Dielectric response in p-type silicon: Screening and band-gap narrowing

L. R. Logan and J. L. Egley
Theoretical Modeling Department, IBM Semiconductor Research and Development Center, East Fishkill Facility, Hopewell Junction, New York 12533
(Received 13 October 1992)

Some physical effects that arise due to large impurity concentrations in p-type semiconductors are discussed. This consists of a detailed analysis of the screening of static impurities and also band-gap narrowing in p-type silicon. A finite-temperature dielectric-response formalism is used together with a 6×6 Hamiltonian operator to describe the hole-energy eigenstates. A comparative study between our results and a Thomas-Fermi model generalized for the same valence-band structure is also presented. It is found that at high doping the Thomas-Fermi approximation overestimates the impurity screening except at very large distances where the Thomas-Fermi potential exhibits a greater long-range tail. Calculations of the band-gap narrowing in p-type silicon are performed using the same dielectric-response function and are shown to be in good agreement with experiment at both 20 and 300 K.

I. INTRODUCTION

The screening of static impurities in solids has long been of interest to researchers in many-body theory. It has also become of importance to the technologist in semiconductor physics through its role in determining ionized-impurity scattering rates. For semiconductors that are highly doped with impurities (say $\gtrsim 10^{19}$ cm^{-3}) the latter is the single most important scattering mechanism in that it has the largest effect on the electron and hole mobilities (e.g., see Ref. 1). The transport properties of charge carriers in highly doped semiconductors may be traced back to the fundamental interaction potential through which the carriers interact with the impurity scattering centers. In this paper we present an analysis of the screening of the impurity potential by holes, choosing silicon as an example. This topic is somewhat rare in the literature due to the inherent complexity in working with realistic valence-band structures. Our analysis therefore includes some direct comparisons between the present work and results obtained from some more frequently used approximations. The screened potential is a fundamental ingredient required for a deeper investigation into the physics of a many-electron system. We will use this information in order to examine the band-gap narrowing that occurs in silicon due to the presence of ionized acceptors and the associated hole gas.

Perhaps the most commonly dealt with screening effect in solids is that associated with valence electron screening (for a review see Ref. 2 and references therein). For the free carriers some work has been done (e.g., see Refs. 3 and 4) with the use of various approximations to simplify what is otherwise a rather laborious calculation. Furthermore, most calculations refer only to simple isotropic band structures which can only be considered suitable for n-type material. Most often the problem is formulated by using an effective dielectric function of the form

$$\epsilon_{\text{eff}}=\epsilon_v(k)+\kappa_0(\beta_s/k)^2 , \tag{1}$$

where $\epsilon_v(k)$ represents the dielectric function characterizing the valence electron screening.[4,5] It contains information on the electron band structure, band gap, etc. and is fundamentally related to the properties of valence–conduction-band transitions. The valence dielectric function has the value κ_0, the static dielectric constant, for its zero wave-vector limit and unity for asymptotically large wave vector. The carrier screening parameter β_s is determined by considering only the long-wavelength behavior of the screening effect. It may be derived from Poisson's equation by assuming a standard exponential screening function. When carried out to first order the calculation produces familiar results: for a nondegenerate semiconductor with spherical-parabolic bands it reduces to the familiar Debye-Huckel screening parameter:

$$\beta_s=\frac{ne^2}{\kappa_0 k_B T} , \tag{2}$$

where n is the carrier concentration, e is the electron charge, k_B is Boltzmann's constant, and T is the absolute temperature. Such a simple prescription is inappropriate for screening in highly doped p-type semiconductors. In fact, even generalized models based on exponential screening are found to be in substantial error when compared to a more realistic treatment as is demonstrated below.

The screened potential described above serves as the net interaction energy for the many-electron system, the physics of which may be described from the interactions that exist among the electrons with themselves and with the impurity ions. The presence of the free carriers that are liberated from the impurities alters the exchange and correlation energies of the electron states. Furthermore, the impurity centers themselves supply an additional perturbing potential to the system. These effects give rise to a modification of the electron energies and thus one describes quasiparticle states having an energy relation depending on wave vector $\mathbf{k}$ and frequency ω given by

$$E(\mathbf{k},\omega)=E^0(\mathbf{k})+\hbar\Sigma(\mathbf{k},\omega)\ , \tag{3}$$

where $E^0(\mathbf{k})$ is the unperturbed energy and $\hbar\Sigma(\mathbf{k},\omega)$ is the self-energy due to the carrier-carrier and carrier-impurity scattering. Thus the band structure in a highly doped material differs from the corresponding intrinsic material due to these effects. One may determine the band-structure shifts through evaluation of the relevant self-energy corrections. Below we shall calculate the band-gap shrinkage in this manner as a function of the hole concentration in silicon. Such calculations are presented for impurity concentrations ranging from the Mott critical density for metal-insulator transitions to an upper limit in the order of 10^{20} cm^{-3}. The perturbation approach that will be employed here to calculate the various contributions to Σ becomes invalid for concentrations outside of these limits as is discussed below.

The remainder of this paper consists of three sections. In Sec. II the basic theory is discussed. Following that, Sec. III presents results, and finally Sec. IV summarizes the work and presents conclusions.

II. THEORY

In this section the basic theory relating to the screening effect is discussed. First, the formalism for obtaining the proper interaction potential is presented. Following that the effective interaction is used to derive the self-energy of the quasiparticle states leading to the calculation of band-gap narrowing.

