# First-order corrections to random-phase approximation *GW* calculations in silicon and diamond

R. T. M. Ummels, P. A. Bobbert, and W. van Haeringen
*Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*
(Received 21 November 1997)

We report on *ab initio* calculations of the first-order corrections in the screened interaction $W$ to the random-phase approximation polarizability and to the $GW$ self-energy, using a noninteracting Green's function, for silicon and diamond. It is found that the first-order vertex and self-consistency corrections to the polarizability largely compensate each other. This does not hold, however, for the first-order corrections to the $GW$ gap. For silicon the compensation between the first-order vertex and self-consistency correction contributions to the gap is only about 35%, while for diamond it is even absent. The resulting gap values are significantly and systematically too large, the direct gaps for silicon and diamond being 0.4 eV and 0.7 eV larger than their $GW$ values, respectively. The success of $GW$ in predicting electronic properties of, e.g., silicon and diamond can therefore apparently not be understood in terms of ''small'' corrections to $GW$ to first order in $W$ using a noninteracting Green's function. [S0163-1829(98)01819-0]

## I. INTRODUCTION

One of the most successful methods of describing exchange-correlation effects in *ab initio* calculations is the random-phase approximation (RPA) $GW$ approach, where the (irreducible) polarizability $P$ is calculated in the RPA and the self-energy $\Sigma$ is calculated to first order in the dynamically screened interaction $W$.[1,2] It is remarkable in this connection that non-self-consistent RPA $GW$ calculations lead to quasiparticle (QP) band gaps that are in excellent agreement with experiment in the case of a large group of semiconductors if the starting point is chosen to be the density-functional theory (DFT) in the local-density approximation (LDA). Henceforth we will call this non-self-consistent RPA $GW$ approach ''standard $GW$.'' The inherent assumption that higher-order corrections can be neglected is, however, far from obvious.

The inclusion of vertex ($V$) and so-called self-consistency (SC) corrections to $\Sigma$ and $P$ has been studied by a number of researchers, mainly for the homogeneous electron gas. Hubbard[3] introduced the corrections to the RPA by means of a local-field factor (not to be confused with the so-called local-field effects to be introduced later on). In the electron gas case much effort has gone into obtaining expressions for local-field factors in the dielectric function; see, for instance, Refs. 2 and 4. DuBois,[5] whose work can be considered as an extension to the work of Gell-Mann and Brueckner,[6,7] stresses the importance of taking into account all polarizability diagrams of the same order in the Wigner-Seitz radius $r_s$. He noticed the significant cancellation between $V$ and SC corrections in the high-density limit. Geldart and Taylor[8] found a similar compensation for the static polarizability. They attempted to construct a local-field factor that includes SC corrections. Mahan and Sernelius,[9] using the local-field factor approach, concluded that the effects of $V$ corrections to the bandwidth of the homogeneous electron gas nearly cancel when added to both the self-energy and the polarizability, as was already predicted by Rice.[10]

In a recent paper of Shirley[11] a self-consistent $GW$ (SCGW) calculation is reported on for the homogeneous electron gas, in which, however, a fixed screened interaction is employed. The resulting bandwidth is found to be appreciably larger than the (assumed) more correct standard $GW$ value. A similar result is obtained by von Barth and Holm,[12] who performed a full SCGW calculation by including the screened interaction in the self-consistency procedure. Shirley[11] furthermore reports on a nearly self-consistent calculation to second order in the screened interaction $W$, i.e., by incorporating, apart from the $GW$ diagram, also the first-order vertex correction diagram to $GW$. The employed Green's function $G$ in this calculation is the one obtained from his SCGW calculation, however. The resulting bandwidth in this latter calculation is found to be very close to the standard $GW$ value. This is indicative of a strong compensation between $V$ and SC corrections, fully in line with arguments put forth in Refs. 13 and 14.

Calculations for inhomogeneous systems are much more difficult to perform, which explains that the situation concerning corrections to standard $GW$ is much less settled for such systems. In Ref. 15 a full SCGW calculation was presented for the relatively simple case of a quasi-one-dimensional semiconducting wire. It resulted in a value for the band gap that is large compared to the standard $GW$ band-gap value. The obtained result appears to be completely at variance with the much smaller band-gap obtained in the quantum Monte Carlo (QMC) calculation for this system presented in Refs. 16 and 17. The QMC value should for reasons of principle lie close to the ''exact'' band-gap value; it appears to be close to the standard $GW$ value. The apparent difference between the SCGW and QMC results strongly points to the need of including vertex corrections in the former type of calculation, like the SCGW result for the electron gas in Refs. 11 and 12. Indeed, in Ref. 18 it was found that the band gap is much closer to the standard $GW$ band gap and the QMC band gap if all first-order corrections to *both* the RPA polarizability and the $GW$ self-energy are systematically included.

As far as calculations on real semiconductors are concerned, we note that Hanke and Sham[19] included $V$ corrections to the RPA polarizability for a covalent crystal (dia-

 

mond), using the bare Coulomb interaction instead of the dynamically screened interaction. Concerning the importance of vertex corrections to the self-energy, Daling and van Haeringen,[20] Daling *et al.*,[21] and Bobbert and van Haeringen[22] conclude that the effect of the first-order vertex correction to the standard $GW$ self-energy on the direct band gap at the $\Gamma$ point of silicon is relatively small. The former results[20,21] were obtained by using the bare Coulomb interaction, the latter[22] by using the dynamically screened interaction. Del Sole, Reining, and Godby[23] have arrived at a similar conclusion for silicon on the basis of a so-called $GW\Gamma$ calculation, which incorporates in an approximate way vertex corrections to the self-energy as well as to the polarizability, by means of the functional derivative of the exchange-correlation potential in DFT with respect to the density. Bechstedt *et al.*[24] have shown that dynamical effects due to vertex corrections and self-consistency corrections to the strength of the optical absorption (and correspondingly $P$) largely cancel for silicon and diamond.

The present work can be seen as an extension of Ref. 18 and also of Ref. 22. The aim is to contribute further to the understanding of the success of standard $GW$. In considering possible improvements to standard $GW$, Hedin[1] argues that one should preferably take $\Sigma$ to $n$th order in $W$ if $P$ is taken to order $n-1$ in $W$, provided both $\Sigma$ and $P$ are taken self-consistently. In view of the complexity of dealing with self-consistency and higher-order corrections, we will restrict ourselves to the investigation of the effect of including the first-order in $W$ corrections to standard $GW$ as well as to the RPA $P$. Not a self-consistent Green's function, but the LDA Green's function will be employed in this investigation. Two aspects concerning the screened interaction $W$ are essential in an accurate evaluation of QP energies for semiconductors: first, its energy dependence, and, second, the off-diagonal matrix elements of $P$ in the employed plane-wave basis set, giving rise to the so-called local-field effects (LFE's). Both dynamical screening and LFE's will be included in our calculations. The energy dependence of the dynamically screened interaction will be modeled by means of the plasmon pole model (PPM) of Engel and Farid.[25] We will consider three kinds of calculations: (i) calculation of the first-order $V+\mathrm{SC}$ correction to the RPA polarizability using the RPA screening, (ii) calculation of the first-order $V+\mathrm{SC}$ correction to the $GW$ self-energy using the RPA screening, and (iii) calculation of the $GW$ self-energy and its first-order corrections using the corrected screening from (i). The calculation of these corrections to standard $GW$ is done for both silicon and diamond, with emphasis on the energy levels around the band gap.

The paper is organized as follows. In Sec. II we will discuss the diagrams that have to be taken into account for $P$ and $\Sigma$ and we will give a short description of some calculational details. In Sec. III we will give the results. In Sec. III A we will focus on the polarizability. In Sec. III B results for the self-energy are given, concentrating on energy levels around the band gap. A few checks are carried through in Sec. IV. Section V contains a further discussion concerning the SC self-energy correction. In Sec. VI we report on a remarkable cancellation between a particular group of second-order corrections to the LDA energy gap. Section VII is devoted to the discussion of our results.

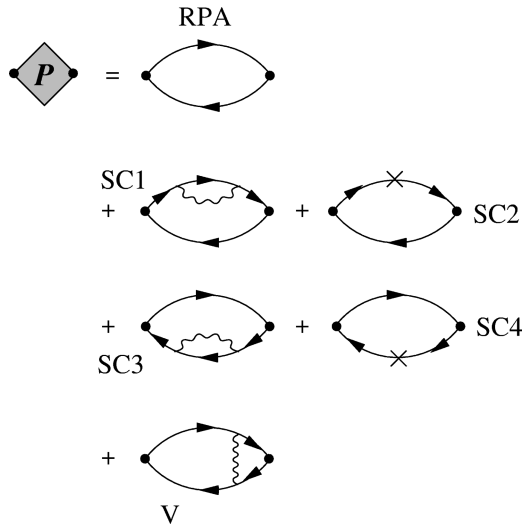


FIG. 1. RPA polarizability plus the first-order corrections to it. SC1–SC4 denote the first-order self-consistency corrections. $V$ denotes the first-order vertex correction. The solid directed line denotes the LDA Green's function. The cross denotes $-V^{\mathrm{xc}}$. The wiggly line denotes the RPA dynamically screened interaction.

## II. THEORY

The RPA polarizability diagram as well as its first-order correction Feynman diagrams are depicted in Fig. 1. The $GW$ self-energy diagram as well as its first-order correction Feynman diagrams are depicted in Fig. 2. Diagrams SC1–SC4 are the first-order self-consistency diagrams since they have self-energy insertions in the Green's functions, taking into account self-consistency effects to first order. The cross in diagrams SC2 and SC4 denotes minus the LDA exchange-correlation potential $-V^{\mathrm{xc}}$. These latter diagrams should be included when the LDA is the starting point because $V^{\mathrm{xc}}$ can be considered as the self-energy in the LDA, which should be canceled out. Diagrams SC3 and SC4 of Fig. 2 are self-energy corrections due to the first-order corrections to the valence charge density. We will henceforth call these latter diagrams SC Hartree diagrams. Diagram $V$ is the first-order vertex correction diagram.

