PHYSICAL REVIEW B VOLUME 24, NUMBER 4 15 AUGUST 1981

Band-gap narrowing in heavily doped many-valley semiconductors

K.-F. Berggren and B. E. Sernelius
Department of Physics and Measurement Technology, Linköping University, S-581 83 Linköping, Sweden

(Received 8 January 1981)

Band-gap shrinkage in impure n-Si and Ge as a function of impurity concentration is studied theoretically at $T = 0$ K. Above the Mott critical density it is assumed that electrons occupy the host conduction band in the form of an electron gas. Interactions among the free carriers and with the ionized point impurities give rise to a downward shift of the conduction band. Interband matrix elements are explicitly included for the valence bands. As a consequence, the two valence bands at the center of the zone are shifted upwards by the same amount. Correlation and impurity scattering both give corrections to the band gap, which are of the same order of magnitude. In particular, one finds that the band gap depends rather sensitively on the arrangement of donor ions. The choice of dielectric screening is investigated, and comparison with previous calculations is made. A comparison between present theoretical results and experimental estimates of the band-gap narrowing is also indicated. One notes an order-of-magnitude agreement, although there is considerable scatter in the experimental data. Shortcomings of the present theory are briefly discussed.

I. INTRODUCTION

The value of the energy gap E_g of heavily doped silicon is an important parameter in transistor design, as well as an intriguing theoretical problem in condensed-matter physics. Keyes's[1] summary of the problems associated with this fundamental parameter is very much to the point; for this reason we wish to cite him at some length:

"The effect of doping with donors or acceptors on the energy gap of semiconductors has both scientific and engineering, theoretical and practical importance. To the physicist it is a manifestation of subtle effects that are difficult to detect and measure. To the applied scientist, the energy gap is an essential parameter needed to design devices and to predict their performance; it determines, for example, the current-voltage characteristics of p-n junctions and the energy of electroluminescent recombination radiation. The value of the energy gap in heavily doped silicon has recently received considerable attention in connection with transistor design. Nevertheless, definitive values for this essential parameter of a well-studied material are still not available.

"The doping levels used in semiconducting devices have tended to increase with the progress of miniaturization. A basic reason for this is to be found in the limitation placed on transistor dimensions by depletion layers. p and n regions of semiconductors are separated by layers depleted of mobile charge but containing the charge of the fixed donors or acceptors. The electric fields produced by this fixed charge must support the potential difference between the n and p regions of a junction; the greater the charge density or doping level, the greater the electric fields, and the thinner the depletion layer needed to accommodate the potential difference. Thus, the increase in doping levels that accompanies miniaturization."

Among other current topics one may also mention solar cells. In this case band-gap narrowing presumably has a strong influence on the highest emitter efficiency that may be obtained for the highly-doped emitter regions in the cell.

Thus the change of the band structure of a semiconductor with doping has attracted considerable interest. Experiments and theory have recently been reviewed by Keyes[1] and Abram, Rees, and Wilson[2] to whom we refer for details and more extensive lists of references to previous work. Energy-gap narrowing of a very similar kind also occurs in intrinsic semiconductors under intense laser irradiation. The states filling in the conduction and valence bands are then strongly controlled by the energy-gap narrowing induced by the optically generated interacting electrons and holes.[3]

The basic feature of most theoretical models is that of an electron gas occupying the conduction band in the case of n-type materials, or a hole gas in the valence band of p-type materials. Such models are thus relevant for concentrations higher than the Mott critical density, $n_c^{1/3} a_H^* \simeq 0.25$ (a_H^* is the effective Bohr radius). The shift of the unperturbed band gap E_g^0 may be subdivided as

$$\Delta E_g = E_g^0 - E_g = (\Delta E_g)_1 + (\Delta E_g)_2 \,. \tag{1.1}$$

With this definition ΔE_g will be a *positive* quantity. $(\Delta E_g)_1$ denotes the shift due to direct influences of impurity centers such as local strain, etc. This term is supposed to be independent of carrier densities n (or at least slowly varying with n), but may change with the chemical nature of the impurity. However, $(\Delta E_g)_1$ seems not to be well understood in detail—in particular it should be sensitive to annealing. Previous theoretical work, as well as this work, rather focuses on the term $(\Delta E_g)_2$, which derives from the Coulomb interactions among the electrons themselves and their interaction with the ionized impurity centers. The particles can then approximately be described as noninteracting quasiparticles. Let the quasiparticle dispersion for the conduction band be

$$E_c(\vec{k},\omega) = \epsilon_c^0(\vec{k}) + \hbar\Sigma_c(\vec{k},\omega) \,, \tag{1.2}$$

where $\epsilon_c^0(\vec{k})$ is the unperturbed band energy and $\hbar\Sigma_c(\vec{k},\omega)$ the self-energy[4] associated with electron-electron and electron-impurity scattering. In the case of the valence band, however, the unperturbed energy $\epsilon_v^0(\vec{k})$ already contains a Hartree-Fock exchange contribution $\hbar\Sigma_{v,0}^{\mathrm{HF}}$ because the band is completely occupied. When adding free carriers, either electrons or holes, one is only interested in the *shifted* self-energy, i.e., the relevant quasiparticle energy is

$$E_v(\vec{k},\omega) = \epsilon_v^0 + \hbar\Sigma_v^*(\vec{k},\omega) \,, \tag{1.3}$$

where

$$\hbar\Sigma_v^*(\vec{k},\omega) = \hbar\Sigma_v(\vec{k},\omega) - \hbar\Sigma_{v,0}^{\mathrm{HF}}(\vec{k}) \,. \tag{1.4}$$

This point was first noted by Inkson.[5] As discussed by him and as will be shown below, the inclusion of $\hbar\Sigma_v^{*,ee}(\vec{k},\omega)$ gives rise to important corrections of the band gap. Since the quasiparticle energies are related to the energy required to remove or add particles, the shift of the band gap relative to the band edges is now

$$\Delta E_g = \hbar\Sigma_v^*(0) - \hbar\Sigma_c(0) \,. \tag{1.5}$$

(Index 2 is omitted here and in the following.) For heavily doped semiconductors the band gap is also frequently measured relative to the Fermi level ϵ_F, giving the "Burstein shift"; in optical absorption and n-type materials excitations must end up in unoccupied states above the Fermi level. This shift is then

$$\epsilon_F + \hbar\Sigma_c(k_F) - \hbar\Sigma_v^*(0) \,. \tag{1.6}$$

The effective *reduction* of the energy gap as measured at the Fermi level is consequently

$$\hbar\Sigma_c(k_F) - \hbar\Sigma_v^*(0) \,. \tag{1.7}$$

With this definition the sign is *opposite* to that one above. Considerable work has been devoted to many-body calculations of the self-energies using various approximations and for various materials (see, e.g., Ref. 2). One common feature to such calculations is that potential fluctuations due to local fluctuations in the density of impurities are neglected. The calculated energy gap is therefore a local energy gap. A subsequent averaging over the statistical fluctuations may be performed at the end of the calculations leading to band tailing.[2]

A simple and frequently used approximation of Eqs. (1.5)–(1.7) is to insert the Hartree-Fock self-energies. For an n-type material one then obtains (κ is the dielectric constant of the host)

$$\Delta E_g = -\Sigma_c^{\mathrm{HF}}(0) = 2e^2 k_F/\kappa\pi \,, \tag{1.8}$$

since in this case there is no shift in the Hartree-Fock energy of the valence band as free carriers are added to the conduction band. Equation (1.8) refers to isotropic conduction-band minima and

$$k_F = (3n\pi^2/\nu)^{1/3} \,, \tag{1.9}$$

if there are ν minima; below we will discuss the influence of anisotropy.[6] The exchange contribution to the energy-gap shrinkage as measured at the Fermi level is, from Eq. (1.7), equal to $-e^2 k_F/\kappa\pi$. Because experiments in some cases indicate an $n^{1/3}$ dependence, this expression and Eq. (1.8) have been used to interpret experimental data. At the same time, though, other many-body effects give changes of the same order of magnitude.[7]

A first step to improve the Hartree-Fock result is obviously to extend it to a correlated homogeneous electron gas in a smeared background of ionized impurities. This is the essence of Inkson's[5] model as well as some previous ones.[2,8,9] The calculations in Refs. 8 and 9 suffer, however, from the same defect. The correlation and exchange energy changes of the appropriate band are considered by treating the carriers as a quasi-free-electron (hole) gas and then applying the appropriate approximations to obtain the energy shift of that gas, the changes in the other bands being neglected. This is, as Inkson stresses, incorrect since the presence of free carriers can change the energies of all of the bands. It is only the difference between valence- and conduction-band shifts which will be measured.

A second step to improve on the Hartree-Fock result is to include the scattering of free carriers from the ionized impurities together with correlation effects. Thus Wolf,[10] Bonch-Bruevich,[11] Rogachev and Sablina,[12] Rimbey and Mahan,[13] and Bergersen *et al.*[14] have performed such calculations. As it seems, these calculations suffer from the same defect as above.

In the present work we will investigate the case of heavily doped *n*-type germanium and silicon[15] with the purpose of theoretically evaluating ΔE_g as closely as possible by means of presently-available many-body techniques. The reason for this is that such detailed calculations for materials of this kind have been lacking in the literature. The purpose is also to investigate the effects of anisotropies of the conduction band and the structure of the valence band as well as the effects of impurity scattering. Inkson's calculations, which ignore these effects, refer to a simple plasmon-pole approximation for the dielectric screening. Since this model grossly exaggerates[2] ΔE_g, we will instead extend the calculations to include the appropriate random-phase (RPA) approximation dielectric screening. For the conduction band we will also investigate the full density of states for the quasiparticles. All calculations are performed for $T = 0$ K. It should be stressed that all the calculations reported below refer to a homogeneous system. For example, band tailing associated with statistical fluctuations has to be considered separately as mentioned above.

