Band-gap shrinkage of semiconductors

D. Auvergne, J. Camassel, and H. Mathieu

Centre d'Etudes d'Electronique des Solides, associé au Centre National de la Recherche Scientifique, Université des Sciences et Techniques du Languedoc, Place E. Bataillon, 34060 Montpellier-Cedex, France

(Received 12 September 1974)

Direct experimental evidence of band-gap shrinkage due to the presence of impurities in semiconductors is given. Experimental results are obtained using a piezomodulation-spectroscopy technique over the (4–300) °K temperature range. They are interpreted in the linear-screened-potential and effective-mass approximations. Coupling spectroscopic and helium-temperature magnetotransport measurements allowed us to obtain in this model a good description of the transition edges over the entire temperaure range. We studied different GaSb samples with impurity concentration allowing us to observe band-gap shrinkage with one and two types of carriers. This is easily obtained in the GaSb case whose band structure presents Γ and L minima located very near in energy. On heavily doped samples, the low-temperature values of the band-gap shrinkage are used to obtain further informations concerning subsidiary minima. On lightly doped semiconductors, non-k-conserving transitions initiating on the residual acceptor level are clearly shown.

I. INTRODUCTION

There has been great interest in the rigidity of the band structure of semiconductors, when a large concentration of impurities is introduced into the lattice. More particularly, the large wavelength shift of band-edge emission in doped semiconductor lasers and its temperature variation have been studied by several authors.[1–3] For example, using a lossless cavity Holonyak *et al.*[3] pointed out that it was possible to obtain stimulated emission at an energy greater or lower than the energy gap, depending on the pumping level.

On the other hand, absorption experiments on doped Ge,[4] InP,[5] and[6] GaSb crystals and modulated reflectivity on[7] GaAs showed that the variation of the transition edge in heavily doped semiconductors cannot be entirely explained by the introduction of the Burstein shift. In particular, experimental values are always lower than theoretical predictions, and obey a law[5]: $E_{gn} = E_{gp} + \alpha E_F$, where E_{gp} is the fundamental edge on pure crystal and α is a constant less than unity. This lowering of the gap on doped materials, as well as the properties of the spontaneous emission edge, have been interpreted by different models. From qualitative consideration the screened potential of randomly distributed donors tends to depress the conduction band and to form tails of the density of states at the band edges. This shrinkage of the energy gap has been described, in terms of Coulomb and exchange interactions,[8–10] to compute the spontaneous emission rate in order to evaluate the properties of injection lasers.

Using three different models,[9] the second-order perturbation method used by Parmenter in the virtual-crystal approximation, the theory of Kane approach, the theory of Halperin and Lax derived from the minimum counting method, C. J. Hwang[10] shows that the number of occupied states in the bandtail is much less than the total number of electrons, and that the Fermi level in the uncompensated n-type crystal is never located in the bandtail for any impurity concentration at which the calculation is valid. This result is in agreement with a number of experimental results such as obtained in[10] GaAs and[11] GaSb injection lasers. Consequently, physical-parameter characteristics of doped semiconductors such as position of electron Fermi level relative to the shifted band edge, carrier screening length, and band shift due to Coulomb interaction can be calculated with sufficient accuracy using a parabolic density of state.

In order to obtain precise evidence of this shrinkage in this paper we shall discuss the variation with temperature of the band shift obtained in doped semiconductors with different impurity concentrations. This band shift is obtained directly from the comparison of direct transition energies determined in pure and doped samples. In order to eliminate uncertainties involved in transmission experiments, we compare, using modulation spectroscopy, the temperature variation of the fundamental edge, obtained in a pure GaSb crystal, to the transition edge measured in Te-doped crystals with different impurity concentration. The low-temperature Fermi energy is determined in doped samples from oscillations of the resistivity in a longitudinal magnetic field. Experimental results are presented in Sec. II. In Sec. III we review the theoretical formulation of the band-gap shrinkage, using the simple model of a screened potential, in order to describe the interaction between all charges. Particular care is taken in the