A. Dielectric screening in p-type silicon

Consider a static impurity located at some origin in a solid. The potential $\varphi(r)$ due to this impurity as a function of distance r may be expressed as

$$\varphi(r)=\int \frac{d^3q}{(2\pi)^3}\frac{Ze^2}{q^2\epsilon(q)}\exp(i\mathbf{q}\cdot\mathbf{r})\ , \tag{4}$$

where Z is the net charge of the impurity in units of $|e|$ and $\epsilon(q)$ is the zero-frequency dielectric function appropriate to that material. In general the dielectric-response function $\epsilon(q)$ contains contributions from the lattice, the valence electrons, and the free carriers. There have been many contributions to the literature on the development of the dielectric function appropriate to semiconductors. Penn[6] has used a simplified band structure in order to facilitate the necessary calculations and there have been several extensions of his work.[7,8] Beyond that some studies involving greater detail in band structure have been reported[9–11] at least for intrinsic material.

In the present work, primary attention is given to the free-carrier contribution to the response function. As we are considering only nonpolar material, the lattice contribution will be ignored.[12] The valence-electron contribution is usually taken to be κ_0, as is done here, although Resta and Resca[5] have performed a more detailed analysis of the problem. The dielectric function used here therefore consists of the static dielectric constant plus a term due to the free carriers. Such a model is discussed in Ref. 12 and is given by

$$\epsilon(q,\omega)=\kappa_0-\frac{e^2}{q^2}\sum_{l,l'}\int\frac{d^3k}{4\pi^3}\Lambda_{l',l}(\mathbf{k}+\mathbf{q},\mathbf{k})\times\frac{f(E_{l'}(\mathbf{q}+\mathbf{k}))-f(E_l(\mathbf{k}))}{E_{l'}(\mathbf{q}+\mathbf{k})-E_l(\mathbf{k})-\hbar\omega}\ , \tag{5}$$

where the sum in the above is understood to run over the free carriers, the valence-electron contribution having been accounted for in κ_0. In the above, and all that follows, we omit the superscript 0 in representing the energies so that $E_n(\mathbf{k})$ shall mean the unperturbed energies. The $f(E_l(\mathbf{k}))$ is the Fermi-Dirac function,

$$f(E_l(\mathbf{k}))=\frac{1}{\exp\{[E_l(\mathbf{k})-\mu]/k_BT\}+1}\ . \tag{6}$$

The function $E_l(\mathbf{k})$ is the energy of a carrier having wave vector $\mathbf{k}$ and residing in band l and μ is the Fermi-energy level. $\Lambda_{l',l}(\mathbf{k}+\mathbf{q},\mathbf{k})$ represents the squared matrix element of the charge-density fluctuation over the two states of interest given by

$$\Lambda_{l',l}(\mathbf{k}+\mathbf{q},\mathbf{k})=|\langle\mathbf{k}+\mathbf{q},l'|e^{i\mathbf{q}\cdot\mathbf{r}}|\mathbf{k},l\rangle|^2\ . \tag{7}$$

The general theory of dielectric response is highly developed and the book by Pines[13] is an excellent overall reference. As indicated in Ref. 13, Eq. (5) represents the dielectric function as determined using the random-phase approximation (RPA). It is beyond first-order perturbation theory and includes spin (Hartree-Fock) and correlation (Coulombic interaction).

For electrons and holes in semiconductors the wave functions have the usual Bloch form, namely, $\psi_{\mathbf{q}}(\mathbf{r})=\exp(i\mathbf{q}\cdot\mathbf{r})u_{q,l}(\mathbf{r})$, where $u_{q,l}(\mathbf{r})$ has the periodicity of the lattice. It is a simple matter to show that the matrix elements are given by

$$\Lambda_{l',l}(\mathbf{k}+\mathbf{q},\mathbf{k})=\left[\frac{1}{\Omega}\int d^3r\,u^{\dagger}_{\mathbf{k},l}(\mathbf{r})u_{\mathbf{k}+\mathbf{q},l'}(\mathbf{r})\right]^2\ , \tag{8}$$

where Ω is the crystal volume (e.g., see Ref. 14). For transitions of holes in the heavy and light bands, which is of primary interest here, Wiley[15] has found the results

$$\Lambda_{l,l}(\mathbf{k}',\mathbf{k})=\tfrac{1}{4}(1+3\cos^2\alpha)\ , \tag{9a}$$

while for $l\neq l'$,

$$\Lambda_{l',l}(\mathbf{k}',\mathbf{k})=\tfrac{3}{4}\sin^2\alpha\ , \tag{9b}$$

where α is the angle between $\mathbf{k}'$ and $\mathbf{k}$.

By far the largest complexity in dealing with holes is the fact that the band structure is warped and highly nonparabolic in contrast to the case of conduction electrons. The model Hamiltonian used in this study is[16]

$$\underline{H}=Ak^2\underline{I}_6+3\underline{I}_2\otimes\sum_i Bk_i^2(\underline{L}_i^2-\tfrac{1}{3}\underline{L}^2)+2\sqrt{3}\underline{I}_2\otimes\sum_{i<j}Dk_ik_j\{\underline{L}_i,\underline{L}_j\}+\frac{\Delta}{3}\sigma\otimes\underline{L}\ , \tag{10}$$

where the indices i,j run over x,y,z, the $\underline{L}_i$ are the angular momentum matrices for total angular momentum one, σ_i are the Pauli spin matrices, and the $\underline{I}_j$ are identity

matrices of rank j. The $\otimes$ operator denotes the direct product of the matrices. The A, B, and C are the standard valence-band constants. For Si we have used the values quoted in Ref. 17. The quantity $\{\underline{L}_i,\underline{L}_j\}$ represents an anticomutator given by $(\underline{L}_i\underline{L}_j+\underline{L}_j\underline{L}_i)/2$. The above Hamiltonian is a 6×6 operator and it is equivalent to the sum of $\underline{I}_2\otimes\underline{H}_{\mathbf{k}\cdot\mathbf{p}}$ ($\underline{H}_{\mathbf{k}\cdot\mathbf{p}}$ is the 3×3 valence-band Hamiltonian that was first obtained by Dresselhaus, Kip, and Kittel[18]) and a 6×6 spin-orbit Hamiltonian written in the same representation.[19] With this approach one can expect that quantitative accuracy will decrease with increasing temperature and carrier concentration. This description is, to our knowledge, the most accurate one that is still amenable to analytic computation. It is therefore employed in the following analysis bearing in mind these limitations.