In Sec. II we will summarize some basic notations and discuss relevant matrix elements. In Sec. III we will give a rather detailed recapitulation of Inkson's model. Because of the simple form of the dielectric function used, the calculations can be carried out with ease in this case and are therefore quite revealing. The inclusion of band-structure effects and impurity scattering in the Inkson model are also discussed. Section IV gives corresponding results for RPA screening as well as for the quasiparticle density of states. Since no definitive experimental values of ΔE_g are available, the main emphasis of the present work is on theory, although some comparison with experimental values is indicated. Section V, finally, contains a short summary and discussion.

After a first version of this manuscript was completed and circulated for comments we become aware of some very recent theoretical work. Thus Mahan[16] has calculated the impurity dependence of the energy gap in Si and Ge with the same purpose as here, i.e., primarily a systematic theoretical examination of all the terms contributing to ΔE_g rather than a confrontation with experimental data. As emphasized by Mahan, previous workers have tended to single out only one particular term, and then claimed agreement with experiments. A more systematic theoretical approach is consequently needed. Although similar in spirit, Mahan's work and the present one differ considerably in detail. The two theories are therefore compared at some length in Sec. IV. In particular, we find that band-gap narrowing depends sensitively on the arrangement of donors. We also wish to mention the recent work on GaAs by Serre, Ghazali, and Leroux Hugon,[17] who focused attention on the relative importance of multiple-impurity scattering and impurity-concentration fluctuations. Their work is briefly commented on in Sec. V. Finally, we mention recent absorption measurements on doped silicon by Schmid.[18] This author also theoretically estimated the band-gap narrowing using arguments rather similar to ours and Mahan's.[16] His results did not include, however, the shift in the valence band.

II. GENERAL NOTATIONS FOR THE GREEN'S FUNCTION AND SELF-ENERGIES

This section is given for convenience. It summarizes notations and states the approximations used for the matrix elements.

If $\{ \psi_j(\vec{x},t) \}$ is the set of one-electron functions of the system, we write the unperturbed Green's function[4] as the diagonal expansion

$$G^0(\vec{x},t_i\vec{x}\,',t') = \int \frac{d\omega}{2\pi} \sum_j \psi_j(\vec{x},t)\psi_j^*(\vec{x}\,',t') \times e^{-i\omega(t-t')} G^0(j,\omega) . \tag{2.1}$$

The transform is

$$G^0(j,\omega) = 1/(\omega - \epsilon_j^0/\hbar + i\delta) , \tag{2.2}$$

where $\delta < 0$ for occupied states and $\delta > 0$ for empty states; ϵ_j^0 is the unperturbed one-particle energy.

We shall consider modifications in the one-electron energies (self-energies) due to electron-electron and electron-donor ion interactions, V_{ee} and V_{ei}, respectively, by means of perturbation theory. Depending on the particular numerical values of the matrix elements various channels are available. For V_{ee} there is generally the exchange-type integral

$$\langle jk \mid V_{ee} \mid kj \rangle = \int d\vec{x} \int d\vec{x}\,'\psi_j^*(\vec{x})\psi_k^*(\vec{x}\,') \times V_{ee}(\vec{x},\vec{x}\,')\psi_k(\vec{x})\psi_j(\vec{x}\,') \; . \tag{2.3}$$

Dealing with particles in the conduction and valence bands of, for example, Si the one-electron functions are subdivided into sets $\{ \psi^n_{\vec{k}}(\vec{x}) \}$ where n is the band index and $\vec{k}$ the crystal momentum. The Green's function is then denoted $G_n^0(\vec{k},\omega)$. The conduction-band electrons occupy ν minima and the associated wave functions are $U_{\vec{k}_\nu}(\vec{r})\exp(i\vec{k}_\nu \cdot \vec{r})$ where $U_{\vec{k}_\nu}(\vec{r})$ is the periodic part which is assumed to vary slowly. We therefore write approximately for the conduction-band matrix elements

$$\langle \vec{k}_\nu,\vec{k}\,'_{\nu'} \mid V_{ee} \mid \vec{k}\,'_{\nu'}\vec{k}_\nu \rangle \simeq \frac{4\pi e^2}{\kappa \mid \vec{k}_\nu - \vec{k}\,'_{\nu'} \mid^2}\delta_{\nu\nu'} \; , \tag{2.4}$$

i.e., transitions between different minima and the variations in $U_{\vec{k}_\nu}(\vec{r})$ are ignored. For the valence band we assume the following simplified band structure.[19,20] At the center of the zone there are two degenerate spherical bands with effective masses m_h and m_l (heavy and light holes). Below these bands there is a third band split off by spin-orbit interactions. This band is sufficiently lower in energy and therefore ignored. Furthermore, matrix elements between valence- and conduction-band states may effectively be included as a background dielectric constant, i.e., $e^2 \rightarrow e^2/\kappa$. Transitions between the light- and heavy-hole bands are somewhat more complicated. Following Combescot and Nozières[19,20] one may approximately include the structure of the valence bands by introducing averaged matrix elements

$$\langle \vec{k}_n,\vec{k}\,'_{n'} \mid V_{ee} \mid \vec{k}\,'_{n'},\vec{k}_n \rangle \simeq v(\vec{k} - \vec{k}\,')\Lambda_{nn'}(\vec{k},\vec{k}\,') \; , \tag{2.4$'$}$$

where

$$v(\vec{q}) = 4\pi e^2/\kappa q^2 \; . \tag{2.5}$$

The "vertex functions" $\Lambda_{nn'}$ are defined as

$$\Lambda_{hh}(\vec{k},\vec{k}\,') = \Lambda_{ll}(\vec{k},\vec{k}\,') = \tfrac{1}{4}(1 + 3\cos^2\theta) \; , \tag{2.6a}$$

$$\Lambda_{lh}(\vec{k},\vec{k}\,') = \tfrac{3}{4}\sin^2\theta \; , \tag{2.6b}$$

where θ is the angle between $\vec{k}$ and $\vec{k}\,'$. Λ_{ll} and Λ_{hh} regulate transitions between light-hole states and heavy-hole states, respectively; Λ_{lh} refer to transitions between light- and heavy-hole states.

With these preliminaries we may now state the self-energies $\hbar\Sigma^{ee}$ due to electron-electron interactions. For the conduction-band states one has

$$\hbar\Sigma_c^{ee}(\vec{k},\omega) = \frac{i}{(2\pi)^4} \int d\vec{q} \int d\omega' v_{\text{sc}}(\vec{q},\omega') \times G_c^0(\vec{k} + \vec{q},\omega + \omega') \; , \tag{2.7}$$

where

$$v_{\text{sc}}(\vec{q},\omega') = \frac{4\pi e^2}{\kappa q^2 \epsilon(q,\omega')} \tag{2.8}$$

is the dielectrically screened interaction $v(q)$ and

$$\epsilon(q,\omega) = 1 - v(q)\sum_\nu \int \frac{d\vec{k}}{(2\pi)^3} \frac{n_\nu(\vec{k})[1 - n_\nu(\vec{k} + \vec{q})]4[\epsilon_\nu^0(\vec{k} + \vec{q}) - \epsilon_\nu^0(\vec{k})]}{(\omega + i\delta)^2 - [\epsilon_\nu^0(\vec{k} + \vec{q}) - \epsilon_\nu^0(\vec{k})]^2} \; . \tag{2.9}$$

Here $\epsilon_\nu^0(\vec{k})$ are the unperturbed one-particle energies associated with the νth conduction-band minimum. For the valence bands one obtains the somewhat more complicated expression

$$\hbar\Sigma_{v,n}^{*,ee}(\vec{k},\omega) = \frac{i}{(2\pi)^4}\sum_{n'} \int d\vec{q} \int d\omega'[v_{\text{sc}}(\vec{q},\omega') - v(q)]\Lambda_{nn'}(\vec{k},\vec{k} + \vec{q})G_{n'}^0(\vec{k} + \vec{q},\omega + \omega') \; . \tag{2.10}$$

The self-energy associated with scattering off the impurity centers $\hbar\Sigma^{ei}$ involves the matrix elements $\langle \vec{k}_n \mid V_{ei} \mid \vec{k}\,'_{n'} \rangle$. As above we may neglect intervalley scattering in the conduction band as well as scattering between the conduction and valence bands. If we assume that

$$V_{ei}(\vec{r}) = \sum_g \omega(\vec{r} - \vec{R}_g) \tag{2.11}$$

where ω is the potential from an impurity ion at $\vec{R}_g$, one obtains from second-order perturbation theory in the screened potential

$$\hbar\Sigma_c^{ei}(\vec{k}) = n\int \frac{d\vec{q}}{(2\pi)^3}\left|\frac{\omega(\vec{q})}{\epsilon(q,0)}\right|^2 \bigg/ [\epsilon_v^0(\vec{k}) - \epsilon_v^0(\vec{k}+\vec{q})]\,, \tag{2.12}$$

where n is the impurity concentration. To arrive at this expression it is assumed that the impurities are distributed at random. For the valence bands the corresponding expression is

$$\hbar\Sigma_{v,n}^{ei}(\vec{k}) = n\sum_{n'}\int \frac{d\vec{q}}{(2\pi)^3}\left|\frac{\omega(\vec{q})}{\epsilon(q,0)}\right|^2 \Lambda_{nn'}(\vec{k},\vec{k}+\vec{q})/[\epsilon_n^0(\vec{k}) - \epsilon_{n'}^0(\vec{k}+\vec{q})]\,. \tag{2.13}$$

Here $\epsilon_n^0(\vec{k})$ are unperturbed one-particle energies associated with the nth valence band.