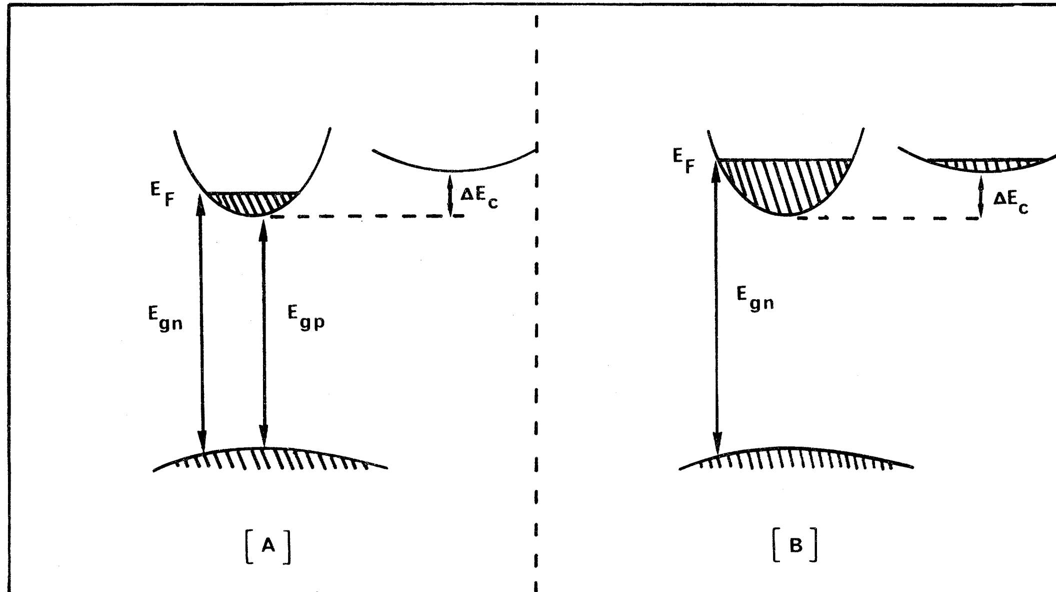

FIG. 1. Schematic band structure of (a) moderately and (b) heavily doped direct semiconductors. E_{gn} and E_{gp} are, respectively, the energy gap in pure and heavily doped samples. E_F is the Fermi level measured from the bottom of the conduction band. In the (b) case, where the subsidiary minimum is also degenerate, the temperature evolution of the population in the conduction band is strongly dependent on the ΔE_c temperature variation.

formulation of the temperature dependence of the carrier distribution in different minima of the conduction band.

Comparison of experimental results with theoretical predictions of the band-gap shrinkage is given over the all-temperature range, in the last part.

II. EXPERIMENTAL RESULTS

In order to obtain precise information about the fundamental edge, we choose a piezomodulation technique described elsewhere.[12] As mentioned, this method permits us to work easily at any temperature over a large energy interval.

As shown in Fig. 1, to interpret the experimental results a knowledge of the Fermi energy is required at each temperature and doping level. For this reason we have studied *Te*-doped GaSb samples because this impurity introduces a donor level, degenerate with the conduction band, at 125 meV above Γ_{1c}.[13] In this case, for sufficient doping the free-carrier concentration is constant and temperature independent. Knowing the Fermi energy at 4 °K, it is then possible to determine its temperature variation.

We have studied four samples with different doping levels (labeled *A*, *B*, *C* and *D* in Table I), in which the Fermi level was determined from longitudinal magnetoresistance oscillations, similar to those shown in Fig. 2.

Approximating the low-temperature Fermi function by a step function we calculated the transition energies in doped samples,[14] corrected for the Burstein shift, by means of

$$E'_{gn} = E_{gp} + E_F\,(1 + m_c/m_v). \qquad (1)$$

The results are given in the three first columns of Table I.

Figures 3 and 4 show low-temperature piezoreflectance spectra measured in these samples and the evolution of transition energies between 4 and 300 °K. Figure 4 gives the comparison of these variations to that obtained in a natural *p*-type sample. The E_{gn} low temperature transition edges are also given on Table I, with the difference $\Delta E = E'_{gn} - E_{gn}$ between experimental and theoretical [Eq. (1)] values.

TABLE I. Comparison of the experimental and theoretical values of the fundamental edge, determined on a *n*-type GaSb samples with different doping level. The theoretical edge E'_{gn} is deduced from Eq. (1), the experimental one; E_{gn}, is obtained from the piezoreflectance spectra.

Sample	n (10^{18} cm^{-3})	E_F(4 °K) (meV)	E'_{gn}(4 °K) (meV)	E_{gn}(4 °K) (meV)	ΔE(4 °K) (meV)
A	2.2	99	924	882	42
B	1.2	86	909	862	47
C	0.46	40	858	774	84
D	0.23	30	847	784	63