In the evaluation of the corrections to the wave-vector-dependent and

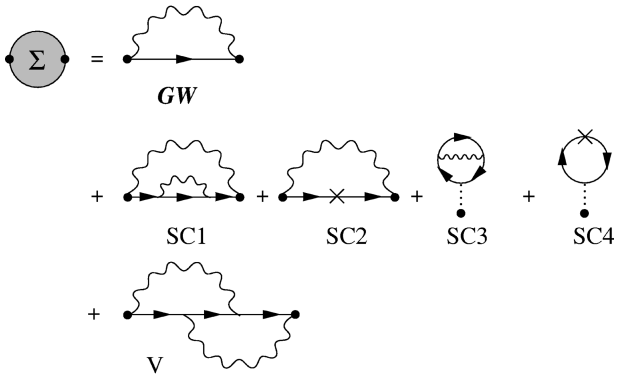


FIG. 2. $GW$ self-energy plus the first-order corrections $\Sigma^{V+\mathrm{SC}}$. The dotted line denotes the bare Coulomb interaction. The wiggly line now denotes the dynamically screened interaction obtained with either the RPA or the RPA$+V+$SC polarizability; see Fig. 1.

energy-dependent polarizability $P_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega)$ and self-energy $\Sigma_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega)$ in a plane-wave basis ($\mathbf{G}$ and $\mathbf{G}'$ are reciprocal lattice vectors, $\mathbf{k}$ is the wave vector, and $\omega$ is the energy) one wave-vector integration can be reduced to an integration over the $\mathbf{k}$-dependent irreducible wedge $\mathcal{I}_{\mathbf{k}}$. The second wave-vector integration, if present, cannot be reduced and has to be performed over the whole first Brillouin zone (1BZ). Furthermore, in our calculations, wave-vector integrations have been replaced by a discrete sum over the wave vectors of the grid

$$\mathbf{q}=(n_1\mathbf{b}_1+n_2\mathbf{b}_2+n_3\mathbf{b}_3)/2N_{\mathrm{gr}} \quad (n_i=-N_{\mathrm{gr}}+1,\ldots,N_{\mathrm{gr}}), \tag{1}$$

with $\mathbf{q}$ reduced to the 1BZ if necessary, where $\mathbf{b}_i$ are the primitive vectors of the reciprocal lattice. A specific $\mathbf{q}$-point set will be identified by giving the number $N_{\mathrm{gr}}$. The integrand pertaining to a particular correction may have a singularity if the wave vector in an interaction line goes to zero. Such singularities are all integrable and are handled analytically in a way described in Appendix B of Ref. 20.

The energy-dependent screened interaction $W_{\mathbf{G},\mathbf{G}'}(\mathbf{q};\omega)$ can be written as (shorthand notation)

$$W=v+v\chi v\equiv v+W^{\mathrm{scr}}, \tag{2}$$

where $\chi$ is the full polarizability matrix and $v$ is the bare Coulomb interaction, $v_{\mathbf{G},\mathbf{G}'}(\mathbf{q})=e^2\delta_{\mathbf{G},\mathbf{G}'}/(\epsilon_0|\mathbf{q}+\mathbf{G}|^2)$. We use SI units: $e$ is the electron charge and $\epsilon_0$ is the vacuum permittivity. The full polarizability $\chi$ is related to the (irreducible) polarizability $P$ by

$$\chi=P(I-vP)^{-1}, \tag{3}$$

where $I$ denotes the unit matrix. We will use a representation for the screening part of $W$, $W^{\mathrm{scr}}$, that is analogous to the Lehmann representation for the noninteracting (LDA) Green's function $G^0$:

$$G^0_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega)=\hbar\sum_l \frac{d_{l,\mathbf{k}}(\mathbf{G})d^*_{l,\mathbf{k}}(\mathbf{G}')}{\omega-\varepsilon_l(\mathbf{k})+i\eta\ \mathrm{sgn}[\varepsilon_l(\mathbf{k})-\mu]}, \tag{4}$$

where the infinitesimally small positive energy $\eta$ ensures the correct causal behavior, $\mu$ is the chemical potential, which in the case of a semiconductor is situated in the energy gap region, and $\varepsilon_l$ and $d_l$ are the energies and plane-wave coefficients, respectively, of the starting point wave functions, for instance, obtained within the LDA. Engel and Farid[25] developed a PPM that provides an analytical approximation for the energy dependence of the dynamically screened interaction $W^{\mathrm{scr}}$, such that energy integrals occurring in expressions for the self-energy and the polarizability can be carried out analytically. The wave-vector-dependent plasmon energies $\omega_m$ in this PPM are obtained from the generalized eigenvalue problem[25] (in matrix notation)

$$\underline{\underline{\chi}}(\mathbf{q};\omega=0)\mathbf{x}_{m,\mathbf{q}}=-\frac{1}{\omega_m^2(\mathbf{q})}\underline{\underline{M}}(\mathbf{q})\mathbf{x}_{m,\mathbf{q}}. \tag{5}$$

If the eigenvectors $\mathbf{x}_{m,\mathbf{q}}$ of Eq. (5), with components $x_{m,\mathbf{q}}(\mathbf{G})$, are normalized according to $\mathbf{x}^\dagger_{m,\mathbf{q}}\underline{\underline{M}}(\mathbf{q})\mathbf{x}_{n,\mathbf{q}}=\delta_{m,n}$ and satisfy the completeness relation $\Sigma_m\mathbf{x}_{m,\mathbf{q}}\mathbf{x}^\dagger_{m,\mathbf{q}}=\underline{\underline{M}}^{-1}(\mathbf{q})$, then the following plasmon pole description for $W^{\mathrm{scr}}$ can be derived:[25]

$$W^{\mathrm{scr}}_{\mathbf{G},\mathbf{G}'}(\mathbf{q};\omega)=\sum_m w_{m,\mathbf{q}}(\mathbf{G})w^*_{m,\mathbf{q}}(\mathbf{G}')\left\{\frac{1}{\omega-\omega_m(\mathbf{q})+i\eta}-\frac{1}{\omega+\omega_m(\mathbf{q})-i\eta}\right\}, \tag{6}$$

where $\mathbf{w}_{m,\mathbf{q}}=\underline{v}(\mathbf{q})\underline{\underline{M}}(\mathbf{q})\mathbf{x}_{m,\mathbf{q}}/\sqrt{2\omega_m(\mathbf{q})}$. Within this particular PPM $\chi$ is approximated by $\tilde{\chi}$,

$$\underline{\underline{\tilde{\chi}}}(\mathbf{q};\omega)=\{\omega^2\underline{\underline{M}}^{-1}(\mathbf{q})+\underline{\underline{\chi}}^{-1}(\mathbf{q};\omega=0)\}^{-1}. \tag{7}$$

For $\omega=0$ the model full polarizability $\tilde{\chi}$ obviously coincides with $\chi$. Further, the correct $\omega\to\infty$ limit can be obtained by inserting a properly chosen matrix $M$. In connection with the application of plasmon pole models it is desirable[26,27] to satisfy the Johnson $f$-sum rule.[28] When using a local one-electron Hamiltonian, this leads to

$$M_{\mathbf{G},\mathbf{G}'}(\mathbf{q})=\frac{\hbar^2}{m}(\mathbf{q}+\mathbf{G})\cdot(\mathbf{q}+\mathbf{G}')\rho_{\mathbf{G}-\mathbf{G}'}, \tag{8}$$

where $\rho_{\mathbf{G}}$ are Fourier components of the valence charge density. The Johnson $f$-sum rule and, accordingly, Eq. (8) are not exact, however, if the Hamiltonian contains a nonlocal ion pseudopotential. This is pointed out in Refs. 26 and 29. The resulting violation of the Johnson $f$-sum rule may very well be of importance in the evaluation of corrections to the $GW$ self-energy.[30] The correct matrix $M$ can easily be obtained by combining Eqs. (3) and (7) for $\omega\to\infty$:

$$M_{\mathbf{G},\mathbf{G}'}(\mathbf{q})=\lim_{\omega\to\infty}\omega^2P_{\mathbf{G},\mathbf{G}'}(\mathbf{q};\omega) \tag{9}$$

and extracting the $\omega\to\infty$ behavior of the polarizability $P$. In this limit the leading term of $P$ is proportional to $1/\omega^2$. In Sec. III B we will investigate the effects of the violation of the Johnson $f$-sum rule. The standard $GW$ band gap of semiconductors calculated with the PPM based on Eq. (8) is in excellent agreement with experiment for silicon and diamond; see Sec. III B.

When splitting the screened interaction $W$ into its static part $v$ and dynamic part $W^{\mathrm{scr}}$ and by taking into account each possible time order of the internal and external points of a diagram, specific subdiagrams of $\Sigma$ and $P$ can be identified. For the polarizability diagram $V$ of Fig. 1 we get a total of 30 subdiagrams. This can be seen as follows: There are 3! possible time orders if the bare Coulomb interaction $v$ is taken and 4! possible time orders if $W^{\mathrm{scr}}$ is taken. For diagrams SC1 plus SC3 of Fig. 1 there are $2\times 30$ subdiagrams, while diagrams SC2 plus SC4 lead to $2\times 3!$ subdiagrams. For the self-energy diagram $V$ of Fig. 2 there are 38 subdiagrams (for more details see Ref. 22). For diagram SC1 of Fig. 2 there are also 38 subdiagrams, for diagram SC2 $2!+3!$ subdiagrams, and for diagrams SC3 and SC4 $2!+3!$ and $2!$ subdiagrams, respectively.

