Effect of heavy doping on the optical properties and the band structure of silicon

L. Viña and M. Cardona
Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, D-7000 Stuttgart 80, Federal Republic of Germany
(Received 14 November 1983)

We have measured by ellipsometry the dielectric constant of pure and heavily doped n- and p-type silicon from 1.8 to 5.6 eV. Both ion-implanted laser-annealed and bulk doped samples were used with concordant results. A red shift of the E_1 and E_2 critical-point energies, together with a decrease in the excitonic interaction at the E_1 energy, has been observed. These results are compared with first- and second-order perturbation-theory calculations of the effect of the impurities on the band structure of silicon.

I. INTRODUCTION

The effect of doping on the band structure of semiconductors has received considerable attention.[1,2] Much effort has been spent studying these effects because of their technological importance since they affect the characteristics and performance of devices. It is also of basic interest to compare the experimental results with theoretical predictions of the change of the band structure. Quantities such as the low-temperature electronic specific heat[3] and Pauli susceptibility[4] are directly proportional to the density of states at the Fermi energy in the degenerate limit and are capable of giving information on the band structure. The nature of the states of individual impurities and of the states at impurity concentrations close to the Mott transition has been obtained through magnetic resonance studies[5] and electronic Raman scattering.[6,7]

Most of the experimental information has been obtained by optical techniques. Two types of distinct phenomena are usually investigated. They are the following.

(a) *Effects of doping on the lowest (fundamental) band edges, both direct and indirect, which have been profusely studied by techniques such as photoluminescence*[8,9] *and optical absorption.*[10–14] The main effect of doping on these edges is the formation of band tails because of the random potential, the Moss-Burstein shift due to the filling of the conduction (valence) band by electrons (holes), and the shrinkage of the gaps by many-body effects where correlation as well as exchange play a very important role. These edges are also suitable for studying the Mott transition.[15]

(b) Effects on higher edges. These edges above the fundamental ones have received much less attention. The studies have been carried out only with optical techniques such as reflectivity[16,17] and ellipsometry measurements.[18–22] In this case, the exchange term plays only a small role due to the fact that carriers are located in a small region of $\vec{k}$ space different from the region where the transitions take place. The Mott transition is also irrelevant to this problem.

In this paper we report on the optical properties of pure and ultraheavily doped silicon in the region from the near infrared (1.8 eV) to the near ultraviolet (5.6 eV). We pay special attention to the higher interband transitions labeled E_1(3.4 eV) and E_2 (4.25 eV). The E_1 edge is due to transitions between the highest valence band and the lowest conduction band along the Λ direction in the Brillouin zone (BZ) in a region from approximately $\vec{k}=(\pi/4a)(1,1,1)$ to the edge of the BZ (L point).[23–25] The region where the E_2 transition takes place is not very well defined. We will use the points X and $(2\pi/a)(0.9, 0.1, 0.1)$ (Ref. 23) as representative of this transition.

The main effect of impurities in the region of measurement is to broaden the critical points and to shift them to lower energies. The contribution of free carriers is only important in the lower part of the spectrum (photon energies below ~ 2.5 eV) and can be neglected at the energies where E_1 and E_2 occur. Thus, direct observation of the effect of high doping on the band structure and a comparison between donors and acceptors is possible. These effects are somewhat larger for E_1 than for E_2 but still important for the latter.

Ion-implanted, laser-annealed (IILA) and bulk-doped samples were used, with the impurity concentration ranging from $N<10^{13}$ cm^{-3} to $N\sim 3\times 10^{21}$ cm^{-3}. The carrier concentration was determined from the minimum in the infrared reflectivity.[26] As pointed out recently,[19] the use of IILA samples can lead to difficulties in measuring the absolute values of the optical constants. These effects, however, are much smaller for the energies and broadenings of the critical point. Modeling of the microscopic structure of the IILA layers and the adjacent substrate can be used to obtain the "true" dielectric constants. However, because of the uncertainties of these models, we have not followed this path. The ellipsometric data presented here have only been corrected for natural SiO_2 overlayers of standard thickness.

In order to obtain the energy of the critical points and their Lorentzian broadening, a line-shape analysis of the data was made by fitting the numerical third-derivative spectra of the complex dielectric constant with respect to the photon energy, $d^3\epsilon/d\omega^3$, with analytical expressions. A two-dimensional minimum critical point with some admixture of a saddle point was used for E_1 and a one-dimensional maximum was used for E_2. The admixture of the saddle point results from exciton interaction and is shown to decrease with increasing carrier concentration. The shifts ΔE, and the enhancement of the Lorentzian broadening parameters, $\Delta\Gamma$, were obtained from the best

fit to the experimental data, taking, as a reference, the fits for pure Si.

The use of IILA samples provides doping levels an order of magnitude higher than the bulk solubility limit.[27] We obtain, for the IILA samples, results which join smoothly with those of the bulk samples, both for the shifts and the broadening parameters, a fact which justifies the validity of the procedure used. A fit of the experimental data to a N^{α} law yields values of $\alpha=0.46$ for ΔE_1, $\alpha=0.36$ for ΔE_2, $\alpha=0.63$ for $\Delta\Gamma_1$, and $\alpha=0.46$ for $\Delta\Gamma_2$.

Theoretical work has focused so far only on the lower edges and mainly on obtaining values for the band-gap narrowing.[1,28–32] A study of impurity-band tails has also appeared recently.[33] The theoretical results differ widely depending on the approximations used, leading to different dependences of gap energies on concentration. For these reasons we have performed perturbation theoretical calculations of the change in the band structure of Si due to substitutional impurities.[34] Terms of first and second order in the impurity potential have been included. The first-order terms, linear in the impurity concentration, turn out to be too small to be of importance. Because of the divergence of the Coulomb potential for $\vec{q}\rightarrow\vec{0}$, the second-order term is rather large and yields a dependence on impurity concentration similar to N^{α} with $\alpha\sim0.5$. The N^{α} dependence is quite close to the $N^{0.33}$ law predicted and observed for the lowest indirect gap of Si,[30] although, in the case of higher gaps, exchange effects are much smaller. This dependence is also in reasonable agreement with the experimental results.

This paper is organized as follows. In Sec. II we describe the experimental details. The results are presented in Sec. III. In Sec. IV we concentrate on the method of calculation, and finally in Sec. V the experimental results are discussed in the light of the theoretical calculations.

II. EXPERIMENTAL SETUP

Pure and n- and p-type bulk doped Si crystals with doping concentrations ranging from 10^{17} to 4×10^{20} cm^{-3} supplied by Wacker Chemitronics (Burghusen, Federal Republic of Germany) were used to prepare samples. The surfaces to be measured were mechanically lapped and polished with 0.75-μm Al_2O_3 powder. A polish with Syton followed. They were finally polished with a bromine-methanol solution.[35] We also used IILA (Ref. 36) crystals with concentrations from 10^{20} to 3×10^{21} cm^{-3}. The range of implantation doses was $(1-5)\times10^{16}$ cm^{-2} at 100–350 keV for the As samples, and $(0.1-3)\times10^{16}$ cm^{-2} at 30–100 keV for the B samples. The crystals were annealed with a XeCl excimer laser at 308 nm. The laser beam was focused to about 0.5×1.5 mm^2. The annealed surface was obtained by multiple overlapping laser spots as described in Ref. 37.

We select five of the 22 samples measured in order to present the complete spectra of ϵ_1 and ϵ_2 and their third derivatives. For the other samples only the critical-point parameters are given. The characteristics of these samples are summarized in Table I. The first sample is very lightly doped ($n<10^{13}$ cm^{-3}), and is taken as reference and labeled as undoped Si in the following. Two samples are bulk doped with phosphorus [n-type Si(P)] and with boron [p-type Si(B I)] with carrier concentrations of $n=8\times10^{19}$ cm^{-3} and $p=4\times10^{20}$ cm^{-3}. The last two samples are IILA crystals implanted with boron [p-type Si(B II); $p=9.2\times10^{20}$ cm^{-3}] and arsenic [n-type Si(As); $n=3\times10^{21}$ cm^{-3}].

Dielectric-function spectra $\epsilon(\omega)=\epsilon_1(\omega)+i\epsilon_2(\omega)$ were measured at room temperature between 1.8 and 5.6 eV with an automatic rotating analyzer ellipsometer similar to that described by Aspnes[38] and by Aspnes and Studna.[39] In brief, it consists of a light source (75-W Xe short-arc lamp), a $\frac{3}{4}$-m Spex monochromator, Rochon prisms of crystal quartz as polarizing elements, a photomultiplier with an $S20$ response and fused-silica windows to extend its operating range to the uv, and mirror optics for collimating and focusing the light. The output of the photomultiplier is digitized and the digital signal analyzed with the help of a Hewlett-Packard model 9845B-HP computer. The spectra can be stored for later processing of data. The measurements were made at an angle of incidence of 67.5°.

The samples were mounted and optically aligned in a windowless cell in flowing N_2 to minimize surface contamination. In the treatment we should distinguish between the bulk and the IILA samples. The bulk samples were etched *in situ* prior to measurement following the prescription in Ref. 35. The procedure was repeated until real-time ellipsometric measurements showed no more changes and the highest values of ϵ_2 at the E_2 singularity were obtained.[40] For the IILA samples a similar pro-

TABLE I. Samples used for the ellipsometric measurements shown in Figs. 1 and 2.