Kim, Cardona, and Rodriguez[16] (KCR) exploited the double degeneracy of the eigenstates of $\underline{H}$ to arrive at analytic formulas for the energy eigenvalues. They reported the heavy hole $E_h(\mathbf{k})$, light hole $E_l(\mathbf{k})$, and split-off hole $E_{s\text{-}o}(\mathbf{k})$ as being

$$E_h(\mathbf{k})=Ak^2+\left[\frac{\mathrm{Tr}\underline{W}^2}{3}\right]^{1/2}\cos(\tfrac{1}{3}\theta)-\frac{\Delta}{3}\ , \tag{11a}$$

$$E_l(\mathbf{k})=Ak^2+\left[\frac{\mathrm{Tr}\underline{W}^2}{3}\right]^{1/2}\cos\left[\frac{1}{3}\theta+\frac{4\pi}{3}\right]-\frac{\Delta}{3}\ , \tag{11b}$$

$$E_{s\text{-}o}(\mathbf{k})=Ak^2+\left[\frac{\mathrm{Tr}\underline{W}^2}{3}\right]^{1/2}\cos\left[\frac{1}{3}\theta+\frac{2\pi}{3}\right]-\frac{\Delta}{3}\ , \tag{11c}$$

where

$$\theta=\cos^{-1}\left[\sqrt{12}\frac{\mathrm{Tr}\underline{W}^2}{(\mathrm{Tr}\underline{W}^3)^{3/2}}\right] \tag{11d}$$

and Δ is the split-off energy. The matrix $\underline{W}$ in Eq. (11) is defined as $\underline{H}-Ak^2\underline{I}_6$, which is just the traceless part of $\underline{H}$. The quantity $\Delta/3$ is subtracted so that the origin is at the top of the heavy-hole band—as will be assumed throughout. The actual formulas derived by KCR are more general than that given above in that they were able to include a linear-elastic strain in the Hamiltonian. Their model Hamiltonian reduces to Eq. (10) when the strain is absent.

In order to carry out the evaluation of Eq. (4) it is necessary to first determine the appropriate value of the Fermi-energy level μ to be used in Eq. (5). For this we equate the specified carrier concentration to the integrals over the Fermi functions that describe the carrier densities in the various bands. In this, and all that follows, we consider only the light- and heavy-hole bands. This is expected to introduce negligible error at low to moderate applied fields since the population of holes in the split-off band is very small compared with that in the heavy and light bands. This if the hole concentration is p, then one may obtain the Fermi energy from

$$p=\int\frac{d^3k}{4\pi^3}f(E_h(\mathbf{k});\mu)+\int\frac{d^3k}{4\pi^3}f(E_l(\mathbf{k});\mu)\ , \tag{12}$$

where we have written $f\equiv f(E;\mu)$ in order to emphasize the parametric dependence of the Fermi-Dirac function on μ. Equation (12) is solved numerically in order to determine a self-consistent value of μ. In doing so one finds the usual limits to hold, namely, as the impurity concentration is reduced the Fermi level becomes large and negative so that the Fermi statistics tends towards classical Boltzmann statistics. In the present case a hole concentration somewhere on the order of 10^{17} cm^{-3} suffices to bring about the classical limit.

It is instructive to examine the limit of $q\to0$ in the dielectric function in Eq. (4). In taking this limit one is essentially studying only the long-range behavior of $\varphi(r)$. Proceeding as such one finds

$$\epsilon(q)=\kappa_0+\frac{\beta_s^2}{q^2}\ , \tag{13}$$

where the β_s is a generalized version of the screening length discussed earlier. For the present case of holes screening an acceptor ion one has

$$\beta_s^2=\beta_s^{(ll)2}+\beta_s^{(hh)2}\ , \tag{14}$$

where

$$\beta_s^{(l,l)2}=\frac{e^2}{4\pi^3\kappa_0k_BT}\int d^3k\,f(E_l(\mathbf{k}))[1-f_l(E(\mathbf{k}))] \tag{15}$$

and similarly for $\beta_s^{(hh)}$. Note that in the $q\to0$ limit the interband term vanishes since $\alpha\to0$ in Eq. (9b). For the case of electrons occupying spherical-parabolic bands the results above simplify to well-known expressions. After using the equivalence of each conduction band one easily obtains

$$\beta_s^2=\frac{e^2n}{\kappa_0k_BT}\frac{F_{-1/2}(\eta)}{F_{1/2}(\eta)}\ , \tag{16}$$

where $\eta=\mu/k_BT$ and F_j is the Fermi-Dirac integral of order j:

$$F_j(\eta)=\frac{1}{\Gamma(j+1)}\int_0^\infty\frac{x^jdx}{\exp(x-\eta)+1}\ . \tag{17}$$

Furthermore, as the impurity concentration decreases, the quotient of the Fermi-Dirac integrals above approaches unity and Eq. (16) reduces to the simple Debye-Huckel screening length [Eq. (2)]. Thus Eq. (15) is a Thomas-Fermi screening length generalized for arbitrary band structures. For convenience, in what follows, we shall simply refer to this as the Thomas-Fermi screening length in our discussion relating to holes.