The proper sign of $\hbar\Sigma_v^{ei}$ may perhaps be confusing. It is therefore instructive to discuss the role of the self-energy in the following elementary way. Let us consider the valence band and for simplicity ignore the band structure by letting $\Lambda = 1$. For the moment we also ignore electron interactions. The total energy associated with the scattering of the valence-band electrons off the impurities is, according to ordinary second-order-perturbation theory,

$$E_v^{(2)} = \sum_{I\neq 0} \frac{|\langle I|\mathscr{V}_{ei}|0\rangle|^2}{(E_0 - E_I)}\,, \tag{2.14}$$

where $\mathscr{V}(\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_N;\vec{R}_1,\ldots,\vec{R}_N)$ denotes the total electrostatic interaction between all valence electrons $\{\vec{r}_i\}$ and donor ions $\{\vec{R}_i\}$; because of the presence of the mobile conduction electrons $\mathscr{V}_{ei}$ is dielectrically screened. As above we assume the donors to be randomly distributed. The Madelung term in the total energy then equals zero. Consequently we only have to bother about the contribution $E_v^{(2)}$ above.

The unperturbed state $|0\rangle$ and the intermediate states $|I\rangle$ in Eq. (2.14) are many electron states in the form of Slater determinants. The corresponding configurations are $\{\vec{k}\}_0$ and $\{\vec{k}\}_I$. Hence, the total energies of the unperturbed and intermediate states are

$$E_0 = \sum_{\{k\}_0} n_{\vec{k}}\epsilon_v^0(\vec{k}) \tag{2.15}$$

and

$$E_I = \sum_{\{\vec{k}\}_I} n_{\vec{k}}\epsilon_v^0(\vec{k})\,, \tag{2.16}$$

where $\epsilon_v^0(\vec{k})$ represent the unperturbed one-electron energies associated with the valence band. The terms in Eq. (2.14) describe the scattering of an electron initially in the one-particle state $|\vec{k}\rangle$ into another one-particle state $|\vec{k}+\vec{q}\rangle$ and back to $|\vec{k}\rangle$ again. Thus

$$E_0 - E_I = \epsilon_v^0(\vec{k}) - \epsilon_v^0(\vec{k}+\vec{q})\,. \tag{2.17}$$

This process is only allowed if state $|\vec{k}+\vec{q}\rangle$ is empty. We thus obtain for the second-order correction

$$E_v^{(2)} = \sum_{\vec{k},\vec{q}} n_{\vec{k}}(1 - n_{\vec{k}+\vec{q}})\frac{|V_{ei}^{\mathrm{sc}}(\vec{q})|^2}{\epsilon_v^0(\vec{k}) - \epsilon_v^0(\vec{k}+\vec{q})}\,, \tag{2.18}$$

where $V_{ei}^{\mathrm{sc}}(\vec{q})$ is the Fourier transform of Eq. (2.11) divided by the screening function. For a completely filled valence band, $E_v^{(2)}$ obviously equals zero because there are no empty states to which the electron can be scattered. If we now, however, make a small change of the occupation numbers by, e.g., the removal of a few electrons from the valence band such that the new occupation numbers become $\{n'_{\vec{k}}\}$, the corresponding second-order correction to the total energy is

$$\begin{aligned} E_v^{(2)}\{n'_{\vec{k}}\} &\simeq E_v^{(2)}\{n_{\vec{k}}\} + \sum_{\vec{k}} \frac{\delta E_v^{(2)}\{n_{\vec{k}}\}}{\delta n_{\vec{k}}}(n'_{\vec{k}} - n_{\vec{k}}) \\ &= E_v^{(2)}\{n_{\vec{k}}\} + \sum_{\vec{k}}\left[\sum_{\vec{q}} |V_{ei}(\vec{q})|^2/[\epsilon_v^0(\vec{k}) - \epsilon_v^0(\vec{k}+\vec{q})]\right](n'_{\vec{k}} - n_{\vec{k}})\,. \end{aligned} \tag{2.19}$$

The expression within the large parentheses is just the self-energy $\hbar\Sigma_v^{ei}$ as defined in Eq. (2.13) in more detail. Thus, if the total energy of the state with a completely filled valence band is $E_v\{n_{\vec{k}}\}$ then the corresponding energy of the state with a single hole in $|\vec{k}\rangle$ is consequently given by

$$E_v\{n'_{\vec{k}}\} = E_v\{n_{\vec{k}}\} - \epsilon_v^0(\vec{k}) - \hbar\Sigma_v^{ei}(\vec{k}) \, . \tag{2.20}$$

We may now argue about the system of conduction-band electrons in a similar way. Let $E_c^{(2)}\{n_{\vec{k}}\}$ be the corresponding second-order correction to the ground state of this system. Furthermore, let the electron removed from the valence band end up at the Fermi level in the conduction band, i.e.,

$$E_c^{(2)}\{n'_{\vec{k}}\} = E_c^{(2)}\{n_{\vec{k}}\} + \hbar\Sigma_c^{ei}(k_F) \, . \tag{2.21}$$

If $E_t\{n_{\vec{k}}\}$ is the ground state of the entire system, valence- as well as conduction-band electrons, the energy of the excited state is thus

$$\begin{aligned} E_t\{n'_{\vec{k}}\} = E_t\{n_{\vec{k}}\} &- \epsilon_v^0(\vec{k}) - \hbar\Sigma_v^{ei}(\vec{k}) \\ &+ \epsilon_c^0(k_F) + \hbar\Sigma_c^{ei}(k_F) \, . \end{aligned} \tag{2.22}$$

The difference between the excited and initial states, $E_t\{n'_{\vec{k}}\}$ and $E_t\{n_{\vec{k}}\}$, respectively, therefore contains the shift $\hbar\Sigma_c^{ei}(k_F) - \hbar\Sigma_v^{ei}(\vec{k})$, which should be compared with expression (1.7). In order to establish the expression for ΔE_g we may now go one step further by letting an electron at the bottom of the conduction band ($k = 0$) recombine with the hole. The total energy of this second excited state is obviously

$$\begin{aligned} E_t\{n'_{\vec{k}'}\} = E_t\{n_{\vec{k}}\} &+ \epsilon_c^0(k_F) + \hbar\Sigma_c^{ei}(k_F) \\ &- \epsilon_c^0(0) - \hbar\Sigma_c^{ei}(0) \, . \end{aligned} \tag{2.23}$$

The difference between Eqs. (2.22) and (2.23) thus defines an effective emission energy as

$$\epsilon_c^0(0) - \epsilon_v^0(\vec{k}) + \hbar\Sigma_c^{ei}(0) - \hbar\Sigma_v^{ei}(\vec{k}) \, . \tag{2.24}$$

Choosing $\vec{k} = 0$ we recover the shift in the band gap, Eq. (1.5), according to its definition in Eq. (1.1).

III. INKSON'S MODEL FOR ΔE_g; EFFECTS OF ANISOTROPY AND IMPURITY SCATTERING

Because of its simplicity it is illuminating to recapitulate Inkson's[5] model for the band-gap narrowing. In this model it is assumed that the impurity ions may be smeared into a uniform neutralizing background and that free electrons occupy the conduction band in the case of n-type materials. The anisotropy of the conduction-band minima and the structure at the top of the valence band are ignored. Below we will therefore consider these effects. We will also show that electron-donor ion interactions may also give rise to a substantial shift of the band gap. All calculations refer to the bottom of the conduction band and the top of the valence band. In the following we have n-Si and n-Ge in mind.

The self-energy correction for the conduction-band electrons due to electron-electron interactions is given by Eq. (2.7). If the plasmon-pole approximation[21] for $\epsilon(q,\omega)$ is used the calculations become elementary. Then

$$\epsilon(q,\omega)^{-1} = 1 + \frac{\omega_p^2}{2\omega_1(q)} \left[\frac{1}{\omega - \omega_1(q) + i\delta} - \frac{1}{\omega + \omega_1(q) - i\delta} \right]_{\delta = 0^+} , \tag{3.1}$$

where the plasmon frequency for a many-valley situation is

$$\omega_p^2 = 4\pi e^2 n / (\kappa m_{op}) \, , \tag{3.2}$$

with

$$m_{op}^{-1} = \frac{1}{3} \left[\frac{2}{m_t} + \frac{1}{m_l} \right] . \tag{3.3}$$

The transverse and longitudinal masses are m_t and m_l, respectively. The plasmon dispersion is approximately

$$\omega_1^2(q) \simeq \omega_p^2 + \frac{2}{3} \frac{\epsilon_F}{m_{op}} q^2 \, , \tag{3.4}$$

where the coefficient in the q^2 term is adjusted such that the correct Thomas-Fermi expression obtains at $\omega = 0$ and $q \to 0$. Inserting Eq. (3.1) and the Green's function

$$G_c^0(\vec{k},\omega) = 1/[\omega - \epsilon_v^0(\vec{k}) + i\delta] \tag{3.5}$$

in the general expression for $\hbar\Sigma_c^{ee}(\vec{k},\omega)$, Eq. (2.7), and letting $\omega \simeq \epsilon_v^0(\vec{k})/\hbar$ one gets

$$\hbar\Sigma_c^{ee}(0,0) = -\frac{2e^2k_F}{\kappa\pi} - \frac{e^2\lambda}{2\kappa}\left[1 - \frac{4}{\pi}\tan^{-1}\left[\frac{k_F}{\lambda}\right]\right], \tag{3.6}$$

where λ is the inverse Thomas-Fermi screening length defined through

$$\lambda^2 = \frac{4\nu}{a_H^*}\left[\frac{3n}{\pi\nu}\right]^{1/3}. \tag{3.7}$$

The effective Bohr radius is

$$a_H^* = \kappa\hbar^2/m_d e^2\,, \tag{3.8}$$

and the density-of-states effective mass is

$$m_d = (m_t^2 m_l)^{1/3}\,. \tag{3.9}$$

In Eq. (3.6) the first term on the right-hand side is simply the Hartree-Fock (HF) self-energy; the second term derives from the poles in $\epsilon(q,\omega)^{-1}$ and the third from the poles in the Green's function. In deriving the expression in Eq. (3.6) it is furthermore assumed that differences between energies $\epsilon_\nu^0(\vec{k})$ can be ignored in comparison with $\hbar\omega_1(k)$. This is equivalent to writing the self-energy in the somewhat more general form

$$\hbar\Sigma_c^{ee}(0,0) = -\int \frac{d\vec{q}}{(2\pi)^3}\frac{v(q)}{\epsilon(q,0)}n_\nu(\vec{q}) - \frac{e^2\lambda}{2\kappa}\,. \tag{3.10}$$

We will return to this expression below.

