Ab initio calculation of $\epsilon_2(\omega)$ including the electron-hole interaction: Application to GaN and CaF_2

Lorin X. Benedict and Eric L. Shirley
Optical Technology Division, Physics Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899
(Received 3 September 1998)

We present a computationally efficient first-principles scheme to calculate $\epsilon_2(\omega)$ for crystalline insulators, including the electron-hole interaction. The effective Hamiltonian for electron-hole pairs contains both the exchange and direct parts of this interaction. An iterative scheme is used in which the ω moments of $\epsilon_2(\omega)$ are computed by repeated action of the Hamiltonian on electron-hole pair states. The scheme is applied to two insulators where there are significant experimental uncertainties in their ultraviolet optical properties: GaN and CaF_2. [S0163-1829(99)08507-0]

I. INTRODUCTION

It is a goal of present day materials theory to relate materials properties to microscopic or atomistic descriptions of matter. First-principles (or ''*ab initio*'') theories, in which the only input is the identity of the atomic constituents and fundamental constants, have succeeded in predicting structural and vibrational properties, as well as ground-state electronic properties (such as the electron density). These properties are determined quite accurately in effective one-electron mean-field theories, like the local-density approximation (LDA).[1] In these theories, each electron feels an orbital-independent potential and is represented by a one-particle wave function, $\psi(\mathbf{r})$. For properties involving electronic excitations, however, mean-field descriptions can fail to give quantitatively accurate results. An example is direct and inverse photoemission, where single electrons or holes are added to the system. Here, it is necessary to dress the added electron or hole with the polarization cloud of the surrounding ''unexcited'' electrons, as in a quasiparticle calculation.[2] In quasiparticle theories, electrons feel an orbital-dependent potential, but each is still described by a one-particle wave function.

For the case of optical properties, the situation is more extreme. Here, a photon interacts with the system to produce an electron-hole pair. In addition to interacting with their respective polarization clouds, the excited electron and hole interact strongly with each other to produce bound states (excitons) and above-threshold resonances. These effects cannot be modeled in an effective one-particle picture. Instead it is necessary to deal with two-particle states: $\psi(\mathbf{r}_e,\mathbf{r}_h)$, where $\mathbf{r}_e$ and $\mathbf{r}_h$ are the positions of the excited electron and hole. Although the importance of the electron-hole interaction in optical absorption has been understood for decades,[3] first-principles calculations of optical properties including the electron-hole interaction have only recently begun to appear.[4–6] In these theories, individual quasielectron and quasihole states are determined in an LDA/quasiparticle approach, and then quasielectron-quasihole pair states are determined by solving the Bethe-Salpeter equation with an appropriately screened Coulomb interaction. These electron-hole pair states are then used to construct the frequency-dependent imaginary part of the dielectric function, $\epsilon_2(\omega)$. Results obtained so far for semiconductors and insulators agree extremely well with reflectivity and ellipsometry data for these materials.[4–6]

Not surprisingly, the computations are quite demanding. This is because each pair state wave function depends upon two variables, $\mathbf{r}_e$ and $\mathbf{r}_h$, so the resulting eigenproblem has a dimension $\propto N^2$, where N is the number of atoms. In a standard approach,[4,6] the Bethe-Salpeter equation is solved for the electron-hole pair state wave functions and energies, which are then used in a sum-over-states expression to determine $\epsilon_2(\omega)$. Such a method requires $\propto (N^2)^3 = N^6$ operations.

In this work, we present the details of our scheme to compute $\epsilon_2(\omega)$ for crystalline materials. Instead of solving the Bethe-Salpeter eigenproblem, we directly determine different ω-moments of $\epsilon_2(\omega)$ by repeatedly acting with the effective Hamiltonian extracted from the Bethe-Salpeter equation.[5] Individual actions of the Hamiltonian are simplified by using two different bases in which to represent electron-hole pair states: the *real-space basis*, and the *one-particle eigenstate basis*. This results in a method that scales as N^4. We apply this method to the crystalline insulators GaN (wurtzite and zinc blende), and CaF_2. Both materials are used in optical technology (GaN for blue/UV optoelectronic devices, and CaF_2 for UV optics), yet neither is fully understood experimentally. We endeavor to resolve some of these difficulties by comparing our results to the available experiments.

The remainder of the paper is organized as follows: Section II contains the basic theory and the details of the computational method. Section III contains the results for wurtzite and zinc blende GaN, and CaF_2, together with a discussion of their application to the interpretation of experimental data. We conclude in Sec. IV.

II. THEORY AND METHOD

In this section, we discuss the theory and computational method used to calculate $\epsilon_2(\omega)$ including the electron-hole interaction. The discussion is divided into five subsections. The first describes the space of singly excited electron-hole pair states, together with the bases we use to represent them. The second contains the equation of motion for the electron-hole pairs. The third presents a detailed treatment of the effective Hamiltonian that is extracted from the equation of motion, and a description of the individual terms. The fourth

presents the iterative scheme to calculate $\epsilon_2(\omega)$ by acting with the effective Hamiltonian. The final subsection describes a computationally efficient way of decomposing $\epsilon_2(\omega)$ into contributions arising from individual interband transitions.

A. Excited electronic states

Our treatment of optical absorption (which follows that of Ref. 3) is based on the following simple picture: Before a photon is absorbed, the solid is in its ground state. We neglect phonons in this work, so the ions are taken to be fixed in their equilibrium lattice positions. Thus, we are unable to treat the decay of excited electron-hole pairs due to electron-phonon scattering, and our calculations are restricted to zero temperature. The solid's ground state, $|0\rangle$, is then the electronic ground state, which we assume to be a spin singlet, and represented by a single Slater determinant of occupied valence orbitals:

$$|0\rangle = \left(\prod_v a^\dagger_{v\uparrow} a^\dagger_{v\downarrow} \right) |\text{vacuum}\rangle. \tag{1}$$

When a photon is absorbed, an electron is promoted from the valence band to the conduction band, leaving a hole behind. An excited state, denoted by $|f\rangle$, is *singly excited* and therefore contains one electron-hole pair: $(a^\dagger_{c\uparrow}a_{v\uparrow} + a^\dagger_{c\downarrow}a_{v\downarrow})|0\rangle = |c,v\rangle$. Henceforth, we suppress spin indices, and consider only spin singlets; any product of the form $a^\dagger a$ is taken to mean $a^\dagger_\uparrow a_\uparrow + a^\dagger_\downarrow a_\downarrow$. Because of the electron-hole interaction, a given (c,v) pair is not an energy eigenstate. Instead, $|f\rangle$ is a superposition of all possible (c,v) pairs:

$$|f\rangle = \sum_{c,v} \psi(c,v) a^\dagger_c a_v |0\rangle = \sum_{c,v} \psi(c,v) |c,v\rangle. \tag{2}$$

The expansion coefficient, $\psi(c,v)$, can be viewed as the electron-hole pair wave function in the *one-particle eigenstate* basis. Another basis which is of use is the *real-space basis*, in which the electron and hole are each located at positions in the solid:

$$|f\rangle = \sum_{\mathbf{r}_e,\mathbf{r}_h} \psi(\mathbf{r}_e,\mathbf{r}_h) a^\dagger_{\mathbf{r}_e} a_{\mathbf{r}_h} |0\rangle = \sum_{\mathbf{r}_e,\mathbf{r}_h} \psi(\mathbf{r}_e,\mathbf{r}_h) |\mathbf{r}_e,\mathbf{r}_h\rangle. \tag{3}$$

Restricting the disscussion to crystalline solids, Eq. (2) can be rewritten as

$$|f\rangle = \sum_{i,j,\mathbf{k}} \psi(i,j;\mathbf{k}) a^\dagger_{i,\mathbf{k}+\mathbf{q}} a_{j,\mathbf{k}} |0\rangle = \sum_{i,j,\mathbf{k}} \psi(i,j;\mathbf{k}) |i,j;\mathbf{k}\rangle, \tag{4}$$

where i and j are electron and hole band indices, and $\mathbf{k}$ is the wave vector of the hole (the wave vector of the electron is $\mathbf{k}+\mathbf{q}$, where $\hbar\mathbf{q}$ is the miniscule photon momentum). Likewise, Eq. (3) can be rewritten as

$$\begin{aligned} |f\rangle &= \sum_{\mathbf{x},\mathbf{y},\mathbf{R}} \psi(\mathbf{x},\mathbf{y};\mathbf{R}) \sum_{\mathbf{R}'} \exp(i\mathbf{q}\cdot\mathbf{R}') a^\dagger_{\mathbf{x}+\mathbf{R}+\mathbf{R}'} a_{\mathbf{y}+\mathbf{R}'} |0\rangle \\ &\equiv \sum_{\mathbf{x},\mathbf{y},\mathbf{R}} \psi(\mathbf{x},\mathbf{y};\mathbf{R}) |\mathbf{x},\mathbf{y};\mathbf{R}\rangle, \end{aligned} \tag{5}$$

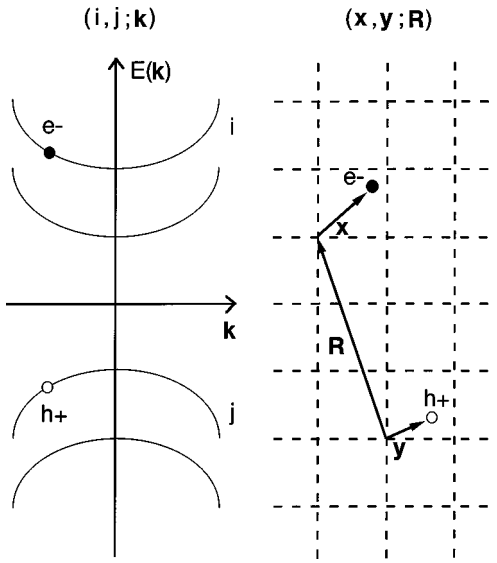

FIG. 1. Pictorial representations of $(i,j;\mathbf{k})$ and $(\mathbf{x},\mathbf{y};\mathbf{R})$ bases. The curves on the left-hand side represent one-electron bands, and the dashed boxes on the right-hand side represent unit cells.

where $\mathbf{x}$ and $\mathbf{y}$ are the positions of the electron and hole in their respective unit cells, and $\mathbf{R}$ is the lattice translation vector separating the cells. In Eq. (5), the indicated Bloch sum over $\mathbf{R}'$ insures that the total crystal momentum of the pair is always $\hbar\mathbf{q}$. Note that the arguments in the triples $(i,j;\mathbf{k})$ and $(\mathbf{x},\mathbf{y};\mathbf{R})$ are closely related to each other. In particular, $\mathbf{R}$ is the Fourier transform conjugate of $\mathbf{k}$. Figure 1 shows a pictorial representation of the two bases.

