mathbf{v}}}{|\mathbf{R}_{\mathbf{v}}|^3} \right)^2 \\ & - \frac{e^2}{2} \left( \frac{N_0}{\epsilon_0} \right)^2 \sum_{\mathbf{R}_{\mathbf{v}}, \mathbf{R}_{\mathbf{v}'} \neq 0} \left( \frac{\mathbf{d}_{\mathbf{v}} \cdot \mathbf{R}_{\mathbf{v}}}{|\mathbf{R}_{\mathbf{v}}|^3} \right) \\ & \times V_{\mathbf{v}\mathbf{v}'} \left( \frac{\mathbf{d}_{\mathbf{v}'} \cdot \mathbf{R}_{\mathbf{v}'}}{|\mathbf{R}_{\mathbf{v}'}|^3} \right) + \dots. \end{aligned} \quad (15)$$

We write  $\mathbf{R}_{\mathbf{v}}$  to identify the dipole location in unit cell  $\mathbf{R}$  at bond number  $v$ ,  $N_0$  and  $\epsilon_0$  as the static limits of Eqs. (A2) and (A10), respectively, and  $\sum_{\text{HF}}^{vv}$  the Hartree-Fock exchange energy of a bonding orbital. The first dipole-field sum in Eq. (15) includes all vectors  $\mathbf{R}_{\mathbf{v}}$  except the zero vector. In the second sum, we include all vectors  $\mathbf{R}_{\mathbf{v}'}$  such that  $\mathbf{R}_{\mathbf{v}'}$  is a neighbor-bond site of  $\mathbf{R}_{\mathbf{v}}$ , neither being zero. (When zero modulus lattice vectors occur, integrals inside the sum are also zero by symmetry.)

The second sum is further manipulated to produce one equivalent to the first, as follows: noting that the  $V_{\mathbf{v}\mathbf{v}'}$  is independent of  $v'$  for the symmetrically related tetrahedral bonds, we write it as a certain fraction of the on-site electron-hole self-interaction (i.e.,  $V_{\mathbf{v}\mathbf{v}'} = \gamma V_{vv}$ );<sup>28</sup>  $\mathbf{R}_{\mathbf{v}'} = \mathbf{R}_{\mathbf{v}} + \mathbf{a}_{\mathbf{v}'}$ , where  $\mathbf{a}_{\mathbf{v}'}$  is a vector from bond  $v$  to bond  $v'$ , so with a small order expansion in  $\mathbf{a}_{\mathbf{v}'} / \mathbf{R}_{\mathbf{v}}$ , the second sum is approximately

$$\gamma V_{vv} \sum_{\mathbf{R}_{\mathbf{v}} \neq 0, \mathbf{v}'} \left( \frac{\mathbf{d}_{\mathbf{v}} \cdot \mathbf{R}_{\mathbf{v}}}{|\mathbf{R}_{\mathbf{v}}|^3} \right) \left( \frac{\mathbf{d}_{\mathbf{v}'} \cdot \mathbf{R}_{\mathbf{v}'}}{|\mathbf{R}_{\mathbf{v}'}|^3} \right), \quad (16)$$

where sums over  $v'$  involving  $\mathbf{a}_{\mathbf{v}'}$  only once have been set to zero owing to the symmetry of near-neighbor bonds in a tetrahedral configuration. Neglecting higher order terms in the expansion leads to errors of  $O(|a|^2/R^4)$ , which is fully consistent with the dipole approximation that introduced errors of  $O(1/R^3)$ . Given a set of dipoles oriented along the tetrahedral bonds in diamond the sum over  $v'$  in Eq. (16) may be completed<sup>29</sup> to give

$$-2 \gamma V_{vv} \sum_{\mathbf{R}_{\mathbf{v}} \neq 0} \left( \frac{\mathbf{d}_{\mathbf{v}} \cdot \mathbf{R}_{\mathbf{v}}}{|\mathbf{R}_{\mathbf{v}}|^3} \right)^2. \quad (17)$$

Following SI we transform the summations into integrals as

$$\sum_{\mathbf{R}_\nu \neq 0} \left( \frac{\mathbf{d}_\nu \cdot \mathbf{R}_\nu}{|\mathbf{R}_\nu|^3} \right)^2 \rightarrow \frac{(\hbar \omega_p)^2}{4E_g} \left( \frac{2\pi}{3} \right)^{1/3} \rho^{1/3}, \quad (18)$$

and recognize that  $V_{\nu\nu} = (\hbar \omega_p)^2 / 4E_g$ . Hence, the valence state self-energy in Eq. (15) becomes

$$\begin{aligned} \Sigma_v = & \Sigma_{\text{HF}}^{vv} + \frac{1}{2} \left( \frac{\epsilon_0 - 1}{\epsilon_0} \right) e^2 \left( \frac{2\pi}{3} \right)^{1/3} \rho^{1/3} \\ & + \frac{1}{2} \left( \frac{\epsilon_0 - 1}{\epsilon_0} \right)^2 \gamma e^2 \left( \frac{2\pi}{3} \right)^{1/3} \rho^{1/3} + \dots, \quad (19) \end{aligned}$$

where the factor of 2 from Eq. (17) has been absorbed into  $\gamma$  keeping only the sign. This is merely a convenience since such factors consistently appear together in higher-order terms in the self-energy expansion. The conduction-band self-energy is achieved similarly.