Some cautionary notes are appropriate regarding the interpretation of length scales over which we can accurately determine the potentials. It was mentioned above that in examining the zero wave-vector limit of the theory that one is effectively limiting attention to long-range interactions. While the generalized description, viz. Eq. (4), relaxes this restriction somewhat it is still not

completely adequate to examine the behavior as $r \rightarrow 0$. Since the impurity itself is modeled as point charge equal to the ionization state of the dopant, one is certainly not able to examine the potential over lengths of atomic dimensions. Indeed, the short-wavelength limit of Eq. (4) yields κ_0, which is certainly not correct at distances that are very close to the nucleus. The calculations of Resta,[20] for example, show that for most materials it takes about three or four Bohr radii before the spatial dielectric function [the antitransform of $\epsilon(q)$] reaches its static value of κ_0. As regards the zero wave-vector limit of Eq. (5) one finds that it diverses as q^{-2}, which is the expected result for screening by free carriers.[5,13] For a sufficiently large distance from the impurity one thus expects nearly complete screening.

B. Self-energies and band-gap narrowing

There have been many discussions of the theory of band-gap narrowing (BGN) in the literature, e.g., Refs. 21–26 to name only a few. We would refer the interested reader to these references for greater detail and a more extensive list of previous publications. In the present paper we shall discuss it only at a brief level that is still sufficient to maintain completeness of the presentation. Essentially, to obtain the BGN, one calculates separately the shift in the band edges due to the electron-electron and electron-impurity self-energies. For the shift in the conduction band one has[23,24]

$$\Delta E_c = \Sigma_c^{ee} + \Sigma_c^{ei} , \tag{18}$$

where Σ_c^{ee} and Σ_c^{ei} represent the self-energy changes due to the electron-electron and electron-impurity scatterings. For the valence bands one has[23,24]

$$\Delta E_v = \Sigma_v^{ee} - \Sigma_v^{\rm int} + \Sigma_v^{ei} , \tag{19}$$

where the meaning of these terms are as above with the subscript v indicating the valence band. The additional term $\Sigma_v^{\rm int}$ represents the Hartree-Fock exchange energy already present in the intrinsic (and filled) valence band. The corresponding term for the conduction band is clearly zero and is absent in Eq. (18).

The self-energy of a quasiparticle in band n with wave vector $\mathbf{k}$ and frequency ω due to the electron-electron interaction is given by

$$\hbar\Sigma_i^{ee}(n;\mathbf{k},\omega) = \frac{i}{2\pi}\int d^3r \int d^3r' \phi^*_{\mathbf{k},n}(\mathbf{r}) \int d\omega' W(\mathbf{r},\mathbf{r}',\omega') G^0(\mathbf{r},\mathbf{r}',\omega-\omega') \exp(i\delta\omega)\phi_{\mathbf{k},n}(\mathbf{r}') , \tag{20}$$

where i represents either a valence or conduction band and $\phi_{\mathbf{k},n}(\mathbf{r})$ is the electron eigenstate in band n with wave vector $\mathbf{k}$. The function $W(\mathbf{r},\mathbf{r}',\omega')$ is the interaction potential given by

$$W(\mathbf{r},\mathbf{r}',\omega') = \frac{1}{(2\pi)^3}\int d^3q\, W(\mathbf{q},\omega)\exp[i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')] . \tag{21}$$

The function $W(\mathbf{q},\omega)$ is the interaction potential which is the frequency-dependent generalization of Eq. (4). viz.

$$W(\mathbf{q},\omega) = \frac{Ze^2}{q^2\epsilon(\mathbf{q},\omega)} . \tag{22}$$

Finally, in Eq. (21), $G^0(\mathbf{r},\mathbf{r}',\omega)$ is the Green function evaluated in the unperturbed system. The quantity δ is an infinitesimal positive quantity used to insure convergence for a contour integral taken in the lower complex plane.[24] The Green function in configuration space is given by

$$G^0(\mathbf{r},\mathbf{r}',\omega) = \sum_{m,\mathbf{q}} \phi^*_{\mathbf{q},m}(\mathbf{r}') G^0(m;\mathbf{q},\omega)\phi_{\mathbf{q},m}(\mathbf{r}) , \tag{23}$$

where

$$G^0(n;\mathbf{k},\omega) = \frac{f(E_n(\mathbf{k}))}{\omega-\omega_{n,\mathbf{k}}-i\delta} + \frac{1-f(E_n(\mathbf{k}))}{\omega-\omega_{n,\mathbf{k}}+i\delta} \tag{24}$$

and $\omega_{n,\mathbf{k}} = E_n(\mathbf{k})/\hbar$. The frequency integral in Eq. (20) is evaluated via contour integration and there are two different types of contributions that occur.[24,25] First, there are contributions from the poles of the Green function which then give rise to the screened exchange (SX) integral over $\mathbf{q}$. Second, the poles in the interaction give rise to an additional contribution that is commonly referred to as the "Coulomb hole" (CH) term. It has often been argued that in the context of BGN the contribution from the CH term may be neglected since it affects the valence and conduction band equally (e.g., see Refs. 21, 26, and 27). Actual calculations[24] have shown, at least for n-type silicon, that the contribution to BGN in not zero but is very small–about 10 meV at $n = 10^{20}$ cm^{-3}. We shall therefore neglect the CH term in our calculation of BGN and in what follows the electron-electron self-energy will be understood to mean the SX term only. Proceeding as such, the SX term is obtained upon substituting Eqs. (21)–(24) into (20). The result is

$$\hbar\Sigma(n;\mathbf{k},\omega) = \frac{1}{16\pi^3}\int d^3q \sum_m \Lambda_{n,m}(\mathbf{k},\mathbf{k}-\mathbf{q}) \times W(\mathbf{q},\omega-\omega^0_{m,\mathbf{k}-\mathbf{q}}) \times f(E_m(\mathbf{k}-\mathbf{q})) . \tag{25}$$