It is often necessary to transform from the $(i,j;\mathbf{k})$ basis to the $(\mathbf{x},\mathbf{y};\mathbf{R})$ basis and vice versa. This is accomplished with the transformation law,

$$\psi(\mathbf{x},\mathbf{y};\mathbf{R}) = \sum_{i,j} \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}+\mathbf{R})} u_{i,\mathbf{k}+\mathbf{q}}(\mathbf{x}) u^*_{j,\mathbf{k}}(\mathbf{y}) \psi(i,j;\mathbf{k}), \tag{6}$$

where $u_{i,\mathbf{k}+\mathbf{q}}(\mathbf{x})$ and $u^*_{j,\mathbf{k}}(\mathbf{y})$ are the periodic parts of Bloch functions for the electron and hole. The inverse transformation is also easily written down.

B. Bethe-Salpeter equation

The equation that the electron-hole pair wave function satisfies can be derived from the equation of motion for the electron-hole pair creation operator:

$$\langle\Psi|[H, a^\dagger_c a_v]|\Psi_0\rangle = E\langle\Psi|a^\dagger_c a_v|\Psi_0\rangle. \tag{7}$$

Here, $|\Psi\rangle$ is an exact excited state of energy E, $|\Psi_0\rangle$ is the exact ground state with energy set to zero, and H is the exact many-body Hamiltonian. If we replace $|\Psi_0\rangle$ by $|0\rangle$ of Eq. (1), and $|\Psi\rangle$ by the $|f\rangle$ of Eq. (4), we obtain the approximate equation of motion for the electron-hole pair in the $(i,j;\mathbf{k})$ basis:

$$[E_i(\mathbf{k}+\mathbf{q})-E_j(\mathbf{k})-E]\psi(i,j;\mathbf{k})$$
$$=-\sum_{i',j'}\int\frac{d^3\mathbf{k}'}{(2\pi)^3}[2\langle i,j;\mathbf{k}|V_{\rm ex}|i',j';\mathbf{k}'\rangle$$
$$-\langle i,j;\mathbf{k}|V_{\rm dir}|i',j';\mathbf{k}'\rangle]\psi(i',j';\mathbf{k}'). \quad (8)$$

E_i and E_j are one-electron band energies. The direct matrix element, $\langle i,j;\mathbf{k}|V_{\rm dir}|i',j';\mathbf{k}'\rangle$, is given by

$$\sum_{\mathbf{R}}\int d^3\mathbf{x}\int d^3\mathbf{y}e^{-i(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{x}-\mathbf{y}+\mathbf{R})}u^*_{i,\mathbf{k}+\mathbf{q}}(\mathbf{x})u_{j,\mathbf{k}}(\mathbf{y})$$
$$\times V(\mathbf{x},\mathbf{y};\mathbf{R})u_{i',\mathbf{k}'+\mathbf{q}}(\mathbf{x})u^*_{j',\mathbf{k}'}(\mathbf{y}),$$

while the exchange matrix element, $\langle i,j;\mathbf{k}|V_{\rm ex}|i',j';\mathbf{k}'\rangle$, is given by

$$\sum_{\mathbf{R}}\int d^3\mathbf{x}\int d^3\mathbf{y}e^{-i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y}+\mathbf{R})}u^*_{i,\mathbf{k}+\mathbf{q}}(\mathbf{x})u_{j,\mathbf{k}}(\mathbf{x})$$
$$\times V(\mathbf{x},\mathbf{y};\mathbf{R})u_{i',\mathbf{k}'+\mathbf{q}}(\mathbf{y})u^*_{j',\mathbf{k}'}(\mathbf{y}).$$

$V(\mathbf{x},\mathbf{y};\mathbf{R})=e^2/|\mathbf{x}-\mathbf{y}+\mathbf{R}|$ is the Coulomb interaction between the electron at $\mathbf{x}+\mathbf{R}$ and the hole at $\mathbf{y}$. Both terms describe the scattering of pair $(i,j;\mathbf{k})$ into pair $(i'j';\mathbf{k}')$.

Note that the interaction V that appears in the matrix elements is unscreened. This is somewhat unphysical. The excited electron and hole move within a sea of all the other electrons in the system. These other electrons form a polarizable medium that should screen the Coulomb interaction between the electron and hole. A diagramatic perturbation theory approach first carried out by Sham and Rice[3] reveals that $\langle i,j;\mathbf{k}|V_{\rm dir}|i',j';\mathbf{k}'\rangle$ should involve the screened Coulomb interaction, W, while $\langle i,j;\mathbf{k}|V_{\rm ex}|i',j';\mathbf{k}'\rangle$ involves the bare interaction, V (if W is used in the exchange term, it turns out that screening is double counted[3]). In principle, W should be time dependent. In practice, however, a statically screened interaction is sufficient for the bulk systems that we study.[4–6] More will be said about the precise form of W that we use in a later subsection.

Equation (8) (with interactions V and W used in exchange and direct terms) is the Bethe-Salpeter equation. It is a secular equation for the electron-hole pair wave function, $\psi(i,j;\mathbf{k})$, and the pair energy E. Two points are worth noting: (1) The exchange term enters with a positive sign, while the direct term has a negative sign. Because both the screened and unscreened interactions are positive, this means that the exchange term is a *repulsion* between the electron and hole, while the direct term is an *attraction*. (2) For a real semiconductor or insulator, the characteristic magnitude of the direct term is much larger than that of the exchange term. This is because the direct integral involves products like $u^*_{i,\mathbf{k}+\mathbf{q}}(\mathbf{x})u_{i',\mathbf{k}'+\mathbf{q}}(\mathbf{x})$, which is large, because i and i' are both conduction bands (so the corresponding wave functions should have appreciable overlap). The exchange integral involves products like $u^*_{i,\mathbf{k}+\mathbf{q}}(\mathbf{x})u_{j,\mathbf{k}}(\mathbf{x})$. Because i is a conduction band and j is a valence band, their corresponding wave functions have little overlap, and this term is small. Thus, for a system with a notable difference between conduction and valence state character (i.e., a system with a gap), the net electron-hole interaction is attractive and favors excitonic binding.

C. Effective Hamiltonian

Because the Bethe-Salpeter equation is a secular equation, one can extract an effective Hamiltonian from it so that Eq. (8) is of the form $H_{\rm eff}\psi=E\psi$. It is important to understand that $H_{\rm eff}$ is different from the exact many-body Hamiltonian, H, which appears in Eq. (7). In particular, $H_{\rm eff}$ acts only on the space of *singly excited states*. From the form of the Bethe-Salpeter equation, $H_{\rm eff}$ has three terms:

$$H_{\rm eff}=H_{1e}+H_{\rm dir}+H_{\rm ex}. \quad (9)$$

We now discuss each term in the order of its importance (magnitude).

1. One-electron term H_{1e}

H_{1e} is most easily defined by the way it acts on a basis vector in the one-particle eigenstate basis:

$$H_{1e}|i,j;\mathbf{k}\rangle=[E_i(\mathbf{k}+\mathbf{q})-E_j(\mathbf{k})]|i,j;\mathbf{k}\rangle. \quad (10)$$

This difference of one-electron band energies is just the excitation energy in the absence of the electron-hole interaction. Because the photon wave vector $\mathbf{q}$ is small compared to the characteristic Brillouin-zone length, we replace $E_i(\mathbf{k}+\mathbf{q})$ by $E_i(\mathbf{k})$ in practice. As stated in the Introduction, the electrons and holes are really *quasielectrons* and *quasiholes*, already renormalized by the presence of their surrounding polarization clouds. Thus, we use accurate quasiparticle band energies for $E_i(\mathbf{k})$ and $E_j(\mathbf{k})$ from GW calculations.[2] These energies compare very well with the results of direct and inverse photoemission for materials that are not srongly correlated. Band energies are computed on a discrete mesh of $\mathbf{k}$ points in the first Brillouin zone, and for a finite number of valence bands j and conduction bands i. LDA u functions and band energies are computed efficiently using the basis set discussed in Ref. 7. When it is appropriate, LDA band energies are updated to include self-energy effects as discussed below.

2. Direct term $H_{\rm dir}$

$H_{\rm dir}$ is an attraction between the electron and hole and involves the screened Coulomb interaction W. Although the exact form of W is impossible to compute, we know that an approximate statically screened interaction should satisfy the following short-range and long-range limits:

$$W(\mathbf{x}\to\mathbf{y};\mathbf{R}=0)\to\frac{e^2}{|\mathbf{x}-\mathbf{y}|},$$

$$W(\mathbf{x},\mathbf{y};\mathbf{R}\to\infty)\to\frac{e^2}{\epsilon|\mathbf{x}-\mathbf{y}+\mathbf{R}|},$$

where ϵ is the IR dielectric constant.[8] In addition, the way in which the two limits are connected should depend on the background electron density.