Let us now consider the valence band which we, like Inkson, first assume to be nondegenerate, isotropic, and completely filled. Thus ignoring the "vertex functions" in Eq. (2.10) we have

$$\hbar\Sigma_v^{*,ee}(0,0) = e^2\lambda/(2\kappa)\,. \tag{3.11}$$

The actual contribution from the poles in $\epsilon(q,\omega)^{-1}$ is here the same as in Eq. (3.10), as pointed out also by Inkson.[5] This expression follows from Eq. (3.6) by omitting the Hartree-Fock term and formally letting $k_F \to \infty$ (λ is small in comparison with the dimensions of the Brillouin zone). In contrast to the conduction band the valence band is thus shifted upwards when electrons are added to the conduction band. The self-energy $\hbar\Sigma_v^{*,ee}$ is positive because it merely describes the fact that free carriers in the conduction band will weaken the exchange interaction among the electrons in the valence band. The shift in the band gap due to electron-electron interactions as defined in Eq. (1.5) is now

$$(\Delta E_g)^{ee} = \frac{2e^2k_F}{\kappa\pi}\left[1 + \frac{\lambda\pi}{2k_F} - \frac{\lambda}{k_F}\tan^{-1}\left[\frac{k_F}{\lambda}\right]\right]. \tag{3.12}$$

This is exactly Inkson's result, which Abram *et al.*[2] have investigated extensively in relation to experiments. One notes that in the derivation of Eq. (3.12) the contributions from the poles in $\epsilon(q,\omega)^{-1}$ cancel exactly. Numerical results for n-type silicon are given in Fig. 1. These results show that the shift due to electron correlation is substantial.

The considerations above refer to isotropic band structures. It is now easy to extend Inkson's result in Eq. (3.6) by integrating over an ellipsoidal Fermi surface in Eq. (3.10). If $\gamma = m_t/m_l \leq 1$, one then obtains

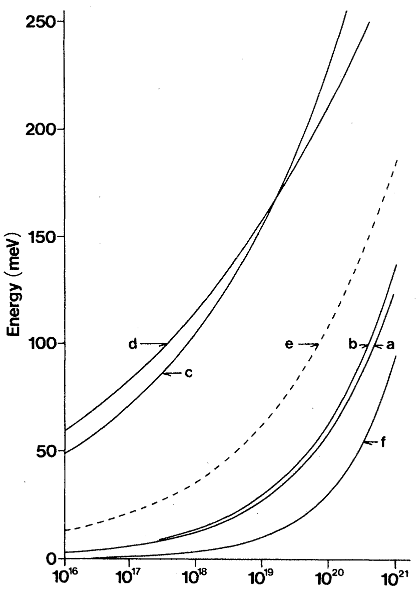

FIG. 1. Band-gap shrinkage in n-type silicon as a function of impurity concentration. ΔE_g is here defined as the difference between the unperturbed and the perturbed band gap; hence $\Delta E_g > 0$. Curves (a) and (b) give the band-gap narrowing for a homogeneous electron gas in the conduction band as computed in the Hartree-Fock approximation; (a) refers to isotropic and (b) to anisotropic conduction-band minima. Curves (c) and (d) refer to a correlated electron gas within the plasmon-pole (Thomas-Fermi) approximation; (c) is for isotropic and (d) for anisotropic minima. Curve (e) is the same electron gas in the RPA approximation as calculated by Abram *et al.* (Ref. 2) for isotropic bands. Curve (f) gives the shift due to electron-impurity ion scattering. The model is only strictly valid for $n > 10^{18}$ cm^{-3}. The numerical results are based on the input parameters $\nu = 6$, $\kappa = 11.40$, $m_l = 0.9163m_0$, $m_t = 0.1905m_0$, $m_{hh} = 0.523m_0$, and $m_{lh} = 0.154m_0$.

$$\hbar\Sigma_c^{ee}(0,0) = -\frac{2e^2}{\kappa\pi}k_F\gamma^{-1/3}\left[\frac{(\lambda/k_F)^2\gamma^{2/3}+\gamma}{1-\gamma}\right]^{1/2}\tan^{-1}\left[\left[\frac{1-\gamma}{(\lambda/k_F)^2\gamma^{2/3}+\gamma}\right]^{1/2}\right]$$
$$-\frac{e^2\lambda}{2\kappa}\left[1-\frac{4}{\pi}\tan^{-1}\left[\frac{k_F}{(\gamma^{1/3}\lambda)}\right]\right], \qquad (3.13)$$

which for $\gamma \to 1$ reduces to Eq. (3.6). If we on the other hand let $\lambda \to 0$ the Hartree-Fock expression for an ellipsoidal band is obtained,

$$\hbar\Sigma_c^{\mathrm{HF}}(0,0) = -\frac{2e^2}{\kappa\pi}k_F\gamma^{1/6}\frac{\sin^{-1}(1-\gamma)^{1/2}}{(1-\gamma)^{1/2}}, \qquad (3.14)$$

an expression considered by Bonch-Bruevich and Rozman.[6] In contrast to the conduction-band edge the valence band at $k = 0$ remains unchanged when the proper expression is considered. At $k = 0$ the vertices simply add to one. The band-gap narrowing due to electron correlation is therefore

$$(\Delta E_g)^{ee} = \frac{2e^2k_F}{\kappa\pi}\left\{\gamma^{-1/3}\left[\frac{(\lambda/k_F)^2\gamma^{2/3}+\gamma}{1-\gamma}\right]^{1/2}\tan^{-1}\left[\left[\frac{1-\gamma}{(\lambda/k_F)^2\gamma^{2/3}+\gamma}\right]^{1/2}\right]\right.$$
$$\left.+\frac{\lambda\pi}{2k_F}-\frac{\lambda}{k_F}\tan^{-1}\left[\frac{k_F}{\gamma^{1/3}\lambda}\right]\right\}. \qquad (3.15)$$

The effects of anisotropy are given numerically in Fig. 1. One notes that such effects are rather small in the case of n-Si. In the next section, which will deal with a more accurate description of the screening, we may then ignore anisotropy.

So far we have only studied the homogeneous interacting electron gas. In the Thomas-Fermi approximation for $\epsilon(q,0)$ one obtains for electron-impurity scattering the conduction-band shift [Eq. (2.12)]

$$\hbar\Sigma_c^{ei}(0) = -\frac{4\pi n e^2}{a_H^*\kappa\lambda^3}. \qquad (3.16)$$

To arrive at this expression we have assumed Coulombic scatterers, i.e., $\omega(q) = -4\pi e^2/\kappa q^2$. For the top of the valence bands one finds in the same way, but with the vertices included as in Eq. (2.13), that the heavy- and light-hole bands (with masses m_{hh} and m_{lh}) are shifted by the same amount

$$\hbar\Sigma_{v,n}^{ei}(0) = -\frac{m_{hh}+m_{lh}}{2m_d}\hbar\Sigma_c^{ei}(0). \qquad (3.17)$$

This contribution to the self-energy for the valence states is thus positive. This can be understood in the following way. An electron filling up the last empty state in the valence band will in doing so prevent the other valence electrons from reducing their energy through polarization and relaxation around the impurity centers. This implies that this electron gets a positive self-energy contribution from the electron-ion interaction and hence the corresponding hole gets a negative one. The total contribution to ΔE_g from electron-impurity scattering is therefore

$$(\Delta E_g)^{ei} = \frac{4\pi n e^2}{a_H^*\kappa\lambda^3}\left[1+\frac{m_{hh}+m_{lh}}{2m_d}\right]. \qquad (3.18)$$

Numerical results are given in Fig. 1. At lower concentrations impurity scattering has evidently a negligible effect on the band-gap narrowing. With increasing concentrations, however, impurity scattering becomes substantial.

All the calculations above make use of a simple plasmon-pole approximation and Thomas-Fermi expression for the dielectric function. The simplified form in Eq. (3.10) suggests, however, that this result may be improved by using, e.g., the Lindhard (RPA) expression for $\epsilon(\vec{q},0)$. It may be improved further by also including the frequency dependence in the screening by the replacement $\epsilon(\vec{q},0) \to \epsilon(q,[\epsilon_v^0(\vec{q})-\epsilon_v^0(0)]/\hbar)$ in Eq. (3.10) and correspondingly in the expression for the valence-band shift. Calculations of this kind have been performed by Abram *et al.*[2] who assumed a single, isotropic valence band. Their results are reproduced in Fig. 1. Clearly the simple form for the static dielectric screening used in this section grossly exaggerates ΔE_g. At the same time a single isotropic band is an oversimplification. In reality, exchange coupling of the light- and heavy-hole bands will change the valence-band shift. In the next section we therefore include this kind of coupling and the frequency dependence in a proper way. Although the expressions then become considerably more complicated the structure is very much the same as in this section.

IV. BAND-GAP NARROWING IN RPA

The calculations in the preceding section give order of magnitudes. We will here investigate the effects of including the full frequency dependence in $\epsilon(q,\omega)$. In doing so we will assume that the conduction-band minima are isotropic, since it was shown above that anisotropy only gives minor corrections.