	Orientation	Doping concentration (cm^{-3})	Doping concentration (at. %)	Implantation dose (cm^{-2})	Implantation energy (keV)	Carrier concentration (cm^{-3})
Undoped Si	[111]					
n-type Si(P)	[111]	8×10^{19}	0.16			8×10^{19} [b]
p-type Si(B I)	polycrystalline	4×10^{20}	0.8			4×10^{20} [b]
n-type Si(As)	[100]	3×10^{21} [a]	6	5×10^{16}	100	2×10^{21} [b]
p-type Si(B II)	[100]	1×10^{21} [c]	2	2×10^{16}	30	9.2×10^{20} [d]

[a]From Rutherford backscattering.
[b]From ir reflectivity minimum, Fig. 2, Ref. 26.
[c]Estimated using data of Ref. 66.
[d]From ir reflection minimum for an average hole mass of 0.40.

cedure was followed. However, only one cycle of etching and stripping was used so as to avoid damage and reduction of the implanted layer. The spectra were measured immediately afterwards.

III. RESULTS

With the use of the measured ellipsometric angles (ψ,Δ) and taking into account the optical activity of the prisms,[41] pseudodielectric functions were calculated from the complex reflectance ratios. The spectra for the five selected samples are shown in Figs. 1(a)–1(d). The dotted lines represent the dielectric function of undoped Si. They have been plotted in the four figures for comparison. In this figure the effect of the impurities is seen: There is an increasing red shift and a lifetime broadening in the two singularities $E_1(\sim 3.4$ eV) and $E_2(\sim 4.25$ eV) with increasing impurity concentration. For the spectra of Figs. 1(a) and 1(b) we have assumed that the ellipsometric data correspond to a simple air–bulk-Si interface; a two-phase model[42] in which the surface is treated as a simple plane boundary between two homogeneous media was used. For the IILA samples an oxide film is still present;

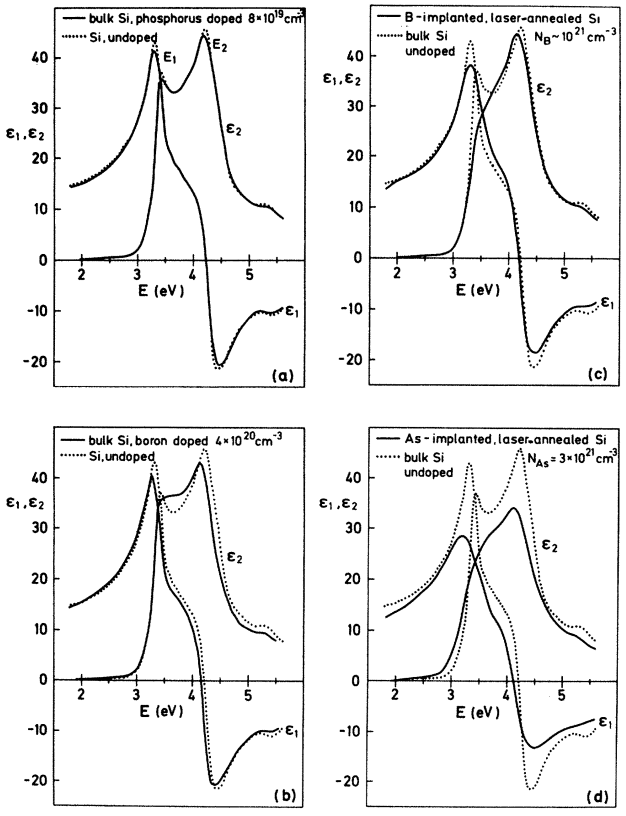

FIG. 1. Solid curves: real (ϵ_1) and imaginary (ϵ_2) parts of the pseudodielectric constant of (a) bulk Si, phosphorus doped with $n=8\times10^{19}$ cm^{-3}; (b) bulk Si, boron doped ($p=4\times10^{20}$ cm^{-3}); (c) boron-implanted, laser-annealed Si ($p\sim10^{21}$ cm^{-3}); (d) arsenic-implanted, laser-annealed Si ($n\sim3\times10^{21}$ cm^{-3}). The pseudodielectric constant for a lightly doped ($n<10^{13}$ cm^{-3}) silicon wafer is also shown as dotted lines in the four figures for comparison.

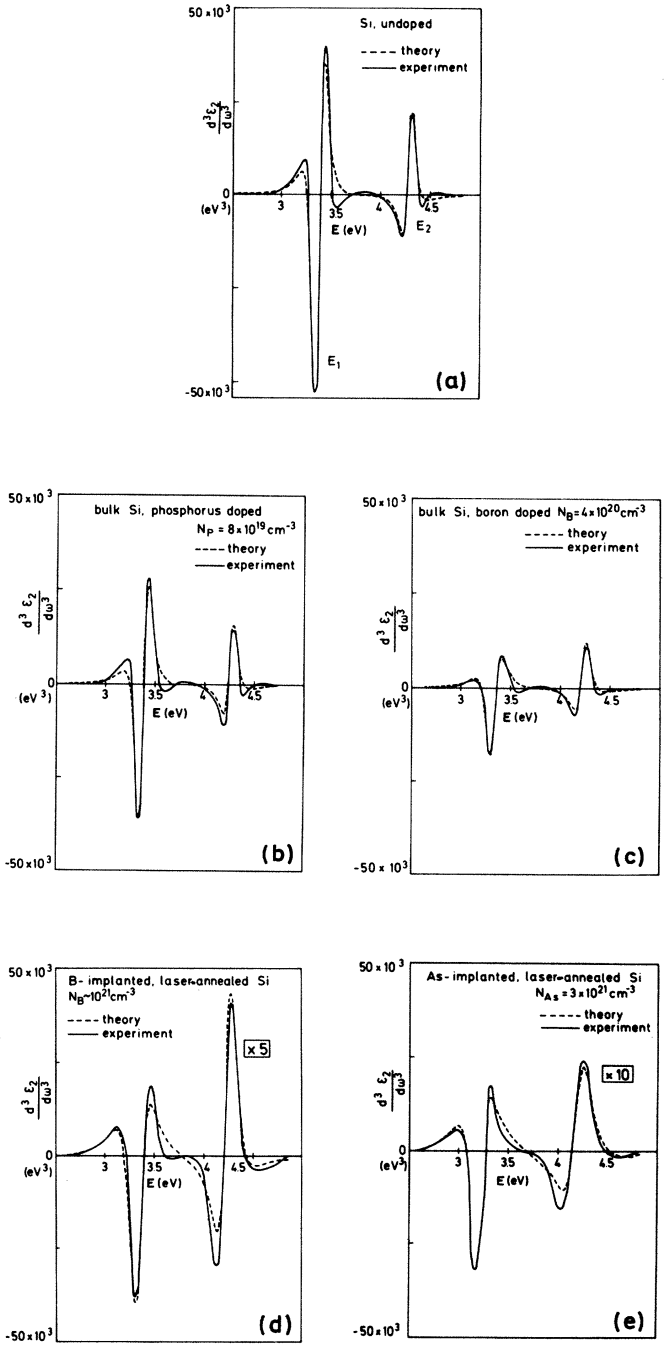

FIG. 2. Solid curves: third derivatives with respect to the photon energy of the imaginary part of the pseudodielectric constant ($d^3\epsilon_2/dw^3$) of the spectra shown in Fig. 1. (a) Undoped Si; (b) n- type Si (P); (c) p-type Si (B I); (d) p-type Si (B II); (e) n-type Si (As) (the notation for the samples is explained in Table I). Dashed curves: best fit of the E_1 and E_2 singularities. A mixture of a two-dimensional minimum and saddle point was used for E_1, and a one-dimensional maximum was assumed for E_2. The derivative spectra of the p-type Si (B II) and n-type Si(As) are enlarged by factors of 5 and 10, respectively.

we estimate the thickness of the film to be about 20 Å (varying slightly from sample to sample). In this case a three-phase model (air-SiO_2-Si) was used for the treatment of the data. It was assumed that the overlayer was SiO_2 with an index of refraction taken from the literature.[43] The equations were solved to obtain $\epsilon(\omega)$ using the two-dimensional Newton method.[44] As a check we have calculated the following integrals which yield well-known sum rules:[45]

$$N_{\text{eff}}=\frac{2m}{4\pi^2e^2N_{\text{at}}}\int_{\omega_m}^{\omega_M}\omega\epsilon_2(\omega)d\omega\ , \tag{1a}$$

$$\tilde{\epsilon}_1(\omega_0)=1+\frac{2}{\pi}\mathrm{P}\int_{\omega_m}^{\omega_M}\frac{\omega\epsilon_2(\omega)}{\omega^2-\omega_0^2}d\omega\ , \tag{1b}$$

where P means the Cauchy principal value of the integral, N_{at} is the atomic density, and m and e are the mass and charge of the electron, respectively.

We take $\omega_m=1.8$ eV, $\omega_M=5.6$ eV, and $\omega_0=2.51$ eV. For $\omega_m\to 0$ and $\omega_M\to\infty$, Eq. (1a) would yield $N_{\text{eff}}=4$ and $\tilde{\epsilon}_1(\omega_0)=\epsilon_1(\omega_0)$. The values obtained for N_{eff} and $\tilde{\epsilon}_1(\omega_0)$ are listed in Table II. N_{eff} for the pure sample is only 2.32, well below the number of valence electrons per atom (four). This well-known fact is due to the poor convergence of Eq. (1a) and the low cutoff value of $\omega_M=5.6$ eV. For n-type Si(P), nearly the same value of N_{eff} is obtained, while for p-type Si(B I), a somewhat lower value, namely $N_{\text{eff}}=2.19$, is found. The latter value of N_{eff} may indicate some transfer of oscillator strength to the region $\omega>5.6$ eV. However, we hesitate to reach this conclusion because of uncertainties in the surface conditions which may depend on the dopant.