These requirements are satisfied by the static random-phase approximation (RPA), although we use an approxima-

tion to RPA in this work: the Hybertsen-Levine-Louie model.[9,10] This model has two parts. First, a function $W_{\rm hom}(r;\rho,\epsilon)$ is defined that represents the statically screened Coulomb interaction between charges separated by r in a fictitious system with *homogeneous* density ρ, and IR dielectric constant ϵ. $W_{\rm hom}$ is constructed to give the RPA result for the electron gas when $\epsilon=\infty$, and is required to satisfy known sum rules when $\epsilon<\infty$. An analytic form for $W_{\rm hom}$ in reciprocal space $[W_{\rm hom}(Q;\rho,\epsilon)]$ is available.[9] Second, an approximate form for the screened interaction in an *inhomogeneous* system is written down as follows:[10]

$$W(\mathbf{r},\mathbf{r}')=\tfrac{1}{2}[W_{\rm hom}(|\mathbf{r}-\mathbf{r}'|;\rho(\mathbf{r}'),\epsilon) + W_{\rm hom}(|\mathbf{r}'-\mathbf{r}|;\rho(\mathbf{r}),\epsilon)], \tag{11}$$

where $\rho(\mathbf{r})$ is the position-dependent charge density in the material. If $\mathbf{r}$ and $\mathbf{r}'$ are both in regions of similar density, and if the region between $\mathbf{r}$ and $\mathbf{r}'$ is at that density, then $W(\mathbf{r},\mathbf{r}')$ should be very close to the RPA result. More importantly, $W(\mathbf{r},\mathbf{r}')$ satisfies the above short- and long-range limits, irrespective of density.

In order to implement Eq. (11), $W_{\rm hom}(r;\rho,\epsilon)$ is computed from $W_{\rm hom}(Q;\rho,\epsilon)$ for a range of densities (and for the IR dielectric constant of the system of interest) by numerical integration:

$$W_{\rm hom}(r;\rho,\epsilon)=\int\frac{d^3Q}{(2\pi)^3}e^{i\mathbf{Q}\cdot\mathbf{r}}W_{\rm hom}(Q;\rho,\epsilon) =\frac{2}{\pi}\left(\frac{1}{r}\right)\int_0^\infty dQ\,\epsilon_{LL}^{-1}(Q;\rho,\epsilon)\left[\frac{\sin(Qr)}{Q}\right], \tag{12}$$

where $\epsilon_{LL}(Q;\rho,\epsilon)$ is the Levine-Louie dielectric function appearing in analytic form in Ref. 9. The ground-state electronic charge density determined from the LDA is used for $\rho(\mathbf{r})$. By using a discrete mesh of vectors $\mathbf{x}$ and $\mathbf{y}$ in the unit cell, and a mesh of lattice vectors $\mathbf{R}$, we can represent W in the $(\mathbf{x},\mathbf{y};\mathbf{R})$ basis:

$$W(\mathbf{x},\mathbf{y};\mathbf{R})=\tfrac{1}{2}[W_{\rm hom}(|\mathbf{x}-\mathbf{y}+\mathbf{R}|;\rho(\mathbf{y}),\epsilon) + W_{\rm hom}(|\mathbf{y}-\mathbf{x}-\mathbf{R}|;\rho(\mathbf{x}),\epsilon)]. \tag{13}$$

$H_{\rm dir}$ is diagonal in this basis, so we have

$$H_{\rm dir}|\mathbf{x},\mathbf{y};\mathbf{R}\rangle=W(\mathbf{x},\mathbf{y};\mathbf{R})|\mathbf{x},\mathbf{y};\mathbf{R}\rangle. \tag{14}$$

Because one has $W_{\rm hom}(r;\rho,\epsilon)\to e^2/r$ for $r\to 0$, $W(\mathbf{x}=\mathbf{y};\mathbf{R}=0)$ diverges. We circumvent this problem by defining $W(\mathbf{x}=\mathbf{y};\mathbf{R}=0)$ to be the average of $W_{\rm hom}$ over a sphere with volume $V_{\mathbf{x}}$ (the volume corresponding to one discrete $\mathbf{x}$ point):

$$W(\mathbf{x}=\mathbf{y};\mathbf{R}=0)=\frac{1}{V_{\mathbf{x}}^2}\int_{V_{\mathbf{x}}}d^3\mathbf{x}_1 d^3\mathbf{x}_2 W_{\rm hom}(|\mathbf{x}_1-\mathbf{x}_2|;\rho(\mathbf{x}),\epsilon). \tag{15}$$

This integral is tabulated numerically for a range of densities ρ.

In practice, we are given the two-particle wave function in the $(i,j;\mathbf{k})$ basis, and it is necessary to transform it to the $(\mathbf{x},\mathbf{y};\mathbf{R})$ basis before Eq. (14) is applied. This transformation is the one of Eq. (6). After the action by $H_{\rm dir}$, the inverse of Eq. (6) is used to transform back to the $(i,j;\mathbf{k})$ basis. The mesh of $\mathbf{R}$ vectors is chosen to be compatible with the mesh of $\mathbf{k}$ points, so fast Fourier transform (FFT) techniques can be used in the $\mathbf{k}\to\mathbf{R}$ and $\mathbf{R}\to\mathbf{k}$ parts of the transformations. We obtain the $u_{i,\mathbf{k}}$ and $u_{j,\mathbf{k}}$ functions [appearing in Eq. (6)] from LDA calculations. In principle, we should use quasiparticle wave functions from *GW* calculations, but because of the computational difficulty of obtaining them for a large set of $\mathbf{k}$ points, we choose not to do this. It has been shown (at least for the few cases tested[2]) that LDA and *GW* wave functions are quite similar, even when the corresponding band energies are quite different.

One subtle point must now be addressed: In order for the $(i,j;\mathbf{k})\to(\mathbf{x},\mathbf{y};\mathbf{R})$ transformation to be unitary, the following conditions must be satisfied:

$$\sum_{\mathbf{x}} u_{i,\mathbf{k}}^*(\mathbf{x})u_{i',\mathbf{k}}(\mathbf{x})=\delta_{i,i'},$$

$$\sum_{\mathbf{y}} u_{j,\mathbf{k}}^*(\mathbf{y})u_{j',\mathbf{k}}(\mathbf{y})=\delta_{j,j'},$$

where we have used $\mathbf{k}+\mathbf{q}\approx\mathbf{k}$. For infinite, continuous sets of $\mathbf{x}$ and $\mathbf{y}$ points, these orthogonality conditions are automatically satisfied. For discrete sets of $\mathbf{x}$ and $\mathbf{y}$ points, however, this is not the case. We enforce these conditions by viewing $u_{i,\mathbf{k}}$ and $u_{j,\mathbf{k}}$ as complex vectors with components $u_{i,\mathbf{k}}(\mathbf{x})$ and $u_{j,\mathbf{k}}(\mathbf{y})$. A Gramm-Schmidt orthogonalization procedure is then applied that enforces orthogonality. This is possible as long as the number of ''vectors'' is less than the ''dimension'' [i.e., the number of bands i (j) is less than the number of real-space grid points $\mathbf{x}$ $(\mathbf{y})$]. Orthogonalization aids the convergence of the calculation with respect to the number of $\mathbf{x},\mathbf{y}$ points. We do not enforce unitarity of the $(\mathbf{x},\mathbf{y};\mathbf{R})\to(i,j;\mathbf{k})$ transformation, because $H_{\rm dir}|\psi\rangle$ may contain contributions from conduction bands not included in the calculation.

3. Exchange term $H_{\rm ex}$

$H_{\rm ex}$ is a repulsion between the electron and hole and involves the bare Coulomb interaction V. Before discussing the way in which this term is evaluated, the use of the bare interaction must be justified. If an infinite number of one-electron bands is used in the calculation, then the work of Sham and Rice[3] shows that it is correct to use the bare interaction in the exchange matrix elements. In practice, however, we use a finite number of conduction bands, and fewer valence bands than there are electrons in the unit cell. Thus, the subspace on which $H_{\rm eff}$ acts is smaller than the full space of singly excited states. The correct interaction to use in $H_{\rm ex}$ is therefore something like $V/\epsilon_{\rm BG}$, where $\epsilon_{\rm BG}$ is the background dielectric constant resulting from all transitions not explicitly included. Instead of calculating $\epsilon_{\rm BG}$, we simply set $V/\epsilon_{\rm BG}=V$, and converge the calculation with respect to the number of bands. When our results become independent of the number of bands we include, we are sure that we are close to the limit, $\epsilon_{\rm BG}\to 1$.

In order to act with H_{ex} on a state vector, we use the form of V in reciprocal space:

$$V(\mathbf{G},\mathbf{G}';\mathbf{q}) = \frac{4\pi}{\Omega}\frac{e^2}{|\mathbf{q}+\mathbf{G}|^2}\delta_{\mathbf{G},\mathbf{G}'} = \hat{V}(\mathbf{G};\mathbf{q})\delta_{\mathbf{G},\mathbf{G}'},$$

where $\mathbf{G}$ and $\mathbf{G}'$ are reciprocal lattice vectors, and Ω is the unit cell volume. If we have

$$H_{ex}\sum_{i,j,\mathbf{k}} \psi(i,j;\mathbf{k})|i,j;\mathbf{k}\rangle = \sum_{i,j,\mathbf{k}} \xi(i,j;\mathbf{k})|i,j;\mathbf{k}\rangle,$$

then we have

$$\xi(i,j;\mathbf{k}) = 2\sum_{i',j',\mathbf{k}'} \psi(i',j';\mathbf{k}')\sum_{\mathbf{G}} \Upsilon^{i',j'}_{\mathbf{k}'}(\mathbf{G})\Upsilon^{*i,j}_{\mathbf{k}}(\mathbf{G})\hat{V}(\mathbf{G};\mathbf{q}), \tag{16}$$

where we have defined

$$\Upsilon^{i,j}_{\mathbf{k}}(\mathbf{G}) = \sum_{\mathbf{G}'} u^*_{i,\mathbf{k}+\mathbf{q}}(\mathbf{G}')u_{j,\mathbf{k}}(\mathbf{G}-\mathbf{G}').$$

Since $\Upsilon^{i,j}_{\mathbf{k}}(\mathbf{G})$ approaches zero for $|\mathbf{G}|\to\infty$ faster than $1/|\mathbf{G}|^2$, and $\hat{V}(\mathbf{G},\mathbf{q})$ goes as $1/|\mathbf{G}|^2$, Eq. (16) converges with very few $\mathbf{G}$ vectors.