In the Appendix the contribution pertaining to one specific subdiagram of the first-order vertex correction polarizability diagram, diagram $V$ in Fig. 1, is given as an example.

We want to emphasize that the head element (HE) ($\mathbf{G}=\mathbf{0}$ and $\mathbf{G}'=\mathbf{0}$) and the wing elements (WE's) ($\mathbf{G}=\mathbf{0}$ or $\mathbf{G}'=\mathbf{0}$) of the polarizability matrix $P_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega)$ have to be treated in a special way for $\mathbf{k}\rightarrow\mathbf{0}$ in the case of semiconductors. In the case of the RPA polarizability the HE has a $|\mathbf{k}|^2$ proportionality and the WE's have a linear $\mathbf{k}$ proportionality for $\mathbf{k}\rightarrow\mathbf{0}$, leading to the correct screening behavior of a semiconductor. Individually, none of the diagrams $V$ and SC in Fig. 1 has the property that the HE is proportional to $|\mathbf{k}|^2$ for $\mathbf{k}\rightarrow\mathbf{0}$. Only the sum of these diagrams fulfills this property. Kohn[31] has proved this for the case that the interaction line represents the bare Coulomb interaction, but it can also rather easily be proved for the screened interaction $W$. We have chosen to tackle the evaluation of the $V+\mathrm{SC}$ polarizability correction in the $\mathbf{k}\rightarrow\mathbf{0}$ limit numerically in the following way: The HE of the $V+\mathrm{SC}$ correction is evaluated for three small $\mathbf{k}$ vectors and then fitted according to

$$\Delta P_{\mathbf{0},\mathbf{0}}(\mathbf{k}\rightarrow\mathbf{0}):=p_0+p_1|\mathbf{k}|+p_2|\mathbf{k}|^2. \tag{10}$$

The WE's of the $V+\mathrm{SC}$ correction are evaluated for four small $\mathbf{k}$ vectors and then fitted according to

$$\Delta P_{\mathbf{0},\mathbf{G}\neq\mathbf{0}}(\mathbf{k}\rightarrow\mathbf{0}):=p_0+p^x k_x+p^y k_y+p^z k_z. \tag{11}$$

Here $p_0$, $p_1$, $p_2$, $p^x$, $p^y$, and $p^z$ are fitting parameters. In Eq. (11) the fitting parameters are $\mathbf{G}$ dependent.

Having obtained the polarizability $P$, the calculation of the dielectric matrix $\epsilon$ can easily be accomplished by evaluating $\epsilon=I-vP$. In the calculation of the dielectric matrix for $\mathbf{k}\rightarrow\mathbf{0}$ the $1/|\mathbf{k}|^2$ singularity of the HE of the bare Coulomb interaction is canceled by the $|\mathbf{k}|^2$ behavior of the HE of the polarizability. Likewise the singularity of the WE's of the bare Coulomb interaction is canceled. The macroscopic response to an applied field is determined by $\epsilon^{-1}$ rather than $\epsilon$. In accordance with Ref. 32, we define a macroscopic dielectric function (MDF) by

$$\epsilon_M(\mathbf{q}+\mathbf{G}):=\frac{1}{\{\epsilon^{-1}(\mathbf{q};\omega=0)\}_{\mathbf{G},\mathbf{G}}}. \tag{12}$$

The effects of the off-diagonal matrix elements of the dielectric function are often referred to as the local-field effects. The macroscopic or static dielectric constant $\epsilon_\infty$ is given by $\epsilon_\infty=\lim_{\mathbf{q}\rightarrow\mathbf{0}}\epsilon_M(\mathbf{q})$. For cubic crystals the static dielectric constant is independent of the direction in which the wave vector goes to zero. In the LDA RPA the static dielectric constant is generally overestimated in the case of semiconductors; see, for instance, Refs. 29 and 32.

Since the $GW$ wave functions in silicon and diamond are practically undistinguishable from the LDA wave functions, it is sufficient to calculate diagonal matrix elements of the $GW$ self-energy in the LDA basis when evaluating QP energies.[26,33] In order to account for the fact that the expectation values should be evaluated at the QP energies, the self-energy is expanded to first order in the difference between the QP and the LDA energies to obtain the desired QP energies $E_l^{\mathrm{QP}}(\mathbf{k})$, leading to

$$E_l^{\mathrm{QP}}(\mathbf{k})=\varepsilon_l(\mathbf{k})+Z_{l,\mathbf{k}}\hbar\langle l,\mathbf{k}|\Sigma(\mathbf{k};\varepsilon_l(\mathbf{k}))|l,\mathbf{k}\rangle, \tag{13}$$

where $Z_{l,\mathbf{k}}$ is the so-called wave-function renormalization factor, given by

$$Z_{l,\mathbf{k}}=\left(1-\hbar\langle l,\mathbf{k}|\frac{\partial\Sigma(\mathbf{k};\omega)}{\partial\omega}\bigg|_{\omega=\varepsilon_l(\mathbf{k})}|l,\mathbf{k}\rangle\right)^{-1}, \tag{14}$$

and $|l,\mathbf{k}\rangle$ indicates a LDA state with band index $l$ and wave vector $\mathbf{k}$. In the procedure of obtaining QP energies we will therefore evaluate both the expectation values $\hbar\langle l,\mathbf{k}|\Sigma(\mathbf{k};\varepsilon_l(\mathbf{k}))|l,\mathbf{k}\rangle$ and their derivatives $\delta$,

$$\delta_l(\mathbf{k})\equiv\hbar\langle l,\mathbf{k}|\frac{\partial\Sigma(\mathbf{k};\omega)}{\partial\omega}\bigg|_{\omega=\varepsilon_l(\mathbf{k})}|l,\mathbf{k}\rangle. \tag{15}$$

Here $\hbar\Sigma$ will equal either $-V^{\mathrm{xc}}+\hbar\Sigma^{GW}$ or $-V^{\mathrm{xc}}+\hbar\Sigma^{GW}+\hbar\Sigma^{V+\mathrm{SC}}$, depending on whether we are calculating $GW$ or $GW+V+\mathrm{SC}$ ($GW$ plus its first order in $W$ corrections) quasiparticle energies. We will refer to the above method of calculating QP energies as the ''expectation value method.'' This method turns out to work well for the $GW$ self-energy and in Sec. III B we will check its validity for the $V+\mathrm{SC}$ self-energy correction. To this end, the result obtained with the expectation value method will be compared with the result of an exact diagonalization of the nonlocal, energy-dependent Hamiltonian $H+\hbar\Sigma$.

## III. RESULTS

In the calculations to be reported on below we used energies and wave functions obtained from a well-converged self-consistent LDA calculation carried through in a plane-wave basis set with a cutoff of 17 and 45 Ry for silicon and diamond, respectively. We used the experimental lattice constants[34] $a=5.43$ Å and $a=3.57$ Å for silicon and diamond, respectively. The implemented parametrization of the *ab initio* nonlocal ionic norm-conserving pseudopotentials is that of Bachelet, Greenside, Baraff, and Schlüter.[35] The exchange-correlation potential $V^{\mathrm{xc}}$ is represented with the Wigner interpolation formula.[36] Unless indicated otherwise, the matrix $M$ of Eq. (8) is used to obtain $W^{\mathrm{scr}}$.

Three cutoffs are to be distinguished in the calculations of $P$ and $\Sigma$: (i) the number of plane waves taken into account in reciprocal lattice vector summations and used for the size of dielectric (and polarizability) matrices $N_{\mathrm{PW}}$, (ii) the number of electron and plasmon bands taken into account in band summations $N_b$, and (iii) the fineness of the $\mathbf{k}$-space grid in the Brillouin zone integrations $N_{\mathrm{gr}}$; see Sec. II. For the polarizability as well as the self-energy $N_{\mathrm{PW}}=137$ is taken for silicon and $N_{\mathrm{PW}}=229$ is taken for diamond. For the RPA polarizability and the $GW$ self-energy, $N_b$ is taken equal to $N_{\mathrm{PW}}$. For the $V+\mathrm{SC}$ polarizability correction this number is $N_b=29$ for silicon and $N_b=30$ for diamond (the choice of $N_b$ is restricted to specific values due to the degeneracy of bands, which has to be properly dealt with for $\mathbf{k}\rightarrow\mathbf{0}$). The contributions to the expectation values and their energy derivatives $\delta$ (evaluated at the LDA energies; see Sec. II) of the $V+\mathrm{SC}$ self-energy correction diagrams have been obtained by using $N_b=65$. In the calculation of the RPA polarizability for $\mathbf{k}\rightarrow\mathbf{0}$, $N_{\mathrm{gr}}=6$ for silicon and $N_{\mathrm{gr}}=4$ for diamond is used. For other $\mathbf{k}$ vectors we use $N_{\mathrm{gr}}=3$ for silicon and $N_{\mathrm{gr}}=2$ for diamond. Also, the $V+\mathrm{SC}$ polarizability correction is calculated with these latter $N_{\mathrm{gr}}$ values for all $\mathbf{k}$ vectors. This holds also for the $GW$ self-energy and its corrections. An exception to this choice of $N_{\mathrm{gr}}$ is made for the $V+\mathrm{SC}$ correction in

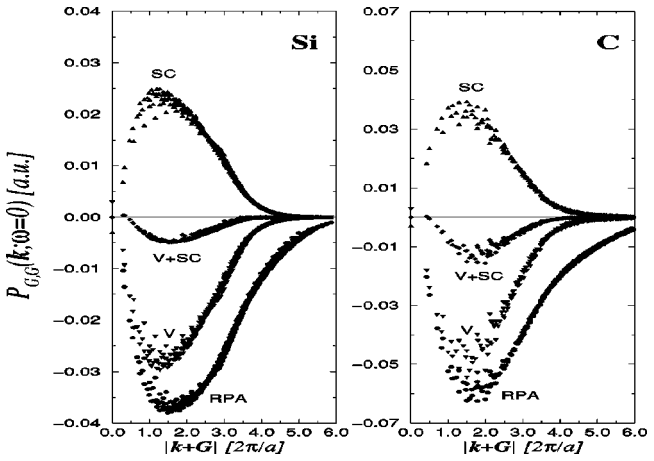


FIG. 3. Compensation between $V$ and SC corrections to the diagonal elements of the static polarizability (in Rydberg a.u.) for silicon and diamond. The constant contributions to the HE of the polarizability matrix for $\mathbf{k}\rightarrow\mathbf{0}$ are also given (the $|\mathbf{k}+\mathbf{G}|=0$ axis).

the case of the $X$ and $L$ points of silicon, where we assumed that $N_{\mathrm{gr}}=2$ would also be sufficient, as it turned out to be sufficient to use $N_{\mathrm{gr}}=2$ for the $V+\mathrm{SC}$ correction in the case of the $\Gamma$ point of silicon.