To obtain the energy shift at the top of the valence band one may set $\mathbf{k}=0$ in the above. The calculation is further simplified by ignoring the energy dependence in the interaction as is commonly done.[21,28] In making such an approximation one is effectively evaluating a statically screened Hartree-Fock exchange potential.[29,30] For the valence band one must subtract the Hartree-Fock contribution that comes with the intrinsic term so that the result is

$$\hbar\Sigma_v^{ee}(n;0) - \hbar\Sigma_v^{\rm int} = \frac{1}{16\pi^3}\int d^3q \left[\frac{e}{q}\right]^2 \left[\frac{1}{\epsilon(\mathbf{q},0)} - 1\right] \times \sum_m \Lambda_{n,m}(0,\mathbf{q}) f(E_m(\mathbf{q})) . \tag{26}$$

For the conduction band a similar expression may be written down, but without the subtraction of the Hartree-Fock term as in Eq. (26). In this case we have $\Lambda_{n,m}=\delta_{nm}$ (e.g., see Ref. 25).

We turn now to the other self-energy terms in Eqs. (18) and (19) that arise from the presence of the charged impurities. This self-energy correction may be derived from second-order perturbation theory as is often done (e.g., see Refs. 23, 24, and 31). One obtains the result

$$\Sigma_j^{ei}(n;0)=\frac{N_I}{(2\pi)^3}\int d^3q \sum_m \Lambda_{n,m}(\mathbf{k},\mathbf{q})\frac{|W(\mathbf{k}-\mathbf{q})|^2}{E_n(\mathbf{k})-E_m(\mathbf{q})} , \tag{27}$$

where j refers to either a conduction or valence band and N_I refers to the concentration of ionized impurities. The form of this term is the same as that obtained at zero temperature except for the temperature dependence of the dielectric-response function,[24,32] and furthermore it assumes a random distribution of impurities. In order to examine the BGN one evaluates the above at the top of the valence band and at the bottom of the conduction band. Note that these two contributions have opposite sign for the conduction and valence bands. Together they give rise to a positive contribution to the BGN.

Having evaluated the relevant self-energy terms given above one readily obtains the BGN from

$$\Delta E_g=\Delta E_v-\Delta E_c , \tag{28}$$

where the two terms on the right are defined in Eqs. (18) and (19). In the following section results are presented for ΔE_g in p-type silicon.

III. RESULTS

While the theory that was discussed in the preceding section was divided into two parts, it is unified by a common element. All of the physical properties that we discuss here are essentially dependent on the nature of the dielectric-response function. In the first part of this section we discuss what is perhaps the more fundamental issue of the screened interaction. In this we do limit ourselves to the static screening of ionized-impurity ions and comment on its implications regarding transport properties. The more general issue of dynamic response would be dealt with by including the frequency variable in the response function. This would lead to a straightforward generalization of the application at hand. Section III B discusses the application of the results presented in Sec. III A to the calculation of BGN.

A. Screening of impurity potentials

The eigenvalues of the Hamiltonian Eq. (10) are given in Eq. (11) for all three valence bands. For the light and heavy bands (the only ones considered here) the intraband components of the response function are evaluated via direct substitution in Eq. (4). The necessary multiple integrals are all evaluated using standard Gaussian-quadrature techniques. Typically, 30-point formulas are used for radial (Laguerre-type) and angular (Legendre-type) integrations. This was found to provide sufficient accuracy when compared with higher-order quadrature formulas. The net impurity potential is obtained from an additional integration over $\mathbf{q}$ as shown in Eq. (4). The oscillatory integrand in this case is segmented into regions of uniform sign. The integral over each such region is evaluated numerically using a 20-point Gauss-Legendre formula.

The Fermi level used in calculation at temperature T and carrier concentration p is determined by Eq. (12). In what follows we make the simplifying assumption that $p=N_A$, where N_A is the concentration of acceptor ions. While this is fine for high temperatures (~ 300 K) it does lose validity as temperature decreases. For this reason we would find it preferable to discuss results in terms of p rather than N_A wherever possible. In Fig. 1 the results of the Fermi-level calculations are shown as a function of hole concentration at both 300 and 20 K. These are the expected results: The μ values become increasingly negative at low concentrations and at higher temperatures implying a transition between classical and quantum statistics.

In Fig. 2 the results for the impurity potential screening are shown along with the corresponding results obtained from the Thomas-Fermi (TF) approximation (dashed lines). In this we define the screening function as $\varphi(r)$ divided by the unscreened potential:

$$\chi(r)=\frac{4\pi\kappa_0 r}{Ze^2}\varphi(r) . \tag{29}$$

These results show that the TF potential overestimates the screening over a substantial region surrounding the impurity. At very large distances the reverse is true—the TF potential exhibits its expected exponential long-range tail whereas that obtained here predicts a much more short-ranged potential. At lower dopings, say $\lesssim 10^{17}$, the RPA value and the TF value are essentially equivalent. At these low concentrations the $q\rightarrow 0$ limit provides the dominant contribution to the potential integral given by Eq. (4). This can be inferred from the behavior of the Fermi level in this limit which yields Boltzmann statistics in Eq. (5) so that $\epsilon(q,0)$ becomes increasingly peaked near $q=0$.

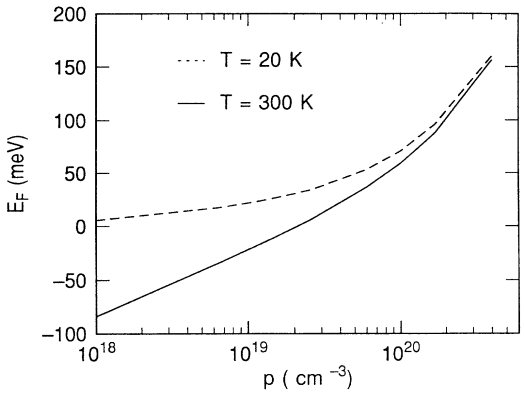

FIG. 1. Calculated Fermi-energy level as obtained from a self-consistent solution of Eq. (12) for silicon at 20 and 300 K.