Given a singly excited state $|S\rangle = \Sigma_{i,j,\mathbf{k}}\psi(i,j;\mathbf{k})|i,j;\mathbf{k}\rangle$, the action of $H_{\rm eff} = H_{1e} + H_{\rm dir} + H_{ex}$ on $|S\rangle$ is computed term by term. H_{1e} is applied with Eq. (10); $H_{\rm dir}$ is applied with Eq. (14) together with the transformation (and corresponding inverse transformation) of Eq. (6); H_{ex} is applied with Eq. (16). The vast majority of the computation time is spent transforming from $(i,j;\mathbf{k})\to(\mathbf{x},\mathbf{y};\mathbf{R})$ and back again, in the evaluation of $H_{\rm dir}|S\rangle$. The number of operations required in this time-consuming part is proportional to $N_v N_c N_{\mathbf{x}}^2 N_{\mathbf{k}}$, where N_v and N_c are the numbers of valence and conduction bands, and $N_{\mathbf{x}}$ $(=N_{\mathbf{y}})$ and $N_{\mathbf{k}}$ $(=N_{\mathbf{R}})$ are the numbers of $\mathbf{x}$ and $\mathbf{k}$ points.

D. Calculation of $\epsilon_2(\omega)$

The imaginary part of the frequency-dependent long-wavelength dielectric function can be expressed in terms of the ground state, $|0\rangle$, the singly excited states (eigenstates of $H_{\rm eff}$), $|f\rangle$, and the component of the macroscopic current operator along the polarization direction, $\hat{\lambda}\cdot\mathbf{J}$:

$$\epsilon_2(\omega) = \frac{4\pi^2}{\omega^2}\sum_f |\langle f|\hat{\lambda}\cdot\mathbf{J}|0\rangle|^2\delta(\omega-E_f) = 4\pi^2\sum_f \left(\frac{1}{E_f^2}\right)|\langle f|\hat{\lambda}\cdot\mathbf{J}|0\rangle|^2\delta(\omega-E_f). \tag{17}$$

E_f is the energy of state $|f\rangle$ (the energy of the ground state is set to zero). Instead of working with this sum-over-eigenstates expression, we choose to eliminate the sum over f altogether by transforming Eq. (17) into a form that involves the ground state only:

$$\epsilon_2(\omega) = -4\pi \times \mathrm{Im}\left[\langle 0|\hat{\lambda}\cdot\mathbf{J}H_{\rm eff}{}^{-1}\left(\frac{1}{\omega-H_{\rm eff}+i\eta}\right)H_{\rm eff}{}^{-1}\hat{\lambda}\cdot\mathbf{J}|0\rangle\right], \tag{18}$$

where η is a positive infinitessimal. We have used the completeness relation $\Sigma_f|f\rangle\langle f| = 1$ which applies in the space in which $H_{\rm eff}$ acts. We are now free to evaluate this ground-state expectation value in any basis we choose. Because we have access to the one-electron band states, we choose the one-particle eigenstate basis:

$$\epsilon_2(\omega) = -4\pi\mathrm{Im}\left[\sum_{i,j,\mathbf{k},i',j',\mathbf{k}'} \langle 0|P^\dagger|i,j;\mathbf{k}\rangle\langle i,j;\mathbf{k}| \times\left(\frac{1}{\omega-H_{\rm eff}+i\eta}\right)|i',j';\mathbf{k}'\rangle\langle i',j';\mathbf{k}'|P|0\rangle\right], \tag{19}$$

where $P = H_{\rm eff}{}^{-1}\hat{\lambda}\cdot\mathbf{J}$ is the polarization operator.[11]

We make the replacement $\langle 0|\hat{\lambda}\cdot\mathbf{J}H_{\rm eff}^{-1}|i,j;\mathbf{k}\rangle = [E_i(\mathbf{k}) - E_j(\mathbf{k})]^{-1}\langle 0|\hat{\lambda}\cdot\mathbf{J}|i,j;\mathbf{k}\rangle$. H_{1e} involves the quasiparticle self-energy corrections for the electron and hole (see Sec. II C 1). As discussed by Levine and Allan,[12] this self-energy operator is nonlocal. If we use LDA single-particle wave functions for $|i,j;\mathbf{k}\rangle$, it is then necessary to make the replacement $\mathbf{J}\to[E_i(\mathbf{k}) - E_j(\mathbf{k}) - H_{1e}^{\rm LDA}(\mathbf{k})]^{-1}[E_i(\mathbf{k}) - E_j(\mathbf{k}) - H_{1e}(\mathbf{k})]\mathbf{J}$, where $H_{1e}(\mathbf{k})$ denotes a $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, $H_{1e}(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{r}}H_{1e}e^{i\mathbf{k}\cdot\mathbf{r}}$.[12] The states $|0\rangle$ and $|i,j;\mathbf{k}\rangle$ represent Slater determinants of single-particle orbitals, so we have

$$\langle i,j;\mathbf{k}|P|0\rangle = \left[\frac{1}{E_i^{\rm LDA}(\mathbf{k}) - E_j^{\rm LDA}(\mathbf{k})}\right] \times \int d^3\mathbf{r}\, u^*_{j,\mathbf{k}}(\mathbf{r})\hat{\lambda}\cdot[\nabla_{\mathbf{k}'}H_{1e}^{\rm LDA}(\mathbf{k}')]|_{\mathbf{k}'=\mathbf{k}}u_{i,\mathbf{k}}(\mathbf{r}). \tag{20}$$

The $\mathbf{k}$-space gradient is computed by finite differences. The integral is then evaluated in reciprocal space, using the Fourier components of the LDA u functions.

Defining $|P\rangle = P|0\rangle = \Sigma_{i,j,\mathbf{k}}|i,j;\mathbf{k}\rangle\langle i,j;\mathbf{k}|P|0\rangle$, Eq. (19) becomes

$$\epsilon_2(\omega) = -4\pi\mathrm{Im}\left[\left\langle P\left|\left(\frac{1}{\omega-H_{\rm eff}+i\eta}\right)\right|P\right\rangle\right]. \tag{21}$$

This can be evaluated iteratively using the Haydock recursion method.[13] For the first iteration, we set $|1\rangle = |P\rangle$, and compute $H_{\rm eff}|1\rangle$ using the results of the previous subsection. The expectation value is defined to be $a_1 = \langle 1|H_{\rm eff}|1\rangle$. We next compute the vector $H_{\rm eff}|1\rangle - a_1|1\rangle$. This vector would be zero if $|1\rangle$ were an eigenvector of $H_{\rm eff}$. The norm of this difference vector is set equal to $b_1 = |H_{\rm eff}|1\rangle - a_1|1\rangle|$. The normalized vector, $b_1^{-1}(H_{\rm eff}|1\rangle - a_1|1\rangle)$, is denoted $|2\rangle$. This completes the first iteration. In the second iteration, a_2 and b_2 are computed from $a_2 = \langle 2|H_{\rm eff}|2\rangle$, and $b_2 = |H_{\rm eff}|2\rangle - a_2|2\rangle - b_1|1\rangle|$. The general form for the nth iteration is

$$a_n = \langle n|H_{\rm eff}|n\rangle,$$

$$b_n = |H_{\rm eff}|n\rangle - a_n|n\rangle - b_{n-1}|n-1\rangle|,$$

$$|n+1\rangle = b_n^{-1}(H_{\rm eff}|n\rangle - a_n|n\rangle - b_{n-1}|n-1\rangle). \quad (22)$$

The expectation value of the Green function can now be written as a continued fraction,

$$\langle P|\left(\frac{1}{\omega - H_{\rm eff} + i\eta}\right)|P\rangle = \cfrac{1}{\omega + i\eta - a_1 - \cfrac{b_1^2}{\omega + i\eta - a_2 - \cfrac{b_2^2}{\omega + i\eta - a_3 \ldots}}}. \quad (23)$$

The number of ''levels'' in this fraction is equal to the number of iterations, $N_{\rm iter}$. The question of how to terminate the calculation is an important one. Methods of varying sophistication have been used; we use the least sophisticated, in which $(a_n, b_n) = (a_{N_{\rm iter}}, b_{N_{\rm iter}})$ for all $n > N_{\rm iter}$. A simple analytic form for the termination can then be derived.[14] Each iteration computes various ω moments of $\epsilon_2(\omega)$. For instance, from Eq. (21) and the definition $a_1 = \langle 1|H_{\rm eff}|1\rangle = \langle P|H_{\rm eff}|P\rangle$, it can be shown that one has $4\pi^2 a_1 = \int_0^{+\infty} d\omega\, \omega \epsilon_2(\omega)$.

The advantage of using an iterative technique is twofold. First, the Hamiltonian matrix never needs to be computed in any *one* basis, since all that is required is to compute the Hamiltonian acting on a vector. This reduces computation time, because $H_{\rm eff}|\psi\rangle$ involves a sum over one expansion coefficient, while $\langle\phi|H_{\rm eff}|\psi\rangle$ involves the simultaneous sum over two coefficients. Second, the number of iterations needed to compute $\epsilon_2(\omega)$ depends mostly on the size of the broadening parameter η and is roughly independent of the dimension of the Hamiltonian. As long as η is greater than the characteristic level spacing of $H_{\rm eff}$, the iterative scheme is more efficient than standard diagonalization. Since $N_{\rm iter}$ is roughly independent of the dimension, the total number of operations is proportional to the number of operations per action of $H_{\rm eff}$. From above, this is $N_{\rm v} N_{\rm c} N_{\mathbf{x}}^2 N_{\mathbf{k}}$. If there are N atoms per unit cell, one has $N_{\rm v}, N_{\rm c}, N_{\mathbf{x}} \propto N$, while one has $N_{\mathbf{k}} \rightarrow$ constant, so the calculation as a whole scales as N^4.