Convergence tests have been performed in order to assess the accuracy of calculated QP gaps: For silicon, the $V+\mathrm{SC}$ correction to the RPA polarizability has also been calculated by taking $N_{\mathrm{PW}}=89$ and $N_b=50$ ($N_{\mathrm{gr}}=2$). This led to the conclusion that the matrix elements of the polarizability corrections were nicely converged, while the effect on the band gap was minor. The $V+\mathrm{SC}$ correction to the $GW$ self-energy for the $\Gamma$ point of silicon and diamond, using RPA screening, has also been calculated by taking $N_{\mathrm{gr}}=2$, together with taking a smaller $N_{\mathrm{PW}}$ ($N_{\mathrm{PW}}=89$ and 169 for silicon and diamond, respectively) or with more plasmon and electron bands ($N_b=89$). For diamond the calculations with RPA screening were also done for $N_{\mathrm{gr}}=3$. We claim on the basis of the above-mentioned convergence tests that the accuracy of reported QP gaps is 0.05 eV for silicon and 0.1 eV for diamond. In Tables II and III we nevertheless express numbers with three decimal places in order to clarify possible cancellation effects.

### A. First-order corrections to the RPA polarizability

Unless stated otherwise (see Sec. III B), we use the PPM described in Sec. II based on the matrix $M$ of Eq. (8) in which case we can restrict ourselves to the calculation of $P_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega=0)$.

In Fig. 3 the RPA, $V$, and SC contributions to the diagonal elements of the static polarizability are shown as a function of the absolute wave vector $|\mathbf{k}+\mathbf{G}|$. The scattering of the points in this figure reflects the anisotropy. One observes that the $V$ and SC contributions compensate each other to a very large degree. The $V$ contribution has the same sign as the RPA polarizability, while the SC contribution has the opposite sign. In absolute value both corrections are roughly 75% of the RPA. The compensation between $V$ and SC is such that the diagonal matrix elements of the RPA$+V+$SC polarizability are in absolute value a little bit larger than the RPA ones (about 15%). Also in Ref. 18 this compensation was seen for the quasi-one-dimensional semiconducting wire, although the compensation was less complete there. In Fig. 3 the constant terms for $\mathbf{k}\rightarrow\mathbf{0}$ ($\mathbf{G}=\mathbf{0}$) for the $V$ and the SC contributions are also given. They are clearly seen to cancel. This is in agreement with Ref. 31, where it has been shown that, for insulating crystals, $P_{\mathbf{0},\mathbf{0}}(\mathbf{k};\omega)$ is proportional to $|\mathbf{k}|^2$ for small $|\mathbf{k}|$.

In Fig. 4 we have plotted the difference between the RPA$+V+$SC and the RPA MDF as a function of the absolute wave vector $|\mathbf{k}+\mathbf{G}|$. For the definition of the MDF, see Eq. (12). The correction is negative for small $|\mathbf{k}+\mathbf{G}|$ and positive for larger $|\mathbf{k}+\mathbf{G}|$. Our obtained RPA$+V+$SC static dielectric constant is $\epsilon_\infty=$ 10.4 and 5.3 in our best calculation for silicon and diamond, respectively, to be compared with our RPA values of 12.8 and 5.6 and with experimental values 11.4,[37,38] 11.7 (Ref. 39) and 5.5,[39] 5.7.[40] So by incorporating the $V+$SC polarizability correction, we find a decrease in the static dielectric constant. This is contrary to previous results in cases in which the local-density-functional formalism is used to go beyond the RPA, which generally show an increase in the static dielectric constant; see, for instance, Ref. 32. In Refs. 32 and 41, the MDF in the local-density-functional formalism was compared to the RPA MDF. In Ref. 32 the QMC exchange-correlation potential as parametrized by Perdew and Zunger[42] was used, while the Slater exchange-correlation potential was used in Ref. 41. Comparing our results for the MDF for silicon with Refs. 32 and 41, we observe that our $V+$SC correction to the RPA MDF is about five times smaller.

To conclude the discussion of the polarizability we can say that the sum of the first-order vertex and self-consistency corrections to the static polarizability is relatively small, with the LDA as the starting point. Our results confirm our previous results for the quasi-one-dimensional semiconducting wire and results of other authors that the $V$ and SC corrections to the polarizability compensate each other to a large

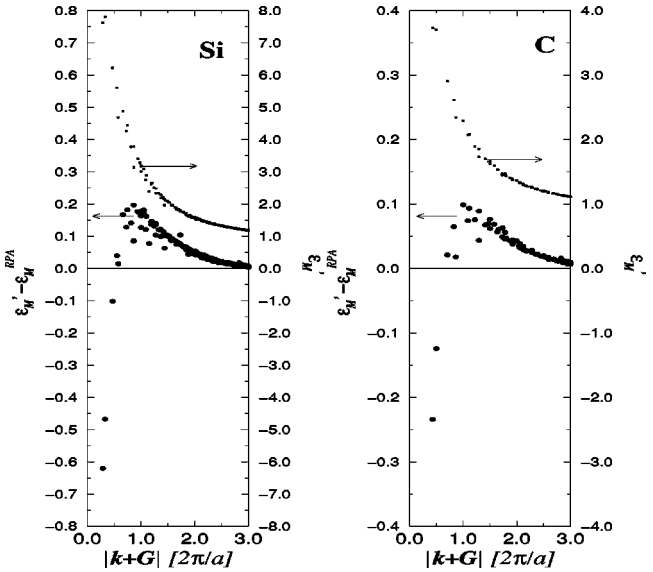


FIG. 4. RPA$+V+$SC macroscopic dielectric function (MDF) as well as the difference between the RPA$+V+$SC and the RPA MDF for silicon and diamond. The definition of the MDF is given in Eq. (12). By $\epsilon_M'$ we mean $\epsilon_M^{\mathrm{RPA}+V+\mathrm{SC}}$. The value of $\epsilon_M'-\epsilon_M^{\mathrm{RPA}}$ for $|\mathbf{k}+\mathbf{G}|\rightarrow 0$ is $-2.4$ for silicon and $-0.3$ for diamond (see the text).

TABLE I. $GW$ and $GW+V+$SC direct gaps (in eV) for RPA screening and RPA$+V+$SC screening for silicon ($\Gamma$, $X$, and $L$ points) and diamond ($\Gamma$ point). The LDA direct gaps and experimental data are also given.

| Method/Source | $\Gamma$ | $X$ | $L$ |
|---|---|---|---|
| | silicon | | |
| $GW$/RPA (standard $GW$) | 3.31 | 4.20 | 3.38 |
| $GW$/RPA$+V+$SC | 3.38 | 4.26 | 3.44 |
| $GW+V+$SC/RPA | 3.58 | 4.53 | 3.65 |
| $GW+V+$SC/RPA$+V+$SC | 3.67 | 4.64 | 3.77 |
| LDA | 2.53 | 3.35 | 2.61 |
| Expt.[a,b] | 3.40 | 4.25 | 3.45 |
| standard $GW$ literature | 3.35[c] | 4.43[c] | 3.54[c] |
| | diamond | | |
| $GW$/RPA (standard $GW$) | 7.63 | | |
| $GW$/RPA$+V+$SC | 7.83 | | |
| $GW+V+$SC/RPA | 8.08 | | |
| $GW+V+$SC/RPA$+V+$SC | 8.36 | | |
| LDA | 5.51 | | |
| Expt.[a] | 7.3 | | |
| standard $GW$ literature | 7.5,[c] 7.63,[d] 7.26[e] | | |

[a]Reference 34.
[b]Reference 43.
[c]Reference 26.
[d]Reference 44.
[e]Reference 33.

degree. Our results can be considered to be complementary to the work of Bechstedt *et al.*,[24] who observed such a compensation at finite frequencies, also for silicon and diamond, but without taking LFE's into account.

### B. First-order corrections to the $GW$ self-energy

Our standard $GW$ results, obtained by applying the expectation value method, are shown in Table I. The gap values for the $\Gamma$, $X$, and $L$ points of silicon of 3.31, 4.20, and 3.38 eV are observed to compare reasonably with the values 3.35, 4.43, and 3.54 eV of Ref. 26 and excellently with the experimental values[34,43] of 3.40, 4.25, and 3.45 eV. The result for the $\Gamma$ point of diamond of 7.63 eV compares well with the 7.5 eV value reported in Ref. 26 and the 7.63 eV value of Ref. 44 and reasonably with the experimental value[34] of 7.3 eV. We remark in this connection that if we take $N_{\rm gr}=3$ instead of $N_{\rm gr}=2$, our standard $GW$ result for diamond changes from 7.63 eV to 7.54 eV. In Table I also the $GW$ gap values with RPA$+V+$SC screening are given. It is observed that the differences between the standard $GW$ and $GW$/RPA$+V+$SC gap values are relatively minor, as could be expected from the closeness of the RPA and RPA$+V+$SC screening (see Sec. III A).

