RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 43, NUMBER 2 15 JANUARY 1991-I

Lowest-order vertex corrections to the energy gap in covalent semiconductors

R. Daling
Department of Physics, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands

P. Unger and P. Fulde
Max-Planck-Institut für Festkörperforschung, D-7000 Stuttgart 80, Federal Republic of Germany

W. van Haeringen
Department of Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
(Received 8 June 1990; revised manuscript received 18 September 1990)

The accuracy of the *GW* approximation is investigated by evaluating the lowest-order vertex contributions to the energy gap of covalent semiconductors. We compare for silicon the energy shifts of a plane-wave calculation with the results of a variational ansatz in local space. The gap corrections at the Γ point are found to be less than 0.1 eV. An analysis of the different diagrammatic contributions shows that there are two counteracting physical processes which lead to partial cancellations.

INTRODUCTION

In recent years band-structure calculations for elemental semiconductors have come to an improved agreement with experimentally determined quasiparticle spectra (see, e.g., Refs. 1–3). An essential ingredient in true calculations has been the *GW* approximation for the electronic self-energy.[4] It corresponds to retaining the first-order term only when an expansion of the self-energy in powers of the screened electron-electron interaction W is made. The electron Green's function G is thereby the one which is obtained from a LDA (local-density approximation) calculation. In order to understand why the *GW* approximation seemingly works so well for calculating quasiparticle excitation energies in semiconductors it is necessary to investigate the contributions of higher-order diagrams which are left out in a *GW* calculation. The most prominent ones are the second-order exchange contributions. In the language of diagrams they correspond to the lowest-order vertex corrections. When discussing the importance of vertex corrections to the *GW* approximation one must, in principle, calculate diagrams of second order in the screened interaction W. That, however, is a formidable if not impossible task. To circumvent this problem W is replaced by the bare interaction V. Within that restriction a recent plane-wave calculation[5] indicates that the influence of the lowest-order vertex or exchange contributions on the direct gap of Si is indeed negligibly small.

It is the aim of our paper to elucidate the physical origin for the smallness of the gap corrections. We therefore consider separately the contributions of the different exchange diagrams of second order evaluated in a plane-wave basis and also in local space. In the calculations in local space a variational ansatz is made for the ground-state wave functions. Thereby a bond-orbital approximation (BOA) is made. Although the BOA limits the accuracy of the results, it has the virtue that analytical expressions can be derived which allow for a simple and transparent interpretation of the vertex contributions. In both calculations we find a significant cancellation of two distinct sets of diagrams representing physical processes with counteracting influence on the gap. Furthermore, we come to the conclusion that the smallness of the lowest-order vertex contributions applies to all covalent semiconductors.

DESCRIPTION OF THE CALCULATIONS

In this section an outline is given of the two different methods mentioned in the Introduction. We start with a brief description of the plane-wave calculation. A detailed discussion can be found in Ref. 5. Figure 1(b) shows the lowest-order exchange diagrams which are of second order in the Coulomb interaction. In evaluating them with respect to a plane-wave basis a local, empirical pseudopotential is used to generate the band energies and wave functions. The idea behind using the empirical pseudopotential is that in this way we meet in an approximate sense

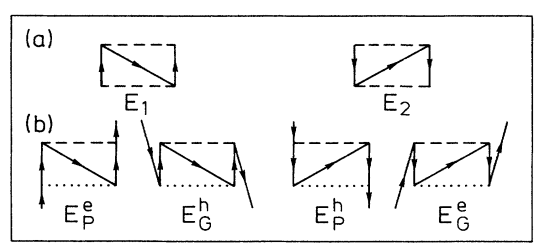

FIG. 1. (a) The two vertex correction diagrams which are of second order in the Coulomb interaction (dashed lines). Upward running lines represent electrons in a conduction-band state and downward running lines are holes in a valence-band state. (b) Energy corrections to conduction-band states denoted by e and to valence-band states denoted by h. In case of the plane-wave calculation the dotted line is a Coulomb interaction and the diagrams represent the expectation values of diagrams E_1 and E_2. In case of the local calculation the dotted lines are the correlation operators S_{IJ}^{π} and S_{IJ}^{η}, respectively.