The sum rule for $\tilde{\epsilon}_1$ [Eq. (1b)] converges much more rapidly than that for N_{eff}: for the undoped sample and for n-type Si(P), $\epsilon_1(2.51\text{ eV})-\tilde{\epsilon}_1(2.51\text{ eV})$ is only 1.2. This value increases to ~ 1.9 for p-type Si(B I), a fact which confirms the conjecture that we are measuring a pseudodielectric constant containing small effects of an uncharacterized surface layer. Surface roughness, decreasing the average Si density, is the most likely form of such a layer.[19] In view of the arbitrariness involved we do not try to model the IILA samples any further.

For sample n-type Si(As) we determined the oxide thickness by fitting the ϵ_2 data below 3 eV to the absorption spectra determined directly for thin IILA silicon-on-sapphire (SOS) samples.[14] The low value of $N_{\text{eff}}=1.94$ obtained for this sample suggests some density defect in the laser-annealed layer. For the p-type Si(B II) sample we assumed the same oxide thickness (20 Å) for the evaluation of ϵ_1 and ϵ_2. This sample seems to be better behaved than n-type Si(As) with regard to the sum rules (see Table II).

The lack of interference fringes below 2.8 eV (they were seen in some samples!) demonstrates that the unimplanted substrate does not affect the measurements. We note that free-carrier effects begin to appear in the ϵ_1 spectra at low frequencies for all of the heavily doped samples of Fig. 1 (ϵ_1 is lower than for the pure samples), especially for n-type Si(As), which has the largest carrier concentration.

In order to compare the shifts and broadenings, we calculate, numerically, the third-derivative spectra, $d^3\epsilon/d\omega^3$, of the complex dielectric function with respect to the photon energy from our ellipsometric data. The results for $d^3\epsilon_2/d\omega^3$ are shown in Fig. 2 for the five samples. The solid lines correspond to the numerical derivatives of the spectra of Fig. 1; the dashed lines are the best fit to the experimental ones. Note that the change in the line shape of Fig. 2(e) compared with Fig. 1 of Ref. 22 is due to the assumption of the 20-Å oxide top layer on the sample which was not taken into account in Ref. 22.

The third-derivative spectra show clearly a red shift of the energy positions of the E_1 and E_2 critical points from the pure to the highest doped sample, with a corresponding increase in their Lorentzian width Γ. These spectra were fitted assuming a two-dimensional minimum[46] for the E_1 singularity and a one-dimensional maximum for the E_2 critical point. Excitonic effects were also taken into account in a standard way by allowing a mixture of two critical points.[46,47] Hence, a two-dimensional minimum and a saddle point were used for the E_1 data. Both the real and imaginary parts of $d^3\epsilon/d\omega^3$ were simultaneously fitted.

The mixture of contiguous two-dimensional critical points can be expressed by[46,47]

$$\epsilon=A-\ln(\omega_0-\omega-i\Gamma)e^{i\varphi}\ , \tag{2}$$

where the angle φ represents the amount of mixture. In Fig. 3 we plot the values of $\tan\varphi$ obtained from our fits versus impurity concentration N_i.

The red shifts of the critical points E_1 and E_2 found from measurements of many bulk and implanted samples are shown in Fig. 4. The points represented by pluses (+) are taken from Ref. 19. They are plotted versus carrier concentration in a double-logarithmic scale. The effect of doping becomes noticeable for concentrations above 10^{19} cm^{-3}. The use of IILA samples allows us to obtain doping concentrations above the solubility limit,[27] and thus increases the range obtained with bulk samples. The results from these samples join smoothly with those of the bulk samples, as can be seen in Figs. 4(a) and 4(b).

TABLE II. Effective number of electrons N_{eff}, real part of dielectric constant obtained through K-K transformation, $\tilde{\epsilon}_1$, and measured real part of dielectric constant ϵ_1, for the five samples of Fig. 1.

	Pure	n-type Si(P)	p-type Si(B I)	p-type Si(B II)	n-type Si(As)
N_{eff}	2.32	2.31	2.19	2.21	1.94
$\tilde{\epsilon}_1(2.51\text{ eV})$	17.7	17.5	17.4	16.5	15.7
$\epsilon_1(2.51\text{ eV})$	19.0	18.7	19.3	18.5	18.0
$\epsilon_1-\tilde{\epsilon}_1$	1.3	1.2	1.9	2.0	2.3

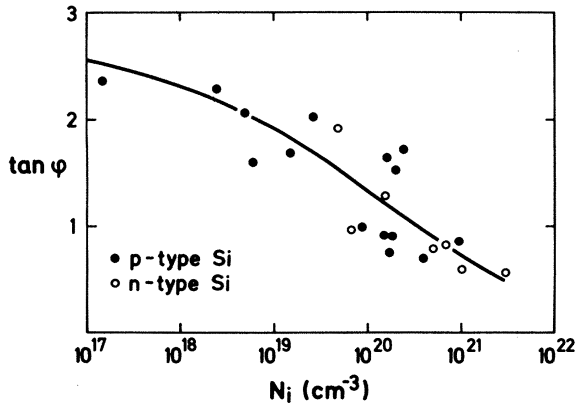

FIG. 3. Dependence on doping concentration of the excitonic parameter $\tan\varphi$, defined in Eq. (2) for the E_1 critical point. Solid circles, p-type silicon; open circles, n-type silicon. Solid line, best fit obtained using Eq. (30).

We have not made separate fits for the n- and the p-type samples since the observed shifts for a given doping level are independent of type within the experimental accuracy.

The increase in the Lorentzian broadening parameter Γ with respect to the undoped sample obtained from the linewidth analysis of the E_1 and E_2 critical points are shown in Figs. 5(a) and 5(b) versus doping. Also here, no distinction was made between p- and n-type samples since no systematic differences were observed. We should mention that these broadenings depend somewhat on the type of critical point chosen. If we take for the E_2 structure an M_2 three-dimensional critical point, we obtain broadenings which are nearly half of those given in Fig. 5(b). In view of this, a more exact determination of the type of critical point, using the Fourier-transform method,[48] would be desirable.

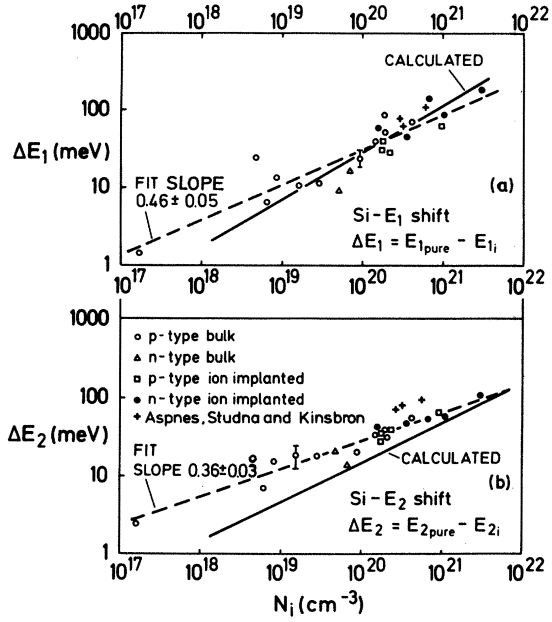

FIG. 4. Dependence on doping concentration of critical point energies: (a) red shifts (ΔE_1) for the E_1 singularity; (b) red shifts (ΔE_2) for the E_2 critical point. Open circles, bulk p-type material; open triangles, bulk n-type material; open squares, p-type IILA samples; solid circles, n-type IILA samples. Pluses, Ref. 19. Dashed lines, best fit to a N_i^{α} law, yielding $\alpha=0.46$ for E_1 and $\alpha=0.36$ for E_2. Solid line, results of the second-order perturbation-theory calculations. The shifts are referred to the undoped sample.

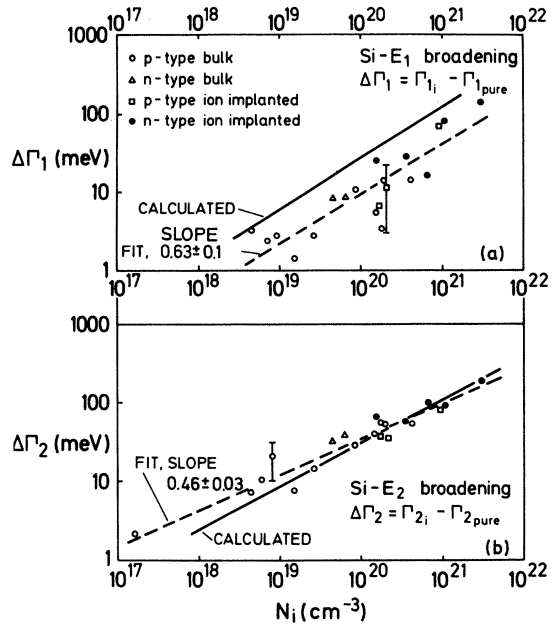

FIG. 5. Dependence on doping concentration of critical point broadening parameter: (a) increasing of broadening parameter ($\Delta\Gamma_1$) for the E_1 singularity; (b) increasing of broadening parameter ($\Delta\Gamma_2$) for the E_2 singularity. The symbols have the same meaning as in Fig. 4. The best fit to a N_i^{α} law yields $\alpha=0.63$ for E_1 and $\alpha=0.46$ for E_2.