This method of computing a spectrum within linear response is completely general; any operator could be used in place of P in Eq. (21). It is even possible to compute the total density of excited states as follows:[15] consider the operator R having the property

$$R|0\rangle = |R\rangle = \sum_{i,j,\mathbf{k}} \exp(\phi_{i,j,\mathbf{k}})|i,j;\mathbf{k}\rangle, \quad (24)$$

where $\phi_{i,j,\mathbf{k}}$ are random numbers between 0 and 2π. Thus, R acts on the ground state to create an *incoherent* (but equal probability) superposition of electron-hole pair excitations. If $|R\rangle$ is inserted in place of $|P\rangle$ in Eq. (21), the resulting spectrum is proportional to the total density of states, $D(\omega) = \Sigma_f \delta(\omega - E_f)$.

E. Analysis of $\epsilon_2(\omega)$ by spectral decomposition

The main advantage of the iterative scheme is that it is not necessary to compute eigenstates and eigenenergies in order to determine the spectrum. However, this is also its major disadvantage. It is often desirable to know which single particle transitions [$(i,j;\mathbf{k})$ pairs] contribute to a peak in $\epsilon_2(\omega)$. It may also be of interest to know the spatial distribution of the electron and hole in this spectral region. Information of this sort is not easily accessed in a scheme in which $\epsilon_2(\omega)$ is determined directly. In this section, we show that it is possible to extract this information by using another iterative technique. The computation time is comparable to that of the scheme discussed above, provided that one is interested in a single narrow-energy window.

Consider the singly excited state $|\Phi(\omega)\rangle$ defined by

$$|\Phi(\omega)\rangle = \frac{2\sqrt{\pi\eta}}{\omega - H_{\rm eff} + i\eta} H_{\rm eff}^{-1} \hat{\lambda}\cdot\mathbf{J}|0\rangle = \frac{2\sqrt{\pi\eta}}{\omega - H_{\rm eff} + i\eta}|P\rangle. \quad (25)$$

It can be shown that $\epsilon_2(\omega)$ is just the square of the norm of this state, $\langle\Phi(\omega)|\Phi(\omega)\rangle$. If we know $|\Phi(\omega)\rangle$ in the $(i,j;\mathbf{k})$ basis, then we can write

$$\begin{aligned}\epsilon_2(\omega) &= \langle\Phi(\omega)|\Phi(\omega)\rangle \\ &= \langle\Phi(\omega)|\left(\sum_{i,j,\mathbf{k}} |i,j;\mathbf{k}\rangle\langle i,j;\mathbf{k}|\right)|\Phi(\omega)\rangle \\ &= \sum_{i,j,\mathbf{k}} |\langle i,j;\mathbf{k}|\Phi(\omega)\rangle|^2. \end{aligned} \quad (26)$$

Each term in the rightmost sum in Eq. (26) involves a particular $(i,j;\mathbf{k})$ configuration. Thus, $\epsilon_2(\omega)$ is decomposed into individual interband transitions.

The state $|\Phi(\omega)\rangle$ can be computed by rewriting Eq. (25) so that $|\Phi(\omega)\rangle$ is the solution to a linear system of equations of the form $Ax = b$:

$$(\omega - H_{\rm eff} + i\eta)|\Phi(\omega)\rangle = 2\sqrt{\pi\eta}|P\rangle. \quad (27)$$

Given $|P\rangle$, $|\Phi(\omega)\rangle$ is determined iteratively by first guessing a solution, and then modifying the guess by computing $(\omega - H_{\rm eff} + i\eta)|\Phi^{\rm guess}(\omega)\rangle$ and comparing to $2\sqrt{\pi\eta}|P\rangle$. As before, the calculation only involves the action of $H_{\rm eff}$ on a vector, rather than the computation of $H_{\rm eff}$ in some basis, followed by inversion. The number of iterations needed depends on η, and the distribution of eigenvalues of $H_{\rm eff}$ in the neighborhood of ω. Care must be taken when dealing with a non-Hermitian operator like $\omega - H_{\rm eff} + i\eta$. The generalized minimal residual method (GMRES) is suited to this task.[16]

Once $|\Phi(\omega)\rangle$ is known in some basis, it can be transformed to any other, to allow for different decompositions of $\epsilon_2(\omega)$. For instance, Eq. (6) can be used to decompose $\epsilon_2(\omega)$ into individual $(\mathbf{x},\mathbf{y};\mathbf{R})$ contributions:

$$\epsilon_2(\omega) = \sum_{\mathbf{x},\mathbf{y},\mathbf{R}} |\langle\mathbf{x},\mathbf{y};\mathbf{R}|\Phi(\omega)\rangle|^2. \quad (28)$$

It should be kept in mind that in order to do a spectral decomposition of this type, an iterative computation must be

done for *each* value of ω. It is usually best to compute $\epsilon_2(\omega)$ first, and then decide on specific spectral regions for which to perform the decomposition.

In addition to being computationally useful, $|\Phi(\omega)\rangle$ also has an important physical interpretation. The interaction of a photon with the electrons in a solid can be viewed as occuring in two steps. First, the photon creates an electron-hole pair by polarizing the material locally. This results in the electron-hole pair state, $P|0\rangle = H_{\text{eff}}^{-1}\hat{\lambda}\cdot\mathbf{J}|0\rangle = |P\rangle$. Second, the localized electron-hole pair decays (or rather, disperses) into more delocalized stationary states (electron-hole pair eigenstates, such as excitons). This decay is governed by the electron-hole propagator, $(\omega - H_{\text{eff}} + i\eta)^{-1}$. The resulting pair state, $(\omega - H_{\text{eff}} + i\eta)^{-1}|P\rangle$, is equal to $|\Phi(\omega)\rangle$ within a constant of proportionality. Because we have $\epsilon_2(\omega) = \langle\Phi(\omega)|\Phi(\omega)\rangle$, this pair state is *directly related* to what is measured in optical absorption, reflection, and ellipsometry experiments. Thus, the physics of optical properties is that of mutually localized electrons and holes spreading apart. Note that $|\Phi(\omega)\rangle$ is not an energy eigenstate, and that no single electron-hole pair eigenstate is sufficient to capture all the features of $|\Phi(\omega)\rangle$.

III. RESULTS AND DISCUSSION

A. GaN

GaN is a tetrahedrally-coordinated material with a direct band gap between 3 eV and 4 eV. It comes in two varieties: wurtzite (hexagonal, four atoms per cell), and zinc blende (cubic, two atoms per cell). The zinc-blende phase is currently the more difficult to grow, and has not been studied as extensively as the wurtzite phase. Because of their band gaps, both phases are being considered for use as optoelectronic devices in the blue/UV spectral range. Efforts to determine their intrinsic optical properties in this range include reflectivity measurements[17] and spectroscopic ellipsometry performed with synchrotron radiation.[18–20]

Although these measurements have led to a better understanding of the optical properties, both suffer from deficiencies. Reflectivity must be coupled with a Kramers-Kronig transform to determine $\epsilon_1(\omega)$ or $\epsilon_2(\omega)$, and this introduces ambiguity into the analysis. Spectroscopic ellipsometry directly determines $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$, but the synchrotron light is contaminated by second-order radiation in certain energy regions.[18,19] Finally, sample quality is always an issue. It has been suggested that surface roughness is responsible for decreased intensity in reflectivity and ellipsometry measurements.[17,19]

In this section, we present the results of our calculations of $\epsilon_2(\omega)$ for wurtzite and zinc-blende GaN in the spectral range $\hbar\omega = 0-20$ eV. We show that the inclusion of the electron-hole interaction shifts oscillator strength to lower frequencies, bringing the shape of the calculated spectrum into better agreement with the experiments. However, as found in previous calculations,[17,21] the overall magnitude of $\epsilon_2(\omega)$ is too large when compared with experiment.

1. Wurtzite

Pseudopotential LDA calculations for the single-particle wave functions, band energies, and ground-state charge density are performed at the experimental lattice constants $a = 3.16$ and $c = 5.13$ Å, and internal bond-length parameter $u = 0.377$.[22] In this and all other calculations presented in this work, spin-orbit coupling is neglected. The conduction bands are shifted up to move the band gap (direct, and at Γ) from the LDA value (2.69 eV) to the *GW* value of 3.5 eV.[23] LDA valence and conduction bands are stretched by 12% and 17% to mimic the *GW* bandwidths. The IR dielectric constant ϵ used in the screening model [Eq. (12)] is taken to be 5.5.[22] It should be noted that, in the wurtzite structure, there are two different principle components of the static dielectric tensor (corresponding to **E** perpendicular to the *c* axis, and **E** parallel to the *c* axis). The difference between the components is roughly 10%, so we take a directionally averaged value to be used in the screening model. Calculations of $\epsilon_2(\omega)$ are performed with $N_{\mathbf{k}} = 600$ ($10\times10\times6$ mesh), $N_{\mathbf{x}} = 45$ ($3\times3\times5$ mesh), $N_{\text{v}} = 8$, $N_{\text{c}} = 8$, and $N_{\text{iter}} = 150$.[24] The broadening parameter, η, is chosen to be 0.2 eV. The $\epsilon_2(\omega)$ has two principle components: **E** perpendicular to *c*, and **E** parallel to *c*. The polarization directions are specified by the direction of $\hat{\lambda}$ in the transition-matrix elements of Eq. 20.

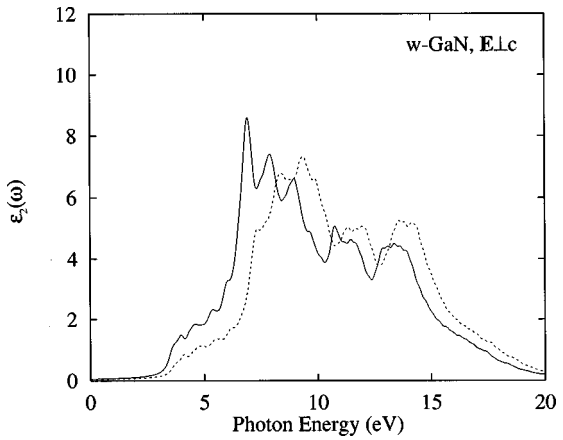

FIG. 2. Calculated $\epsilon_2(\omega)$ for **E** perpendicular to *c* for wurtzite GaN. The solid line includes the electron-hole interaction, while the dashed line is the result of one-electron theory.