 © 1991 The American Physical Society

the requirement of self-consistency since the empirical pseudopotential yields a band structure which is close to the exact one. The calculations involve the numerical computation of a six-dimensional integral with an integrand containing summations over three band indices and over four sets of reciprocal-lattice vectors. Due to those summations the time needed to evaluate the integrand at one point in the integration volume and thus the time needed to calculate the integral rises roughly as the fourth power of the number of bands. Moreover, if we increase the number of bands the number of independent matrix elements needed to specify the diagram with respect to a plane-wave basis also increases. Because of the two effects the calculation was limited to 15 plane waves. Although this number is too small in order to give a very accurate value for the vertex corrections, the calculations are nevertheless expected to give semiquantitative results.

The second method uses a local ansatz for the treatment of correlations combined with a BOA. For detailed discussion see Refs. 6 and 7. Consider a lattice of diamond structure built up from orthogonalized sp^3 hybrids. The latter form bonding and antibonding states with corresponding creation operators $B_{I\sigma}^\dagger$ and $A_{I\sigma}^\dagger$, respectively. The index I denotes the bond. The Hamiltonian when expressed in terms of the sp^3 basis set is divided into a self-consistent field part $H_{\rm SCF}$ and a remaining part which contains the residual interactions, i.e., $H=H_{\rm SCF}+H_{\rm res}$. In the BOA the ground state $|\Phi_0\rangle$ of $H_{\rm SCF}$ is of the form

$$|\Phi_0\rangle=\prod_{I,\sigma}B_{I\sigma}^\dagger|0\rangle\,. \tag{1}$$

The operator $A_{c\sigma}^\dagger(\mathbf{k})$ adds an electron with wave vector $\mathbf{k}$ to the conduction band, i.e.,

$$A_{c\sigma}^\dagger(\mathbf{k})|\Phi_0\rangle=\frac{1}{\sqrt{N}}\sum_I \alpha_{\mathbf{k}c\sigma}(I)A_{I\sigma}^\dagger|\Phi_0\rangle\,. \tag{2}$$

The coefficients α follow from a self-consistent-field (SCF) calculation. N is the number of bonds. Correlations are included by making the following ansatz for the wave function:

$$\begin{aligned}|\Psi_{c\sigma}(\mathbf{k})\rangle&=e^S A_{c\sigma}^\dagger(\mathbf{k})|\Phi_0\rangle\,,\\ S&=-\sum_{I,J}(\pi_{IJ}S_{IJ}^\pi+\eta_{IJ}S_{IJ}^\eta)\,.\end{aligned} \tag{3}$$

The operators

$$\begin{aligned}S_{IJ}^\pi&=\tfrac{1}{4}\sum_{\sigma,\sigma'}A_{J\sigma'}^\dagger B_{J\sigma'}A_{I\sigma}^\dagger A_{I\sigma}\,,\\ S_{IJ}^\eta&=\tfrac{1}{4}\sum_{\sigma,\sigma'}A_{J\sigma'}^\dagger B_{J\sigma'}A_{I\sigma}^\dagger B_{I\sigma}\,,\end{aligned} \tag{4}$$

describe the polarization cloud of the added electron and the reduction (by blocking) of the ground-state correlations. The parameters π_{IJ} and η_{IJ} are chosen as to minimize the energy. After some calculations the quasiparticle energy $\varepsilon_{c\sigma}(\mathbf{k})$ to lowest order in π_{IJ} and η_{IJ} is found to be [Ref. 7, Eqs. (37–39)]:

$$\begin{aligned}\varepsilon_{c\sigma}(\mathbf{k})&=\varepsilon_{c\sigma}^{\rm HF}(\mathbf{k})+\varepsilon_{c\sigma}^\pi(\mathbf{k})+\varepsilon_{c\sigma}^\eta(\mathbf{k})\,,\\ \varepsilon_{c\sigma}^\pi(\mathbf{k})&=-\sum_{I,J}\pi_{IJ}\langle A_{c\sigma}^\dagger(\mathbf{k})|H_{\rm res}S_{IJ}^\pi|A_{c\sigma}^\dagger(\mathbf{k})\rangle_{\rm con}\,,\\ \varepsilon_{c\sigma}^\eta(\mathbf{k})&=-\sum_{I,J}\eta_{IJ}\langle A_{c\sigma}^\dagger(\mathbf{k})|H_{\rm res}S_{IJ}^\eta|A_{c\sigma}^\dagger(\mathbf{k})\rangle_{\rm con}\,.\end{aligned} \tag{5}$$

The contribution $\varepsilon_{c\sigma}^{\rm HF}(\mathbf{k})$ is the quasiparticle energy in Hartree-Fock (HF) approximation. The expectation values can be represented by diagrams (Ref. 7). The index "con" implies that only connected contractions are taken when they are evaluated. They include all second-order processes and therefore also the exchange shown in Fig. 1(b). The correlation operators $\pi_{IJ}S_{IJ}^\pi$ and $\eta_{IJ}S_{IJ}^\eta$ may be considered as describing effective two-particle interactions, the strengths of which are proportional to the variational parameters π_{IJ} and η_{IJ}, respectively. The same calculations can be repeated by adding a hole instead of an electron to the system.

The blocking of ground-state correlations by the added electron is described by $\varepsilon_{c\sigma}^\eta(\mathbf{k})$ and is dominated by the intrabond parameter $\eta_0(=\eta_{II})$. The effect of the other η_{IJ} is to describe blocking of van der Waals correlations between bond I and bond J due to the added electron in bond I. It is small and can be neglected. The parameters $\pi_{IJ}(=\pi_{|I-J|})$ describe the effect of polarization of bond J when the added electron is in bond I. They fall off slowly with increasing bond separation. However, the exchange diagram contributions to $\varepsilon_{c\sigma}^\pi(\mathbf{k})$ contain as a factor a Coulomb interaction integral $V_{A_IB_JA_JA_I}$ (in standard notation involving bonding and antibonding states in bonds I and J). It depends on the overlap between those states in different bonds and is very small for nearest-neighbor bonds and more distant ones. Because of the minimal basis set which is used there is no polarization within the bond in which the added electron is positioned. The parameters η_0 and π_1 are determined according to Eq. (34) of Ref. 6 and Eq. (49) of Ref. 7. Thereby third-order processes and the dependence of π_1 on the quasiparticle energy $\varepsilon_{c\sigma}(\mathbf{k})$ are neglected. The following result is found for the change in the energy gap at the Γ point, $\Delta E_{\rm gap}$ due to the exchange diagrams:

$$\Delta E_{\rm gap}=-2\left[\frac{V_{A_0B_0A_0B_0}^2}{2(E_c-E_v)}-6\frac{(V_{A_0A_0A_1B_1}-\frac{1}{2}V_{A_0B_1A_1A_0})V_{A_0B_1A_1A_0}}{E_c-E_v}\right]. \tag{6}$$

The E_c and E_v are mean values of the energy of the conduction and valence bands, respectively. They may be obtained from a calculated HF band structure. The first term in Eq. (6) is the exchange contribution to the blocking of ground-state correlations within the bond in which the extra particle (hole) is positioned. The remaining term describes the effect of exchange on the polarization cloud of the added electron (hole) for the six neighboring bonds. The two contributions cancel each other partially.