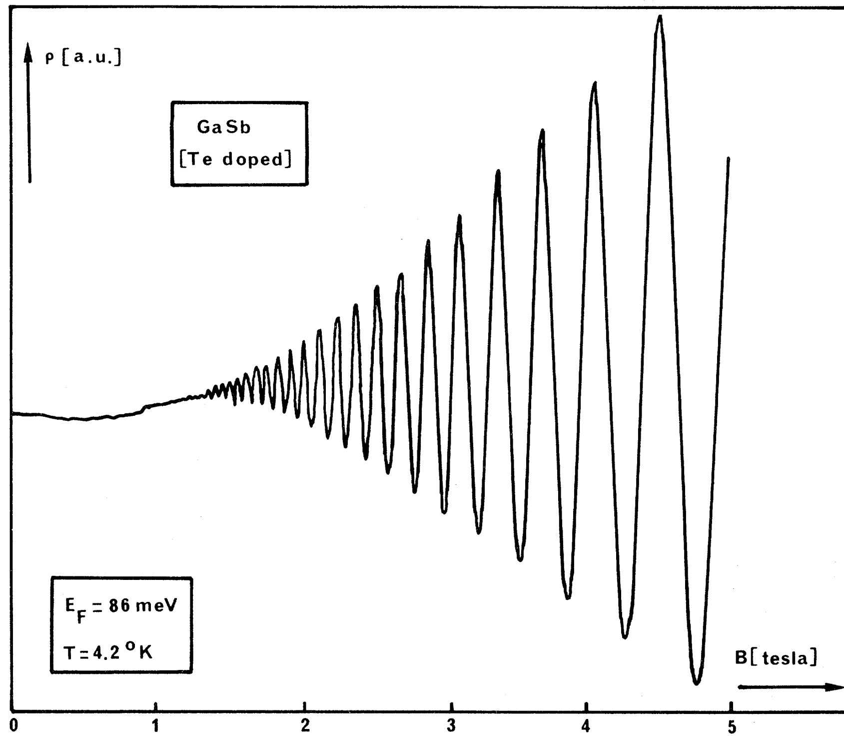

FIG. 2. Illustration of the longitudinal magnetoresistance spectrum obtained in the B sample at helium temperature. The oscillations of the resistivity correspond to the crossing of the Landau level through the Fermi level: $E_F = (n_i + \frac{1}{2})\, \hbar e H_i / m_c$, where H_i represents the value of the field at the n_i oscillation. This gives $E_F = [\Delta n / \Delta(1/H)]\,(e\hbar/m_c)$.

At high temperature the transition energy for doped samples, is lower than that obtained in a natural one (E_p). On C and D samples the transition edge is always lower than E_p. This denotes an apparent non-k-conserving excitation acceptor-level → Fermi-level, as this can be obtained from the k extension of the impurity wave function in compensated semiconductors.[15] The ΔE lowering can be interpreted in terms of the interaction between all the charges contained in the system. We

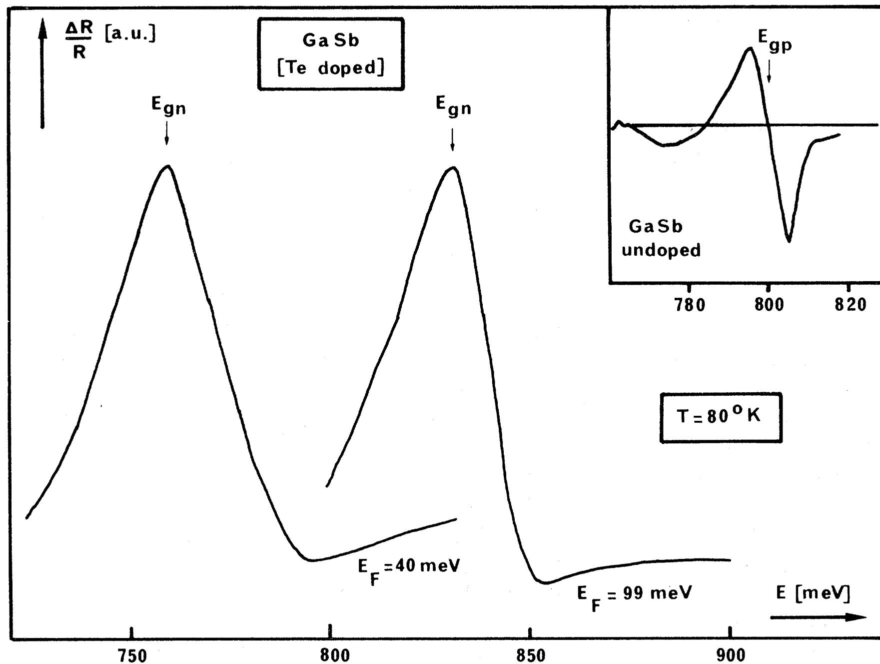

FIG. 3. Examples of low-temperature piezoreflectance spectra obtained on Te-doped GaSb samples.