We first consider the conduction band. It is convenient to define the variables

$$\overline{\omega}(\overline{k}) = \epsilon_v^0(k)/4\epsilon_F = \overline{k}^2 , \tag{4.1}$$

where $\overline{k} = k/2k_F$ and

$$\overline{\omega} = \hbar\omega/4\epsilon_F . \tag{4.2}$$

The self-energy associated with electron correlation is then

$$\hbar\Sigma_c^{ee}(k,\omega) = -\frac{4\pi e^2}{\kappa}2k_F \int \frac{d^3\overline{q}}{(2\pi)^3} \int \frac{d\overline{\omega}'}{2\pi i}\frac{1}{\overline{q}^2}\left[\frac{1}{\widetilde{\epsilon}(\overline{q},\overline{\omega}')} - 1\right]\overline{G}_c^0(\overline{k}+\overline{q},\overline{\omega}+\overline{\omega}') , \tag{4.3}$$

where

$$\overline{G}_c^0(\overline{k},\overline{\omega}) = 1/[\overline{\omega} - \overline{\omega}(\overline{k}) + i\delta] , \tag{4.4}$$

and $\widetilde{\epsilon}$ is the dielectric function in the reduced variables $\overline{k},\overline{\omega}(\overline{k})$ and $\overline{\omega}$. In Eq. (4.3) we have for reasons of convergency subtracted the HF contribution. In order to compute the integral it is now convenient to deform the path of integration as described by Rice[22] and as illustrated in Fig. 2. As a result one obtains

$$\begin{aligned}\hbar\Sigma_c^{ee}(k,\omega) = -\frac{4\pi e^2}{\kappa}2k_F &\left[\int \frac{d^3\overline{q}}{(2\pi)^3}\frac{1}{\overline{q}^2}\int_{-\infty}^{+\infty}\frac{d\overline{\omega}'}{2\pi}\overline{G}_c^0(\overline{k}+\overline{q},\overline{\omega}+i\overline{\omega}')\left[\frac{1}{\widetilde{\epsilon}(\overline{q},i\overline{\omega}')} - 1\right]\right.\\ &+\int \frac{d^3\overline{q}}{(2\pi)^3}\frac{1}{\overline{q}^2}\{n_v(\overline{k}+\overline{q})\eta(\overline{\omega}(\overline{k}+\overline{q})-\overline{\omega}) - [1-n_v(\overline{k}+\overline{q})]\eta(\overline{\omega}-\overline{\omega}(\overline{k}+\overline{q}))\}\\ &\left.\times\left[\frac{1}{\widetilde{\epsilon}(\overline{q},-\overline{\omega}+\overline{\omega}(\overline{k}+\overline{q}))} - 1\right]\right] ,\end{aligned} \tag{4.5}$$

where $\eta(x) = 1$ for $x > 0$ and zero otherwise. This expression is properly convergent. The last integral derives from the poles within the closed loops in Fig. 2. In order to elaborate the expression further we put $\overline{\omega} = \overline{\omega}(\overline{k})$. Performing the integration over the angle and adding the HF contribution we get the final result for the real part:

$$\begin{aligned}\hbar\Sigma_c^{ee}\left[k,\frac{\epsilon_v^0(k)}{\hbar}\right] = \frac{e^2k_F}{2\kappa\pi^2}&\left[-4\overline{k}\pi + \int_0^\infty d\overline{q}\int_0^\infty du\left[\frac{1}{\widetilde{\epsilon}(\overline{q},iu\overline{q})} - 1\right]\frac{1}{\overline{k}}\ln\left|\frac{u^2+(\overline{q}+2\overline{k})^2}{u^2+(\overline{q}-2\overline{k})^2}\right|\right]\\ &-\eta(\tfrac{1}{2}-\overline{k})\frac{2e^2k_F}{\kappa\pi}\int_{\overline{k}}^{1/2}d\overline{q}\,\overline{q}^2\int_{-1}^{+1}dz\frac{1}{\overline{q}^2+\overline{k}^2-2\overline{k}\overline{q}z}\\ &\times\mathrm{Re}\left[\frac{1}{\widetilde{\epsilon}((\overline{k}^2+\overline{q}^2-2\overline{k}\overline{q}z)^{1/2},\overline{q}^2-\overline{k}^2)}\right]\\ &+\eta(\overline{k}-\tfrac{1}{2})\frac{2e^2k_F}{\kappa\pi}\int_{1/2}^{\overline{k}}d\overline{q}\,\overline{q}^2\int_{-1}^{+1}dz\frac{1}{\overline{q}^2+\overline{k}^2-2\overline{k}\overline{q}z}\\ &\times\mathrm{Re}\left[\frac{1}{\widetilde{\epsilon}((\overline{k}^2+\overline{q}^2-2\overline{k}\overline{q}z)^{1/2},\overline{k}^2-\overline{q}^2)}\right] .\end{aligned} \tag{4.6}$$

The expression for $\widetilde{\epsilon}(\overline{q},iu\overline{q})$ is

$$\widetilde{\epsilon}(\overline{q},iu\overline{q}) = 1 + \frac{1}{2}\left[\frac{\lambda}{2k_F}\right]^2 \frac{1}{\overline{q}^2}\left\{1 + \frac{1-\overline{q}^2+u^2}{4\overline{q}}\ln\left[\frac{(1+\overline{q})^2+u^2}{(1-\overline{q})^2+u^2}\right] - u\left[\tan^{-1}\left[\frac{1+\overline{q}}{u}\right] + \tan^{-1}\left[\frac{1-\overline{q}}{u}\right]\right]\right\}. \tag{4.7}$$

If we in Eq. (4.6) replace $\widetilde{\epsilon}$ with the static Thomas-Fermi (TF) expression and set $\overline{k}=0$, Eqs. (3.6) and (3.10) follow.

The self-energy associated with the electron-impurity scattering is for the conduction band

$$\hbar\Sigma_c^{ei}(k) = -\frac{\nu}{6\pi^2}\frac{e^2}{\kappa a_H^*}\int_0^\infty d\overline{q}\frac{1}{\overline{k}\overline{q}^3\widetilde{\epsilon}^2(\overline{q},0)} \times \ln\left|\frac{\overline{q}+2\overline{k}}{\overline{q}-2\overline{k}}\right|, \tag{4.8}$$

where the expression for $\widetilde{\epsilon}(\overline{q},0)$ follows from Eq. (4.7). If the TF expression is used for $\widetilde{\epsilon}(\overline{q},0)$, Eq. (4.7) reduces to Eq. (3.16) in the limit $\overline{k}\rightarrow 0$.

If we now consider the valence band, the calculations of the self-energy $\hbar\Sigma_v^{*,ee}$ becomes considerably more complicated because of the band structure. For this reason we shall only consider a particle at $k=0$. Subtracting the HF contribution as in Eq. (2.10) we have

$$\hbar\Sigma_{v,l}^{*,ee}(k,\omega) = -\frac{4\pi e^2}{\kappa}2k_F\int\frac{d^3\overline{q}}{(2\pi)^3}\int\frac{d\overline{\omega}'}{2\pi i}\frac{1}{\overline{q}^2}\left[\frac{1}{\widetilde{\epsilon}(\overline{q},\overline{\omega}')}-1\right] \times [\overline{G}_{v,l}^0(\overline{k}+\overline{q},\overline{\omega}+\overline{\omega}')\Lambda_{ll} + \overline{G}_{v,h}^0(\overline{k}+\overline{q},\overline{\omega}+\overline{\omega}')\Lambda_{lh}], \tag{4.9}$$

where

$$\overline{G}_{v,l(h)}^0 = 1/[\overline{\omega}-\overline{\omega}_{l(h)}(\overline{k})+i\delta], \tag{4.10}$$

with $\overline{\omega}_{l(h)}(\overline{k}) = \epsilon_{l(h)}(k)/4\epsilon_F$. At $\overline{k}=0$, Eq. (4.9) can be evaluated as above. In this limit there will be no contribution from the residues and the final result is for both light and heavy masses;

$$\hbar\Sigma_v^{*,ee}(0,0) = -\frac{2e^2}{\kappa\pi^2}k_F\int_0^\infty d\overline{q}\int_0^\infty du\left[\frac{1}{\widetilde{\epsilon}(\overline{q},iu\overline{q})}-1\right]\left[\frac{\gamma_l\overline{q}}{u^2+\gamma_l^2\overline{q}^2}+\frac{\gamma_h\overline{q}}{u^2+\gamma_h^2\overline{q}^2}\right], \tag{4.11}$$

where $\gamma_l = m_d/m_{lh}$ and $\gamma_h = m_d/m_{hh}$. If we omit the frequency dependence in $\widetilde{\epsilon}$ and replace it with the Thomas-Fermi expression we recover Eq. (3.11).

Finally we consider the shift associated with impurity scattering for a hole in the valence band. From Eq. (2.13) we obtain for both masses at $k=0$

$$\hbar\Sigma_v^{ei} = \frac{e^2}{\kappa a_H^*}\frac{\nu}{3\pi^2}\frac{m_{lh}+m_{hh}}{m_d}\int_0^\infty\frac{d\overline{q}}{[\widetilde{\epsilon}(\overline{q})\overline{q}^2]^2} \tag{4.12}$$

which reduces to Eq. (3.17) in the TF limit.

The interactions among the particles themselves and with the impurity ions give rise to two effects. Thus the conduction band is pulled down in energy and at the same time the corresponding density of states is distorted. The valence band, on the other hand, appears shifted upwards in energy. Figure 3 illustrates the concentration dependence of the

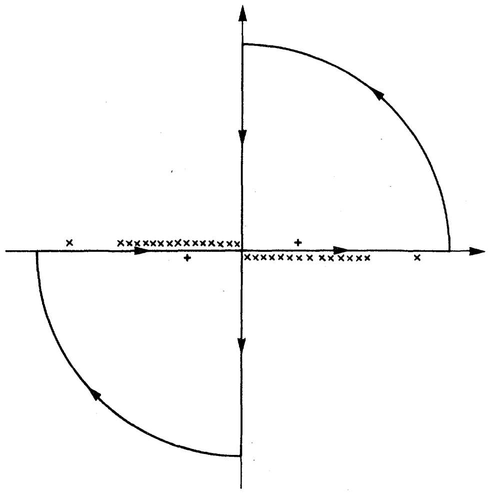

FIG. 2. Contour used to evaluate the self-energy due to electron correlation. Symbols ($\times$) denote the poles in $\epsilon(q,\omega)^{-1}$ and (+) denote the poles in Green's function.