IV. THEORY

A. General

In order to evaluate the renormalization of the semiconductor band structure in the presence of impurities, we need a Hamiltonian that describes the coupled electron-impurity system. Let us explicitly write the Hamiltonian as follows:

$$\hat{H}=\hat{H}_e+\hat{H}_{ei}\ , \tag{3}$$

where

$$\hat{H}_e=\sum_{\vec{k},n} e_{\vec{k},n}c^{\dagger}_{k,n}c_{\vec{k},n} \tag{4}$$

describes the band structure of pure Si. $c^{\dagger}_{\vec{k},n}$ and $c_{\vec{k},n}$ are the creation and annihilation operators for an electron with energy $e_{\vec{k},n}$ and wave vector $\vec{k}$ in band n, respectively. The carrier-impurity part of the Hamiltonian is[34,49]

$$\hat{H}_{ei}=\sum_{\vec{k}n;\vec{k}',n'}\langle\vec{k}',n'\mid V_I\mid\vec{k},n\rangle c^{\dagger}_{\vec{k}',n'}c_{\vec{k},n}\ , \tag{5}$$

where V_I represents the difference between the crystal potential with and without impurities present.

We assume that N impurities are randomly distributed at sites $\vec{R}_i$, and that V_I can be written as a sum of identical localized potentials V^{imp} centered at the impurities sites $\vec{R}_i$,

$$V_I=\sum_i V_i^{\text{imp}} . \tag{6}$$

For the $e_{\vec{k},n}$ we use the pseudopotential band structure obtained with the local pseudopotentials of Cohen and Bergstresser[50] (see Table III).

Using Bloch's theorem, the matrix element in Eq. (5) can be written

$$\langle \vec{k}',n' \mid V_I \mid \vec{k},n \rangle = S(\vec{k}'-\vec{k})\langle \vec{k}',n' \mid V^{\text{imp}} \mid \vec{k},n \rangle , \tag{7}$$

with

$$S(\vec{k}'-\vec{k})=\sum_i e^{-i(\vec{k}'-\vec{k})\vec{R}_i} ,$$

the structure factor of the impurities, i.e., the Fourier transform of the density function $\sum_i \delta(\vec{r}-\vec{R}_i)$ for the scattering centers.

In order to take into account the randomness of the impurity locations, an ensemble average over these locations must be performed. The $\vec{k}$ vector, BZ, periodicity in $\vec{k}$ space, etc., which were lost under the random distribution of impurities, remain clearly defined because periodicity in $\vec{r}$ space is restored after this averaging.

With the assumption that there is no correlation for the location of the substitutional impurities, the impurity form factor has the following ensemble averages:

$$\langle S(\vec{k}'-\vec{k})\rangle = N\delta_{\vec{k},\vec{k}'} , \tag{8a}$$

$$\langle S(\vec{q})S(-\vec{q}')\rangle = N\delta_{\vec{q},\vec{q}'}+\langle S(\vec{q})\rangle\langle S(\vec{q}')\rangle . \tag{8b}$$

The energy bands renormalized by the presence of the impurities are given by the zeros of[34]

$$\det[(z-e_{\vec{k},n})\delta_{nn'}-\Sigma(\vec{k},n,n';z)]=0 , \tag{9}$$

where $\Sigma(\vec{k},n,n';z)$ is the self-energy matrix which depends on the band indices n and n'.

If the self-energy Σ is small it is realistic to use Rayleigh-Schrödinger perturbation theory to find the zeros of Eq. (9). Up to second order in the impurity potential, the result is a complex energy $E_{\vec{k},n}$,

$$E_{\vec{k},n}=e_{\vec{k},n}+\Delta_{\vec{k},n}+\frac{i}{2\tau_{\vec{k},n}}=e_{\vec{k},n}+\Sigma(\vec{k},n;E_{\vec{k},n})$$

$$=e_{\vec{k},n}+\Sigma^{(1)}+\Sigma^{(2)} . \tag{10}$$

Equation (10) neglects correlated scattering effects from different impurities, i.e., it assumes that the impurities act independently of each other.

B. First-order perturbation terms

The term $\Sigma^{(1)}$ in Eq. (10) represents the average first-order effect and involves the structure factor $\langle S(\vec{k}-\vec{k}'=\vec{0})\rangle=N$. From Eq. (7) we find

$$\Sigma^{(1)}(\vec{k},n;z)=N\langle \vec{k},n \mid V^{\text{imp}} \mid \vec{k},n \rangle . \tag{11}$$

This gives a real self-energy correction which depends only on the $\vec{k}$-diagonal matrix elements of V^{imp}. It takes into account the fact that the average periodic potential has been slightly altered.

We have used empirical pseudopotential form factors for evaluating Eq. (11) for group-V donors and group-III acceptors. These form factors are estimated from the corresponding values for the III-V compounds. The atomic form factors v_G can be evaluated from the symmetric and antisymmetric form factors v_G^{S} and v_G^{A} of a III-V compound with the expression

$$v_G^{(\text{III})}=v_G^{\text{S}}+v_G^{\text{A}} , \tag{12a}$$

$$v_G^{(\text{V})}=v_G^{\text{S}}-v_G^{\text{A}} . \tag{12b}$$

v_G^{S} and v_G^{A} are only known at a few values of G ($\sqrt{3}$, $\sqrt{4}$, $\sqrt{8}$, and $\sqrt{11}$ in units of $2\pi/a$). The $v_G^{(\text{III})}$ and $v_G^{(\text{V})}$ needed for calculating the effect of the impurities in silicon

TABLE III. Pseudopotential form factors V_3, V_8, and V_{11} (in Ry) used for the calculations of the first-order effect of impurities on the band structure of silicon. In parentheses are the III-V compounds from which the pseudopotentials were extracted.

Pseudopotential	V_3	V_8	V_{11}
Si[a]	−0.21	+0.04	+0.08
N (AlN,[b] GaN[b])	−0.40(1)	−0.182(3)	−0.09(1)
P (GaP,[a] AlP,[b] InP[a])	−0.33(1)	+0.02(2)	+0.06(1)
As (GaAs,[a] InAs[a])	−0.318(2)	+0.009(2)	+0.04(2)
Sb (GaSb,[a] InSb,[a] AlSb[a])	−0.31(1)	+0.03(3)	+0.04(1)
B (BN,[b] BP[b])	−0.285(5)	−0.06(3)	+0.00(5)
Al (AlSb,[a] AlP,[b] AlN[b])	−0.13(5)	+0.08(2)	+0.08(2)
Ga (GaP,[a] GaSb,[a] GaAs,[a] GaP[b])	−0.14(4)	+0.05(2)	+0.08(1)
In (InP,[a] InAs,[a] InSb[a])	−0.17(1)	+0.06(2)	+0.06(3)

[a]From Ref. 50.
[b]From Ref. 67.

TABLE IV. Linear coefficients $\Delta E_g/dN_i$ (in units of 10^{-24} eV cm^3) of the dependence of optical gaps on substitutional dopant concentration calculated by first-order perturbation theory. A positive $-dE_g/dN_i$ corresponds to a *red shift* upon doping. The limits of error correspond to calculations using pseudopotentials obtained for different III-V compounds. $E_{\rm ind}(X)$ and $E_{\rm ind}(L)$ represent the lowest gaps at the X and L points, respectively.

	$-\frac{dE_g}{dN_i}$						
(10^{-24} eV cm^3)	E_0	E_0'	E_1	E_2	E_1'	$E_{\rm ind}(X)$	$E_{\rm ind}(L)$
N	433±10	−23± 8	192± 8	42± 6	−15±13	28±10	177±10
P	62±18	−38± 4	4± 8	−34±6	−36± 4	−56±10	−8±8
As	88±16	−26± 6	22±10	−20± 4	−26± 6	−37± 8	13±10
Sb	48±4	−22± 8	4± 5	−32± 6	−21± 9	−41± 6	3± 2
B	198±20	−18±14	89±40	24±16	−10±10	20±14	84±40
Al	−60± 6	30± 8	−9± 6	10±15	28± 8	20±14	−8±15
Ga	− 34±15	28±10	−3± 8	16±12	26±10	24±12	4±10
In	−24±14	26± 8	−6±11	4± 6	24± 8	16±10	−4±11

correspond to slightly different values of G because of the fact that the lattice constant of Si is not the same as that for a given III-V compound. We thus interpolate the v_G's by using the general functional dependence of v_G on G given in Ref. 51. These v_G's so obtained must also be renormalized to take into account the change in volume of the unit cell,

$$V_G^{(\rm Si)}=\frac{\Omega_{\rm III\text{-}V}}{\Omega_{\rm Si}}v_G^{(\rm III\text{-}V)}, \tag{13}$$

where Ω represents the volume of the unit cell. The pseudopotential form factors so obtained are listed in Table III. The values in this table represent the average obtained for all III-V compounds which contain a given atom. The estimated error bars represent the differences between the various possible compounds. The average pseudopotentials v_0 are not known with certainty. While they could induce changes in the absolute energies of the various states, they shift all states by the same amount and are thus of no relevance for the shifts of energy gaps.

A more accurate method than the direct evaluation of Eq. (11) is the "virtual crystal approximation" (VCA, i.e., the exact diagonalization of the Hamiltonian matrix with averaged pseudopotentials,

$$V(G)=(1-x)V_G^{(\rm Si)}+x\,V_G^{(\rm II\ or\ V)}, \tag{14}$$

for a small value of x ($x<0.01$). By using two values of x one can convince oneself of the fact that we are in the linear region, i.e., that the energy shifts are proportional to x. We have used the VCA, but the linearity of our results implies that Eq. (11) is equally good for the small values of x under consideration.