We now turn to the first-order vertex and self-consistency self-energy corrections to the $GW$ self-energy. We recall Hedin's argument[1] to take $\Sigma$ to $n$th order in $W$ when $P$ is taken to order $n-1$ in $W$ and our restriction of using the LDA Green's function. In following this line of reasoning when calculating the $GW+V+$SC self-energy, the resulting gap values obtained with RPA$+V+$SC screening should obviously be preferred. From Table I it is observed that both the $GW+V+$SC/RPA and the $GW+V+$SC/RPA$+V+$SC gap values, obtained by applying the expectation value method, differ considerably from the standard $GW$ values. The differences with the standard $GW$ values appear to be largest if we take the screening to be RPA$+V+$SC. For this type of screening the differences amount to 0.36, 0.44, and 0.39 eV for the $\Gamma$, $X$, and $L$ points of silicon, while the difference for the $\Gamma$ point of diamond is even 0.73 eV. All corrections apparently have the same sign. If RPA screening is used instead (see also Table I), the differences from the standard $GW$ values reduce to roughly 0.3 eV in the case if silicon and to roughly 0.4 eV for diamond. The obvious conclusion is that, if vertex and self-consistency corrections are included to first order, the compensation between them (see also further on) is clearly incomplete.

In Table II details of our calculational results are presented. It is seen that we have concentrated on QP energies of the highest valence band (HVB) and the lowest conduction band (LCB). The highest valence state and lowest conduction state at the $\Gamma$ point are denoted by $\Gamma'_{25v}$ and $\Gamma_{15c}$, respectively. At the $X$ and $L$ points these states are denoted by $X_{4v}$, $X_{1c}$, $L'_{3v}$, and $L_{1c}$, respectively. We have given for the $GW$ self-energy diagram, minus $V^{\rm xc}$, as well as for the respective SC1, SC2, SC3, SC4, and $V$ self-energy diagrams the expectation values together with the related energy derivatives $\delta$. Results are given for both RPA and RPA$+V+$SC screening for both silicon and diamond. The calculation of the contribution of the $V$ self-energy correction to the difference in expectation value for the LCB and HVB at the $\Gamma$ point for silicon had already been done in Ref. 22. However, due to an error in the program code, the result given in Ref. 22 of 0.12 eV is incorrect. The correct value is $-0.26$

TABLE II. Expectation values of the first-order self-consistency and vertex self-energy corrections, with RPA and RPA+$V$+SC screening, for the HVB and LCB of the $\Gamma$, $X$, and $L$ points of silicon and for the HVB and LCB of the $\Gamma$ point of diamond (in eV). Their energy derivatives $\delta$ are given in parentheses (— means zero energy derivative). The expectation values and energy derivatives of the $GW$ self-energy minus $V^{xc}$ are also given; see column 2.

| $l$,**k** | $GW-V^{xc}$ | SC1 | SC2 | SC3 | SC4 | Total SC | $V$ | Total $V$+SC |
|---|---|---|---|---|---|---|---|---|
| | | | | silicon: RPA screening | | | | |
| $\Gamma'_{25v}$ | −1.211 | −3.652 | 3.513 | 1.287 | −1.152 | −0.004 | 0.076 | 0.072 |
| | (−0.288) | (0.231) | (−0.147) | (—) | (—) | (0.084) | (−0.088) | (−0.004) |
| $\Gamma_{15c}$ | −0.203 | −2.180 | 2.734 | 0.441 | −0.396 | 0.599 | −0.181 | 0.418 |
| | (−0.286) | (−0.130) | (0.197) | (—) | (—) | (0.067) | (−0.079) | (−0.012) |
| $X_{4v}$ | −1.283 | −3.744 | 3.424 | 0.792 | −0.706 | −0.234 | 0.158 | −0.075 |
| | (−0.315) | (0.338) | (−0.237) | (—) | (—) | (0.101) | (−0.079) | (0.021) |
| $X_{1c}$ | −0.159 | −1.688 | 2.174 | −0.077 | 0.068 | 0.477 | −0.162 | 0.314 |
| | (−0.261) | (−0.101) | (0.161) | (—) | (—) | (0.060) | (−0.070) | (−0.010) |
| $L'_{3v}$ | −1.254 | −3.675 | 3.468 | 1.074 | −0.957 | −0.090 | 0.130 | 0.040 |
| | (−0.299) | (0.284) | (−0.194) | (—) | (—) | (0.091) | (−0.089) | (0.002) |
| $L_{1c}$ | −0.246 | −2.211 | 2.720 | 0.587 | −0.524 | 0.572 | −0.193 | 0.378 |
| | (−0.273) | (−0.123) | (0.187) | (—) | (—) | (0.064) | (−0.076) | (−0.011) |
| | | | | silicon: RPA+$V$+SC screening | | | | |
| $\Gamma'_{25v}$ | −1.310 | −3.654 | 3.526 | 1.292 | −1.152 | 0.012 | 0.017 | 0.030 |
| | (−0.290) | (0.226) | (−0.140) | (—) | (—) | (0.086) | (−0.092) | (−0.006) |
| $\Gamma_{15c}$ | −0.217 | −2.209 | 2.755 | 0.444 | −0.396 | 0.594 | −0.181 | 0.413 |
| | (−0.288) | (−0.125) | (0.193) | (—) | (—) | (0.068) | (−0.084) | (−0.016) |
| $X_{4v}$ | −1.365 | −3.710 | 3.407 | 0.805 | −0.706 | −0.204 | 0.073 | −0.131 |
| | (−0.317) | (0.325) | (−0.222) | (—) | (—) | (0.103) | (−0.082) | (0.020) |
| $X_{1c}$ | −0.161 | −1.692 | 2.163 | −0.078 | 0.068 | 0.461 | −0.130 | 0.331 |
| | (−0.263) | (−0.090) | (0.150) | (—) | (—) | (0.060) | (−0.072) | (−0.012) |
| $L'_{3v}$ | −1.347 | −3.643 | 3.455 | 1.091 | −0.957 | −0.054 | 0.036 | −0.019 |
| | (−0.301) | (0.271) | (−0.179) | (—) | (—) | (0.093) | (−0.091) | (0.001) |
| $L_{1c}$ | −0.257 | −2.215 | 2.708 | 0.597 | −0.524 | 0.566 | −0.168 | 0.398 |
| | (−0.275) | (−0.111) | (0.176) | (—) | (—) | (0.065) | (−0.078) | (−0.013) |
| | | | | diamond: RPA screening | | | | |
| $\Gamma'_{25v}$ | −2.083 | −3.666 | 3.524 | 1.725 | −1.543 | 0.040 | 0.109 | 0.149 |
| | (−0.191) | (0.128) | (−0.085) | (—) | (—) | (0.043) | (−0.036) | (0.007) |
| $\Gamma_{15c}$ | 0.445 | −2.305 | 2.853 | 1.055 | −0.956 | 0.647 | 0.019 | 0.666 |
| | (−0.187) | (−0.062) | (0.099) | (—) | (—) | (0.036) | (−0.032) | (0.005) |
| | | | | diamond: RPA+$V$+SC screening | | | | |
| $\Gamma'_{25v}$ | −2.321 | −3.696 | 3.567 | 1.776 | −1.543 | 0.104 | −0.013 | 0.091 |
| | (−0.194) | (0.125) | (−0.081) | (—) | (—) | (0.044) | (−0.041) | (0.003) |
| $\Gamma_{15c}$ | 0.459 | −2.356 | 2.889 | 1.091 | −0.956 | 0.668 | 0.041 | 0.709 |
| | (−0.191) | (−0.058) | (0.096) | (—) | (—) | (0.039) | (−0.037) | (0.002) |

eV; see Table II. Concerning the SC self-energy corrections, one can argue that a $GW$ insertion and a $V^{xc}$ insertion have much in common, such that SC1 and SC2 are likely to compensate partially. It is observed that the SC1 and SC2 self-energy diagrams individually lead to relatively large corrections to the absolute energies of the HVB and LCB and have indeed the tendency to compensate each other, though not completely. The SC3 and SC4 self-energy diagrams are self-energy corrections to the Hartree diagram, in which the valence charge density is corrected to first-order in $W$. As the valence charge density in DFT is equal to the exact density, the first-order corrections to the LDA valence charge density are expected to be minor.[33] It is observed that the contribution to the difference in expectation value for the LCB and HVB due to the SC Hartree diagrams (SC3+SC4) is very small indeed as compared to the contribution due to the SC1 +SC2 self-energy diagrams. The difference in expectation value for the LCB and HVB of the $V$ and of the total SC correction can also easily be obtained from the entries in Table II. By doing so, the $V$ correction appears to compensate the SC correction to about 45% in the case of RPA screening and to about 35% in the case of RPA+$V$+SC screening for silicon. For diamond these percentages are about 15% and −10%, respectively, so that the term compensation is not even appropriate.

The above-mentioned $GW$+$V$+SC/RPA+$V$+SC result is puzzling in a certain sense: If the sum of all first-order corrections to the standard $GW$ gap does not appear to be negligibly small, the question arises which group of diagrams then have to be considered in order to ''justify'' the

standard $GW$ result. Before trying to answer this question it is necessary, however, to be as certain as possible that our calculations do not contain weaknesses of whatever kind. This has led us to the performance of a few checks, which are outlined in the next section.