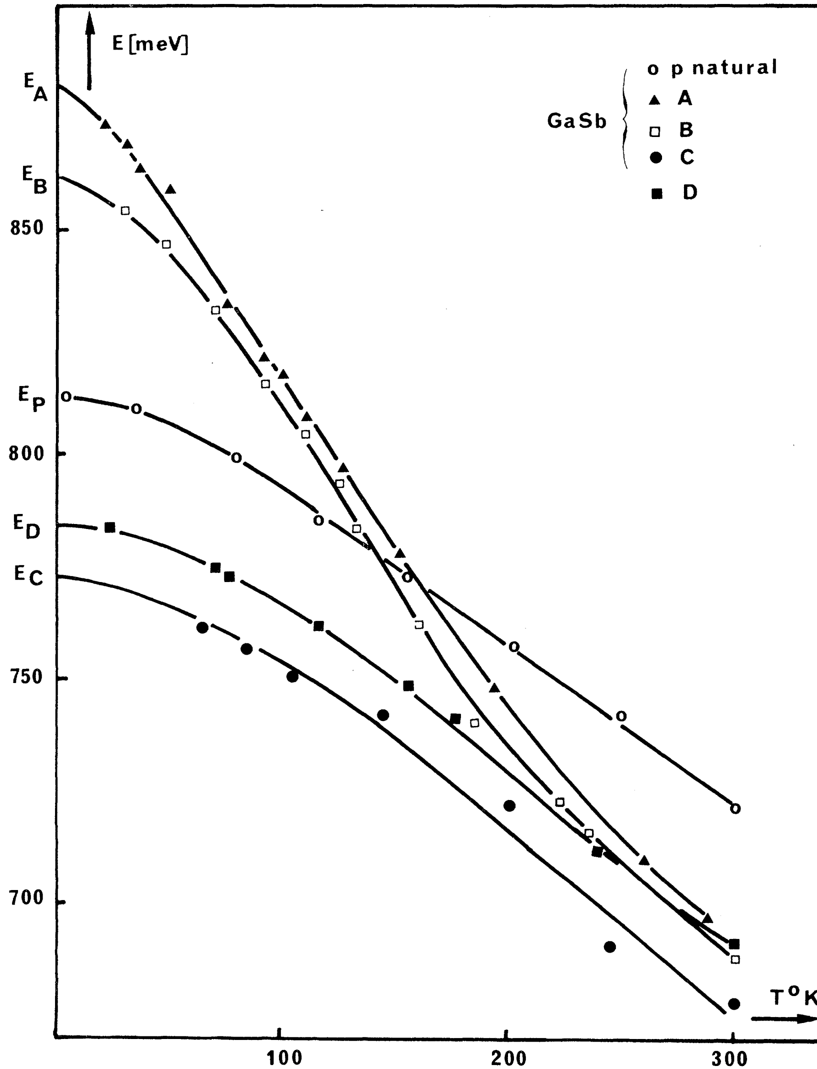

FIG. 4. Comparison of transition energy edge between pure (p type) and n-doped GaSb samples. At helium temperature $E_F(A)=99$ meV, $E_F(B)=86$ meV, $E_F(C)=40$ meV, and $E_F(D)=30$ meV.

are going to review the different terms contributing to ΔE in order to describe the experimental results over the entire temperature range by means of

$$E_{gn}=E_{gp}+E_F\,(1+m_c/m_v)-\Delta E. \tag{2}$$

III. BAND-GAP SHRINKAGE

In order to calculate stimulated and spontaneous emission spectra of GaAs lasers, Hwang[10] determined in the effective-mass approximation the conduction and valence-band density of states in doped semiconductors, modified by the core potential of the impurities.

Using Wolff's[8] results he shows that the system of all electrons in the conduction band can be approximated by single electrons moving in the field of a screened impurity potential, provided the whole conduction band is shifted downward by a quantity equal to E_c^e. E_c^e represents the electron-electron interaction energy.

In this case the one-electron Hamiltonian can be written as

$$H=-\frac{\hbar^2\nabla^2}{2m_c^*}+E_c^0-E_c^e-E_c^c+V(x), \tag{3}$$

where E_c^0 is the energy of the $k=0$ conduction-band state in the crystal, E_c^c is the average impurity potential producing the shift of the conduction band, and $V(x)$ represents the potential fluctuation about E_c^c.

When $V(x)=0$ the solutions of H are plane waves, and the density of states is parabolic at energies greater than $E_c^0-E_c^e-E_c^c$. The fluctuations $V(x)$ produce localized band states below the conduction band: the tail of the density of states at energies lower than $E_c^0-E_c^e-E_c^c$. The self-consistent calculation by Hwang[10] of the density of states in the

tail shows that the number of electrons in tail states is always smaller than the electron concentration. So the Fermi energy relative to the bottom of the band can be calculated with a good accuracy in the formalism of a parabolic density of states.

In these conditions the valence–conduction-band transition edge measured in doped materials can be developed by means of an expression similar to Eq. (2), in which ΔE is the band-gap shrinkage produced by the charges contained in the crystal.

For an n-type semiconductor ΔE is composed of two terms:

$$A: E_c^e = \frac{e}{2\pi \epsilon_0 \epsilon_r}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3}\ (\text{eV}), \tag{4}$$

which represents the electron-electron interaction term.[8] Equation (4) is only applicable for heavily doped semiconductors. This corresponds to the condition that the average electron separation is lower than the Bohr radius in the crystal. This gives a condition on the carrier concentration:

$$n^{1/3} > 11.7\times 10^7 (m_c/m_0)(1/\epsilon_r)\ (\text{cm}^{-1}), \tag{5}$$

$$B: E_c^c = (e/\epsilon_0 \epsilon_r Q^2) \sum_a N_a Z_a\ (\text{eV}), \tag{6}$$

which represents the average screened potential produced by N_a impurities of charge Z_a.

$$Q^2 = \sum_\alpha Q_\alpha^2 \tag{7}$$

defines the reciprocal screening length of the carriers moving in the band α. In the linear screening approximation of Thomas-Fermi[16] this quantity is given by

$$Q_\alpha^2 = \frac{e^2}{\epsilon_r \epsilon_0} \int \rho_c(E) \ -\left(\frac{\partial f(E)}{\partial E}\right) dE\ , \tag{8}$$

where $\rho_c(E)$ is the density of states and $f(E)$ the Fermi function in the conduction band. In the parabolic approximation[10] we obtain

$$Q_\alpha^2 = \frac{4\pi^2 e^2 m_\alpha^{3/2} (2\pi kT)^{1/2}}{h^3 \epsilon_r \epsilon_0}\ \mathfrak{F}_{-1/2}\ \frac{E_F - E_c}{kT}\ , \tag{9}$$