In the Appendix we show that each dipole-field summation may be approximated in the manner of Eqs. (16) and (17), and the series is convergent for any value of  $\epsilon_0$  greater than unity. Deducing a prefactor  $\gamma^{(N-1)} [(\epsilon_0 - 1)/\epsilon_0]^N$  at every order  $N$  (where Hartree-Fock terms are order zero) leads to  $N$ th-order terms given by

$$\delta \Sigma_N = \frac{1}{2} \left( \frac{\epsilon_0 - 1}{\epsilon_0} \right)^N \gamma^{(N-1)} e^2 \left( \frac{2\pi}{3} \right)^{1/3} \rho^{1/3}. \quad (20)$$

Summing to all orders we obtain

$$\begin{aligned} \delta \Sigma & \equiv \sum_N \left( \delta \Sigma_N \right) \\ & = \frac{1}{2} e^2 \left( \frac{2\pi}{3} \right)^{1/3} \left( \frac{\rho^{1/3}}{1-\gamma} \right) \left( \frac{\epsilon_0 - 1}{\epsilon_0} \right) \left[ 1 + \frac{\gamma}{(1-\gamma)\epsilon_0} \right]^{-1}. \quad (21) \end{aligned}$$

Where before we had an expansion in terms of the polarization  $[(\epsilon_0 - 1)/\epsilon_0]$ , valid for  $\epsilon_0 \rightarrow 1$ , we now expand in terms of small  $\gamma/\epsilon_0$  and keep only terms to order  $1/\epsilon_0^2$ . The self-energies then may be written

$$\begin{aligned} \Sigma_v = & \Sigma_{\text{HF}}^{vv} + \frac{1}{2} e^2 \left( \frac{2\pi}{3} \right)^{1/3} \left( \frac{\rho^{1/3}}{1-\gamma} \right) \\ & \times \left[ 1 - \frac{1}{(1-\gamma)\epsilon_0} + \frac{\gamma}{(1-\gamma)^2 \epsilon_0^2} \right], \quad (22) \end{aligned}$$

$$\begin{aligned} \Sigma_c = & \Sigma_{\text{HF}}^{vc} - \frac{1}{2} e^2 \left( \frac{2\pi}{3} \right)^{1/3} \left( \frac{\rho^{1/3}}{1-\gamma} \right) \\ & \times \left[ 1 - \frac{1}{(1-\gamma)\epsilon_0} + \frac{\gamma}{(1-\gamma)^2 \epsilon_0^2} \right], \quad (23) \end{aligned}$$

and the effort of summing terms to all orders in the polarization leads to an expansion in the inverse dielectric constant rather than the polarization.

Originally, SI omitted near-neighbor interactions (i.e.,  $\gamma \equiv 0$ ) and kept only the first-order term in the inversion of the dielectric matrix. By approximately replacing  $\Sigma_{\text{HF}}^{vv}$  with  $-(2\pi/3)e^2\rho^{1/3}$  they were led to a correction to LDA band gaps of the form  $\Delta E_g = \alpha e^2 \rho^{1/3} / \epsilon_0$ ,<sup>30</sup> which is valid for the

case of a system of highly localized orbitals. We obtain, of course, the same result in the  $\gamma=0$  limit, but a more general form resulting from our extended ETBM can now be proposed, which is also valid for systems of delocalized orbitals. In Eqs. (22) and (23) it can be seen that the introduction of interactions renormalizes the magnitude of the correlation energy and dielectric constant; the effective dielectric strength [given by  $(1-\gamma)\epsilon_0$ ] is reduced by the near-neighbor interactions as the cost of polarizing the medium is increased. It is now apparent that we can effectively introduce these interactions into the model by making the simple replacements

$$\left( \frac{2\pi}{3} \right)^{1/3} \left( \frac{\rho^{1/3}}{1-\gamma} \right) \rightarrow \alpha \rho^{1/3}$$

and

$$(1-\gamma)\epsilon_0 \rightarrow \epsilon_0$$

where  $\alpha$  modifies the exchange-correlation magnitude and  $\epsilon_0$  now represents the dielectric constant of a real material. We can again make replacements of SI (replacing  $\Sigma_{\text{HF}}^{vv}$  with  $-\alpha e^2 \rho^{1/3}$ ) to obtain a self-energy difference between a conduction- and valence-band state with the form

$$\Delta \equiv \Sigma_c - \Sigma_v = \alpha e^2 \rho^{1/3} \left[ \frac{1}{\epsilon_0} - \gamma \frac{1}{\epsilon_0^2} \right]. \quad (24)$$

Essentially, local-field effects are now incorporated in an average way through  $\gamma$ . By spherically averaging the dipole sums in Eq. (15), an assumption of a diagonal dielectric matrix is made, and the interaction represented by  $\gamma$  simply allows for the approximate inclusion of the nonzero off-diagonal terms in the dielectric matrix, which contribute the local-field effects.

A great simplification of the self-energy occurs by way of the primary approximations, the dipole and nonoverlapping bond approximation. Making the dipole approximation first in Eq. (13) and performing the spherical average of the summations reduces the nonlocal self-energy to a local one with an exchange-correlation potential difference for the valence- and conduction-band states given by

$$\Delta v_{\text{xc}} = -\alpha e^2 \rho^{1/3} \left[ \frac{1}{\epsilon_0} - \gamma \frac{1}{\epsilon_0^2} \right]. \quad (25)$$

Assuming for the moment that we complete a calculation of the conduction-band eigenvalues utilizing the valence-band exchange-correlation potential, then it becomes necessary to add  $\Delta v_{\text{xc}}$  as a perturbation to produce the correct conduction-band eigenvalues.<sup>31</sup> This corrective potential (applied as a perturbation) simply reproduces the the model discontinuity  $\Delta$  in Eq. (24).

If the LDA adequately represents the exchange-correlation effects in the valence-band states of semiconductors (or at least the highest occupied valence-band state), then we might include  $\Delta v_{\text{xc}}$  as a perturbation to DFT conduction-band states calculated by the KS method within the LDA. First, however, we make the assumption (as in the LDA) that the average electron density of the model can be replaced by the local-electron density of the semiconductor

[i.e.,  $\rho \rightarrow \rho(\mathbf{r})$  in Eq. (25)]. Hence, the resulting correction as given by first-order perturbation theory is

$$\Delta \epsilon_c = -\alpha e^2 \langle \rho(\mathbf{r})^{1/3} \rangle_c \left[ \frac{1}{\epsilon_0} - \gamma \frac{1}{\epsilon_0^2} \right], \quad (26)$$

where

$$\langle \rho(\mathbf{r})^{1/3} \rangle_c = \int d\mathbf{r} \Psi_c^*(\mathbf{r}) \rho(\mathbf{r})^{1/3} \Psi_c(\mathbf{r})$$

and  $\Psi_c(\mathbf{r})$  is the conduction-band eigenstate being perturbed. This then is our primary result for band-gap corrections to DFT band gaps. The model correction is derived from the difference between an ‘‘average’’ valence-band and conduction-band state, and the exact location in  $\mathbf{k}$  space to which these average states correspond in a real system remains unknown. Also, currently undetermined are the parameters  $\alpha$  and  $\gamma$ .