EQUIVALENCE OF THE DIAGRAMS AND THEIR INTERPRETATION

The diagrams used in the two computational methods seem rather different at first sight. In the plane-wave calculation they always contain two Coulomb interaction lines. In the local calculations one line stands for the Coulomb interaction, while the second line stands for a correlation operator. The nature of this operator depends on the particular diagram. In case of diagrams E_p^e and E_p^h it is proportional to π_{IJ} while in the case of E_g^h and E_g^e it is proportional to η_{IJ}. As pointed out before, the π_{IJ} fall off slowly while the η_{IJ} fall off rapidly with increasing bond separation. In order to see how these different properties are reflected in the plane-wave calculations we study the behavior of the diagrams in the limit of small momentum transfer $\mathbf{q}$ by the Coulomb interaction line which replaces the correlation operator in the local diagrams [dotted line in Fig. 1(b)]. To be more specific we redraw in Fig. 2 the lower left vertices of diagrams E_p^e and E_g^h in Fig. 1(b) together with the momentum transfer through the incoming and outgoing lines. For $\mathbf{q}\rightarrow 0$ each vertex contributes a factor which is the inner product between the incoming and outgoing state. This implies for the vertex in Fig. 2(b) the inner product between an incoming hole state and an intermediate particle state at the same value of $\mathbf{k}$. This product vanishes because the valence-band state of the hole and the conduction-band state of the particle are orthogonal to each other. Consequently the q^{-2} singularity of the Coulomb interaction is suppressed and with it its long-range behavior. This corresponds to the fast decrease of η_{IJ} with increasing bond distance. The situation is different when the vertex in Fig. 2(a) is considered. Here the inner product between the identical conduction-band states does not vanish and the q^{-2} singularity reflects the slow decrease of π_{IJ} with increasing bond distance.

The signs of the energy contributions of the different diagrams follow from the physical interpretation which one can associate with each of them. For that purpose we show in Fig. 3 the remaining second-order diagrams. It is noticed that the diagrams E_p^e and E_p^h in Fig. 1(b) are the exchange versions of the diagrams D_p^e and D_p^h of Fig. 3(a). The latter correspond to processes in which the added particle (hole) polarizes its surroundings. They lower the energy of the added electron or hole, respectively. A lowering of the hole energy causes an increase in energy of the electron state in which the hole is situated. Therefore D_p^e is negative while D_p^h is positive. This implies that E_p^e and E_p^h are positive and negative, respectively, since they are exchange versions of the former diagrams. The diagrams E_g^e and E_g^h are the exchange corrections of the diagrams D_g^e and D_g^h in Fig. 3(a). In order to understand the sign of the energy contribution of the latter, consider the second-order diagram shown in Fig. 3(b). It describes the negative contribution to the ground-state energy which arises from van der Waals interactions between different bonds. When a particle or hole is added to the system, the van der Waals interactions are partially blocked due to Pauli's principle (diagrams D_g^e and D_g^h). This results in an energy increase for the particle or hole, respectively. By the same reasoning as before D_g^e is positive, E_g^e and E_g^h are therefore negative and positive, respectively.

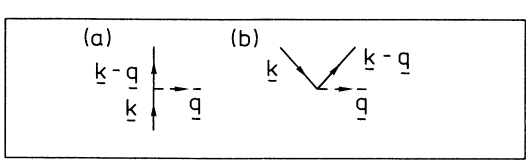

FIG. 2. Lower left vertices of E_p^e and E_g^h of Fig. 1(b). The incoming line represents the state for which the energy correction is determined. The outgoing line is an intermediate particle state.

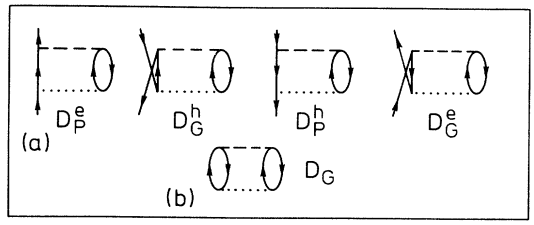

FIG. 3. Direct second-order diagrams. Their interpretation is used in order to discuss the signs of the various vertex corrections. Symbols have the same meaning.