## IV. CHECKS

In this section we report on a few checks that we have performed concerning the $V+\mathrm{SC}$ self-energy corrections. First, for RPA screening, we improved upon the PPM concerning the matrix $M$ in accordance with a proposal of Farid,[30] in which he points to the violation of Johnson's $f$-sum rule if one sticks to the matrix $M$ of Eq. (8). Apart from $P_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega=0)$ we now also have to calculate the leading term of $P_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega\rightarrow\infty)$; see Eq. (9). In agreement with Ref. 29, we find that the diagonal elements of the matrix $M$ of Eq. (9) are smaller than those given by the Johnson $f$-sum rule and that the off-diagonal elements deviate even more. When applying the correct matrix $M$, the standard $GW$ direct band gap of silicon at the $\Gamma$ point becomes only 0.015 eV smaller, however. In fact, this is not unexpected, as the use of Johnson's $f$-sum rule (see, for instance, Refs. 26 and 27) generally leads to excellent agreement between standard $GW$ and experimental gap values. Furthermore, it is also found that the $GW+V+\mathrm{SC/RPA}$ direct band gap of silicon at $\Gamma$, when calculated with the correct matrix $M$, stays practically the same: The value becomes only 0.011 eV smaller. It can therefore safely be concluded that the violation of the Johnson $f$-sum rule yields only insignificant deviations in corrected gap values.

A second point of possible concern is the assumed closeness of LDA wave functions and $GW+V+\mathrm{SC}$ wave functions. Though it is demonstrated in Refs. 26 and 33 that the LDA wave functions and the $GW$ wave functions are close, it is not *a priori* certain that this also holds in the presence of the $V+\mathrm{SC}$ self-energy corrections. This, however, has to be fulfilled in order to safely apply the expectation value method. We therefore carried through an exact diagonalization procedure for the HVB and LCB at the $\Gamma$ point of silicon, using RPA screening. In this procedure only coupling between states of equal symmetry needs to be considered, leading, among 65 electron bands, to a $6\times6$ matrix in the LDA basis to be diagonalized only. In doing so, both the standard $GW$ and the $GW+V+\mathrm{SC/RPA}$ direct band gap become larger by an amount of only 0.001 eV compared to the values obtained within the expectation value method. It can therefore also safely be concluded that the LDA wave functions and the $GW+V+\mathrm{SC}$ wave functions are sufficiently similar.

A third check concerns the LDA starting point. Though the LDA wave functions are to a large extent similar to the $GW$ wave functions (and to the $GW+V+\mathrm{SC}$ wave functions as shown above), the conduction-band energy levels in the LDA are significantly lower than those in $GW$ (and $GW+V+\mathrm{SC}$). Though this is generally not thought to be an important issue, we nevertheless would like to investigate whether a LDA input in which the quasiparticle shift is included could improve the results. In a sense such an altered input could be considered to be closer to a ''$GW$ set of wave functions with accompanying energy levels'' than the usual LDA input. We therefore applied the so-called scissors operator to the LDA by changing the LDA exchange-correlation potential $V^{\mathrm{xc}}(\mathbf{k})$ into $V^{\mathrm{xc}}(\mathbf{k})+\Delta\Sigma_c|c,\mathbf{k}\rangle\langle c,\mathbf{k}|$, where $c$ is meant to indicate conduction bands and $\Delta=0.8$ eV (for silicon) being about the standard $GW$ conduction-band shift. Leaving the LDA wave functions unaltered, we can now construct another Green's function and insert it into the RPA polarizability, the $GW$ self-energy, and the $V+\mathrm{SC}$ self-energy correction. In doing so, the (RPA) dielectric constant changes and becomes equal to 10.9 ($N_{\mathrm{gr}}=6$). The new standard $GW$ direct gap at $\Gamma$ (of silicon) appears to be about 0.2 eV larger than the value obtained with the LDA as the starting point, while the new $GW+V+\mathrm{SC/RPA}$ direct gap is about 0.1 eV larger than before. It therefore shows that this kind of change in the LDA starting point in the $GW$ type of calculations does not improve things. Incidentally, the dependence on starting point Hamiltonians for $GW$ or related types of calculations and, more specifically, the apparent preference for the LDA starting point are interesting in themselves and not sufficiently settled in our opinion.

## V. DISCUSSION ON THE SELF-CONSISTENCY SELF-ENERGY DIAGRAMS

If we, in spite of the preceding discussion, nevertheless pursue the issue of LDA wave functions and $GW$ wave functions being highly similar, it is tempting to subdivide the SC1+SC2 self-energy subdiagrams. To this end we give the expression for the insertion, in the LDA basis, occurring in the SC1+SC2 self-energy correction to the LDA energy $\varepsilon_l(\mathbf{k})$ (the electron bands $l'$ and $l''$ are summation variables and the energy $\omega$ and the wave vector $\mathbf{q}$ are integration variables):

$$\langle l',\mathbf{q}+\mathbf{k}|\hbar\Sigma^{GW}(\mathbf{q}+\mathbf{k};\omega+\varepsilon_l(\mathbf{k}))-V^{\mathrm{xc}}(\mathbf{q}+\mathbf{k})|l'',\mathbf{q}+\mathbf{k}\rangle. \tag{16}$$

The SC1+SC2 self-energy subdiagrams can be subdivided into the ''diagonal SC'' (DSC) and ''nondiagonal SC'' (NDSC) groups of subdiagrams of Fig. 5. The reason for doing this is that the DSC group of subdiagrams consists of subdiagrams in which the insertion, given by Eq. (16), is taken between *the same* conduction ($c$) or valence ($v$) states only. The NDSC group of subdiagrams can further be subdivided into the NDSCA and NDSCB group of subdiagrams; see Fig. 5. The NDSCA and NDSCB groups of subdiagrams are initially expected to be small, the reason being that in these latter subdiagrams the insertion, given by Eq. (16), is taken between *different* states. If we simply ignore the NDSC contribution and calculate the difference between the $GW+\mathrm{DSC/RPA}$ and the standard $GW$ gap value, applying the expectation value method using standard $GW$ energy derivatives $\delta$, we obtain the values 0.32, 0.49, 0.40, and 0.34 eV for the $\Gamma$, $X$, and $L$ points of silicon and the $\Gamma$ point of diamond, respectively. Considering the fact that the vertex correction $V$ to the self-energy yields a gap correction of about $-0.3$ eV for silicon and about $-0.1$ eV for diamond (see Table II), we would then find that the totality of the $V$ plus DSC correction appears to be minor for silicon, while the compensation is less pronounced for diamond. Unfortunately, however, our *actual* results on the group of NDSC subdiagrams do not confirm the above reasoning at all. It is

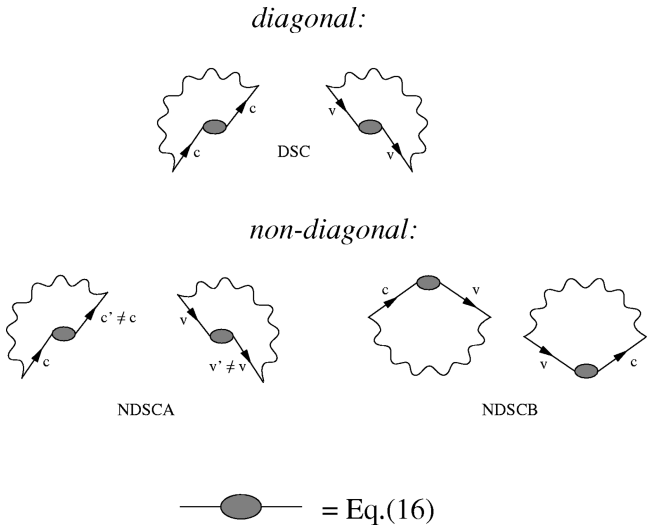


FIG. 5. SC1+SC2 self-energy correction subdiagrams with diagonal and nondiagonal $GW$ self-energy insertions in the LDA basis. The former subdiagrams are indicated by DSC and the latter subdiagrams are indicated by NDSCA and NDSCB. Time increases from bottom to top. $v$ stands for a valence-band index and $c$ stands for a conduction-band index (which have to be summed over). In the case of DSC the band indices on both sides of the ellipse are identical.

true that we find the contribution to the gap due to the NDSCA group of Fig. 5 to be very small, but, unfortunately, this does not hold for the remaining group of subdiagrams, NDSCB. The discrepancy between the actual contribution of the NDSC subdiagrams and our above reasoning (expecting them to be insignificant), however, can be understood as follows. Taking the square of the matrix elements $\langle l,\mathbf{k}|\hbar\Sigma^{GW}(\mathbf{k};\varepsilon_l(\mathbf{k}))-V^{\rm xc}(\mathbf{k})|l',\mathbf{k}\rangle$, the nondiagonal ($l\neq l'$) values are found to be about two orders of magnitude smaller than the diagonal ($l=l'$) ones. This is one of the reasons for the correction[33]

$$\sum_{l'\neq l}\frac{|\langle l,\mathbf{k}|\hbar\Sigma^{GW}(\mathbf{k};\varepsilon_l(\mathbf{k}))-V^{\rm xc}(\mathbf{k})|l',\mathbf{k}\rangle|^2}{\varepsilon_l(\mathbf{k})-\varepsilon_{l'}(\mathbf{k})} \tag{17}$$

to the QP energies to be very small, which is related to the high similarity of the $GW$ and LDA wave functions. On the other hand, in order to calculate the SC1+SC2 self-energy contribution the matrix element itself is required, instead of its square. Furthermore, the energy denominators pertaining to the NDSCB subdiagrams contain one energy difference consisting only of electron energies, while in the case of the DSC (and NDSCA) subdiagrams each energy difference contains a plasmon energy (which is relatively large).