where $\mathfrak{F}_{-1/2}$ is the Fermi integral.[16] Equation (9) can be simplified in two limiting cases: (a) For a degenerate electronic system ($E_F - E_c > 0$), $\mathfrak{F}_k(n) = n^{k+1}/(k+1)!$ and Eq. (9) becomes

$$Q_\alpha^2 = \frac{4\pi e^2}{\epsilon_0 \epsilon_r h^2}\left(\frac{3}{\pi}\right)^{1/3} m_\alpha n_\alpha^{1/3}\ , \tag{10}$$

where n_α is the free-carrier concentration in the band α. (b) For a nondegenerate system ($E_F - E_c < 0$), $\mathfrak{F}_k(\eta) \simeq e^\eta$ and

$$Q_\alpha^2 = \frac{e^2}{\epsilon_0 \epsilon_r kT}\ n_\alpha\ . \tag{11}$$

In the case of interest, in heavily doped GaSb samples (A and B) the carriers are distributed in the Γ and L conduction-band minima. Labeling by m_1, n_1 and m_2, n_2 the effective mass and carrier concentration of the Γ and L minima, the screening length of a degenerate sample takes the simple form

$$\frac{1}{Q^2} = \frac{\epsilon_0 \epsilon_r h^2}{4\pi e^2}\left(\frac{\pi}{3}\right)^{1/3} \frac{1}{m_1 n_1^{1/3} + m_2 n_2^{1/3}}\ . \tag{12}$$

This gives

$$E_c^c(\text{eV}) = \frac{h^2}{4\pi e}\left(\frac{\pi}{3}\right)^{1/3} \frac{n_1 + n_2}{m_1 n_1^{1/3} + m_2 n_2^{1/3}}\ . \tag{13}$$

Figure 5 gives an illustration of E_c^c variation vs doping level in the GaSb case at 4 °K. The values of the parameters used are $m_1 = 0.047 m_0$, $\Delta E_c = 85$ meV and various values of m_2. In the case of one kind of carrier ($n_2 = 0$), E_c^c is a linear function of $n_1^{2/3}/m_1$. For $m_1 n_1^{1/3} = m_2 n_2^{1/3}$ the variation of E_c^c shows a discontinuity corresponding to the advent of the carriers of heavier mass. For these carriers the screening length is lower than that of the first minimum, so this decreases E_c^c. When $m_2 n_2^{1/3} > m_1 n_1^{1/3}$ we obtain the $n_2^{1/3}$ variation which is this time inversely proportional to the second-minimum density-of-state mass.

As a verification we have plotted in Fig. 6 the difference $E_{gp} - E_{gn}$ corresponding to the direct edge of As-doped germanium, as determined by Pankove[4] at helium temperature. In this particular case free carriers are only located in L minima and we can write

$$E_{gp} - E_{gn} = E_c^c \sim \frac{n_2^{2/3}}{m_2}\ . \tag{14}$$

With $m_2 = 0.53\ m_0$,[17] we obtain $E_c^c(\text{meV}) = 4.6\, n^{2/3}$ in excellent agreement with the value of 4.7 representing the slope of the variation given in Fig. 6. We have to note that the straight line of this figure does not intercept the axis origin because Eq. (10), (13), and (14) can only be used in the degenerate case.

We are now going to show that using this definition of the band-gap shrinkage, we are able to reproduce the variations of transition energies observed in doped semiconductors.

IV. RESULTS AND DISCUSSION

We want to show that using the preceding formulation of the band-gap shrinkage it is possible to explain the variations E_A, E_B, E_C, and E_D plotted in Fig. 4 by using the E_{gp} variation modified by the conservation equation (2). This equation permits us to compute the energy at which interband transitions occur in a doped sample, in terms of the energy gap determined in a pure crystal and the