In Table IV we show the results of this procedure for the $E_0(\Gamma_{25'}\rightarrow\Gamma_{2'})$, $E_0'(\Gamma_{25'}\rightarrow\Gamma_{15})$, $E_{\rm ind}(\Gamma_{25'}\rightarrow X_1)$, $E_{\rm ind}(\Gamma_{25'}\rightarrow L_1)$, $E_1(\Lambda_{3'}\rightarrow\Lambda_1)$, $E_1'(\Lambda_{3'}\rightarrow\Lambda_3)$, and E_2 gaps of Si. For the E_2 gap we took the point $\vec{k}=(2\pi/a)(0.9,0.1,0.1)$.[23] Rather similar results were obtained at the X point. For the E_1 transitions we have calculated the shifts for all points along the Λ direction $\vec{k}=(\pi x/a)(1,1,1)$ (see Fig. 6) and averaged them for $\frac{1}{4}\le x\le 1$. At E_1' we used the same procedure while averaging for $\frac{1}{2}\le x\le 1$.

In Table IV we also show linear shifts of Si gaps with doping calculated as described above. Here and in the rest of the article positive shifts mean *red* shifts with increasing doping. The shifts in Table IV are usually about an order of magnitude smaller than the second-order shifts

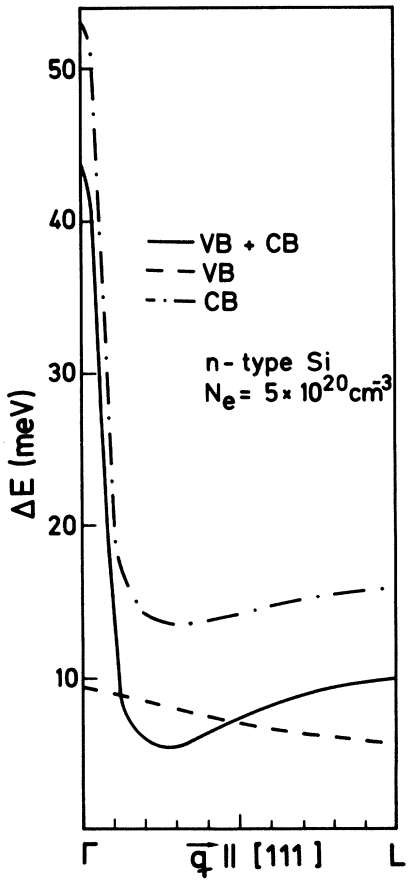

FIG. 6. Self-energy of As-doped silicon with a carrier concentration $N_e=5\times10^{20}$ cm^{-3} obtained by first-order perturbation theory along the Λ direction. Dashed line red shift of the highest valence band (VB) [$\Delta E_{\rm VB}=E_{\rm VB}$ (pure Si)-$E_{\rm VB}$(doped)]; dotted-dashed line, shift of the lowest conduction band (CB) [$\Delta E_{\rm CB}=E_{\rm CB}$ (pure Si)-$E_{\rm CB}$(doped)]; solid line, red shift of the band gap (VB + CB) ($\Delta E_T=\Delta E_{\rm CB}-\Delta E_{\rm VB}$). The calculation was made with pseudopotential form factors for As in Si listed in Table III.

due to the Coulomb potential and are calculated in Sec. IV C. The only exceptions are found for some of the gaps in the case of N and B doping. For these atoms, however, the pseudopotentials are rather ill defined. This is the only effect considered here which depends on the nature of the dopant; the second-order effect of Sec. IV C does not. For a given impurity some gaps exhibit red shifts and some blue shifts without a clear systematics. However, gaps which exhibit red shifts for donors usually show blue shifts for acceptors, as expected. In some cases where this rule seems to be violated (e.g., As vs Ga), it actually still holds when the margins of error are considered.

C. Second-order perturbation terms

The next term in the perturbation series can be written[34]

$$\Sigma^{(2)}(\vec{k},n,n')=\frac{N\sum_{\vec{q},m}\langle \vec{k}n \mid V^{\rm imp}\mid \vec{k}+\vec{q},m\rangle\langle \vec{k}+\vec{q},m \mid V^{\rm imp}\mid \vec{k},n'\rangle}{e_{\vec{k},n}-e_{\vec{k}+\vec{q}m}+i\delta} . \tag{15}$$

This term corresponds to a second-order perturbation via a virtual intermediate state (scattering of an electron initially in the one-particle state $\mid \vec{k},n\rangle$ into another one-particle state $\mid \vec{k}+\vec{q},m\rangle$ and back to $\mid \vec{k},n\rangle$ again). In addition to the shift in the band structure given by

$$\Delta^{(2)}_{\vec{k},n}=N\sum_{\vec{q},m}\frac{\mid\langle \vec{k},n \mid V^{\rm imp}\mid \vec{k}+\vec{q},m\rangle\mid^2}{(e_{\vec{k},n}-e_{\vec{k}+q,m})} , \tag{16}$$

this term produces an energy broadening $\Gamma_{\vec{k}n}$ (lifetime $\tau_{\vec{k}n}$),

$$\Gamma_{\vec{k},n}=\frac{1}{2\tau_{\vec{k},n}}=\pi N\sum_{\vec{q},m}\mid\langle \vec{k},n \mid V^{\rm imp}\mid \vec{k}+\vec{q},m\rangle\mid^2 \times\delta(e_{\vec{k},n}-e_{\vec{k}+\vec{q}m}) . \tag{17}$$

In this case we assume screened "hydrogenic" Coulomb scatterers. The Fourier transform of the screened impurity potential is (atomic units $\hbar=m=e=1$ will be used in the following)

$$[V^{\rm imp}(q)]=S(q)V^{\rm imp}(q)=\frac{4\pi}{\epsilon(q)q^2}S(q) , \tag{18}$$

where $S(q)$ is the impurity structure factor in the unit cell.

The calculation should be performed separately for both donors and acceptors. Both cases differ because of the details of the corresponding band edges. Here we perform the calculation for acceptors. The case of donors will be published elsewhere. The results are not very different from those for acceptors.

In the case of acceptors, and in the random-phase approximation (RPA) the dielectric function $\epsilon(q)$ is

$$\epsilon(q)=\epsilon_L(q)-\frac{4\pi}{q^2}\frac{1}{V}[F_l(q_l^*)+F_h(q_h^*)] , \tag{19}$$

where $\epsilon_L(q)$ is the dielectric function of the pure host semiconductor, and $F_l(q_l^*)$ $[F_h(q_h^*)]$ is the Lindhard polarizability[52] for the light (heavy) holes. The renormalized q, labeled q^*, is defined in Eq. (20c) below.

For the Lindhard polarizabilities F_l and F_h we use the following set of equations valid for $k_BT\ll E_F$:

$$E_F=\frac{1}{2}\left[\frac{3\pi^2N_h}{m_l^{3/2}+m_h^{3/2}}\right]^{2/3} , \tag{20a}$$

$$[D(E_F)]_{l\,(h)}=\frac{(2m_{l(h)})^{3/2}}{2\pi^2}E_F^{1/2} , \tag{20b}$$

$$q^*_{l\,(h)}=\frac{q}{(2E_Fm_{l\,(h)})^{1/2}} , \tag{20c}$$

$$g(q)=\frac{1}{2}\left\{1-\frac{1}{q}\left[1-\left[\frac{q}{2}\right]^2\right]\ln\frac{1-q/2}{1+q/2}\right\} , \tag{20d}$$

$$F_{l\,(h)}(q)=-V[D(E_F)]_{l\,(h)}g(q_{l\,(h)}) , \tag{20e}$$

where E_F is the Fermi energy, $[D(E_F)]_{l\,(h)}$ is the density of states at the Fermi energy, and $m_{l\,(h)}$ of the effective mass of light (heavy) holes. We use the values $m=0.154$ and 0.523.[53] N_h is the hole density whereby we assume all the impurities to be ionized.

For the q-dependent dielectric constant of the pure semiconductor we have used Penn's interpolation formula,[54]

$$\epsilon_L(q)=1+\left[\frac{\omega_p}{E_g}\right]^2D\left[1+\frac{E_F}{E_g}\left[\frac{q}{k_F}\right]^2D^{1/2}\right]^{-2} ,$$

where $D=1-\frac{1}{4}(E_g/E_F)$, ω_p is the plasma frequency ($\omega_p=4\pi N_v$, with N_v denoting the valence-electron concentration), k_F is the Fermi wave vector, and E_g is a parameter representing an average energy gap (the Penn gap). In our case (Si) the following values were used: $\omega_p=16.6$ eV, $E_F=12.5$ eV, $k_F=0.96$ bohr, and $E_g=4.8$ eV.

The second-order terms can be divided roughly into two categories. They are more clearly seen when one takes the Thomas-Fermi (TF) approximation for the dielectric constant. In this case, Eqs. (16) and (17) take the form

$$\Delta^{(2)}_{\vec{k},n}(\Gamma_{\vec{k},n})\sim\frac{N_h}{(q^2+q_{\rm TF}^2)^2}\sim\frac{N_h}{(q^2+CN_h^{1/3})^2} , \tag{21}$$

where C is a constant.

Terms with a small q transfer to the intermediate state yield a contribution proportional to $N_h^{1/3}$. On the other hand, terms with a large q transfer yield a contribution

proportional to N_h. The total result depends on which of both processes is dominant, but it will lie between $N_h^{1/3}$ and N_h.