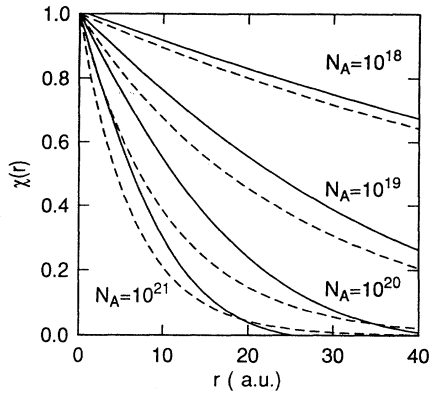

FIG. 2. The screening function $\chi(r)$ given by Eq. (29) as a function of distance from the impurity center. The plot separately shows the result obtained for several values of the acceptor concentration N_A. The calculations correspond to silicon at 300 K and it is assumed that all impurities are ionized.

Finally, it is of interest to examine how the present calculations for holes screening an acceptor compare with that of an electron-gas screening a donor ion. This calculation is essentially the same as above except now one must sum over the six conduction valleys once having a chosen band-structure model. For the present comparison the standard nonparabolic model for the electron-band structure is used, namely

$$\frac{\hbar^2 k^2}{2m^*} = E(k)[1+\alpha E(k)] , \tag{30}$$

where α is the nonparabolicity parameter and m^* is the density of states effective mass. For silicon these have values of 0.5 eV^{-1} and 0.32 times the electron rest mass, respectively. Figure 3 shows a comparison between the

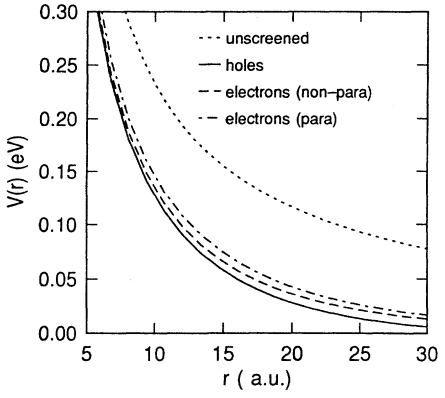

FIG. 3. Comparison of the screened potentials about an acceptor impurity in a hole gas of concentration 10^{20} cm^{-3} and a donor impurity in an electron gas having the same concentration. Both impurities are singly charged. The results for the electrons are obtained using Eq. (30) for the band structure. The dashed line corresponds to a nonparabolic band structure [$\alpha=0.5$ eV^{-1} in Eq. (30)] and the dash-dotted line corresponds to a parabolic-band structure [$\alpha=0$ in Eq. (30)]. All results are for silicon at 300 K. The unscreened interaction is a Coulomb potential with $\kappa_0=11.7\epsilon_0$.

screened donor potential in an electron gas having a concentration of 10^{20} cm^{-3}, and a screened acceptor potential in a hole gas of the same concentration. It is seen from this figure that the holes are somewhat more effective in screening the impurity potential than are the electrons. This is as one might expect due to the heavier-hole effective mass. The effect of nonparabolicity in the electron-band structure is seen to increase the screening slightly over the corresponding parabolic band structure (i.e., setting $\alpha=0$) in Eq. (30).

B. Band-gap narrowing

Using the expressions given in Sec. II for the various self-energy terms we derive the BGN using Eq. (28). First of all, one notes that for p-type material the SX term in the conduction band is zero due to the absence of free carriers in that band. Thus the electron-electron contribution to the BGN comes entirely from the valence band and its effect is to shift the band upwards, thus narrowing the gap. For the case of the electron-impurity self-energy, both the conduction and valence bands have finite contributions that serve to narrow the band gap. Thus the electron-impurity self-energy is negative for the conduction while it is positive for the valence band. As mentioned above, for the study of BGN, the CH terms for the valence and conduction bands nearly cancel and hence are not included here. Figure 4 shows a plot of the self-energy terms that contribute to the BGN for p-type silicon. We believe that these calculations are novel as we are unaware of any literature reporting self-energy calculations for p-type material at finite temperature.

There have been several experimental investigations of BGN in silicon.[33–35] In the present work the BGN calculations were done at two temperatures, 20 and 300 K, chosen in order to make comparison with experimental data.[33,34] In Fig. 5 we show the BGN as calculated at 20 K. The experimental data does show some scatter and an error of ±10 meV is quoted by the author of Ref. 33. One finds the overall agreement between the present

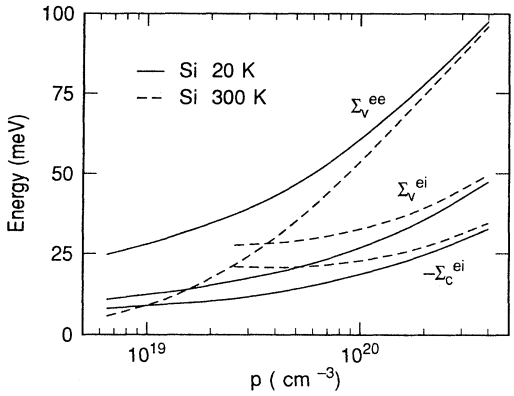

FIG. 4. Calculated values of self-energy terms vs hole concentration. The dashed curve refers to silicon at 300 K whereas the solid curves refer to silicon at 20 K. The subscripts c and v refer to the conduction and valence bands, respectively. The superscripts ee and ei refer to the carrier-carrier and carrier-impurity self-energy, respectively.