Figure 2 shows the calculated $\epsilon_2(\omega)$ for **E** perpendicular to *c* for wurtzite GaN. The solid line includes the electron-hole interaction, while the dashed line is the result of one-electron theory [obtained by neglecting H_{dir} and H_{ex} in Eq. (9)]. The net attraction of the electron-hole interaction moves oscillator strength to lower frequencies. The one-electron results are very similar to those of previous workers.[17,21] Our results including the electron-hole interaction look very similar to $\epsilon_2(\omega)$ deduced from a Kramers-Kronig transform of reflectivity data.[17] Figure 3 shows our results plotted together with the reflectivity data. Peak positions and relative peak heights are reproduced well,[25] but the average magnitude of our calculated $\epsilon_2(\omega)$ is $\sim 50-100\,\%$ too high compared to experiment. This is also true when comparing to the spectroscopic ellipsometry experiments of Logothetidis *et al.*[18] and Wethkamp *et al.*[19] Following the reasoning in Lambrecht *et al.*,[17] we suspect that this may be an experimental artifact. The GaN films may have appreciable surface roughness, which scatters light, resulting in decreased intensity in reflectivity and ellipsometry measurements. In Wethkamp *et al.*, it is explicitly demonstrated that increased surface roughness

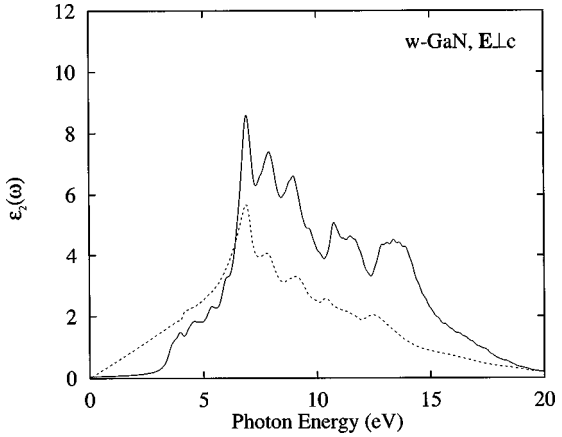

FIG. 3. Calculated $\epsilon_2(\omega)$ for **E** perpendicular to c for wurtzite GaN including the electron-hole interaction (solid line), and measured $\epsilon_2(\omega)$ from Ref. 17.

gives rise to a decreased magnitude for the measured $\epsilon_2(\omega)$.[19] We stress that our previous calculations of $\epsilon_2(\omega)$ including the electron-hole interaction for diamond, Si, Ge, and GaAs (Ref. 5) gave results in excellent agreement with experiments, both for peak positions and average peak heights.

Figure 4 shows $\epsilon_2(\omega)$ for **E** parallel to c for wurtzite GaN. The solid line includes the electron-hole interaction, and the dashed is one-electron theory. The major absorptive feature is a large peak at 6.9 eV. This is to be compared with the three distinct peaks of lower intensity between 6 eV and 10 eV in the $\mathbf{E}\perp c$ case. Similar polarization-dependent spectral features have been observed in other wurtzite structure semiconductors, such as CdS and CdSe.[26] To the best of our knowledge, the UV optical properties of wurtzite GaN with $\mathbf{E}||c$ have yet to be measured.

2. Zinc blende

We perform calculations at the experimental lattice constant $a=4.50$ Å.[22] Conduction bands are shifted up to move the band gap (direct, and at Γ) from the LDA value (2.1 eV) to the GW result of 3.1 eV.[23] The LDA and GW bandwidths are quite close, so we choose not to stretch conduction or valence bands. The IR dielectric constant used in the screened interaction is taken to be 5.5.[22] We use $N_{\mathbf{k}}=512$ ($8\times8\times8$ mesh), $N_{\mathbf{x}}=27$ ($3\times3\times3$ mesh), $N_{\mathrm{v}}=4$, $N_{\mathrm{c}}=8$, $N_{\mathrm{iter}}=150$, and $\eta=0.2$ eV.[24]

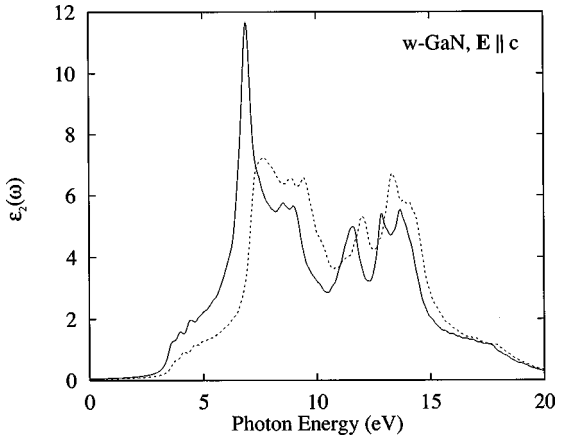

FIG. 4. Calculated $\epsilon_2(\omega)$ for **E** parallel to c for wurtzite GaN. The solid line includes the electron-hole interaction, while the dashed line is the result of one-electron theory.

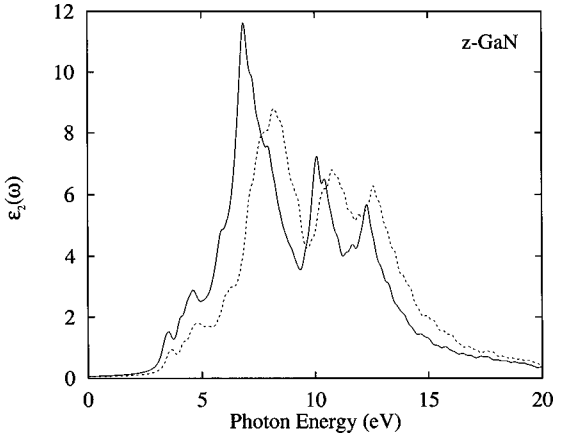

FIG. 5. Calculated $\epsilon_2(\omega)$ for zinc-blende GaN. The solid line includes the electron-hole interaction, while the dashed line is the result of one-electron theory.

Figure 5 shows the calculated $\epsilon_2(\omega)$ for zinc-blende GaN including the electron-hole interaction (solid line), and neglecting it (dashed line). As for wurtzite GaN with $\mathbf{E}||c$, the result of the interacting theory shows a large peak at ≈7 eV. Note that the electron-hole interaction shifts the peak positions as well as the peak heights. Noninteracting peaks at 8.2 eV, 10.8 eV, and 12.6 eV are shifted to 6.8 eV, 10.2 eV, and 12.3 eV. The peak heights decrease with increasing energy in both noninteracting and interacting results, but the decrease is more pronounced in the interacting case. The net effect of the interaction is to shift oscillator strength to lower energies, as for wurtzite GaN. The $\epsilon_2(\omega)$ for zinc-blende GaN has been measured by spectroscopic ellipsometry.[20] These authors obtain three distinct peaks at 7 eV, 10.6 eV, and 13 eV, with the 13-eV peak being lower in intensity than the first two. Although the electron-hole interaction moves the lowest-energy peak into better agreement with experiment, the other two peak positions seem to be in worse agreement. As for wurtzite GaN, the overall magnitude of the calculated $\epsilon_2(\omega)$ is larger than that of the experimental results.

With the electron-hole interaction having such an appreciable effect, it is natural to ask: What gives rise to such a large shift in oscillator strength? From Eq. (17), we see that two factors determine the spectral features: the excitation energies E_f and the transition-matrix elements $\langle f|\hat{\lambda}\cdot\mathbf{J}|0\rangle$. Using the discussion surrounding Eq. (24), we can calculate the total density of excited states, $D(\omega)=\Sigma_f\delta(\omega-E_f)$, with and without the electron-hole interaction. Figure 6 shows the results for zinc-blende GaN. Interacting and noninteracting $D(\omega)$ are almost identical.[27] Thus, the excitation energies are essentially unchanged. This is reasonable, for the scale of the electron-hole interaction is set by the exciton binding energy in the material. For GaN, this is on the order of tens of meV,[28] which is small on the scale of many eV. The difference between interacting and noninteracting $\epsilon_2(\omega)$ is therefore due to the transition-matrix elements. Although the electron-hole interaction does not change the electron-hole pair energy by much, the pair wave function is changed

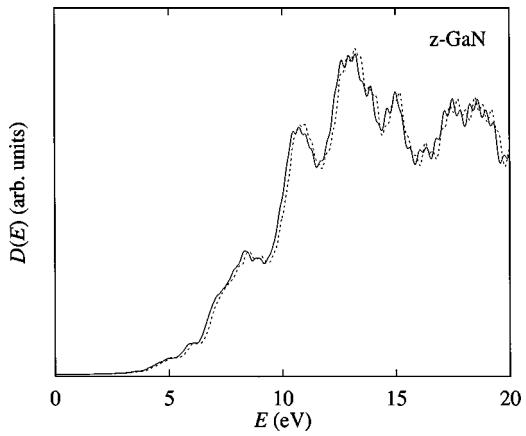

FIG. 6. Total density of (singly) excited states for zinc-blende GaN. The solid line includes the electron-hole interaction, while the dashed line is the result of one-electron theory.