RESULTS AND CONCLUSIONS

Table I contains the results for Si of the two calculations described before and also the gap corrections for diamond and Ge by the local ansatz. A 15 band model, to which we have to restrict ourselves, is not realistic enough for Ge and diamond. Therefore the corresponding values in the plane wave basis are missing. In the local calculation for Si the set of parameters listed in Table II is used. Displayed are the energy corrections due to the second-order exchange contributions for both the lowest conduction-band state and the highest valence-band state at the Γ point. The next to last column contains the net correction to the gap. If these corrections are compared with the values of the direct gaps in the last column it is seen that these corrections are small in all investigated covalent semiconductors. The introduction of a screened interaction will certainly lower their values still more. One notices that the various corrections have the expected sign in both calculations. The results of the plane-wave calcu-

TABLE I. Energy corrections to the lowest conduction-band state (E_g^e,E_p^e) at $\mathbf{k}=0$ and the highest valence-band state at $\mathbf{k}=0$ (E_g^h,E_p^h). All energies are in eV. The last two columns show the energy-gap correction and the direct gap (Ref. 7).

	Method	E_g^e	E_p^e	E_g^h	E_p^h	ΔE_{gap}	E_{gap}
Si	Plane wave	−0.46	0.73	0.49	−0.25	0.03*	3.4
	Local ansatz	−0.18	0.15	0.18	−0.15	−0.06	
C	Local ansatz	−0.30	0.23	0.30	−0.23	−0.14	7.4
Ge	Local ansatz	−0.14	0.11	0.14	−0.11	−0.06	0.9

TABLE II. Coulomb integrals and mean values for the HF conduction and valence bands of silicon (Ref. 7). The Coulomb integrals are scaled according to Ref. 6 from the corresponding values of diamond.

$V_{A_0B_0A_0B_0}=2.6$ eV	$E_c=11.4$ eV
$V_{A_0A_0A_1B_1}=1.7$ eV	$E_v=-6.9$ eV
$V_{A_0B_1A_1A_0}=0.3$ eV	

lation exceed those of the local ansatz by a factor of 2–5. This is not very surprising. On one hand bond orbitals of Si are sufficiently localized so that a calculation with 15 plane waves can give semiquantitative results only. On the other hand they are rather extended so that the BOA is relatively crude for that system. In the simplified version of the local ansatz particle-hole symmetry has been assumed. This restriction leads to distinct deviations in the results for the polarization diagrams E_p^e, E_p^h of both calculations.

In each covalent semiconductor the exchange contribution of the blocking of ground-state correlations is significantly canceled by the corresponding polarization term. Both processes are of the same order of magnitude because they arise from the direct neighborhood of the additional particle as Eq. (6) shows. As a result the gap correction is one order-of-magnitude smaller. The interplay between polarization and blocking also appears, for example, in the series of direct diagrams in Fig. 3(a), which are included in the *GW* approximation. However, in that case the contributions of the long-ranged polarization diagrams D_p^e, D_p^h clearly exceed the energy shifts due to the short-ranged blocking diagrams D_g^e, D_g^h. It is worth pointing out that the ground-state energy is considerably modified by lowest-order vertex corrections.[6] There is only one type of second-order exchange diagram and no cancellation takes place.

In conclusion both computational methods indicate the smallness of the lowest-order vertex contribution to the energy gap in covalent semiconductors as a consequence of the partial cancellation of different physical processes. The nature of the processes has been clarified. The fact that two independent calculations which are based on two completely different descriptions of the electron band structure came essentially to the same results provides a firmer basis for the findings of Ref. 5.

ACKNOWLEDGMENTS

One of us (R.D.) would like to acknowledge the hospitality of the Max-Planck-Institut für Festkörperforschung.

[1]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[2]W. v. d. Linden and P. Horsch, Phys. Rev. B **37**, 8351 (1988).

[3]R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

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

[5]R. Daling and W. van Haeringen, Phys. Rev. B **40**, 11659 (1989).

[6]W. Borrmann and P. Fulde, Phys. Rev. B **31**, 7800 (1985).

[7]W. Borrmann and P. Fulde, Phys. Rev. B **35**, 9569 (1987).