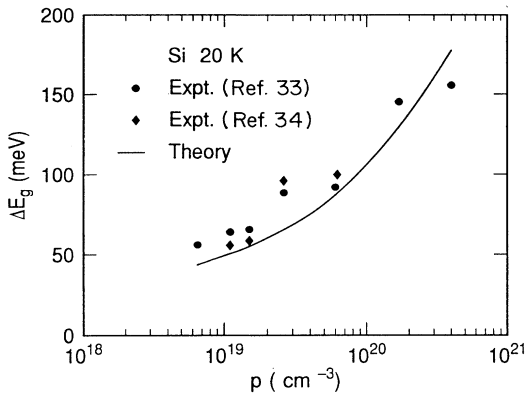

FIG. 5. Calculated and experimental values of band-gap narrowing in silicon at 20 K. The experimental values are from Refs. 33 and 34 for the solids circles and diamonds, respectively.

theory and experiment to be quite satisfactory. The same calculations are carried out for a temperature of 300 K and these results are shown in Fig. 6. Again one finds reasonable agreement with experiment. The most notable discrepancy occurs at the upper limit of the carrier-concentration range where the present theory overshoots the experimental results.

As regards the range of validity of the present calculations one can make a few general remarks. Below the Mott critical density the present treatment is invalid due to the breakdown in the perturbative treatment of the electron-impurity self-energy.[24] For silicon at 300 K the Mott density is about 3×10^{19} cm^{-3},[24] so that the curve in Fig. 6 stops at about this lower limit. At 0 K the Mott density is about 6×10^{18} cm^{-3},[24] so that we expect that the 20-K results are valid throughout the range that is presented in Fig. 5. Furthermore, at high enough impurity concentrations the average interimpurity separation $[(2\pi N_I)^{-1/3}]$ becomes comparable to the lattice spacing. In such a case the impurity potential becomes too large for a perturbative treatment. In silicon this occurs at an impurity concentration of about 10^{21} cm^{-3}. Thus the portion of the curves in Figs. 5 and 6 that show the largest discrepancy with experiment lie in a region where the validity of the theory is becoming questionable.

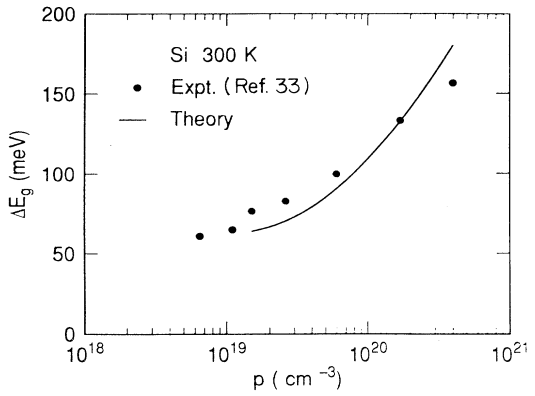

FIG. 6. Calculated and experimental values for the band-gap narrowing in silicon at 300 K. The experimental data is from Ref. 34.

Regarding the numerical details we use the same integration scheme described above. The present calculations for the BGN require a negligible amount of computation. Typically, using 30-point formulas for all angular and radial integrations, a BGN calculation takes about 5 min on an IBM RISC 6000/530. This, of course, includes a self-consistent solution of Eq. (12) for the appropriate Fermi level.

It should be mentioned at this point that we have neglected the coupling between the heavy- and light-hole bands to the split-off band. This is not expected to be a major source of error in the context of BGN since the matrix elements for the transition are approximately zero.[15,25] Furthermore, the BGN calculations have been obtained by considering the shift in the heavy-hole band only. This choice is arbitrary since at $\mathbf{k}=0$ the shift in both the heavy- and light-hole bands are identical (e.g., see Refs. 23–25).

IV. CONCLUSION

We have presented a detailed examination of some aspects of the screening effects in p-type silicon. The formalism that has been used incorporates what is probably the most accurate band structure and response function that still permits facile computation. The present work makes use of a fully temperature-dependent RPA dielectric-response function. For holes the energies are determined from the eigenvalues of the 6×6 Hamiltonian given by Eq. (10). While the union of these makes the methodology appear formally complicated, it is well within reach of efficient computational evaluation as evidence by our reported computer resources. We believe that this enables a state-of-the-art determination of the dielectric response, which then can be used to examine some of the physical properties of the hole gas in the presence of charged impurities. Hence the present work has really hinged upon that theme.

The screened impurity potentials for acceptors in silicon immersed in a hole gas have been calculated and compared with the more traditional Thomas-Fermi approach. One finds upon doing so that the Thomas-Fermi potential gives nearly the same results as the more general RPA for concentrations of about 1×10^{18} cm^{-3} and lower. Above that, the RPA calculated screening is significantly different from that predicted from the Thomas-Fermi theory. Figure 2 shows that in the "near" region (that within $\sim N_A^{-1/3}$) the potential is stronger than the predicted Thomas-Fermi result. This would mean, for example, that the charge carriers would scatter more strongly from the impurity centers and consequently one would obtain a lower net mobility as compared with the Thomas-Fermi prediction. Essentially, the Thomas-Fermi screening gives rise to the well-known Brooks-Herring mobility (in the Born approximation). The latter is notorious for overestimating the mobility in the high doping regime (say $\gtrsim10^{17}$ cm^{-3}) (e.g., see Refs. 1 and 2). One might assume that a more realistic screened interaction would improve this situation some-

what. It should be mentioned, however, that this point is not worth overemphasizing since the actual calculation of mobility involves a great deal more that just electron-impurity interactions. Nevertheless a comparison of the actual potentials does shed some light on the situation.