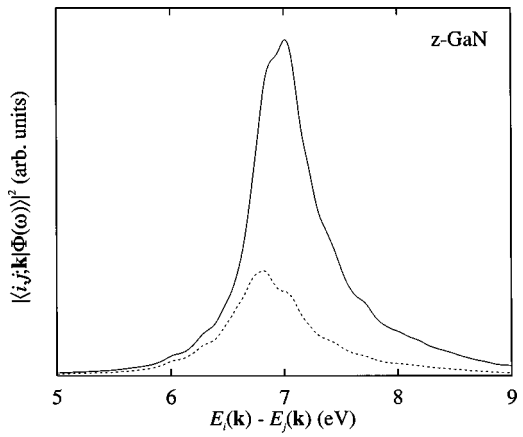

FIG. 7. $|\langle i,j;\mathbf{k}|\Phi(\hbar\omega=6.7\ \text{eV})\rangle|^2$ vs $E_i(\mathbf{k})-E_j(\mathbf{k})$ for zinc-blende GaN for noninteracting (dashed) and interacting (solid) cases. Each $|\langle i,j;\mathbf{k}|\Phi(\hbar\omega=6.7\ \text{eV})\rangle|^2$ has been broadened by a Lorentzian with a full width at half-maximum of 0.2 eV so that a smooth curve is obtained.

enough to affect these matrix elements a great deal. This subtle but important effect is responsible for much of the discrepancy between one-electron theories of optical properties of semiconductors and experiments.[3–6] More will be said about the origin of this effect (as pertaining to insulators) in the next subsection.

To get insight into the origins of peaks in $\epsilon_2(\omega)$, we can use the results of Sec. II E to decompose $\epsilon_2(\omega)$ into individual interband transitions. We focus on the first major peak in zinc-blende GaN. Choosing $N_{\mathbf{k}}=512$ ($8\times8\times8$ mesh), $N_{\mathbf{x}}=27$ ($3\times3\times3$ mesh), $N_{\mathrm{v}}=4$, $N_{\mathrm{c}}=8$, and $\eta=0.2$ eV, $|\Phi(\hbar\omega=6.7\ \text{eV})\rangle$ is computed. By analyzing its components in the $(i,j;\mathbf{k})$ basis, we find that $\epsilon_2(\hbar\omega=6.7\ \text{eV})$ is composed primarily of transitions between bands 3,4 (admixture of Ga 4p and N 2p) $\rightarrow$ band 5 (admixture of Ga 4s and N 2s). Since excitation energies are almost unchanged by the electron-hole interaction, this conclusion could have been reached by decomposing the noninteracting $\epsilon_2(\omega)$.[17] However, valuable insight into the effects of the interaction can be obtained. Figure 7 shows $|\langle i,j;\mathbf{k}|\Phi(\hbar\omega=6.7\ \text{eV})\rangle|^2$ plotted against the *noninteracting* excitation energy, $E_i(\mathbf{k})-E_j(\mathbf{k})$, for noninteracting (dashed) and interacting (solid) cases. Each ordinate in the plots represents the square of a particular $(i,j;\mathbf{k})$ component of $|\Phi(\omega)\rangle$, i.e., a term in the sum of Eq. (26). The nonzero width of the distribution for the noninteracting case is due to the nonzero η. Note that for the interacting case, weight is shifted to higher *noninteracting* energy. Thus, states of higher noninteracting energy are used to create states that take better advantage of the attractive electron-hole interaction. This is perfectly reasonable, for the noninteracting energy plays the role of the kinetic energy of the electron-hole pair in the effective-mass approximation. Because kinetic energy must increase in order for the pair wave function to localize, the presence of the interaction leads to an increase in this noninteracting excitation energy.

B. CaF_2

CaF_2 is a highly ionic material that crystallizes in a cubic structure with one Ca and two F atoms per cell (called the calcium fluoride structure). Because of its ionicity, the band gap is large [$\approx$11.8 eV (Ref. 29)] so its optical absorption edge is well in the UV. CaF_2 occurs naturally, so measurements of optical properties have been available for decades. One of the earliest studies of the UV optical properties is the reflectivity measurement of R. Tousey[30] using the technique of photographic photometry. This work, together with later studies,[31] reveals the presence of an exciton at 11.2 eV. More recent work of Barth *et al.*[32] uses spectroscopic ellipsometry with light from a synchrotron source to measure $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$ directly. The exciton is not clearly resolved in their work, because the peak is in an energy range that is contaminated by second-order diffracted light. In this section, we present our calculated $\epsilon_2(\omega)$ for CaF_2 including the electron-hole interaction. Results are compared to those of the two experiments.

We perform calculations at the experimental lattice constant $a=5.46$ Å.[22] The LDA and *GW* values for the band gap are 6.77 eV and 11.38 eV, respectively.[33] Because the *GW* gap is roughly 0.4 eV smaller than the gap inferred from experiment,[29] we shift the LDA conduction bands up to agree with the experimental minimum gap of 11.8 eV. This is justified below when comparing our results to experiment. LDA valence bands are stretched by 18% (conduction bands are unstretched) to agree with *GW* bandwidths. The IR dielectric constant used in the screened interaction is taken to be 2.03.[22] We use $N_{\mathbf{k}}=512$ ($8\times8\times8$ mesh), $N_{\mathbf{x}}=64$ ($4\times4\times4$ mesh), $N_{\mathrm{v}}=8$, $N_{\mathrm{c}}=8$, $N_{\mathrm{iter}}=150$, and $\eta=0.3$ eV.[24]

Figure 8 shows the calculated $\epsilon_2(\omega)$ for CaF_2 including the electron-hole interaction (solid line), and neglecting it (dashed line). The difference between interacting and noninteracting results is far more pronounced than for GaN. This is because the electron-hole interaction is much stronger in insulating systems (due to less screening). An important manifestation of the strength of the interaction is the peak at 11.1 eV. This is an exciton peak. The large height of the peak results from the large transition-matrix element that couples the photon to a localized electronic excitation. Since the band gap is 11.8 eV, the calculated exciton binding energy is 11.8–11.1 eV=0.7 eV. Figure 9 shows our results including the interaction (solid line), together with $\epsilon_2(\omega)$ deduced from the reflectivity measurement of *R* Tousey.[30]

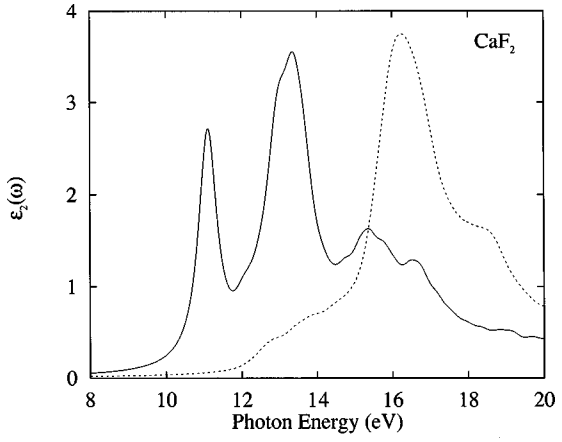

FIG. 8. Calculated $\epsilon_2(\omega)$ for CaF_2. The solid line includes the electron-hole interaction, while the dashed line is the result of one-electron theory.

The agreement is good overall, despite the fact that there are very few experimental points. One notable discrepancy is the amount of oscillator strength in the exciton peak. The area under the peak is larger in the experiment than it is in the theoretical results. This has been observed before for LiF.[5] It should be noted that our model for the screened electron-hole interaction is expected to break down somewhat for small electron-hole separations in inhomogeneous systems. The exciton in CaF_2 has charge-transfer character, so the average electron-hole separation should be on the order of a few bond lengths. We suspect that using the full RPA screening should lead to better agreement. The calculated exciton peak position, 11.1 eV, is very close to the observed peak position of 11.2 eV.[30,31] We stress that this close correspondence between theory and experiment is facilitated by the use of the experimentally determined band gap. If our *GW* gap is used, the exciton peak position is 10.7 eV, 0.5 eV lower than the experimental result. Figure 10 shows our calculated results (solid line) together with the spectroscopic ellipsometry data of Barth *et al.*[32] Theory and experiment do not agree well below ~ 14 eV, but seem to agree at higher energies. Note that the one-electron results shown in Fig. 8 do not even agree qualitatively with either experiment.

Because the electron-hole interaction in CaF_2 is large enough to lead to a substantial exciton peak, the spatial distribution of the electron and hole in this spectral region is of interest. Using $N_{\mathbf{k}}=512$ ($8\times8\times8$ mesh), $N_{\mathbf{x}}=64$ ($4\times4\times4$ mesh), $N_v=8$, $N_c=8$, and $\eta=0.3$ eV, $|\Phi(\omega)\rangle$ is calculated in the $(i,j;\mathbf{k})$ basis. Eq. (6) is then used to transform $|\Phi(\omega)\rangle$ into the $(\mathbf{x},\mathbf{y};\mathbf{R})$ basis, from which we determine an average electron-hole separation:

$$d(\omega)=\langle\Phi(\omega)||\mathbf{x}-\mathbf{y}+\mathbf{R}||\Phi(\omega)\rangle.$$

We find that $d(\hbar\omega=11.1\ \text{eV})=4.0$ Å. This is significantly larger than the Ca-F nearest-neighbor distance of 2.4 Å, and is very close to the Ca-Ca nearest-neighbor distance of 3.9 Å. However, it is still somewhat smaller than a typical electron-hole separation for an excitation in the continuum; we find $d(\hbar\omega=13.0\ \text{eV})=4.7$ Å.[34] It should be kept in mind that $|\Phi(\omega)\rangle$ is not an electron-hole pair eigenstate, but is a pair state created by the perturbing radiation field. Eigenstates with energies in the continuum are expected to be very delocalized. Another point of interest is that while continuum eigenstates are relatively unaffected by the electron-hole interaction, continuum $|\Phi(\omega)\rangle$ pair states are very affected [the noninteracting electron-hole separation, $d_0(\hbar\omega=13.0\ \text{eV})$, equals 7.7 Å as compared to 4.7 Å for $d(\hbar\omega=13.0\ \text{eV})$].[34] This is to be expected, because $\epsilon_2(\omega)=\langle\Phi(\omega)|\Phi(\omega)\rangle$ is strongly affected by the interaction at all ω. Finally, we note that the magnitude of $d(\omega)$ for all ω is largely governed by the localized nature of the state $|P\rangle$ [see Eq. (25)]. We find that $\langle P||\mathbf{x}-\mathbf{y}+\mathbf{R}||P\rangle=2.75$ Å, only slightly greater than the Ca-F nearest-neighbor distance.