### III. APPLICATION TO LDA BAND GAPS

As a test of this band-gap correction we apply Eq. (26) to a set of semiconductors from the group IV, III-V, and II-VI classes. The electronic contributions to the static-dielectric constant are given by the value of  $\epsilon_\infty$ , which is taken directly from published experimental data<sup>32</sup> but may in principle be calculated within DFT. The perturbation is applied to the lowest conduction-band state at the Brillouin-zone center ( $\Gamma$ ), which locates the direct optical gap for the tetrahedrally bonded systems considered and allows a consistent application to both direct and indirect gap semiconductors. The resulting correction is then taken as a scissors translation to the entire conduction-band complex (i.e., all bands are rigidly shifted upward by an amount  $\Delta \epsilon_c$ ).

We comment on one potential concern with materials that have tetrahedral bonding but lack the diamond or zincblende symmetry (e.g., hexagonal structures). These materials have anisotropic dielectric tensors, and clearly our model assumes a scalar dielectric *constant* associated with an overall average band gap. In the case of a hexagonal system, bonding along the perpendicular  $c$  axis leads to bands of a different average energy and bandwidth compared with those along the  $a$  axes; therefore, a different average gap and dielectric tensor component can be associated with each principal direction. To obtain a single *approximate* dielectric constant for use here, we consider each principal direction in  $\mathbf{k}$  space separately, envisioning a fictitious cubic system with a dielectric constant given by the corresponding tensor component. This implies the existence of an average gap to which we can now apply the model correction. With three of these fictitious systems we obtain three separate corrections, and the final gap correction for the real system is reasonably a simple average of these. We observe then that the averaging of three first-order corrections [i.e., keeping only terms to order  $1/\epsilon_0$  in Eq. (26)] provides an effective inverse dielectric constant given by

$$\frac{1}{\epsilon_{\text{eff}}} = \frac{1}{3} \left( \frac{1}{\epsilon_1} + \frac{1}{\epsilon_2} + \frac{1}{\epsilon_3} \right). \quad (27)$$

This will be our method for finding a single dielectric constant for the anisotropic materials considered here; it is certainly satisfactory for relatively small anisotropies.

Determination of the parameter  $\alpha$  can be accomplished by invoking the known LDA exchange-correlation potential. By considering the limiting case of a homogeneous electron gas, which has a diverging static-dielectric constant [ $\epsilon_0 = 1 - (\omega_p/\omega)^2$ ], and taking the same limit in Eqs. (22) and (23), we find the discontinuity  $\Delta$  is zero; there is no gap correction, and the same exchange-correlation potential applies to both the conduction- and valence-band states. Assuming then that this model remains valid for a metallic system, it follows that [by comparison of the valence-band self-energy in Eq. (22) with the expectation value of the LDA exchange potential for some eigenstate  $\langle v_x \rangle = -(3/\pi)^{1/3} \langle \rho(\mathbf{r})^{1/3} \rangle$ ],  $\alpha = 2(3/\pi)^{1/3}$ . Clearly there must be some small correction for correlation effects, but the exchange-only component is a reasonable choice for this example. Now given that silicon has an LDA gap of 0.45 eV but an experimental gap of 1.17 eV, and taking  $\alpha$  determined by resort to the LDA exchange limit above, the first-order correction from the model can be found. Returning to Eq. (26) and keeping only the first-order result (i.e., first order in  $1/\epsilon_0$ ), the band-gap correction is calculated to be 1.42 eV;  $\langle \rho \rangle_c^{1/3}$  for silicon can be found in Table I along with  $\epsilon_0$ . This correction is nearly twice the magnitude of the necessary gap correction of 0.72 eV. Note that we do not expect this model to be accurate in the regime of metallic systems because we originally considered highly localized orbitals and only nearest-neighbor interactions. (This is certainly not a usual description of a metal.) Accordingly then, we choose to determine  $\alpha$  as with a Slater  $X$ - $\alpha$  method and require the model correction to reproduce the experimental gap of silicon.

In what follows, the calculation of semiconductor energy bands, wave functions, and the electronic density are determined using *ab initio* DFT methods within the LDA using a conjugate gradient technique in a plane-wave formalism.<sup>33</sup> High quality norm- and hardness-conserving pseudopotentials are used throughout. Energy cutoffs are selected to produce well-converged eigenstates,<sup>34</sup> and the electron density is determined at a set of 10 special  $\mathbf{k}$  points for the diamond and zinc-blende structures and a six-point set for the wurtzite structures.<sup>35</sup> Experimental lattice parameters have been chosen for all materials. Spin-orbit (SO) corrections have been introduced by subtracting one third of the experimental valence-band maximum SO splitting from the band gap. This has almost zero effect on the overall results (i.e., the average accuracy of the model correction is neither reduced nor enhanced).