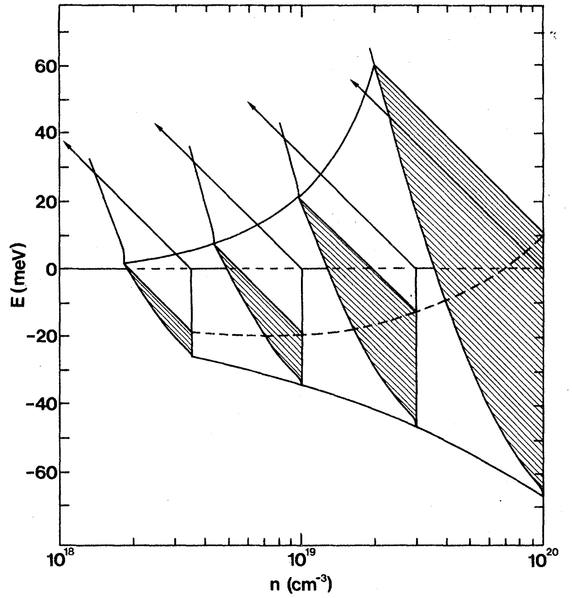

FIG. 3. Quasiparticle density of states for n-type silicon, as a function of concentration. The shaded areas indicate occupied regions for four different concentrations. The position of the bottom of the unperturbed conduction band is at zero energy while that of the perturbed band is given by the full curve at the bottom of the figure. The position of the perturbed chemical potential is given by the dashed curve. The calculations are based on the parameters listed in Fig. 1.

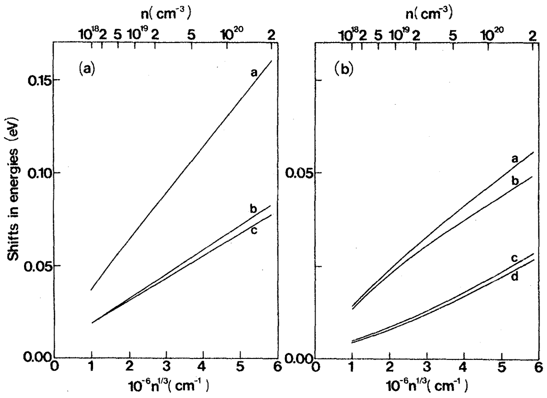

FIG. 4. Various contributions (absolute magnitudes) to ΔE_g in n-Si at different impurity concentrations. In Fig. (a) curve (a) is the total band-gap narrowing whereas curve (b) gives the *downward* shift of the bottom of the conduction band and curve (c) the *upward* shift of valence band at the center of the zone. In Fig. (b) the shifts of the two bands are separated into contributions from electron interactions and impurity scattering. Thus curve (a) is the shift in the conduction band due to electron interactions [$-\hbar\Sigma_c^{ee}(0,0)$, Eq. (4.6)], and curve (b) the corresponding shift in the valence band [$\hbar\Sigma_v^{*,ee}(0,0)$, Eq. (4.11)]. Impurity scattering shifts the valence band according to (c) [$\hbar\Sigma_v^{ei}(0)$, Eq. (4.12)], and the conduction band edge according to (d) [$-\hbar\Sigma_c^{ei}(0)$, Eq. (4.8)]. The calculations are based on the input parameters listed in Fig. 1.

quasiparticle density of states associated with the conduction band in n-Si. At the bottom of the band one notes that the density of states is strongly reduced into a small tail. At the Fermi level, on the other hand, the density of states is enhanced in relation to the unperturbed case. As a consequence, the occupied region is somewhat modified by the interactions. At $n = 3.5 \times 10^{18}$ cm^{-3} the occupied region (including the tail) is reduced by $\sim 10\%$. At $n = 10^{20}$ cm^{-3} the same region is, on the other hand, increased by $\sim 5\%$ relative to the case of noninteracting particles in a smeared positive background. The enhancement at E_F is a real effect; it gives rise to an enhancement of the extrinsic specific heat, as observed recently by Kobayashi *et al.*[23] in the case of Si-P. The experimentally observed enhancement is in good numerical agreement with the type of theory discussed here.[24]

Figure 4(a) shows the computed band-gap narrowing for n-Si as a function of concentration. Over the region displayed one notes that the computed data closely follow an $n^{1/3}$ dependence. Figure 4(a) also shows that the downward shift of the conduction band is almost equal to the upward shift of the valence band. Because of the coupling between the valence bands, the upward shift at $k = 0$ is the same for both light and heavy masses. In Fig. 4(b) the separate contributions to ΔE_g are given. Curves (a) and (b) in Fig. 4(b) refer to interaction effects in the jellium system. Added together (a) and (b) coincide rather closely with the results of Abram *et al.*[2] in Fig. 1. Such an agreement is, however, only fortuitous since the present calculation explicitly includes also the valence band couplings. Figure 4(b) also shows that impurity scattering increases in importance when a more accurate dielectric screening function is used. In fact, Fig. 4(b) demonstrates that the impurity scattering is of such a magnitude that it cannot be dismissed. As a consequence the total band-gap narrowing in Fig. 4(a) is larger than the previous estimates of Abram *et al.*[2] On the other hand, it is much smaller than Inkson's[5] results, also based on the jellium model and the simple plasmon-pole approximation for the dielectric function.

At this point it is appropriate to discuss Mahan's work at some length, and to compare his results with ours. As mentioned in the Introduction, Mahan's work and the present one are rather similar in spirit, but differ considerably in details. Actually Mahan has discussed two approaches. In the first one Mahan thus considered an electron gas occupying the conduction band and proceeded to compute its total energy as the sum of the kinetic, exchange, and electron-donor ion-interaction energies. To begin with, the correlation energy was, however, omitted since Mahan found it to be small. The electron-donor ion-interaction potential was assumed to be of Yukawa type with the screening length given by the Thomas-Fermi expression [Eq. (3.7)], and the corresponding interaction energy was elaborated in different ways. The final choice consisted in locating all the electrons as screening charge and evaluating the corresponding potential energy. Thus having established the expression for the total energy the chemical potential μ was obtained by differentiating this expression with respect to the density, keeping the volume fixed. The chemical potential may be subdivided as $\mu = \mu_0 + \delta\mu$ where μ_0 corresponds to the free gas and $\delta\mu$ the corrections due to exchange and electron-donor ion interactions. In our notations $\delta\mu = \hbar\, \Sigma_c^{ee}(k_F) + \hbar\Sigma_c^{ei}(k_F)$. To obtain the position of the conduction-band minimum E_0, Mahan now argued that the k dependence of the self-energies was weak. Hence he could approximately write $E_0 \simeq \delta\mu$. As we have seen above this approximation is good to $\sim 5-10\,\%$ depending on the concentration. In summary, Mahan's and the present work differ considerably in the treatment of the electron-donor interaction. In addition we include the correlation energy as well as the k dependence of the self-energies.

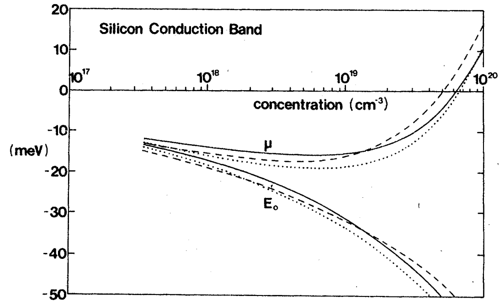

FIG. 5. Concentration dependence in silicon of the chemical potential μ and the conduction-band minimum E_0. The reference energy is the bottom of the unperturbed conduction band. Dotted curves are present results. The solid and dashed curves refer to Mahan's results (Ref. 16). The solid curve is the result of a variational calculation while the dashed curve refers to an approximate treatment of the electron-impurity interaction as described in the text. Input parameters to the present calculations are listed in Fig. 1.

In a second approach Mahan proceeded to improve the electron-donor ion contribution to the total energy arguing that the major theoretical uncertainty is associated with the inhomogeneity in the electron screening charge around the impurities. For this purpose a variational calculation was done, in the Hartree-Fock approximation, for the ground-state energy of the electron-donor system. The ionized donor was taken as a point charge, and the assembly of ionized donors was arranged in a regular fcc lattice. Once more Mahan found that the conduction band minimum could be approximated by $E_0 \simeq \delta\mu$ where $\delta\mu$ is the correction to the free-electron value μ_0 due to inhomogeneities and exchange. The results for μ and E_0 in silicon and germanium, as obtained from the two approaches, are shown in Figs. 5 and 6 together with present results. The results from the three different calculations are indeed quite similar. The difference between the various curves are at most some millielectron volts. This relatively small spread establishes the present accuracy of the theoretical predictions for μ and E_0.

Having found an overall agreement with Mahan's results for the conduction band we now focus attention on the valence band. We shall then note some important differences. The first one is that we include the valence band structure. Mahan, on the other hand, chooses a single isotropic band with $m_{lh(hh)} = m_d$. The second difference consists in the treatment of the impurity scattering. In this case

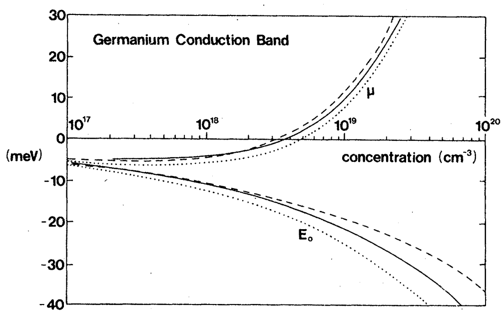

FIG. 6. Concentration dependence in germanium of the chemical potential μ and the conduction-band minimum E_0. The reference energy is the bottom of the unperturbed conduction band. The notations are the same as in Fig. 4. The input parameters for the present calcuation are $\nu = 4$, $\kappa = 15.36$, $m_l = 1.58 m_0$, and $m_t = 0.082 m_0$.