## VI. A REMARKABLE CANCELLATION

We note that it is possible to view upon the orders in $W$ in a different way by expanding the wave-function renormalization factor $Z_{l,\mathbf{k}}$ occurring in Eq. (13). $Z$ can formally be written as

$$Z_{l,\mathbf{k}}=1+\sum_{n=1}^{\infty}\left\{\hbar\langle l,\mathbf{k}|\frac{\partial\Sigma(\mathbf{k};\omega)}{\partial\omega}\bigg|_{\omega\,=\,\varepsilon_l(\mathbf{k})}|l,\mathbf{k}\rangle\right\}^n \tag{18}$$

such that, in fact, an infinite number of higher-order terms in $W$ are involved via the energy derivative of $\Sigma$. In doing so, we observe a remarkable cancellation between corrections to the band gap due to the $V$ and SC corrections to the self-energy $\Sigma$ on the one hand and corrections to the band gap due to the energy dependence of $\Sigma^{GW}$ on the other hand. A cancellation of this particular kind has been reported by van Haeringen for both the Bloch-Nordsieck model describing electron-photon coupling[45,46] and the Fröhlich polaron model describing electron-phonon coupling.[47] A cancellation between self-consistency corrections to $\Sigma$ and the energy dependence of $\Sigma$ was indicated by DuBois[5] and also mentioned by Rice.[10]

In order to be able to present the cancellation effect, it is necessary to return to the expectation value method that was introduced in Sec. II. We expand the right-hand side (RHS) of Eq. (13) to second order in the screened interaction $W$ by making a Taylor expansion of the denominator containing the energy derivative and the resulting expression for the quasiparticle energy $E_l^{\rm QP}(\mathbf{k})$ is

$$\begin{aligned}E_l^{\rm QP}(\mathbf{k})\approx{}&\varepsilon_l(\mathbf{k})+\langle l,\mathbf{k}|\hbar\Sigma^{GW}(\mathbf{k};\varepsilon_l(\mathbf{k}))-V^{\rm xc}(\mathbf{k})|l,\mathbf{k}\rangle\\&+\langle l,\mathbf{k}|\hbar\Sigma^{GW}(\mathbf{k};\varepsilon_l(\mathbf{k}))-V^{\rm xc}(\mathbf{k})|l,\mathbf{k}\rangle\,\delta_l^{GW}(\mathbf{k})\\&+\hbar\langle l,\mathbf{k}|\Sigma^{V+{\rm SC}}(\mathbf{k};\varepsilon_l(\mathbf{k}))|l,\mathbf{k}\rangle.\end{aligned} \tag{19}$$

We have regrouped our calculational results in accordance with Eq. (19), taking care of correction terms in the ''appropriate order'' for both RPA and RPA+$V$+SC screening. In doing so, note that the convention of viewing upon orders in $W$ is now different from before: The $GW$ expectation value is of ''first order'' in $W$, the $V$+SC self-energy expectation value is of ''second order'' in $W$, and the $GW$ expectation value times its energy derivative is also of ''second order'' in $W$. The above-mentioned cancellation effect concerns the last two terms on the RHS of Eq. (19). In Table III we present the values of these terms for the HVB and LCB of the $\Gamma$, $X$, and $L$ points of silicon and for the $\Gamma$ point of diamond, which can be produced with the data given in Table II. It is observed that no cancellation occurs for the LCB and HVB separately. A remarkable cancellation is seen to occur, however, for the band-gap values, in all cases leading to a gap contribution smaller than 0.1 eV. It should be noted in this connection that a similar result has already been obtained in the case of the quasi-one-dimensional semiconducting wire; see Ref. 48.

It will be clear that this result does not as yet contribute to a deeper understanding of the celebrated standard $GW$ result. This particular cancellation causes the $GW$+$V$+SC energies to be equal to the LDA energies plus the $GW$ self-energy expectation values calculated at the LDA energy. This is puzzling since in the case of the $GW$ gap, calculating the $GW$ self-energy expectation value at the LDA energy instead of at the $GW$ energy, the energy derivative was absolutely required to obtain agreement with experiment. The cancella-

TABLE III. Correction contributions due to the last two terms on the RHS of Eq. (19) for HVB and LCB of silicon and diamond (in eV), in the case of RPA screening and RPA+$V$+SC screening (for the latter case the values are given in parentheses). With $\langle GW-V^{\rm xc}\rangle$ we mean $\langle l,\mathbf{k}|\hbar\Sigma^{GW}(\mathbf{k};\varepsilon_l(\mathbf{k}))-V^{\rm xc}(\mathbf{k})|l,\mathbf{k}\rangle$ and with $\langle V+{\rm SC}\rangle$ we mean $\hbar\langle l,\mathbf{k}|\Sigma^{V+{\rm SC}}(\mathbf{k};\varepsilon_l(\mathbf{k}))|l,\mathbf{k}\rangle$.

| $l,\mathbf{k}$ | $\langle GW-V^{\rm xc}\rangle\delta_l^{GW}(\mathbf{k})$ | $\langle V+{\rm SC}\rangle$ | Total | Gap contribution |
|---|---|---|---|---|
| | | silicon | | |
| $\Gamma'_{25v}$ | 0.349 (0.380) | 0.072(0.030) | 0.421 (0.410) | |
| $\Gamma_{15c}$ | 0.058 (0.062) | 0.418(0.413) | 0.476 (0.475) | 0.055 (0.065) |
| $X_{4v}$ | 0.404 (0.433) | −0.075(−0.131) | 0.329 (0.302) | |
| $X_{1c}$ | 0.041 (0.042) | 0.314(0.331) | 0.355 (0.373) | 0.026 (0.071) |
| $L'_{3v}$ | 0.375 (0.405) | 0.040(−0.019) | 0.415 (0.386) | |
| $L_{1c}$ | 0.067 (0.071) | 0.378(0.398) | 0.445 (0.469) | 0.030 (0.083) |
| | | diamond | | |
| $\Gamma'_{25v}$ | 0.398 (0.450) | 0.149(0.091) | 0.547 (0.541) | |
| $\Gamma_{15c}$ | −0.083 (−0.088) | 0.666(0.709) | 0.583 (0.621) | 0.036 (0.080) |

tion rather points to the existence of an apparent ''sum rule,'' which unfortunately is unexplained as yet. There is a possible connection to Ward identities,[13] but no reference to a specific Ward identity has been discussed in the literature. If, in our case, an identity of the above kind indeed exists, the above results could presumably be considered as an internal check on the correctness of our calculations rather than contributing to the identification of the relevant group of diagrams that leads to the same gap results as standard $GW$.

## VII. DISCUSSION AND CONCLUSIONS

The present work was motivated by the idea that the apparent success of standard $GW$ in predicting electronic properties could possibly be supported by a compensation between first-order vertex and self-consistency corrections to the band gap of silicon and diamond. Compensations of this type are occasionally reported on in the literature, mainly in the case of the homogeneous electron gas, but also in the case of a quasi-one-dimensional semiconducting wire. It seemed of interest to investigate to what extent such a compensation can also be found for a completely realistic case. The starting point has been a fully converged LDA calculation for both silicon and diamond. Our effort has been to calculate the contribution of complete sets of subdiagrams contributing to both the polarizability $P$ (to first order in $W$) and the self-energy $\Sigma$ (to second order in $W$). We found large compensations between the first-order vertex and self-consistency corrections to $P$, but the result concerning $\Sigma$ and the related gap value is disappointing in the sense that there appears to be only a 35% compensation between the $V$ and SC self-energy corrections (to the difference in their expectation value for the LCB and HVB) in the case of silicon, while such a compensation is in fact absent in the case of diamond. The resulting corrected gap values appear to be about 0.4 eV and 0.7 eV larger than the standard $GW$ values for silicon and diamond, respectively. This result therefore does not give the expected help in understanding the success of the standard $GW$ approach. In view of this more or less unexpected result, much effort has been put in checking the correctness of a large number of computational steps, such that we are convinced of the correctness of our final results. Furthermore, by expanding the wave-function renormalization function, we have found a cancellation of particular correction terms occurring for the $GW+V+$SC gap. It is therefore worthwhile to speculate on other more refined compensating mechanisms that possibly could explain the ''correctness'' of standard $GW$. In this connection we recall the work of Shirley,[11] briefly discussed in earlier sections, from which it could be deduced that a possibly more complete compensation could be obtained if we would be able to evaluate self-consistently the sum of the $GW$ self-energy dia-

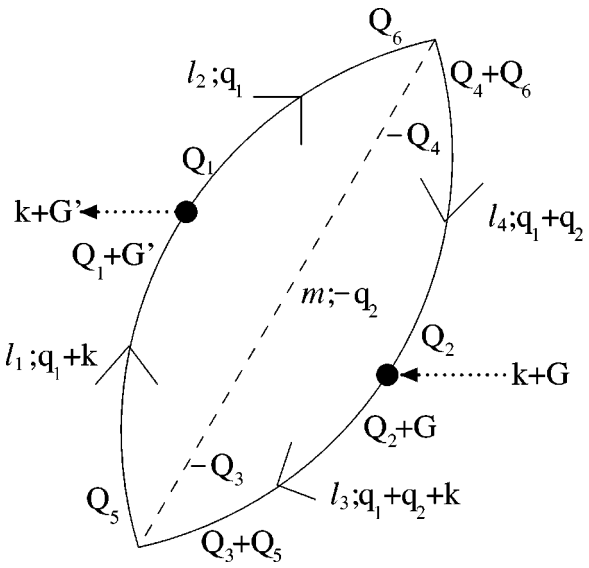


FIG. 6. Vertex correction subdiagram included in the correction to the polarizability $P_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega)$. A directed line denotes the LDA Green's function $G^0$ and the dashed line stands for $W^{\rm scr}$ [see Eq. (2)]. The two dotted, arrowed lines with $\mathbf{k}+\mathbf{G}$ and $\mathbf{k}+\mathbf{G}'$ indicate the two crystal momenta for which the correction $\Delta P$ is taken. The meaning of the labels is explained in the text.

gram and the vertex correction diagram. In 1965 Hedin[1] already put forth that corrections to standard $GW$ should preferably be included by using a self-consistent $G$. Unfortunately, this is a tremendous task, even for the homogeneous electron gas, but in our opinion its performance for silicon and diamond is considered to be crucial as it could very well contribute to a better understanding of the success of the standard $GW$ approach for these latter materials.