Figure 1 summarizes the LDA band gaps for all systems considered here (24 in all) and provides a measure of the variation of errors inherent in the LDA. The deviation from the theory-experiment line can be quite large and no simple constant correction appears appropriate. Results obtained for a first-order Sterne-Inkson-type correction, i.e., keeping only the first term in Eq. (26), are given in Table I as  $E_{\text{gap}}^{\text{Si}}$  and also in Fig. 2(a). In this case,  $\alpha$  is adjusted to give the correct gap for silicon and it is found to have the value 27.1 eV bohr<sup>3</sup>. In general, the first-order results agree moderately well with experiment as seen in the figure, but the notable exceptions are a few wide-band semiconductors, e.g., C and BN, where

TABLE I. Material parameters utilized in the model  $GW$  correction along with resulting gaps.  $\langle \rho^{1/3} \rangle_c$  represents the expectation value of  $\rho^{1/3}(r)$  for the lowest conduction-band state at the Brillouin-zone center,  $\epsilon_\infty$  is the static-dielectric constant, and BW is the measured LDA bandwidth of the upper  $p$ -like valence bands taken from the *ab initio* results. Numbers in parentheses indicate the spin-orbit correction (one-third of which estimates the band-gap decrease applied to the model band gaps). The notion 3C, 2H, and WZ indicate cubic, hexagonal, and wurtzite for those semiconductors that have a number of recognized structures (e.g., GaN has also been grown in the zinc-blende phase).

	$\langle \rho^{1/3} \rangle_c$	$\epsilon_\infty$	BW (eV)	$E_{\text{gap}}^{\text{da}}$ (eV)	$E_{\text{gap}}^{\text{st}}$ (eV) <sup>a</sup>	$E_{\text{gap}}^{\text{model}}$ (eV)	$E_{\text{gap}}^{\text{expt}}$ (eV)
C	0.519	5.7	13	4.09	6.56	5.48 <sup>b</sup>	5.48
Si	0.321	12.1	7	0.45	1.17 <sup>b</sup>	1.17 <sup>b</sup>	1.17
Ge	0.347	15.9	8	0.00	0.57	0.58 (0.18)	0.74
Sn	0.399	24.1	7	-0.30	-0.12	-0.08 (0.80)	0.10
SiC (3C)	0.434	6.5	11	1.30	3.11	2.59 (0.00)	2.39
SiC (2H)	0.402	6.6	9	2.05	3.68	3.37 (0.00)	3.30
BN	0.456	4.5	10	4.34	7.09	5.93 (0.00)	6.10 <sup>a</sup>
BP	0.391	10.8	11	1.31	2.33	2.18 (0.00)	2.05
AlN	0.421	4.8	7	4.18	6.55	6.01 (0.00)	6.28
AlP	0.327	7.5	6	1.44	2.62	2.55 (0.00)	2.51
AlAs	0.313	8.2	6	1.29	2.22	2.18 (0.33)	2.23
AlSb	0.376	10.2	6	1.16	1.94	1.93 (0.67)	1.67
GaN (WZ)	0.429	5.6	8	2.45	4.54	4.07 (0.00)	3.42
GaP	0.344	9.1	7	1.50	2.52	2.44 (0.00)	2.35
GaAs	0.331	10.6	7	0.42	1.16	1.17 (0.34)	1.52
GaSb	0.384	14.4	6	0.00	0.51	0.52 (0.70)	0.81
InN	0.608	9.3	5	0.00	1.77	1.79 (0.00)	1.89
InP	0.397	9.6	6	0.43	1.55	1.54 (0.00)	1.42
InAs	0.399	12.3	6	0.00	0.76	0.78 (0.38)	0.42
ZnO	0.4469	3.7	5	0.60	3.87	3.23 (0.00)	3.44
ZnS	0.3977	5.1	5	1.64	3.75	3.64 (0.00)	3.78
ZnSe	0.3867	5.4	5	0.82	2.63	2.46 (0.40)	2.82
CdS	0.3624	5.5	4	0.77	2.62	2.63 (0.00)	2.58
CdSe	0.3533	6.2	4	0.24	1.65	1.63 (0.42)	1.83

<sup>a</sup>Experimental gaps for BN range from 6.0 to 6.5 eV. A recent result from Ref. 37 is used here for comparison.

<sup>b</sup>The semiconductor was used as a reference for the model parameters.

a nearly 1-eV discrepancy remains. The standard deviation of this set is about 0.35 eV compared with the original deviation in LDA band gaps of greater than 1 eV.

Proceeding further, we examine the consequences including of the second-order correction. The second-order term in Eq. (26), however, includes the interaction ( $\gamma$ ) between neighboring bonds. As stated previously, this interaction is a constant for the near-neighbor bonds but nevertheless must be realized by appeal to some physical property in the *ab initio* calculation. Here we shall associate it with the bandwidth of the upper valence-band  $p$ -type states because these states produce the prototypical bonding orbitals in the tetrahedrally bonded semiconductors, and their near-neighbor interactions are then directly related to the tight-binding bandwidth. We also recognize that a number of semiconductors require inclusion of local-field effects in their dielectric descriptions to obtain accurate  $GW$  corrections. These tend to be the wide-band semiconductors C, BN, etc., so, on this basis,  $\gamma$  is reasonably expected to be related to the bandwidth. Incorporation of this effect at the level of a linear dependence then introduces a single additional parameter leading to a subsequent correction of the form

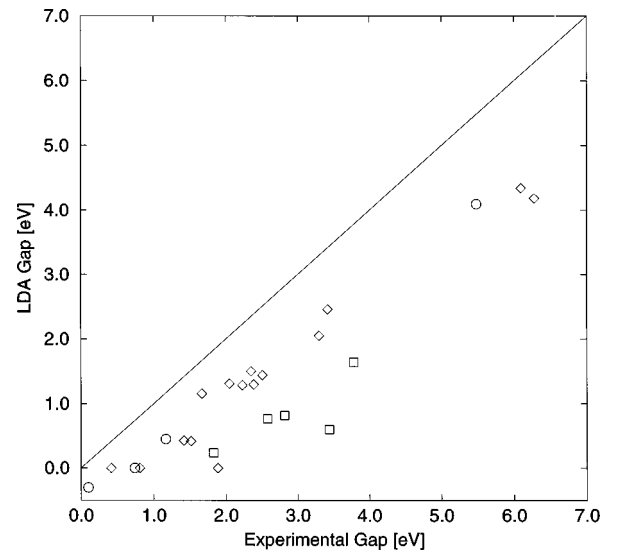


FIG. 1. Theoretical bands gaps calculated within the LDA of the semiconductors listed in Table I. Circles, diamonds, and squares represent group IV, group III-V, and group II-VI semiconductors, respectively.