Mahan resorts to the type variational calculation for a fcc lattice described above. We, on the other hand, use perturbation theory for a random system of donors.

Mahan's as well as our treatment include two distinct contributions to the shift of valence band, namely the shifts due to correlation and impurity scattering. To begin with correlation, Mahan calculated the self-energy of the hole in very much the same way as we have done, but for an isotropic valence band as mentioned above. Furthermore the dielectric function was evaluated in the plasmon-pole approximation in which the q^4 term was retained [cf. Eq. (3.4)]. In this way he should closely reproduce ordinary RPA. Indeed we obtain close numerical agreement with his results. For silicon with $n \simeq 4\times10^{19}$ cm^{-3} Mahan finds that the self-energy of a hole at the center of the zone is -33 meV. We have -35 meV. [The self-energy of the *hole* is $-\hbar\Sigma_v^{*,ee}(0)$ in Eq. (2.10).] At $n = 10^{18}$ cm^{-3} the corresponding numbers are -13 and -14 meV. In spite of the different approximations we thus achieve close numerical agreement for the self-energy associated with correlation. The results for impurity scattering, however, differ considerably at a first sight. In fact we obtain almost the same magnitude for the impurity-induced shift as Mahan does, but the sign differs. Therefore in Mahan's calculation the contributions from impurity scattering of conduction and valence states effectively cancel in the final expression for ΔE_g. The argument would be that a hole and the impurity ions repel, hence its self-energy is positive. We, on the other hand, have argued on the basis of second-order perturbation theory. In the case of a completely filled band no scattering can take place because all states are occupied. Hence the contribution from impurity scattering to the total energy is zero. Now, if a hole is created at, for example, the top of the valence band, the remaining valence electrons will relax around the impurity ions. This new flexibility of the system makes the system lower its total energy by an amount $-\hbar\Sigma_v^{ei}(0)$. By definition the self-energy of the hole then has to be negative. When computing ΔE_g we therefore find that the effects of impurity scattering add rather than subtract. One should note, however, that Mahan has assumed an fcc lattice while we have considered a random array of donors. This difference does, in fact, explain the apparent controversy. Mahan writes the self-energy of the hole as [his Eq. (26)]

$$\Sigma_h(k=0) = \Sigma_{h,c} + 0.481e^2n^{1/3}/\kappa + \delta\Sigma_h \ , \tag{4.13}$$

where $\Sigma_{h,c}$ is associated with electron correlation [$-\hbar\Sigma_v^{*,ee}(0,0)$ in our notations]. The second term on the right-hand side is an electrostatic Madelung-type term which arises because of the fcc lattice. This term is, like $\Sigma_{h,c}$, associated with a uniform hole charge. [The electrostatic interaction of a hole (electron) with a donor ion and uniform electron charge in a Wigner-Seitz sphere is $(\pm)$ $0.483e^2n^{1/3}/\kappa$.] The last term, finally, is a correction term which relates to the inhomogeneity in the distribution of charge around the impurities. This term, which thus takes into account the relaxation of the valence electrons must therefore be negative and is just our $-\hbar\Sigma_v^{ei}(0)$. Because we have chosen a random arrangement of donor ions the Madelung-type term does not appear in our case. Mahan finds that $\delta\Sigma_h$ is small ($|\delta\Sigma_h| \lesssim 3.5$ meV in silicon) and therefore neglects it. In fact, we know from previous experience that $\delta\Sigma_h$ should be small for a regular lattice, but large in magnitude for a disordered system.[24] To prove this point we have recalculated $-\hbar\Sigma_v^{ei}(0)$ for an fcc lattice and find a value of -2.1 meV at $n = 10^{18}$ cm^{-3} and -2.5 meV at $n = 10^{19}$ cm^{-3} in the case of silicon. For germanium the corresponding values are ~ -1 meV. Thus the electrostatic term $0.481e^2n^{1/3}/\kappa$ which we now also should add to our expression for the self-energy of the hole, dominates. Consequently, one may ignore $-\hbar\Sigma_v^{ei}(0)$ as Mahan did. To make the comparison with Mahan's work complete we may treat the conduction electrons in the same way. For example, our expression for the chemical potential should then be for an fcc lattice

$$\mu = E_F + \hbar\Sigma_c^{ee}(k_F) - 0.481e^2n^{1/3}/\kappa + \delta\mu_c \ , \tag{4.14}$$

where $\hbar\Sigma_c^{ee}$ is the self-energy associated with a homogeneous system and $\delta\mu_c$ represent corrections due to inhomogeneities [cf. Mahan's Eq. (21)]. Calculating $\delta\mu_c$ for an fcc lattice we generally find it to be much smaller (~ -2 meV in the case of silicon) than for the disordered case. When creating an electron-hole pair the electrostatic terms in Eqs. (4.13) and (4.14) cancel. So they should do, of course, because the system remains charge neutral. As a result impurity scattering contributes very little to ΔE_g if the donors are ordered. In fact, this is the way it should be! In general a disordered system scatters more than an ordered one.

In Figs. 7 and 8 we compare our results for ΔE_g with Mahan's. Obviously the large discrepancy at large concentrations derives mainly from the different treatments of the electron-ion interactions, as discussed above. In Fig. 7 we also give our results

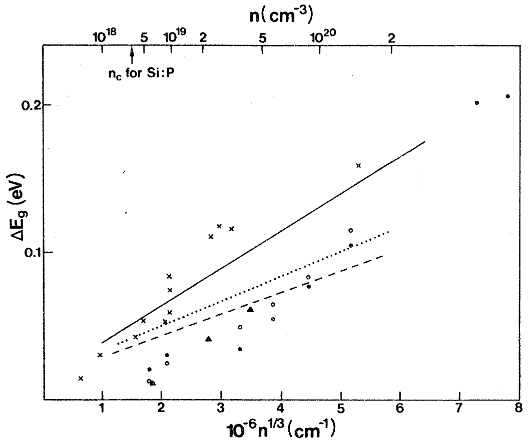

FIG. 7. Comparison of experimental and theoretical values for the band-gap narrowing in n-type Si at different impurity concentrations. The solid line refers to the present calculation for the case of a random arrangement of donor ions. The dotted line shows the same results for the case of the donors arranged in an fcc lattice. The dashed line gives the results of Mahan's variational calculation (Ref. 16), also for the fcc lattice. Circles show the results for the band-gap shrinkage as deduced by Balkanskii *et al.* (Ref. 25) from optical measurements (open circles refer to 35 K and filled to 300 K). Filled triangles refer to optical measurements by Schmid (Ref. 18) at 4 K. Crosses show the data compiled by Lanyon and Tuft (Ref. 9) for different devices. (The compilation excludes all data which assume in their interpretation that the band-gap narrowing is independent of temperature.) The Mott critical density for Si:P is denoted n_c. Input parameters for the present calculations are listed in Fig. 1.

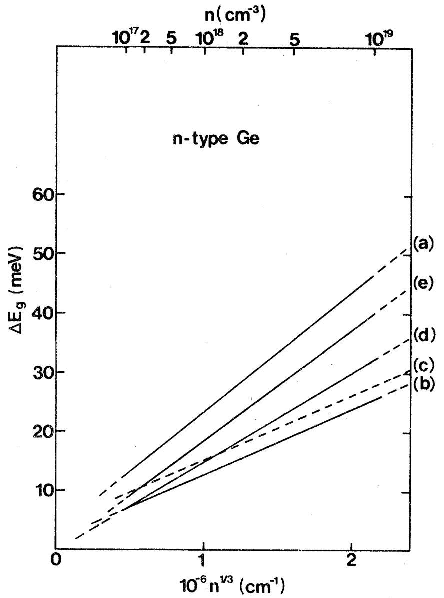

FIG. 8. Effective band-gap shrinkage in n-type Ge as a function of impurity concentration; (a) shows the total narrowing ΔE_g and (b) the magnitude of downward shift of the bottom of the conduction band $[-\hbar\Sigma_c^{ee}(0,0) - \hbar\Sigma_c^{ei}(0)]$ according to the present model. Donors are assumed to form a random system. The dotted line (c) gives Mahan's (Ref. 16) variational results for an fcc lattice of donors. Experimental estimates of Rogachev and Sablina (Ref. 12) and Shotov and Murashov (Ref. 27) are given as the straight lines (d) and (e), respectively. The present theoretical values are based on the input parameters $\nu = 4$, $\kappa = 15.36$, $m_l = 1.58m_0$, $m_t = 0.082m_0$, $m_{lh} = 0.042m_0$, and $m_{hh} = 0.347m_0$.

for the case that the donors form an fcc lattice. In accordance with our previous arguments we then obtain smaller values for ΔE_g and find quite a good agreement with Mahan's results. We thus conclude that the actual distribution of donors has significant effects on the actual value of ΔE_g. Of course, the choice of a random arrangement should correspond more to real cases.