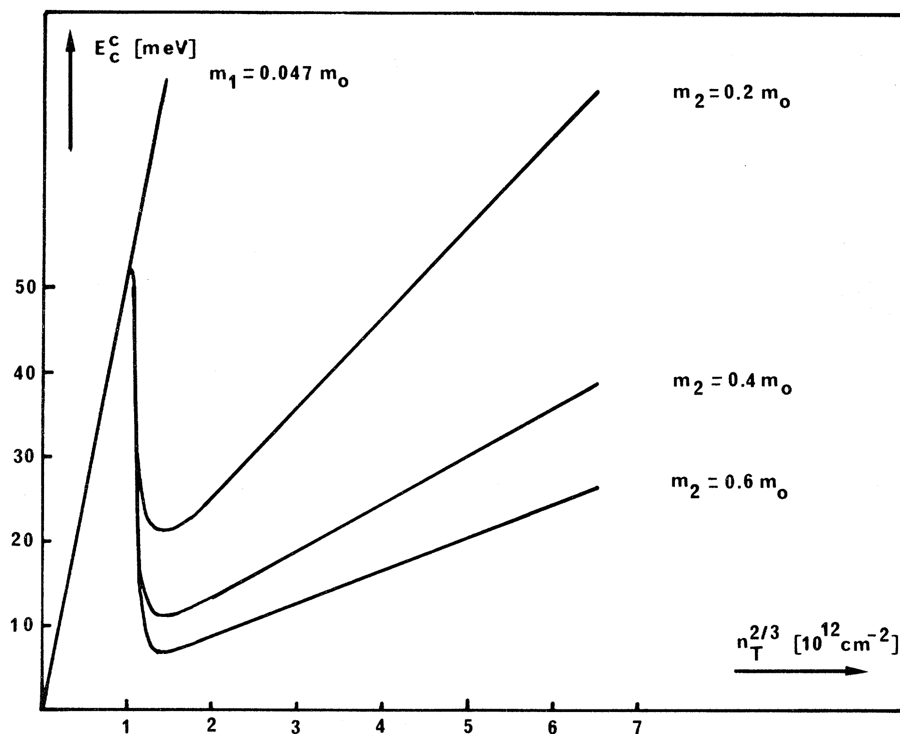

FIG. 5. Electron impurity interaction vs electron concentration, in the multivalley-conduction-band case. The GaSb example is given with $m_1 = 0.047 m_0$, $\Delta E_c = 85$ meV, and different values of the L_{1c} density-of-state mass.

Γ_{1c}, L_{1c} band population. The Fermi energy is determined at helium temperature from the oscillations of the magnetoresistance (Fig. 2). The total concentration of free carriers is then obtained from

$$n_1 + n_2 = A(kT)^{3/2}\left[\left(\frac{m_1}{m_0}\right)^{3/2} \mathfrak{F}_{1/2}\left(\frac{E_F - E_c}{kT}\right) + \left(\frac{m_2}{m_0}\right)^{3/2} \mathfrak{F}_{1/2}\left(\frac{E_F - (E_c + \Delta E_c)}{kT}\right)\right], \quad (15)$$

where ΔE_c is the $L_{1c} - \Gamma_{1c}$ energy difference.

On Te-doped samples, between 4 °K and 300 °K, $n_1 + n_2$ is constant. Knowing the $\Delta E_c(T)$ variation we can then calculate E_F at each temperature and consequently n_1, n_2, and E_{gn}.

A. Low-temperature results

1. A and B samples

At helium temperature the relation (15) can be simplified by

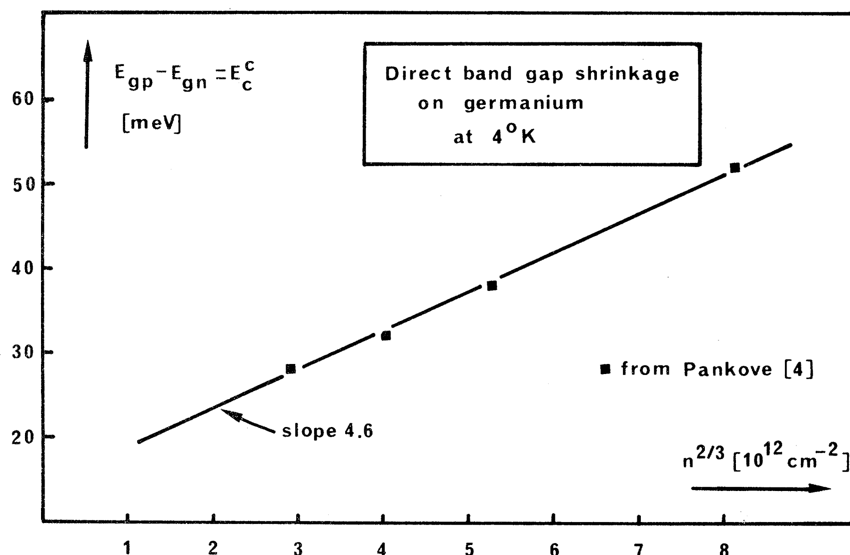

FIG. 6. Relative variation of the direct edge of As-doped Ge vs carrier concentration, at helium temperature (Ref. 4).