## ACKNOWLEDGMENTS

It is a pleasure to thank B. Farid and E. Shirley for very valuable discussions. This work was supported by the National Computing Facilities Foundation (NCF) for the use of supercomputer facilities, with financial support from the Netherlands Organization for Scientific Research (NWO).

## APPENDIX: THE EXPRESSION FOR A POLARIZABILITY CORRECTION SUBDIAGRAM

In this appendix we give the worked-out algebraic expression for a vertex correction subdiagram of the polarizability in the momentum representation; see Fig. 6. The momentum flow through the subdiagram can be deduced using momentum conservation at the interaction vertices. This particular correction entails summations over conduction bands ($l_1$ and $l_2$), valence bands ($l_3$ and $l_4$), plasmon bands ($m$), and reciprocal lattice vectors $\mathbf{Q}_1-\mathbf{Q}_6$. The presence of reciprocal lattice vectors is due to the interaction of an electron or hole with the ion lattice. Furthermore, this correction involves an integration over the reduced wave vectors $\mathbf{q}_1$ and $\mathbf{q}_2$. As mentioned in Sec. II, one integration can be performed over $\mathcal{I}_{\mathbf{k}}$. The expression is

$$\begin{aligned}\Delta P_{\mathbf{G},\mathbf{G}'}(\mathbf{k};\omega)=-2\int_{1\mathrm{BZ}}\frac{d^3q_1}{(2\pi)^3}\int_{1\mathrm{BZ}}\frac{d^3q_2}{(2\pi)^3}\sum_{l_1\in c}\sum_{l_2\in c}\sum_{l_3\in v}\sum_{l_4\in v}\sum_{m}&\left(\sum_{\mathbf{Q}_1} d_{l_1,\mathbf{q}_1+\mathbf{k}}(\mathbf{Q}_1+\mathbf{G}')d^*_{l_2,\mathbf{q}_1}(\mathbf{Q}_1)\right)\\
\times\left(\sum_{\mathbf{Q}_2} d_{l_4,\mathbf{q}_1+\mathbf{q}_2}(\mathbf{Q}_2)d^*_{l_3,\mathbf{q}_1+\mathbf{q}_2+\mathbf{k}}(\mathbf{Q}_2+\mathbf{G})\right)&\left(\sum_{\mathbf{Q}_5} d^*_{l_1,\mathbf{q}_1+\mathbf{k}}(\mathbf{Q}_5)\sum_{\mathbf{Q}_3} d_{l_3,\mathbf{q}_1+\mathbf{q}_2+\mathbf{k}}(\mathbf{Q}_3+\mathbf{Q}_5)w_{m,-\mathbf{q}_2}(-\mathbf{Q}_3)\right)\\
\times\left(\sum_{\mathbf{Q}_6} d_{l_2,\mathbf{q}_1}(\mathbf{Q}_6)\sum_{\mathbf{Q}_4} d^*_{l_4,\mathbf{q}_1+\mathbf{q}_2}(\mathbf{Q}_4+\mathbf{Q}_6)w^*_{m,-\mathbf{q}_2}(-\mathbf{Q}_4)\right)&\{[\omega_m(-\mathbf{q}_2)+\varepsilon_{l_1}(\mathbf{q}_1+\mathbf{k})-\varepsilon_{l_3}(\mathbf{q}_1+\mathbf{q}_2+\mathbf{k})]\\
\times[\omega_m(-\mathbf{q}_2)+\varepsilon_{l_2}(\mathbf{q}_1)-\varepsilon_{l_4}(\mathbf{q}_1+\mathbf{q}_2)][\omega-\omega_m(-\mathbf{q}_2)-\varepsilon_{l_1}(\mathbf{q}_1+\mathbf{k})+\varepsilon_{l_4}(\mathbf{q}_1+\mathbf{q}_2)]\}^{-1}.&\qquad\text{(A1)}\end{aligned}$$

Here $\varepsilon_l$ are LDA energies and $d_l$ are LDA plane-wave coefficients; $\omega_m$ are PPM energies and $w_m$ are PPM coefficients. $c$ and $v$ denote the conduction bands and valence bands, respectively. The factor 2 originates from summation over spins. In this particular example given by Eq. (A1), the head element ($\mathbf{G}=\mathbf{0}$, $\mathbf{G}'=\mathbf{0}$) for $\mathbf{k}\to\mathbf{0}$ has a constant contribution if $l_1=l_2$ and $l_3=l_4$ because then we have twice an inner product of a wave function with itself, which is unity. The HE also has linear contributions in $\mathbf{k}$ for $\mathbf{k}\to\mathbf{0}$ if $l_1\neq l_2$ or $l_3\neq l_4$. Also the wing elements ($\mathbf{G}=\mathbf{0}$ or $\mathbf{G}'=\mathbf{0}$) have a constant contribution. The energy denominator contains products of energy differences. For this particular correction to the polarizability given by Eq. (A1) the energy denominator consists of three such energy differences. For $\omega=0$ there are no vanishing energy denominators for any subdiagram contributing to the polarizability.

[1] L. Hedin, Phys. Rev. **139**, A796 (1965).
[2] L. Hedin and B.I. Lundqvist, J. Phys. C **4**, 2064 (1971).
[3] J. Hubbard, Proc. R. Soc. London, Ser. A **243**, 336 (1957).
[4] F. Brosens, L.F. Lemmens, and J.T. Devreese, Phys. Status Solidi B **74**, 45 (1976).
[5] D.F. DuBois, Ann. Phys. (N.Y.) **7**, 174 (1959); **8**, 24 (1959).
[6] M. Gell-Mann and K.A. Brueckner, Phys. Rev. **106**, 364 (1957).
[7] M. Gell-Mann, Phys. Rev. **106**, 369 (1957).
[8] D.J.W. Geldart and R. Taylor, Can. J. Phys. **48**, 155 (1970); **48**, 167 (1970).
[9] G.D. Mahan and B.E. Sernelius, Phys. Rev. Lett. **62**, 2718 (1989).
[10] T.M. Rice, Ann. Phys. (N.Y.) **31**, 100 (1965).
[11] E.L. Shirley, Phys. Rev. B **54**, 7758 (1996).
[12] U. von Barth and B. Holm, Phys. Rev. B **54**, 8411 (1996); B. Holm and U. von Barth, *ibid.* **57**, 2108 (1998).
[13] J.C. Ward, Phys. Rev. **78**, 182 (1950).
[14] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 287 (1961); G. Baym, *ibid.* **127**, 1391 (1962).
[15] H.J. de Groot, P.A. Bobbert, and W. van Haeringen, Phys. Rev. B **52**, 11 000 (1995).
[16] W. Knorr and R.W. Godby, Phys. Rev. Lett. **68**, 639 (1992).
[17] W. Knorr and R.W. Godby, Phys. Rev. B **50**, 1779 (1994).
[18] H.J. de Groot, R.T.M. Ummels, P.A. Bobbert, and W. van Haeringen, Phys. Rev. B **54**, 2374 (1996).
[19] W. Hanke and L.J. Sham, Phys. Rev. B **12**, 4501 (1975).
[20] R. Daling and W. van Haeringen, Phys. Rev. B **40**, 11 659 (1989).
[21] R. Daling, P. Unger, P. Fulde, and W. van Haeringen, Phys. Rev. B **43**, 1851 (1991).
[22] P.A. Bobbert and W. van Haeringen, Phys. Rev. B **49**, 7564 (1994).
[23] R. Del Sole, L. Reining, and R.W. Godby, Phys. Rev. B **49**, 8024 (1994).
[24] F. Bechstedt, K. Tenelsen, B. Adolph, and R. Del Sole, Phys. Rev. Lett. **78**, 1528 (1997).
[25] G.E. Engel and B. Farid, Phys. Rev. B **47**, 15 931 (1993).
[26] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **34**, 5390 (1986).
[27] W. von der Linden and P. Horsch, Phys. Rev. B **37**, 8351 (1988).
[28] D.L. Johnson, Phys. Rev. B **9**, 4475 (1974).
[29] G.E. Engel and B. Farid, Phys. Rev. B **46**, 15 812 (1992).

[30] B. Farid (private communication).
[31] W. Kohn, Phys. Rev. **110**, 857 (1958).
[32] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **35**, 5585 (1987).
[33] R.W. Godby, M. Schlüter, and L.J. Sham, Phys. Rev. B **37**, 10 159 (1988).
[34] *Semiconductors: Physics of Group IV Elements and III-V Compounds*, edited by K.-H. Hellwege and O. Madelung, Landolt-Börnstein, New Series, Group III, Vol. 17, Pt. a (Springer-Verlag, Berlin, 1982).
[35] G.B. Bachelet, H.S. Greenside, G.A. Baraff, and M. Schlüter, Phys. Rev. B **24**, 4745 (1981).
[36] E.P. Wigner, Phys. Rev. **46**, 1002 (1934).
[37] R.A. Faulkner, Phys. Rev. **184**, 713 (1969); H.W. Icenogle, B.C. Platt, and W.L. Wolfe, Appl. Opt. **15**, 2348 (1976).
[38] H.H. Li, J. Phys. Chem. Ref. Data **9**, 561 (1980).
[39] C. Kittel, *Introduction to Solid State Physics*, 5th ed. (Wiley, New York, 1976), p. 309.
[40] J.C. Phillips, Phys. Rev. Lett. **20**, 550 (1968).
[41] P.E. van Camp, V.E. van Doren, and J.T. Devreese, Phys. Rev. B **24**, 1096 (1981).
[42] J.P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[43] D.E. Aspnes and A.A. Studna, Phys. Rev. B **27**, 985 (1983).
[44] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **48**, 17 791 (1993).
[45] F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).
[46] W. van Haeringen, Physica (Amsterdam) **26**, 289 (1960).
[47] W. van Haeringen, Phys. Rev. **137**, A1902 (1965).
[48] H.J. de Groot, Ph.D. thesis, Eindhoven University of Technology, 1996.