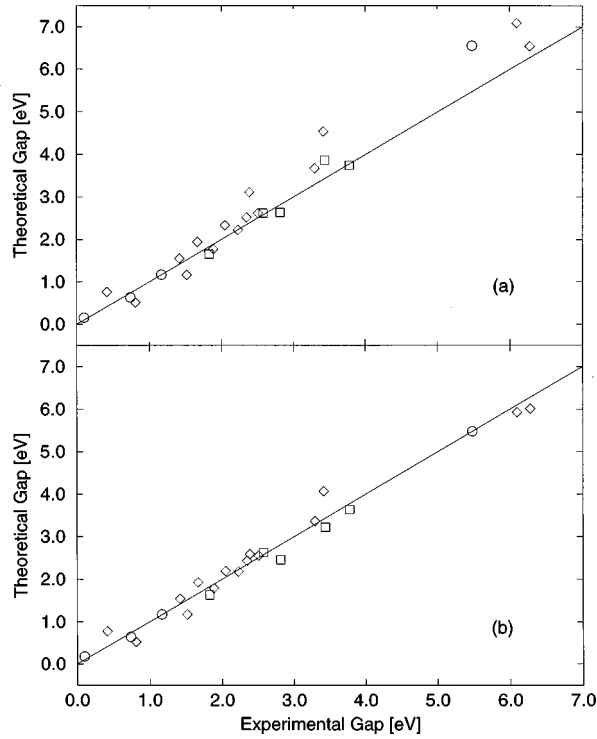


FIG. 2. Theoretical band gaps resulting from the LDA band gaps in Table I corrected with (a) the first-order (Sterne-Inkson)  $GW$  result [Eq. (26) with  $\gamma=0$ ] and (b) the second-order  $GW$  result [Eq. (26) with  $\gamma$  chosen as described in the text]. Circles, diamonds, and squares represent group IV, group III-V, and group II-VI semiconductors, respectively.

$$\Delta \varepsilon_c = \alpha \langle \rho^{1/3}(\mathbf{r})^{1/3} \rangle_c \left[ \frac{1}{\epsilon_0} - \beta \frac{(W)}{\epsilon_0^2} \right],$$

where we have replaced  $\gamma$  as  $\beta W$  with  $W$  representing the bandwidth taken from the *ab initio* calculation.

Because the second-order term now includes this bandwidth, we have chosen to use both silicon and carbon as reference materials; silicon as a prototypical narrow gap, narrow bandwidth material and carbon corresponding to a large bandwidth, wide-gap material. The parameters  $\alpha$  and  $\beta$  then take on the values  $31.1 \text{ eV bohr}^3$  and  $0.225/\text{eV}$ , respectively. Numerical results for a range of semiconductors corrected with these choices are listed in Table I under  $E_{\text{gap}}^{\text{model}}$  and plotted versus experimental values in Fig. 2(b). We see that the second-order correction now considerably reduces the error in the wide-band materials while leaving the other semiconductor gaps relatively unchanged. Applying Eq. (26) to the entire set, we find the corrected gaps now generally fall within  $0.1\text{--}0.3 \text{ eV}$  of the experimental values with a standard deviation of  $0.24 \text{ eV}$ .

The gap in GaN is least well given with an error of about  $0.65 \text{ eV}$ . This discrepancy may be explained by the model assumption of a single near-neighbor bond interaction introduced to simplify higher order terms in the series expansion of Eq. (15). For a semiconductor with inversion symmetry, this assumption is valid because there is no distinction between bonds in a unit cell; however, the III-V and II-VI semiconductors are partially ionic, and therefore two distinct bond interactions exist.

#### IV. DISCUSSION AND CONCLUSION

We have obtained a correction to Kohn-Sham LDA band gaps by extending the Stern-Inkson first-order result to include a second-order term that includes local-field effects in an approximate manner. Experimental band gaps are reproduced to within  $0.1\text{--}0.3 \text{ eV}$  for an extensive range of semiconducting materials. This band-gap correction is at the level of a scissors translation and could be useful for further calculations of the dynamical dielectric properties of semiconductors in addition to providing an estimate for the band gap of novel semiconductor systems.

In principle, a simple method for applying the same form of correction to nonhomogeneous systems such as layered superlattice systems exists. From the microscopic picture of the ETBM in which the semiconductor is composed of a set of localized bonds with some excitation energy between valence- and conduction-state orbitals. We can then propose a spatially dependent static-dielectric function whose value within the region of any bond is given by the equivalent homogeneous semiconducting system composed of those bonds (i.e., near a GaAs bond we use the dielectric constant of GaAs). This possibility will be considered in further work when full *ab initio*  $GW$  results are available for nonhomogeneous systems. Such test cases will also determine whether the choice of  $\alpha$  is universal and whether the near-neighbor approximation is a significant limitation. The *ad hoc* nature of the conduction-band corrections must also be tested in these novel systems because it is based on the assumption that the LDA is a reasonable approximation for the valence-band exchange-correlation potential. It may be appropriate to introduce corrections to valence states also.

#### ACKNOWLEDGMENTS

We thank the Semiconductor Research Corporation for funding this work under Task 069.23. We also thank Professor Michael Teter for the use of the Corning plane-wave code needed to obtain the *ab initio* electronic structure and charge density for the semiconductors.