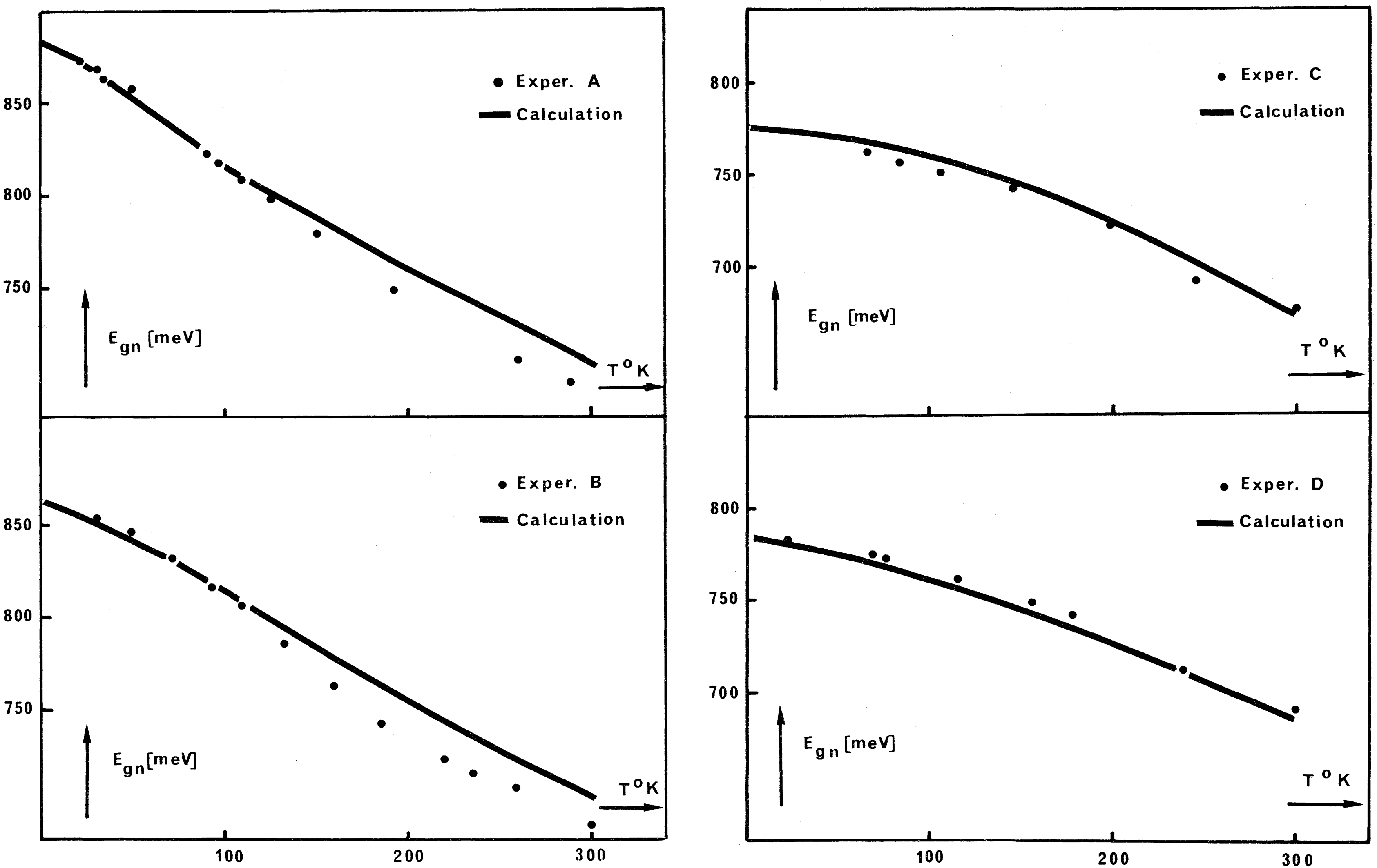

FIG. 7. Comparison of experimental with calculated E_{gn} values, using linear-screening and parabolic-density-of-state approximations.

$$n_1+n_2=B\left[\left(\frac{m_1}{m_0}\right)^{3/2}(E_F-E_c)^{3/2}+\left(\frac{m_2}{m_0}\right)^{3/2}[E_F-(E_c+\Delta E_c)]^{3/2}\right]. \quad (16)$$

In this case Eq. (2) takes a simple form and the $E_{gn}-E_{gp}$ values (respectively, 69 and 49 meV) determined at 4 °K on A and B samples can be used to obtain further information on the value of the L_{1c} density-of-state mass and of the $L_{1c}-\Gamma_{1c}$ gap. Inserting the experimental values of Fig. 4 in Eq. (2), (4), (13), and (16), we obtain two equations of the form

$$m_2^{2/3}(a-\Delta E_c)^{1/2}[(b-c(a-\Delta E_c)]=d, \quad (17)$$

which gives $m_2=0.22m_0$ and $\Delta E_c=85$ meV, in good agreement with recently published determinations.[18,19]

2. *C and D samples*

We noted that the energy variations obtained on these samples, always lower than E_{gp}, could be interpreted in terms of a transition initiating in the residual-acceptor level always present in natural GaSb samples.[20]

$$E_{gn}=E_{gp}+E_F-\Delta E_g-E_A, \quad (18)$$

where E_A is the ionization energy of the acceptor.

Using the helium-temperature E_{gn} (C and D) and E_{gp} values (Table 1) we obtain:

sample C: $E_A=36\pm5$ meV,

sample D: $E_A=30\pm5$ meV.

These values are in good agreement with the determination of the residual-acceptor ionzation energy determined on GaSb by Mathieu *et al.*[21] ($E_A=30\pm2$ meV).

B. Temperature evolution of E_{gn}

Having specified the ΔE_c (4 °K), m_2 and E_A values it is now possible using the $E_{gp}(T)$ variations, to calculate [Eqs. (2), (4), (13), and (18)] the$E$ $E_{gn}(T)$ variation obtained in the samples under consideration and to compare these values to the experimental results. The basis of the calculation is to determine, at each temperature, the carrier distribution in the Γ and L conduction-band minima. This permits to obtain the quantities E_F, E_c^c, E_c^e, and then E_{gn}.