We now proceed to compute the two expressions $\Delta^{(2)}_{\vec{k},n}$ and $\Gamma_{\vec{k},n}$. Equation (15) can be rewritten as

$$\Sigma^{(2)}_{\vec{k},n}(E)=\lim_{\delta\to 0^+}\int \frac{dE'}{E-E'+i\delta}A_{\vec{k},n}(E') , \tag{22}$$

where $A(E)$, the spectral electron-impurity operator, is defined as

$$A(E)=N_h\sum_{\vec{k}',m}[V^{\rm imp}]\,|\,\vec{k}',m\rangle \times\delta(E-e_{\vec{k}',m})\langle\vec{k}',m\,|\,[V^{\rm imp}] . \tag{23}$$

Its matrix elements are

$$A_{\vec{k},n}(E)=N_h\sum_{\vec{k}',m}\langle\vec{k},n\,|\,[V^{\rm imp}]\,|\,\vec{k}'m\rangle \times\langle\vec{k}',m\,|\,[V^{\rm imp}]\,|\,\vec{k},n\rangle\delta(E-e_{\vec{k}',m}) . \tag{24}$$

Using the Dirac identity, Eq. (22) can be written

$$\Sigma^{(2)}_{\vec{k},n}(E)=\mathrm{P}\int\frac{A_{\vec{k},n}(E')}{E-E'}dE'-i\pi A_{\vec{k},n}(E) , \tag{25}$$

where P means principal value.

In this way,

$$\Delta^{(2)}_{\vec{k},n}=\mathrm{Re}\Sigma^{(2)}_{\vec{k},n}(e_{\vec{k},n})=\mathrm{P}\int\frac{A_{\vec{k},n}(E')}{e_{\vec{k},n}-E'}dE' , \tag{26a}$$

$$\Gamma_{\vec{k},n}=-\mathrm{Im}[\Sigma^{(2)}_{\vec{k},n}(e_{\vec{k},n})]=\pi A_{\vec{k},n}(e_{\vec{k},n}) . \tag{26b}$$

Here we calculate $\Gamma_{\vec{k},n}$ through direct integration over the BZ and $\Delta^{(2)}_{\vec{k},n}$ by performing the Hilbert transform of $\Gamma_{\vec{k},n}$.

We restrict ourselves to the case of the diamond structure (Si); the impurities are substitutionally located at places $R(l,\kappa)$, where l labels the primitve cells and κ labels the two sublattices. The origin is taken midway between the atoms so that their coordinates are $\tau=+\tau_1=-\tau_2=(a/8)(1,1,1)$.

The electronic wave functions are eigenvectors of the secular equation

$$\sum_{\vec{G}'}\{[(\vec{k}+\vec{G})^2-e_{\vec{k},n}]\delta_{\vec{G},\vec{G}'} + V(\vec{G}-\vec{G}')S(\vec{G}-\vec{G}')\}c_{\vec{k},n}(\vec{G}')=0 ,$$
$$\psi_{\vec{k},n}=\Omega^{-1/2}\sum_{\vec{G}}c_{\vec{k},n}(\vec{G})e^{i(\vec{k}+\vec{G})\cdot\vec{r}} , \tag{27}$$
$$\sum_{\vec{G}}c^2_{\vec{k},n}(\vec{G})=1 ,$$

where $V(\vec{G})$ is the local pseudopotential form factor and $S(\vec{G})=\cos(\vec{G}\vec{\tau})$. The coefficients $c_{\vec{k},n}(\vec{G})$ are chosen to be real.

The matrix elements of Eq. (24) are easily worked out to yield

$$\langle\vec{k},n\,|\,[V^{\rm imp}]\,|\,\vec{k}',m\rangle=\phi(\vec{k},n;\vec{k}',m)+i\theta(\vec{k},n;\vec{k}',m) , \tag{28a}$$

$$\phi(\vec{k}n;\vec{k}',m)=\sum_{\vec{G},\vec{G}'}c_{\vec{k},n}(\vec{G})c_{\vec{k}',m}(\vec{G}') \times V^{\rm imp}(\vec{k}-\vec{k}'+\vec{G}-\vec{G}') \times\cos[(\vec{k}-\vec{k}'+\vec{G}-\vec{G}')\cdot\vec{\tau}] . \tag{28b}$$

θ is identical to ϕ except that the cosine is replaced by the sine of the same argument.

We now evaluate the spectral function $A_{\vec{k},n}(E)$ for the conduction and valence states which correspond to the E_1 and E_2 gaps in Si. The matrix elements of Eq. (28a) are calculated using pseudo-wave-functions. The values of $V(\vec{G})$ are taken from Ref. 50 (see Table III). Because of the large time required, most of our calculations were made for 15 plane-wave energy bands. The calculation with this 15×15 Hamiltonian gives results similar to those performed with a 59×59 secular equation[55,56] (to within 20%). The sum over the BZ was performed over a discrete mesh of 89 $\vec{k}$ points in the irreducible $\frac{1}{48}$th wedge of the BZ. The spectral functions were computed with the use of the tetrahedron method.[57–59] The calculations take ~5 min of CPU (central processing unit) time for a state at the Γ point on a Honeywell-Bull 66-80P computer.

1. E_1 gap

The Γ point case is the simplest because Eq. (28a) then has full cubic symmetry: sums over the BZ can be restricted to the irreducible zone. For a general point $\vec{k}$ along Λ (E_1 transitions), the $\vec{k}$ sum is also carried out over the irreducible part of the BZ. A sum is then performed over all vectors in the star of $\vec{k}$ in order to obtain $A_{\vec{k},n}(E)$. Since the sum was carried out in the irreducible $\frac{1}{48}$th wedge of the BZ, the contributions corresponding to the eight vectors of the star of $|\vec{k}|\{1,1,1\}$ were calculated and added to obtain $A_{\vec{k},n}(E)$.

The calculated spectral functons $\pi A_{\vec{k},n}(E)$ are shown in Fig. 7 for the states at the L point, i.e., $L_{3'}$ and L_1, for a concentration of 5×10^{20} cm^{-3}, together with the density of states (DOS). In the case of the valence band ($L_{3'}$), this function is strongly peaked close to the band edge and the other bands in the DOS are strongly suppressed. The spectral function calculated for the conduction band (L_1) follows the curve of the DOS much more closely; the major contribution also comes close to the band edge, but the other contributions are not as strongly reduced as in the

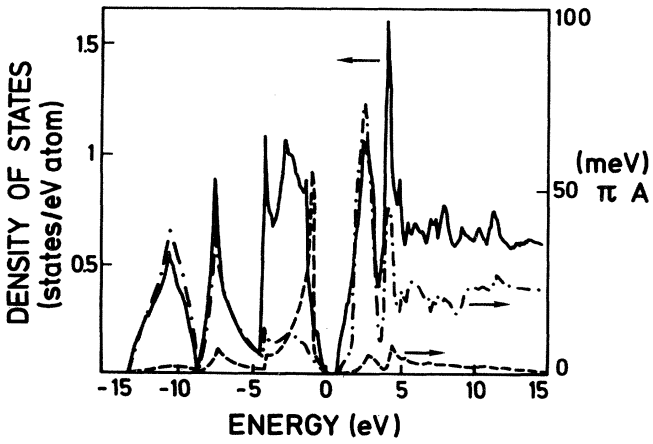

FIG. 7. Dashed line: spectral function πA for a valence-band state at the $L_{3'}$ point. Dotted-dashed line: spectral function for a conduction-band state at the L_1 point. These curves correspond to a carrier concentration $N_h=5\times10^{20}$ cm^{-3}. The DOS of Si is also shown as a solid line for comparison.

valence-band case. The spectral function for the mainly s-like L_1 conduction band is particularly large at the energies of the lowest valence bands (also s-like). Similar observations, which can be explained on the basis of orthonormality, can be made for the $L_{3'}$ valence band. Keeping in mind that the shifts in the bands are obtained through a Hilbert transform of the spectra functions, Fig. 7 shows the importance of a complete band-structure calculation using real Bloch states instead of plane waves.

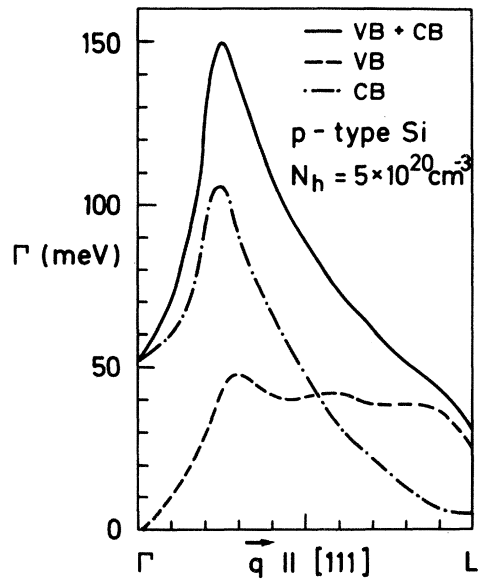

FIG. 8. Lorentzian-broadening parameter along the Λ direction for holes in Si with a carrier concentration $N_h=5\times10^{20}$ cm^{-3}. Dashed line: broadening of states at the highest valence band (Γ_{VB}). Dotted-dashed lines: broadening of states at the lowest conduction band (Γ_{CB}). Solid line: total broadening parameter, $\Gamma=\Gamma_{CB}+\Gamma_{CB}$.

The Lorentzian-broadening parameters of the valence and conduction bands (VB and CB, respectively) together with the total $\Gamma=\Gamma_{VB}+\Gamma_{CB}$ along the Λ direction are shown in Fig. 8 for holes in Si with a carrier concentration of 5×10^{20} cm^{-3}. This parameter vanishes at the Γ point for valence-band states, as expected from the DOS, but has a finite value of $\sim$50 meV for conduction-band states. The major contribution is seen to arise from the conduction-band states and from $\vec{k}$ points close to $(\pi/4a)(1,1,1)$. For $|\vec{k}|>|(\pi/2a)(1,1,1)|$ the major contribution arises from the valence-band states.