#### APPENDIX: DERIVATION OF THE SCREENED INTERACTION

To obtain the screened interaction, it is first necessary to invert the dielectric matrix (8) within the ETBM. This was accomplished in Ref. 26, and we briefly repeat the steps required to obtain higher order terms employing the same notation where possible. The polarization for a system with a single optical gap  $E_g$  (here assumed to be the average ETBM gap discussed in Sec. II) is given by

$$P(\mathbf{r}, \mathbf{r}'; \omega) = N(\omega) \sum_{\nu \mathbf{R}} A_{\nu}(\mathbf{r} - \mathbf{R}) A_{\nu}^*(\mathbf{r}' - \mathbf{R}), \quad (\text{A1})$$

where

$$N(\omega) = \frac{4E_g}{\omega^2 - E_g^2} \quad (\text{A2})$$

and

$$A_\nu(\mathbf{r}-\mathbf{R}) = \phi_\nu^v(\mathbf{r}-\mathbf{R}) \phi_\nu^{c*}(\mathbf{r}-\mathbf{R}) \quad (\text{A3})$$

is an approximate electron-hole pair state in the cell  $\mathbf{R}$  at bond site  $\nu$ . The dielectric matrix in Eq. (8) is then written

$$\epsilon(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r}, \mathbf{r}') - N(\omega) \sum_{\nu \mathbf{R}} D_\nu(\mathbf{r}-\mathbf{R}) A_\nu^*(\mathbf{r}'-\mathbf{R}), \quad (\text{A4})$$

with

$$D_\nu(\mathbf{r}-\mathbf{R}) = \int d\mathbf{r}' \phi_\nu^v(\mathbf{r}'-\mathbf{R}) v(\mathbf{r}-\mathbf{r}') \phi_\nu^{c*}(\mathbf{r}'-\mathbf{R}) \quad (\text{A5})$$

being the Coulomb field due to an electron-hole pair at lattice site  $\mathbf{R}$  and bond site  $\nu$ .

Inversion of the dielectric matrix is carried out using the relation

$$\epsilon^{-1} = (1 - \mathbf{D}\mathbf{A})^{-1} = [1 + \mathbf{D}(1 - \mathbf{A}\mathbf{D})^{-1}\mathbf{A}]. \quad (\text{A6})$$

Identifying  $\mathbf{D}_{\nu, \nu}^{\mathbf{R}} = N(\omega) D_\nu(\mathbf{r}-\mathbf{R})$  and  $\mathbf{A}_{\nu, \nu}^{\mathbf{R}} = A_\nu^*(\mathbf{r}-\mathbf{R})$  we can write

$$\begin{aligned} \epsilon^{-1}(\mathbf{r}, \mathbf{r}'; \omega) &= \delta(\mathbf{r}, \mathbf{r}') + \sum_{\nu \mathbf{R}, \nu' \mathbf{R}'} N(\omega) D_\nu(\mathbf{r}-\mathbf{R}) \\ &\quad \times \mathbf{Q}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} A_{\nu'}^*(\mathbf{r}'-\mathbf{R}'), \end{aligned} \quad (\text{A7})$$

where

$$(\mathbf{Q}^{-1})_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} = \delta_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} - (\mathbf{A}\mathbf{D})_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} = \delta_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} - N(\omega) \mathbf{V}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} \quad (\text{A8})$$

with the definition

$$\mathbf{V}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} = \int d\mathbf{r} \int d\mathbf{r}' A_\nu^*(\mathbf{r}-\mathbf{R}) v(\mathbf{r}-\mathbf{r}') A_{\nu'}(\mathbf{r}'-\mathbf{R}'). \quad (\text{A9})$$

Now, SI assumed only on-site interactions leading to  $\mathbf{V}_{\nu \nu}^{\mathbf{R} \mathbf{R}} = (\hbar \omega_p)^2 / 4E_g$  and a diagonal matrix with elements

$$(\mathbf{Q}^{-1})_{\nu \nu}^{\mathbf{R} \mathbf{R}} = \left( 1 + \frac{(\hbar \omega_p)^2}{E_g^2 - \omega^2} \right) \equiv \epsilon(\omega) \quad (\text{A10})$$

given by a simple, single-pole approximation for the dielectric function with the average optical gap  $E_g$  being the pole.<sup>36</sup>

Here, however, we shall include the off-diagonal components as a small perturbation and write  $\mathbf{Q}^{-1}$  as  $\mathbf{Q}_0^{-1} + \mathbf{Q}'$ , where the diagonal elements of  $\mathbf{Q}_0^{-1}$  are given in Eq. (A10) and the off-diagonal elements of  $\mathbf{Q}'$  are  $(\mathbf{Q}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'})^{-1} = -N(\omega) \mathbf{V}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'}$ . We now write the inverse of  $\mathbf{Q}^{-1}$  as

$$\mathbf{Q} = [\mathbf{Q}_0^{-1} + \mathbf{Q}']^{-1} = \mathbf{Q}_0 [1 - \mathbf{Q}' \mathbf{Q}_0 + (\mathbf{Q}' \mathbf{Q}_0)^2 + \dots], \quad (\text{A11})$$

with matrix elements

$$\mathbf{Q}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} = \frac{1}{\epsilon(\omega)} \delta_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} + \frac{N(\omega)}{\epsilon(\omega)^2} \mathbf{V}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'} + \dots \quad (\text{A12})$$

We further consider only first-neighbor electron-hole pair interactions in  $\mathbf{V}_{\nu \nu'}^{\mathbf{R} \mathbf{R}'}$ , which is now written as  $V_{\nu \nu'}$ , where  $\nu'$  indicates a neighbor of  $\nu$ . Finally, the the screened interaction becomes

$$\begin{aligned} W(\mathbf{r}, \mathbf{r}'; \omega) &= v(\mathbf{r}-\mathbf{r}') + \left[ \frac{N(\omega)}{\epsilon(\omega)} \right] \sum_{\nu \mathbf{R}} D_\nu(\mathbf{r}-\mathbf{R}) D_\nu^*(\mathbf{r}'-\mathbf{R}) \\ &\quad + \left[ \frac{N(\omega)}{\epsilon(\omega)} \right]^2 \sum_{\nu \mathbf{R}, \nu' \mathbf{R}'} D_\nu(\mathbf{r}-\mathbf{R}) V_{\nu \nu'} D_{\nu'}^*(\mathbf{r}'-\mathbf{R}) \\ &\quad + \dots \end{aligned} \quad (\text{A13})$$