Finally, we may attempt a comparison between theory and experiments. Figure 7 gives the case of n-type silicon. In this comparison we have only included theoretical values approximately below the solubility limit. The experimental data refer to optical measurement and devices. Let us first consider the optical data. Thus the data of Balkanskii *et al.*[25] were obtained from absorption measurements at two different temperatures (35 and 300 K). Schmid's[18] absorption data were taken at 4 K. Although there is an order-of-magnitude agreement between theory and experiments, there is a considerable numerical disagreement, particularly at low concentrations. One should, perhaps, act with due caution when comparing with optical data. The reduction of ΔE_g from such data seems to involve some rather drastic simplifications, and it is therefore hard to estimate the accuracy of the final results. The same caution should, of course, be exercised in connection with our theoretical results. In our model we rely on a perturbation expansion, which in principle should be accurate only at high densities. Figure 7 indicates that theory and experiments come close together in such a limit. This trend may perhaps be taken seriously. The second set of data in Fig. 7 is a compilation due to Lanyon and Tuft[9] of ΔE_g in different transistor devices. At first sight one may be pleased that this set of data is quite close to our own results. The interpretation of device data seems, however, even more inaccurate than in the case of optical properties.[16] The nice agreement in Fig. 7 is therefore not be taken too seriously.

Theoretical results for n-type germanium are given in Fig. 8. As in the case of silicon one notes an approximate $n^{1/3}$ dependence of ΔE_g. The total, computed band-gap narrowing is given by (a). Curve (b) is the downward shift of the conduction-band edge. As in the case of n-Si the upward shift of the valence band is about the same. Rogachev and Sablina[12] have determined the effective band-gap narrowing from electroluminescence experiments at 80 K to be (in units of eV)

$$\Delta E_g = 1.5 \times 10^{-8} n^{1/3} , \tag{4.15}$$

which is shown as (d) in Fig. 8. As in the case of silicon there is an order-of-magntiude agreement between theory and experiments. There is, however, considerable scatter among experimental results. Thus Pankove and Aigrain[26] found $\Delta E_g \sim 40$ meV at 4×10^{18} cm^{-3} from interband absorption. This value is considerably higher than what Eq. (4.15) predicts, and, in fact, rather close to present theory. Shotov and Murashov[27] have also found an empirical relation like Eq. (4.13), but with a coefficient of 1.85 instead of 1.5 [shown as (e) in Fig. 8]. These authors also point out that ΔE_g is the same at 10 and 77 K and independent of the nature of the impurity. Figure 8 also shows Mahan's variational results as the dashed line (c). As in the case of silicon his values for ΔE_g are smaller than ours, presumably depending on his choice of a regular lattice of donor ions.

V. SUMMARY AND DISCUSSION

We have considered the energy gap E_g of heavily doped n-type germanium and silicon as a function of impurity concentration. The purpose of the calculation is to study the inclusion of full RPA screening, impurity scattering, and interband matrix elements. It is assumed that we deal with shallow impurities and that the effective-mass approximation is valid. The model used consists in carriers in the host conduction band and randomly distributed ionized point impurities. Such a model is relevant for impurity concentrations above the Mott critical concentration n_c. The model is unorthodox in the sense that no separate impurity band is assumed to exist above the critical density.

The addition of free carriers causes, because of their mutual exchange and Coulomb interactions, a downward shift of the conduction band. This downward shift is further accentuated by the attractive impurity scattering. The valence-band states are affected in the opposite way. Owing to the added free carriers the interactions among particles in the valence bands are weakened; the initial Hartree-Fock exchange is replaced by dynamically screened interactions. As a result the valence band is shifted upwards in energy. Alternatively, we say that the holes acquire a negative self-energy because of electron interactions. The creation of a hole in the valence band allows the other valence electrons to relax around the impurity centers and thus gain in energy. Consequently impurity scattering also leads to a negative self-energy for the holes. Because interband couplings are explicitly introduced for the valence bands, heavy- and light-hole states are shifted by the same amount. The contributions to ΔE_g from the different processes considered here are all of the same order of magnitude. It is also found that the introduction of the RPA dielectric screening function leads to substantial corrections compared to our simple plasmon-pole and the Thomas-Fermi approximation. Finally, the size of ΔE_g was found to depend in a sensitive way on the actual arrangement of donor ions.

As it seems the present calculation is, together with Mahan's,[16] the most ambitious theoretical treatment so far of the band-gap shrinkage in impure n-Si and Ge. Some shortcomings and simplifications should, however, be recapitulated. All calculations have been performed at zero temperature, although an extension to finite temperatures is possible in principle.[4] The effects of the simplified matrix elements are not known. Although the present matrix elements are expected to represent a first good estimate, it would be desirable to obtain more accurate values from realistic wave functions which include also the periodic component of the Bloch functions. Furthermore, the present calculation is based on perturbation theory and the elementary Born approximation for the impurity scattering. The model is therefore basically a high-density model, i.e., it is expected to be accurate, within the limitations mentioned above, in the regime $\sim 10^{19}$–10^{20} cm^{-3} in the case of n-Si. Below this regime corrections to Born scattering may be anticipated as well as effects of multiple scattering. The model assumes a perfectly uniform distribution of impurity centers. It therefore remains to statistically average the band gap over possible fluctuations in the density. A possible way to improve the present calculations would therefore be to extend them in a way suggested by the recent work of Serre *et al.*[17] To include multiple scattering the Dyson equation was solved numerically by these authors, followed by a statistical averaging. The calculations were performed for a Thomas-Fermi screening of the electron-donor ion interactions, neglecting correlation effects in the free gas. The inclusion of multiple scattering and statistical effects in the present case would obviously make our calculations numerically more cumbersome. Another approach would be to extend Mahan's[16] variational type of calculations to a random arrangement of donors. Such a procedure would, however, be rather demanding from a numerical point of view.

In spite of the simplifications listed above a comparison between theory and experiments indicates an order-of-magnitude agreement. One notes, however, considerable scatter among experimental values, particularly in the case of germanium. Any comparison of this kind should therefore be viewed with caution. As pointed out the experimental deduction of ΔE_g involves several simplifying assumptions. For example, Schmid[18] points out that the value of the gap shrinkage deduced from optical-absorption measurements depends strongly on the assumed value of the Fermi energy. A more satisfactory procedure would therefore be to consider measured quantities theoretically, rather than to use experimental estimates of ΔE_g itself. In the present work we have, however, focused attention on a comparison with other theories. Therefore such considerations are outside our present scope. As it seems, theory and experiments both have some way to go before they can be reconciled in a detailed manner. Isn't that a challenge?

Note added in proof. Further comments on Mahan's work (Ref. 16) are made by P. Sterne and J. C. Inkson [J. Appl. Phys. (in press)].

ACKNOWLEDGMENTS

We are indebted to Dr. Frank Stern for drawing the problem of band-gap narrowing to our attention and for helpful correspondence. We also want to thank R. A. Abram, J. C. Inkson, P. Leroux Hugon, R. W. Keyes, G. D. Mahan, and P. E. Schmid for useful comments on a first draft of this article. Partial financial support from the Swedish Natural Science Research Council is gratefully acknowledged.

---

[1] R. W. Keyes, Comments Solid State Phys. 7, 149 (1977).

[2] R. A. Abram, G. J. Rees, and B. L. H. Wilson, Adv. Phys. 27, 799 (1978).

[3] D. K. Ferry, Phys. Rev. B 18, 7033 (1978), and references therein.

[4] See, e.g., A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).

[5] J. C. Inkson, J. Phys. 9, 1177 (1976).

[6] V. L. Bonch-Bruevich and R. Rozman, Fiz. Tverd. Tela (Leningrad) 6, 2535 (1964) [Sov. Phys.—Solid State 6, 2016 (1965)].

[7] For a review for the case of GaAs see H. C. Casey and F. Stern, J. Appl. Phys. 47, 631 (1976).

[8] G. Mahler and J. L. Birman, Phys. Rev. B 13, 3661 (1976) (band-gap narrowing in connection with electron-hole drops in doped Ge and Si.)

[9] H. P. D. Lanyon and R. A. Tuft, IEEE Trans. Electron Devices, 26, 1014 (1979).

[10] P. A. Wolff, Phys. Rev. 126, 405 (1962).

[11] V. L. Bonch-Bruevich, *The Electronic Theory of Heavily Doped Semiconductors* (Elsevier, New York, 1966).

[12] A. A. Rogachev and N. I. Sablina, Fiz. Tverd. Tela (Leningrad) 8, 866 (1966) [Sov. Phys.—Solid State 8, 691 (1966)].

[13] P. R. Rimbey and G. D. Mahan, Phys. Rev. B 10, 3419 (1974) (specifically GaAs).

[14] B. Bergersen, P. Jena, and A. J. Berlinsky, J. Phys. C 8, 1377 (1975) (electron-hole liquid in heavily doped *n*-type Ge and Si).

[15] A preliminary report is found in B. Sernelius and K.-F. Berggren, *1980 Annual Conference of the Condensed Matter Division, European Physical Society, Antwerpen, 1980* (Plenum, New York, in press).

[16] G. D. Mahan, J. Appl. Phys. 15, 2634 (1980).

[17] J. Serre, A. Ghazali, and P. Leroux Hugon, Phys. Rev. B 23, 1971 (1981).

[18] P. E. Schmid (unpublished).

[19] M. Combescot and P. Nozières, Solid State Commun. 10, 301 (1972).

[20] M. Combescot and P. Nozières, J. Phys. C 5, 2369 (1972).

[21] L. Hedin and S. Lundqvist, in *Solid State Physics,* edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), Vol. 23, p. 1.

[22] T. M. Rice, Ann. Phys. (New York) 31, 100 (1965).

[23] N. Kobayashi, S. Ikehata, S. Kobayashi, and W. Sasaki, Solid State Commun. 32, 1147 (1979).

[24] B. E. Sernelius and K.-F. Berggren, Philos. Mag. B 43, 115 (1981).

[25] M. Balkanskii, A. Aziza, and E. Amallzag, Phys. Status Solidi 31, 323 (1969).

[26] J. I. Pankove and P. Aigrain, Phys. Rev. 126, 956 (1962).

[27] A. P. Shotov and M. S. Murashov, Fiz. Tekhn. Poluprovodn. 1, 573(1967) [Sov. Phys.—Semicond. 1, 476 (1967)].