In Fig. 9 the shifts (real part of second-order self-energy) are shown also for p-type Si with $N_h=5\times10^{20}$ cm^{-3} along the Λ direction. The total red shift arises mostly from a down shift of the conduction band and up shift of the valence band ($\Delta E_T=\Delta E_{CB}-\Delta E_{VB}$). The contribution of each of both bands to the total shift depends on the particular $\vec{k}$ point of the BZ, but both contributions are of the same order of magnitude over a large part of the $\Gamma-L$ points. The directly measurable quantities, the total shifts, and the increases of Lorentzian broadenings, for the E_1 singularity, are taken to be the average of those from $\vec{k}=(\pi/4a)(1,1,1)$ to the L point [a mesh of 0.02 $(2\pi/a)$ was used]. The calculated averages are shown by a solid line in Figs. 4(a) (shifts) and 5(a) (broadenings) as a function of carrier concentration together with the experimental points and the best fit (dashed line) to a N^{α} law. The agreement between theory

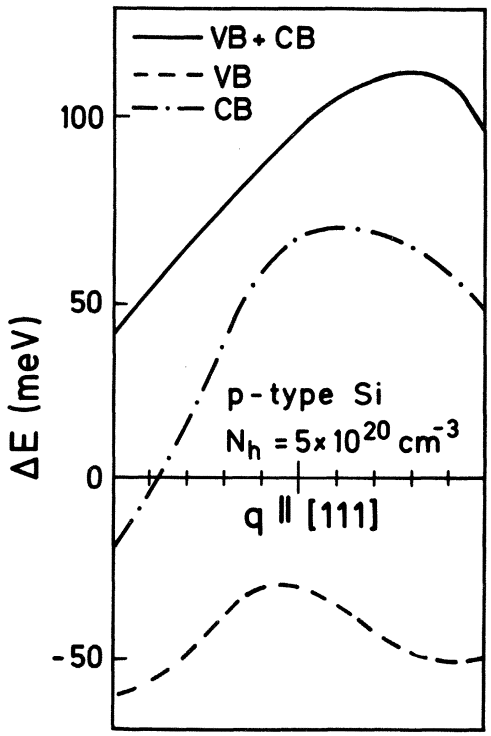

FIG. 9. Energy shifts obtained by second-order perturbation theory for a p-type silicon with a carrier concentration $N_h=5\times10^{20}$ cm^{-3} along the Λ direction. Dashed line: real part of the self-energy for states at the highest valence band (VB). Dotted-dashed line: real part of the self-energy for states at the lowest conduction band (CB). Solid line: total second-order red shift of the band gap (VB + CB).

and experiment is satisfactory. The measured $\Delta\Gamma_1$ depend somewhat on the type of critical point chosen for the fit.[22] They may decrease if exciton effects are properly taken into account.

2. E_2 gap

For the E_2 critical point we have made calculations at the point $(2\pi/a)(0.9,0.1,0.1)$ (here labeled P). In this case the star of $\vec{k}$ contains 24 points; the calculation took ~ 2 h of CPU time. The region of $\vec{k}$ space where most of the E_2 transitions take place is not very well determined. Hence, we have also calculated the self-energy at the X point of the BZ, i.e., the X_4 and X_1 points for the valence and conduction bands, respectively. The results are summarized in Table V for two concentrations, 0.01 and 1 at. % and for both points X and P. In view of the better agreement between theory and experiment for the P point, we have taken this point as being a better representative of the E_2 singularity. The $\vec{k}$-space multiplicity of this point also supports this choice. In Figs. 4(b) (shifts) and 5(b) (broadenings), we show the theoretical results together with experimental points and their best fit with N^{α}. Again, the agreement between theory and experiment is satisfactory.

V. DISCUSSION

Perhaps the most conspicuous effect of the doping seen in Fig. 1 is the blurring of the E_1 peak which has become a shoulder for $N_{As}=3\times10^{21}$ cm^{-3}. Similar effects have also been noted in Refs. 19 and 60. In order to describe the effect quantitatively, in Fig. 3 we have plotted $\tan\varphi$ [see Eq. (2)] vs N_i. $\tan\varphi$ represents, within the simple contact-interaction model, the amount of excitonic interaction in the E_1 structure.[46,47] Figure 3 clearly shows the decrease of excitonic interaction with increasing N_i. We assume that this interaction is screened by the free carriers according to the expression

$$\bar{V}(q)=\frac{Vq^2}{q^2+q_{TF}^2} . \tag{29}$$

Contact-interaction theory corresponds to a q-independent V. In order to approximately extend this theory to the potential $V(q)$ of Eq. (29), we replace q by an average value $q\simeq r_{ex}^{-1}$ where r_{ex} is an effective exciton radius. With this assumption we obtain

$$\tan\varphi=A\left[1+\frac{4\times3^{1/3}m_d^{*}N_i^{1/3}r_{ex}^2}{\epsilon_0\pi^{1/3}}\right]^{-1} , \tag{30}$$

where m_d^{*} is DOS mass ($m_d^{*}=1.06$ for n-type Si and 0.58 for p-type Si). We have fitted all the points of Fig. 3 with an average value of the mass, $m_d^{*}=0.82$ and obtained $r_{ex}\simeq 9$ Å, a reasonable result in view of the value of 10 Å, for the E_1 exciton of GaAs.[61]

We have fitted the shifts of Figs. 4(a) and 4(b) with the function CN_i^{α} ($C=$const) and find values of $\alpha=0.46$ for E_1 and 0.36 for E_2, nearly the same in view of the fit confidence of α ($\sim\pm0.05$). The first-order perturbation shifts given in Table IV are negligible except for $N_i>10^{21}$ cm^{-3} (the only exception may be the E_1 gap for B doping). However, even for the few samples with these high values of N_i we find no evidence for these linear shifts and thus neglect them. The prediction of the second-order perturbation calculation is quite satisfactory for E_1 although the value of α found theoretically (0.64) is somewhat larger than the experimental one. For the E_2 gap the theoretical α is 0.53. The calculated shifts are smaller than the experimental ones, a fact which may be due to the extended nature (in $\vec{k}$ space) of these transitions. The $\vec{k}$ points chosen for the calculation may not be fully representative of the actual transitions.

In this paper we have discussed the electronic effects due to the screened pseudopotentials. A few words regarding possible effects of the hard cores of the impurities are in order. These effects manifest themselves as changes in the lattice constant which should only be important in the case of boron ($\Delta a/a_0\simeq-5\times10^{-24}N_i$ for B; $\lesssim10^{-24}$ for P and As).[62,63] The effect of this strain on energy gaps can be estimated with the deformation potentials of the E_1 and E_2 gaps given in Ref. 64. It amounts to a blue shift of 30 meV for E_1 and 20 meV for E_2 in the case of $N_i=4\times10^{20}$ B atoms cm^{-3}, the maximum bulk doping investigated. Since the electronic effects are sublinear in N_i, this effect can be neglected for most of the bulk samples measured. For the ion-implanted samples the hydrostatic strain should be half of that estimated above because of the lateral constraint of the substrate. This would make the strain effect negligible within experimental scatter except for the maximum boron concentration measured (10^{21} cm^{-3}), at which blue hydrostatic shifts of 40 meV for E_1 and 25 meV would be expected for E_2. The uniaxial component of the strain should be negligible for the [100] surfaces measured. These shifts may be somewhat larger if surface microcracks release, in part, the lateral constraint. The rather uncertain red shifts estimated in Table IV for boron by first-order perturbation theory are likely to offset the strain shift discussed above, which, in any case, is not clearly observed in Fig. 4 (note, however, that the measured point for $N_i=10^{21}$ B atoms cm^{-3} lies slightly below the fitted line, as would be expected for a blue shift due to strain).

TABLE V. Self-energies and broadening calculated for the gaps at X ($X_4\rightarrow X_1$) and $\vec{k}=(2\pi/a)(0.9,0.1,0.1)$ for two impurity concentrations. These $\vec{k}$ points are representative of the E_2 transitions.

N_i (cm^{-3})	$\vec{k}$ point	E_2 shift (meV)	E_2 broadening (meV)
5×10^{18}	(0.9,0.1,0.1)	3	6
	X	3	4
5×10^{20}	(0.9,0.1,0.1)	37	78
	X	30	59

Similar considerations may be made for the broadenings of Fig. 5. The experimental and theoretical slopes for the E_1 transitions are the same (0.63). The calculated $\Delta\Gamma_1$, however, are nearly twice as large as the measured ones. Although this difference is not very critical, a possible explanation may be the neglect of excitonic resonances implicit in our fits. This may also explain why the increases in the widths calculated for the E_2 structures [Fig. 5(b)] agree rather well with experiment (measured slope 0.46, calculated slope 0.56). These structures are not severely modified by exciton interaction.[25]

We have performed our first- and second-order perturbation calculations neglecting multiple scattering up to relative impurity concentrations of 1 at. %. In the case of short-range impurity potentials these calculations should certainly be a good approximation. Because of the relatively long range of the screened Coulomb potential some questions are left as to the validity of the approximation. More elaborate calculations, such as those based on the coherent-potential-approximation,[65] would be desirable. Ultimately, one would also like to compute the complete dielectric function in the presence of the impurities.[34] Such calculations, however, are beyond the scope of this work.