The  $V_{\nu \nu'}$ 's may be approximated by expanding the Coulomb potential to the quadrupole level, which is the first nonzero term in the integral. This leads to

$$V_{\nu \nu'} \equiv e \frac{\mathbf{d}_\nu \cdot \mathbf{d}_{\nu'}}{|\mathbf{a}_{\nu'}|^3} - 3e \frac{(\mathbf{d}_\nu \cdot \mathbf{a}_{\nu'})(\mathbf{d}_{\nu'} \cdot \mathbf{a}_{\nu'})}{|\mathbf{a}_{\nu'}|^5},$$

i.e., the dipole-dipole interaction energy for two neighboring bonds. Given a pair of dipoles oriented along the tetrahedral bonds (and symmetry requires we need only examine any pair of our choice),  $\mathbf{d} = (1, 1, 1)\alpha$  and  $\mathbf{d}' = (-1, -1, 1)\alpha$  for instance, and the vector between them,  $\mathbf{s} \equiv \mathbf{a}_{\nu, \nu'} = (1, 1, 0)a_0/4$ , we have  $V_{\nu \nu'} = e5\alpha^2/s^3 = e5\sqrt{2}|\mathbf{d}|^2\rho/6$  after substituting for the average electron density,  $\rho = 32/a_0^3 = \sqrt{2}/s^3$ . Now with  $|\mathbf{d}|^2 = 3\hbar^2 e^2 / 4mE_g$  we find

$$V_{\nu \nu'} = \frac{5\sqrt{2}}{8\pi} \frac{(\hbar \omega_p)^2}{4E_g} = \frac{5\sqrt{2}}{8\pi} V_{\nu \nu}. \quad (\text{A14})$$

Continuing, we consider the summations over  $\nu, \nu'$ , etc. in Eq. (A13) directly. At each order we introduce sums over longer chains of bonds through near-neighbor interactions. The first-order sum includes on-site terms, and the second-order sum includes near-neighbor terms that reduce approximately to Eq. (17). We can separate the third-order sum into those chains that return to the same site, a first-neighbor site, or a second-neighbor site. For a tetrahedral configuration each bond has six near neighbors, so six possible sequences exist for a chain of two interactions leading from site  $\nu$  to a first neighbor  $\nu'$  and back to  $\nu$ . If we consider the number of paths to get to any first neighbor by two near-neighbor interactions, we find only two. Only one path to each second neighbor through two near-neighbor interactions exists. The third-order sum is proportional to

$$\begin{aligned} 6V^2 \sum_{\nu} \left( \frac{\mathbf{d}_\nu \cdot \mathbf{R}_\nu}{|\mathbf{R}_\nu|^3} \right)^2 + 2V^2 \sum_{\nu, \nu'} \left( \frac{\mathbf{d}_\nu \cdot \mathbf{R}_\nu}{|\mathbf{R}_\nu|^3} \right) \left( \frac{\mathbf{d}_{\nu'} \cdot \mathbf{R}_{\nu'}}{|\mathbf{R}_{\nu'}|^3} \right) \\ + V^2 \sum_{\nu, \nu''} \left( \frac{\mathbf{d}_\nu \cdot \mathbf{R}_\nu}{|\mathbf{R}_\nu|^3} \right) \left( \frac{\mathbf{d}_{\nu''} \cdot \mathbf{R}_{\nu''}}{|\mathbf{R}_{\nu''}|^3} \right), \end{aligned} \quad (\text{A15})$$

where the number of primes indicates the neighbor distance,  $V$  is the constant near-neighbor interaction, and we immediately apply the dipole approximation to the electron-hole field terms. Making the small-order approximations to the bond vectors  $\mathbf{R}_{\nu''}$ , as in Eq. (16), with the second sum already considered in Eq. (17), the third-order term is

$$4V^2 \sum_{\nu} \left( \frac{\mathbf{d}_\nu \cdot \mathbf{R}_\nu}{|\mathbf{R}_\nu|^3} \right)^2. \quad (\text{A16})$$



Examining the fourth-order sum using the same approximations we find

$$-8V^3 \sum_{\mathbf{R}_\nu} \left( \frac{\mathbf{d}_\nu \cdot \mathbf{R}_\nu}{|\mathbf{R}_\nu|^3} \right)^2, \quad (\text{A17})$$

and the pattern continues. Hence, we may identify a factor of  $(-2V)^{N-1}$  appearing at each order  $N$ , so that with Eqs. (A14) and (18) we can write the  $N$ th-order expansion of the static screened interaction as  $C(5\sqrt{2}/4\pi)^{N-1}[(\epsilon_0 - 1)/\epsilon_0]^N \rho^{1/3}$ , which converges for  $(5\sqrt{2}/4\pi)[(\epsilon_0 - 1)/\epsilon_0] < 1$  or any  $\epsilon_0 > 1$ .

<sup>1</sup>P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

<sup>2</sup>W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

<sup>3</sup>J. F. Janak, Phys. Rev. B **18**, 7165 (1978).

<sup>4</sup>J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Jr., Phys. Rev. Lett. **49**, 1691 (1982).

<sup>5</sup>There is some residual  $\mathbf{k}$  dependence found in many-body corrections to the excited states of semiconductors, but a simple scissors-type correction often provides sufficient agreement with experiment.

<sup>6</sup>L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

<sup>7</sup>J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

<sup>8</sup>L. Hedin, Phys. Rev. **139**, A796 (1965).

<sup>9</sup>L. Hedin and S. Lundqvist, Solid State Phys. **23**, 1 (1969).

<sup>10</sup>M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

<sup>11</sup>R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10 159 (1988).