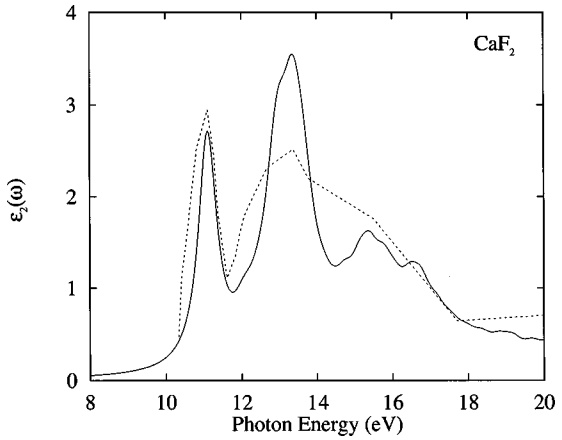

FIG. 9. Calculated $\epsilon_2(\omega)$ for CaF_2 including the electron-hole interaction (solid line), and measured $\epsilon_2(\omega)$ from Ref. 30.

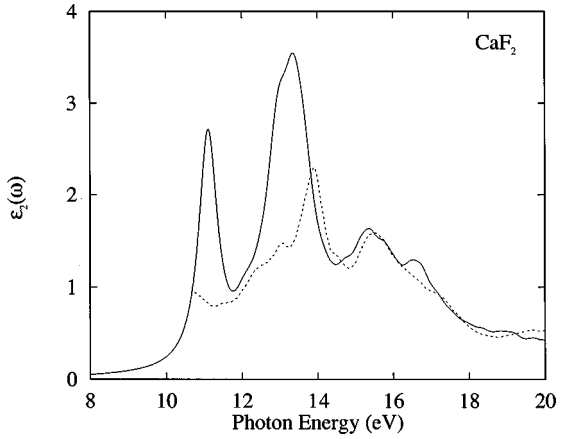

FIG. 10. Calculated $\epsilon_2(\omega)$ for CaF_2 including the electron-hole interaction (solid line), and measured $\epsilon_2(\omega)$ from Ref. 32.

IV. CONCLUSIONS

We have presented a computationally efficient scheme for calculating $\epsilon_2(\omega)$ for crystalline semiconductors and insulators, which includes the electron-hole interaction. The efficiency results from an iterative method of computing the spectrum, which does not require the calculation of energies or wave functions. Each iteration involves an action of the effective Hamiltonian (including the electron-hole interaction) on an electron-hole pair state. Actions are simplified by representing these states in bases that diagonalize different terms in the Hamiltonian.

The scheme has been applied to the wide-gap semiconductor GaN (both wurtzite and zinc blende), and the ionic

insulator CaF_2. In both cases, the inclusion of the electron-hole interaction greatly improves the agreement between theory and the available experiments. By focusing on the discrepancies, we are led to suggest that optical experiments on GaN may be strongly affected by surface roughness and/or noncrystallinity. Also, we suspect that the Hybertsen-Levine-Louie model for screening the electron-hole interaction suffers from inaccuracies when applied to very insulating materials. Finally, our results support the experimentally inferred band gap of 11.8 eV for CaF_2.

Note added in proof: We are now aware of more recent ellipsometry measurements on GaN (wurtzite and zinc blende) which show a larger overall magnitude for the dielectric function, closer to our theoretical results (see Ref. 35).

ACKNOWLEDGMENTS

We thank M. Rohlfing, S.G. Louie, Z.H. Levine, B. Segall, A.J. Fischer, T. Wethkamp, R. Gupta, R.B. Bohn, C.W. Clark, and B. Schneider for helpful discussions. Many thanks go to J. Rife for providing us with the experimental data appearing in Fig. 3. We extend special thanks to W.R.L. Lambrecht and M. Cardona for their interest and insight during many stages of this work.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).

[2] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **34**, 5390 (1986); R.W. Godby, M. Schluter, and L.J. Sham, *ibid.* **37**, 10 159 (1988); S.B. Zhang, D. Tomanek, M.L. Cohen, S.G. Louie, and M.S. Hybertsen, *ibid.* **40**, 3162 (1989); M. Rohlfing, P. Kruger, and J. Pollmann, *ibid.* **48**, 17 791 (1993); E.L. Shirley, X. Zhu, and S.G. Louie, *ibid.* **56**, 6648 (1997).

[3] L.J. Sham and T.M. Rice, Phys. Rev. **144**, 708 (1965); W. Hanke and L.J. Sham, Phys. Rev. Lett. **33**, 582 (1974); W. Hanke and L.J. Sham, Phys. Rev. B **21**, 4656 (1980).

[4] M. Rohlfing and S.G. Louie, Phys. Rev. Lett. **80**, 3320 (1998); M. Rohlfing and S.G. Louie, *ibid.* **81**, 2312 (1998).

[5] L.X. Benedict, E.L. Shirley, and R.B. Bohn, Phys. Rev. B **57**, 9385 (1998); L.X. Benedict, E.L. Shirley, and R.B. Bohn, Phys. Rev. Lett. **80**, 4514 (1998).

[6] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, Phys. Rev. Lett. **80**, 4510 (1998).

[7] E. L. Shirley, Phys. Rev. B **54**, 16 464 (1996).

[8] By ''IR dielectric constant,'' we mean $\epsilon_1(\omega_0)$, where ω_0 is much less than an electronic transition frequency, but much greater than a typical phonon frequency. This is often refered to as ϵ_∞.

[9] Z.H. Levine and S.G. Louie, Phys. Rev. B **25**, 6310 (1982).

[10] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **37**, 2733 (1988).

[11] $\mathbf{J}$ is current per area, so $H_{eff}^{-1}\hat{\lambda}\cdot\mathbf{J}$ has units of dipole moment per volume (polarization) if $\hbar$ is set to unity.

[12] Z.H. Levine and D.C. Allan, Phys. Rev. Lett. **63**, 1719 (1989).

[13] R. Haydock, Comput. Phys. Commun. **20**, 11 (1980).

[14] It should be noted that this is not the scheme currently recommended by R. Haydock for use in multiband systems. However, we have found that for all cases tested, our converged results are essentially identical to those of exact diagonalization. We thank R. Haydock and J.D. Joannopoulos for their cautionary remarks in this regard.

[15] Random phases were employed to calculate the average density of one-electron states for a disordered two-dimensional system in R. Haerle, R. Haydock, and R.L. Te, Comput. Phys. Commun. **90**, 81 (1995).

[16] Y. Saad and M.H. Schultz, SIAM (Soc. Ind. Appl. Math) J. Sci. Stat. Comput. **7**, 856 (1986).

[17] W.R.L. Lambrecht, B. Segall, J. Rife, W.R. Hunter, and D.K. Wickenden, Phys. Rev. B **51**, 13 516 (1995). Measurements were performed at room temperature.

[18] S. Logothetidis, J. Petalas, M. Cardona, and T.D. Moustakas, Phys. Rev. B **50**, 18 017 (1994).

[19] T. Wethkamp, K. Wilmers, N. Esser, W. Richter, O. Ambacher, H. Angerer, G. Jungk, R.L. Johnson, and M. Cardona, Thin Solid Films **313-314**, 745 (1998).

[20] C. Janowitz, M. Cardona, R.L. Johnson, T. Cheng, T. Foxon, O. Gunther, and G. Jungk, BESSY Jahresbericht Report No. 230, 1994 (unpublished).

[21] N.E. Christensen and I. Gorczyca, Phys. Rev. B **50**, 4397 (1994).

[22] *Numerical Data and Functional Relationships in Science and Technology*, edited by K.H. Hellwege, Landolt-Bornstein Tables (Springer, New York, 1982).

[23] A. Rubio, J.L. Corkill, M.L. Cohen, E.L. Shirley, and S.G. Louie, Phys. Rev. B **48**, 11 810 (1993).

[24] For all calculations we use a $\mathbf{k}$ point shift, so that the Γ point ($\mathbf{k}=0$) is shifted to $\mathbf{k}=\frac{1}{64}\mathbf{b}_1+\frac{1}{32}\mathbf{b}_2+\frac{3}{64}\mathbf{b}_3$. Such symmetry breaking is not problematic, because it becomes irrelevant when results are converged.

[25] On close inspection, it is apparent that the peak positions in the theoretical results are slightly off from the measured positions. This may be related to the use of a simple $\mathbf{k}$-independent shift and stretch of the LDA bands. Use of the full *GW* band energies may lead to better agreement.

[26] M. Cardona, Solid State Commun. **1**, 109 (1963); M. Cardona, J. Appl. Phys. **36**, 2181 (1965).

[27] M. Rohlfing and S.G. Louie have obtained similar results for the total density of excited states of GaAs with and without the electron-hole interaction, M. Rohlfing and S.G. Louie (unpublished).

[28] A.J. Fischer, W. Shan, J.J. Song, Y.C. Chang, R. Horning, and B. Goldenberg, Appl. Phys. Lett. **71**, 1981 (1997).

[29] 11.8 eV is the indirect gap inferred from the 12.1-eV direct gap (at room temperature) proposed in G.W. Rubloff, Phys. Rev. B **5**, 662 (1972).

[30] R. Tousey, Phys. Rev. **50**, 1057 (1936). No mention of temperature appears here; we assume the measurements were performed at room temperature.

[31] T. Tomiki and T. Miyata, J. Phys. Soc. Jpn. **27**, 658 (1969). Measurements were performed at room temperature.

[32] J. Barth, R.L. Johnson, M. Cardona, D. Fuchs, and A.M. Bradshaw, Phys. Rev. B **41**, 3291 (1990). The data we reproduce here (Fig. 10) are from a room-temperature measurement. This paper also presents data taken at 90 K; however, the differences between the 90 K and 300 K data are significantly smaller than the differences between either and our theoretical results.

[33] E.L. Shirley, Phys. Rev. B **58**, 9579 (1998).

[34] Note that the value of $d(\omega)$ for $\hbar\omega$ in the continuum (above the band gap) is sensitive to the choice of η.

[35] T. Wethkamp and M. Cardona (unpublished).