VI. CONCLUSIONS

We have measured the effect of heavy doping (up to 3×10^{21} cm^{-3}), both n and p type, on the optical critical points of silicon. Both IILA and bulk samples give concordant results. Band-structure calculations of these effects, i.e., of the self-energies of the critical points, agree with experimental results. An erosion of the excitonic effects on the E_1 critical point has also been observed and quantitatively studied. It can be attributed to screening of the excitonic interaction by the free carriers.

ACKNOWLEDGMENTS

We would like to thank A. Compaan and C. Umbach for help with some of the measurements, G. Kisela and A. Breitschwert for sample preparation and characterization, H. Bleder and H. Birkner for help with construction of the ellipsometer, and M. Polo for her contribution to the numerical calculations. We thank P. B. Allen for help with the theory and providing computer programs and Dr. A. Axmann for the ion-implantation of the IILA samples.

[1]R. A. Abram, G. J. Rees, and B. L. H. Wilson, Adv. Phys. 27, 799 (1978).
[2]R. W. Keyes, Comments Solid State Phys. 7, 149 (1977).
[3]N. Kobayashi, S. Ikehata, S. Kobayashi, and W. Sasaki, Solid State Commun. 24, 67 (1977).
[4]W. Sasaki and J. Kinoshita, J. Phys. Soc. Jpn. 25, 1622 (1968).
[5]D. F. Holcomb, in *The Metal Non-Metal Transition in Disordered Systems,* edited by L. R. Friedman and D. P. Tunstall (Scottish Universities Summer School Program, Edinburgh, 1978).
[6]J. Doehler, Phys. Rev. B 12, 2917 (1975).
[7]K. Jain, S. Lai, and M. V. Klein, Phys. Rev. B 13, 5448 (1976).
[8]P. E. Schmid, M. L. Thewalt, and W. P. Dumke, Solid State Commun. 38, 1091 (1981).
[9]J. Wagner, Phys. Rev. B 29, 2002 (1984).
[10]W. Spitzer and H. Y. Fan, Phys. Rev. 108, 268 (1957).
[11]A. A. Vol'fson and V. K. Subashiev, Fiz. Tekh. Poluprovodn. 1, 397 (1967) [Sov. Phys.—Semicond. 1, 327 (1967)].
[12]M. Balkanski, A. Aziza, and E. Amzallag, Phys. Status Solidi 31, 323 (1969).
[13]P. E. Schmid, Phys. Rev. B 23, 5531 (1981).
[14]L. Viña, C. Umbach, M. Cardona, A. Compaan, and A. Axmann, Solid State Commun. 48, 457 (1983).
[15]H. Fritzsche, in *The Metal Non-Metal Transition in Disordered Systems,* Ref. 5.
[16]M. Cardona and H. S. Sommers, Jr., Phys. Rev. 122, 1382 (1961).
[17]E. Vigil, J. A. Rodríguez, and R. Pérez-Alvarez, Phys. Status Solidi B 90, 409 (1978).
[18]D. E. Aspnes, G.K. Celler, J. M Poate, G. A. Rozgonyi, and T. T. Sheng in *Laser and Electron Beam Processing of Electronic Materials,* edited by C. L. Anderson, G. K. Celler, and G. A. Rozgonyi (Electrochemical Society, Princeton, N.J., 1980) p. 414.
[19]D. E. Aspnes, A. A. Studna, and E. Kinsbron Phys. Rev. B 29, 768 (1984).
[20]G. E. Jellison, Jr., F. A. Modine, C. W. White, R. F. Wood, and R. T. Young, Phys. Rev. Lett. 46, 1414 (1981).
[21]L. Viña and M. Cardona, Physica 117&118B, 356 (1983).
[22]L. Viña, C. Umbach, A. Compaan, M. Cardona, and A. Axmann, in Proceedings of the Materials Research Society, Symposium on Laser Processing of Materials, Strasbourg, 1983 [J. Phys. (Paris) Colloq. 44, C5-203 (1983)].
[23]J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B 14, 556 (1976).
[24]M. Cardona, in *Atomic Structure and Properties of Solids,* edited by E. Burstein (Academic, New York, 1972).
[25]W. Hanke and L. J. Sham, Phys. Rev. B 21, 4656 (1980); H. J. Mattausch, W. Hanke, and G. Strinati, *ibid.* 27, 3735 (1983).
[26]M. Miyao, T. Motooka, N. Natsuaki, and T. Tokuyama, in *Laser and Electron-Beam Solid Interactions and Materials Processing,* edited by J. F. Gibbons, L. D. Hess, and T. W. Sigmon (North-Holland, New York, 1981).
[27]C. W. White, P. P. Pronki, S. R. Wilson, B. R. Appleton, J. Narayan, and R. T. Young, J. Appl. Phys. 50, 3261 (1979).
[28]J. C. Inkson, J. Phys. C 9, 1170 (1970).
[29]G. D. Mahan, J. Appl. Phys. 51, 2634 (1980).
[30]K.-F. Berggren and B. E. Sernelius, Phys. Rev. B 24, 1971 (1981).
[31]A. Selloni and T. Pantelides, Phys. Rev. Lett. 49, 586 (1982).
[32]H. Van Cong, Phys. Status Solidi B 117, 575 (1983).
[33]M. Takeshima, Phys. Rev. B 27, 2387 (1983).
[34]P. B. Allen, Phys. Rev. B 18, 5217 (1978); B. Chakraborty and P. B. Allen, *ibid.* 18, 5225 (1978).
[35]D. E. Aspnes, Appl. Phys. Lett. 39, 316 (1981).
[36]The crystals were implanted by Dr. A. Axmann at the Fraunhofer-Institut für angewandte Festkörperphysik, Freiburg, Federal Republic of Germany.
[37]A. Compaan, G. Contreras, M. Cardona, and A. Axmann, in Proceedings of the Material Research Society, Symposium on Laser Processing of Materials, Strasbourg, 1983, Ref. 22.
[38]D. E. Aspnes, Opt. Commun. 8, 222 (1973).

[39]D. E. Aspnes and A. A. Studna, Appl. Opt. 14, 220 (1975).
[40]D. E. Aspnes, Physica 117&118B, 359 (1983).
[41]D. E. Aspnes, J. Opt. Soc. Am. 64, 812 (1974).
[42]N. M. Bashara and R. M. Azzam in *Ellipsometry and Polarized Light* (North-Holland, Amsterdam, 1977).
[43]*American Institute of Physics Handbook,* 3rd ed., edited by D. E. Gray (McGraw-Hill, New York, 1953).
[44]A. S. Householder, in *Principles of Numerical Analysis* (McGraw-Hill, New York, 1953).
[45]D. L. Greenaway and G. Harbeke, *Optical Properties and Band Structure of Semiconductors* (Pergamon, London, 1968).
[46]M. Cardona, *Modulation Spectroscopy,* Suppl. 11 of *Solid State Physics,* edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969).
[47]Y. Toyozawa, M. Inoue, T. Inui, M. Okazaki, and E. Hanamura, J. Phys. Soc. Jpn. Suppl. 21, 133 (1967); J. E. Rowe and D. E. Aspnes, Phys. Rev. Lett. 25, 162 (1970).
[48]D. E. Aspnes, Surf. Sci. 135, 284 (1983).
[49]S. Doniach and E. H. Sondheimer, *Green's Functions for Solid State Physicists* (Benjamin, Reading, Mass., 1974).
[50]M. L. Cohen and T. K. Bergstresser, Phys. Rev. 141, 789 (1966).
[51]M. L. Cohen and V. Heine, in *Solid State Physics,* edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1970), Vol. 24.
[52]J. Lindhard, Kgl. Dansk. Videnskab. Selskab. Mat.-Fys. Medd. 28, No. 8 (1954).
[53]J. C. Hensel and G. Feher, Phys. Rev. 129, 1041 (1963).
[54]D. R. Penn, Phys. Rev. 128, 2093 (1962).
[55]M. Cardona and F. H. Pollak, Phys. Rev. 142, 530 (1966).
[56]P. B. Allen and M. Cardona, Phys. Rev. B 23, 1495 (1981).
[57]G. Lehmann and M. Taut, Phys. Status Solidi B 54, 469 (1972); 57, 815 (1973).
[58]J. Rath and A. J. Freeman, Phys. Rev. B 11, 2109 (1975).
[59]O. Jepsen and O. K. Andersen, Solid State Commun. 9, 1763 (1971).
[60]G.-J. Jan and F. H. Pollak, Sol. Energy Mater. 8, 241 (1982).
[61]J. E. Rowe, F. H. Pollak, and M. Cardona, Phys. Rev. Lett. 22, 933 (1969).
[62]J. A. Vergés, D. Glötzel, M. Cardona, and O. K. Andersen, Phys. Status Solidi B 113, 519 (1982).
[63]B. C. Larson and J. F. Barhorst, J. Appl. Phys. 51, 3181 (1980).
[64]Landolt-Börnstein Tables, edited by O. Madelung, H. Schulz, and H. Weiss (Springer, Berlin, 1982), Vol. 17, p. 45.
[65]B. Velický, S. Kirkpatrick, and H. Ehrenreich, Phys. Rev. 175, 747 (1968); S. Kirkpatrick, B. Velický, and H. Ehrenreich, *ibid.* B 1, 3250 (1970).
[66]D. H. Lowndes, J. W. Cleland, W. H. Christie, R. E. Eby, G. E. Jellison, Jr., J. Narayan, R. D. Westbrook, R. F. Wood, J. A. Nilson, and S. C. Dass, Appl. Phys. Lett. 41, 938 (1982).
[67]J. A. Sanjurjo, E. López-Cruz, P. Vogl, and M. Cardona, Phys. Rev. B 28, 4579; (1983).