<sup>12</sup>M. Rohlfing, P. Krüger, and J. Pollman, Phys. Rev. B **48**, 17 791 (1993).

<sup>13</sup>Z. H. Levine and S. G. Louie, Phys. Rev. B **25**, 6310 (1982).

<sup>14</sup>G. Cappellini, R. Del Sole, L. Reining, and F. Bechstedt, Phys. Rev. B **47**, 9892 (1993).

<sup>15</sup>M. Palumbo, R. Del Sole, L. Reining, F. Bechstedt, and G. Cappellini, Solid State Commun. **95**, 393 (1995); F. Bechstedt, M. Fiedler, C. Kresse, and R. Del Sole, *ibid.* **89**, 669 (1994).

<sup>16</sup>M. S. Hybertsen and S. G. Louie, Phys. Rev. B **37**, 2733 (1988) [C, Si, Ge].

<sup>17</sup>X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14 142 (1991) [Si, AlP, AlAs, AlSb, GaP, GaAs, GaSb, InP, InAs].

<sup>18</sup>F. Bechstedt, R. Del Sole, G. Cappellini, and L. Reining, Solid State Commun. **84**, 765 (1992).

<sup>19</sup>M. P. Surh and S. G. Louie, Phys. Rev. B **43**, 9216 (1991) [BN]; A. Rubio, J. L. Corkill, M. L. Cohen, E. L. Shirley, and S. G. Louie, *ibid.* **48**, 11 810 (1993) [AlN, GaN]; O. Zakharov, A. Rubio, X. Blase, M. L. Cohen, and S. G. Louie, *ibid.* **50**, 10 780 (1994) [group II-VI].

<sup>20</sup>P. A. Sterne and J. C. Inkson, J. Phys. C **17**, 1497 (1984).

<sup>21</sup>A. E. Carlsson, Phys. Rev. B **31**, 5178 (1985).

<sup>22</sup>W. Borrmann and P. Fulde, Europhys. Lett. **2**, 471 (1986).

<sup>23</sup>W. Hanke, Th. Gölzer, and H. J. Mattausch, Solid State Commun. **51**, 23 (1984); W. Hanke and L. J. Sham, Phys. Rev. B **38**, 13 361 (1988).

<sup>24</sup>F. Bechstedt and R. Del Sole, Phys. Rev. B **38**, 7710 (1988).

<sup>25</sup>F. Gygi and A. Baldereschi, Phys. Rev. Lett. **62**, 2160 (1989).

<sup>26</sup>M. Ortuno and J. C. Inkson, J. Phys. C **12**, 1065 (1979).

<sup>27</sup>Writing out the frequency-dependent part of the screened interaction in Eq. (A13), we see that the difference here is found to be a modification of the plasmon-pole frequency ( $\omega_p$  becoming  $\sqrt{\omega_p^2 + E_g^2}$ ), physically reflecting the presence of the crystalline potential which gives rise to the optical gap  $E_g$ . This shifted pole later leads to the straightforward identification of the static-dielectric constant in the final self-energy and a considerable

subsequent simplification in terms of the number of physical parameters for our proposed gap correction.

<sup>28</sup>This is strictly true only for the ETBM in which all bonds are identical. For a real system this can be true only for nonpolarized bonds (i.e., semiconductors with inversion symmetry) because polarized systems (those with some ionic character) contain two distinct bond-bond interactions. There is the bond-bond interaction in which the cation is the shared atom and similarly that interaction in which the anion is shared.

<sup>29</sup>A factor of  $-2$  appears by making the replacement  $\sum_{\nu'} \mathbf{d}_{\nu'} = \sum_{\nu'} \mathbf{d}_{\nu'} + \mathbf{d}_\nu - \mathbf{d}_\nu = \sum_{\nu'} \mathbf{d}_{\nu'} - \mathbf{d}_\nu$ , where  $\nu'$  in the final sum includes the entire set of dipoles in the unit cell (not only the near neighbors of  $\nu$ ). Summation over the whole set of dipoles gives zero, leaving  $-\mathbf{d}_\nu$ , and the existence of two equivalent sets of near-neighbors for each bond site (i.e., each bond site is shared between two atoms each of which has three neighbor bonds) then contributes the factor of 2.

<sup>30</sup>The on-site exchange term in the conduction bands can generally be ignored as originally argued in Ref. 20. For a semiconductor composed of extended orbitals this term goes to zero while for a more localized system of orbitals (e.g., carbon) it may approach 1–2 eV. This, however, is still quite small compared to the on-site valence-band term, which, on examination of the magnitude  $-(2\pi/3)^{1/3} e^2 \rho^{1/3}$ , is seen to range from 15 to 20 eV for any typical semiconductor.

<sup>31</sup>Applying the perturbation to the conduction-band eigenvalues assumes, of course, that the LDA is a valid exchange-correlation potential for the valence-band eigenvalues. One could claim equivalently that the LDA is appropriate for the conduction-band states and apply the correction to valence-band eigenvalues.

<sup>32</sup>O. Madelung, *Semiconductors: Group IV and III-V Compounds*, edited by R. Poersckey (Springer, Berlin, 1991).

<sup>33</sup>M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12 255 (1989); M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

<sup>34</sup>Large energy cutoffs were used with 60 hartrees being sufficient to achieve well-converged eigenvalues and eigenfunction  $s$  for nearly all materials. Semiconductor properties with Zn and Cd constituents were calculated with larger cutoffs of 120 and 90 hartrees, respectively, to assure accurate results. In all cases these energies were large enough to converge the valence band and low-lying conduction-band states for the materials considered here.

<sup>35</sup>D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973).

<sup>36</sup>D. L. Johnson, Phys. Rev. B **9**, 4475 (1974).

<sup>37</sup>N. Miyata, K. Moriki, O. Mishima, M. Fujisawa, T. Hattori, J. Nordgren, R. Singh, and T. D. Moustakas, Phys. Rev. B **40**, 12 028 (1989).