The second part of our paper has dealt with the related problem of band-gap narrowing in p-type semiconductors. We would find it useful at this point to briefly discuss some previous results that are of related interest. Although they are somewhat rare, there have been some notable contributions to the literature on the topic of BGN in p-type material. Sernelius[31] has calculated the BGN for p-type GaAs. With the use of various approximations including those relating to band structure and matrix elements Sernelius found excellent agreement with results obtained from luminescence spectra over an impurity concentration range of 2×10^{18} cm^{-3} to about 2×10^{19} cm^{-3}. Beyond this upper limit the calculated values exceed the experimental ones rather significantly. All calculations were carried out for zero temperature. Also for p-type GaAs, Bardyszewski and Yevick[36] (BY) have examined the accuracy of several theoretical results by using various approximation for the response function to calculate the BGN. The most accurate response function chosen for study here was that obtained in the RPA. It was found that a damped plasmon-pole approximation provided an excellent result compared with RPA for concentrations up to $\sim10^{20}$ cm^{-3} with plasmon-pole results giving somewhat poorer agreement. Between $\sim10^{20}$ and 10^{21} cm^{-3} both damped as well as undamped plasmon-pole calculations differed dramatically for the RPA results. For the impurity-induced shifts, BY compared the RPA results with those obtained from the Thomas-Fermi approximation for the response function. They observed that the Thomas-Fermi approximation overestimated the impurity contribution by roughly 10% uniformly from 10^{18} to about 10^{21} cm^{-3}. Again, all calculations were carried out for the zero-temperature case.

Making use of the work in Sec. III A we have calculated the band-gap narrowing in p-type silicon as a function of the free-carrier density. Having set up the problem for finite temperatures using the RPA we have calculated the BGN for temperatures of 20 and 300 K. In both cases we have observed reasonably good agreement with experiment except at the very high doping range ($\sim5\times10^{20}$ cm^{-3}) where the limits of applicability start to become an issue. Although from a formal viewpoint the finite-temperature calculation pursued here is somewhat novel, the actual variation of the BGN with temperature is weak. Indeed, in comparison of the results for these two temperatures one concludes that there is very little variation with temperature. This observation is also borne out with experiment as noted in Ref. 34 within a ±10-meV uncertainty.

[1]M. V. Fischetti, Phys. Rev. B **44**, 5527 (1991).
[2]D. Chattopadhyay and H. J. Queisser, Rev. Mod. Phys. **53**, 745 (1981).
[3]N. Takimoto, J. Phys. Soc. Jpn. **14**, 1142 (1959).
[4]Raffaele Resta, Phys. Rev. B **19**, 3022 (1979).
[5]R. Resta and L. Resca, Phys. Rev. B **20**, 3254 (1979).
[6]David R. Penn, Phys. Rev. **128**, 1093 (1962).
[7]G. Srinivasan, Phys. Rev. **178**, 1244 (1969).
[8]R. D. Grimes and E. R. Cowley, Can. J. Phys. **53**, 2549 (1975).
[9]H. Nara, J. Phys. Soc. Jpn. **20**, 778 (1965); **20**, 1097 (1965).
[10]P. K. W. Vinsome and M. Jaros, J. Phys. C **3**, 2140 (1970); **4**, 1360 (1971); P. K. W. Vinsome and D. Richardson, *ibid.* **4**, 2650 (1971).
[11]J. P. Walter and M. L. Cohen, Phys. Rev. B **2**, 1821 (1970).
[12]D. K. Ferry and R. O. Grondin, *Physics of Submicron Devices* (Plenum, New York, 1991).
[13]D. Pines, *Elementary Excitations in Solids* (Addison-Wesley, New York, 1963).
[14]H. Ehrenreich and M. L. Cohen, Phys. Rev. **115**, 786 (1959).
[15]J. D. Wiley, Phys. Rev. B **4**, 2485 (1971).
[16]C. K. Kim, M. Cardona, and S. Rodriguez, Phys. Rev. B **13**, 5429 (1976).
[17]J. M. Hinckley and J. Singh, Phys. Rev. B **41**, 2912 (1990).
[18]G. Dresselhaus, A. F. Kip, and C. Kittel, Phys. Rev. **98**, 368 (1955).
[19]M. Tiersten, IBM J. Res. Dev. **5**, 122 (1961).
[20]Raffaele Resta, Phys. Rev. B **16**, 2717 (1977).
[21]J. C. Inkson, J. Phys. C **9**, 1177 (1976).
[22]G. D. Mahan, J. Appl. Phys. **51**, 2634 (1980).
[23]K.-F. Berggren and B. E. Sernelius, Phys. Rev. B **24**, 1971 (1981).
[24]P. A. Saunderson, Ph.D. thesis, Durham University, 1983.
[25]R. A. Abrams, G. N. Childs, and P. A. Saunderson, J. Phys. C **17**, 6105 (1984).
[26]S. C. Jain and D. J. Roulston, Solid State Electron. **34**, 453 (1991).
[27]E. O. Kane, Phys. Rev. B **5**, 1493 (1972).
[28]W. Brinkman and B. Goodman, Phys. Rev. **149**, 597 (1966).
[29]J. C. Inkson, J. Phys. C **6**, L181 (1973).
[30]R. A. Abrams, G. J. Rees, and B. L. H. Wilson, Adv. Phys. **27**, 799 (1978).
[31]B. E. Sernelius, Phys. Rev. B **34**, 5610 (1986).
[32]F. Thuselt and M. Rosler, Phys. Status Solidi B **130**, 661 (1985).
[33]J. Wagner and J. A. Del Alamo, J. Appl. Phys. **63**, 425 (1988).
[34]J. Wagner, Phys. Rev. B **32**, 1323 (1985).
[35]J. W. Slotboom and H. C. DeGraaff, Solid State Electron. **19**, 857 (1976).
[36]W. Bardyszewski and D. Yevick, Phys. Rev. B **35**, 619 (1987).