For a given sample Eq. (15) gives

$$n=Ct=A\mathfrak{F}_{1/2}\frac{E_F-E_c}{kT}+B\mathfrak{F}_{1/2}\frac{E_F-E_c-\Delta E_c(T)}{kT}. \quad (19)$$

$\Delta E_c(T)$ is deduced from the helium-temperature value, assuming a linear temperature variation with a coefficient of -1.7×10^{-4} eV/°K.[22] It is then possible to obtain the value of E_F-E_c satisfying (19) and to deduce $E_{gn}(T)$. The theoretical values of E_{gn} we have calculated, using this method, are compared in Fig. 7 (a) and (b) to the experimental data.

The good agreement obtained confirms the model of linear screening and of parabolic density of states. At high temperature, the E_{gn} value is lightly overestimated in the A and B samples. This is due to the electron-electron interaction term whose validity conditions permit only to take account of the Γ_{1c} electron. This gives for E_c^e lower values than we could obtain in a more complete calculation. This shows that it is not the bandtail but the carrier-carrier and the carrier-impurity interactions which really reduces the band gap.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Professor J. L. Robert and D. Barjon for their assistance during the transport experiments. It is also a pleasure to thank Professor M. Averous for helpful comments.

[1] J. A. Rossi, D. L. Keune, N. Holonyak, P. D. Dapkus, and R. D. Burnham, J. Appl. Phys. 41, 312 (1970).
[2] A. G. Aleksanian, I. A. Poluektov, Yu.M. Popov, IEEE J. Quantum Electron. QE-10, 297 (1974).
[3] N. Holonyak, D. R. Schifres, H. M. Macksey, and R. D. Dupuis, J. Appl. Phys. 43, (1972).
[4] C. Haas, Phys. Rev. 125, 1965 (1962); J. I. Pankove and P. Aigrain, Phys. Rev. 126, 956 (1962).
[5] W. P. Dumke, M. R. Lorenz, and G. D. Pettit, Phys. Rev. B 1, 4668 (1970).
[6] W. M. Becker, A. K. Ramdas, and H. Y. Fan, J. Appl. Phys. Sup. 32, 2094 (1961).
[7] M. Cardona, K. L. Shaklee, and F. H. Pollak, Phys. Rev. 154, 696 (1967).
[8] P. A. Wolff, Phys. Rev. 126, 405 (1962).
[9] B. I. Halperin and M. Lax, Phys. Rev. 148, 722 (1966); R. H. Parmenter, Phys. Rev. 97, 587 (1955); 104, 22 (1956); E. O. Kane, *ibid.* 131, 79 (1963).
[10] C. J. Hwang, J. Appl. Phys. 41, 2668 (1970); Phys. Rev. B 2, 4117 (1970); 2, 4126 (1970).
[11] H. Mathieu, J. Phys. Chem. Solids 31, 67 (1970).
[12] D. Auvergne, J. Camassel, H. Mathieu, and A. Joullie, J. Phys. Chem. Solids 35, 133 (1974).
[13] B. Pistoulet, J. L. Robert, and D. Barjon, Solid State Commun. 8, 897 (1970).
[14] It can easily be shown (Ref. 12) that the line shape corresponding to the structures observed in Fig. 3 characterizes interband transitions occurring at the M_0 critical point whose energy (noted E_{gp} in Fig. 1) is directly determined by the maximum of the structure. In the degenerate case we consider that the only available final state for the transition is located at the

Fermi level. Depending on the temperature, two cases must be considered. At helium temperature the Fermi function can be approximated by a step function, and the energy of the maximum (E_{gn} in Fig. 1) gives the critical-point energy, increased by the Fermi energy E_F, corrected in order to satisfy the k-conserving selection rule. At greater temperatures it is no longer possible to approach the Fermi function by a step function. In all the related equations we then have to replace E_F by E_m, defined as the energy in the band at which the transition really occurs. E_m is determined from the maximum value of the product (occupation probability) $\times$ (state density). This gives $E_F = E_m + kT \ln (2 E_m/kT) - 1)$.

[15]W. P. Dumke, Phys. Rev. 132, 1998 (1961).

[16]R. B. Dingle, Philos. Mag. 831 (1955); V. Fistul, *Heavily Doped Semiconductors* (Plenum, New York, 1969).

[17]R. L. Aggarwal and B. Lax, Phys. Rev. Lett. 19, 236 (1967).

[18]J. L. Robert and D. Barjon, Phys. Status Solidi 3, 421 (1970).

[19]G. Bastide, B. Pistoulet, J. L. Robert, and C. Roustan, Solid State Commun. 11, 835 (1972).

[20]We have to point out that this residual acceptor which has been attributed to lattice defects (Ga vacancy and Ga in Sb site) has never been observed on heavily doped samples. This result is in relation with the fact that the density of acceptor states can be reduced by an antimony-excess increase [D. Effer and P. J. Etter, J. Phys. Chem. Solids 25, 451 (1964), see also W. Jakowetz, W. Ruhle, K. Breuminger, and M. Pilkuhm, Phys. Status Solidi A 12, 169 (1972)]. Taking into account the similarity between the tetrahedral radii of Sb and Te [J. C. Phillips, *Bonds and Bands in Semiconductors* (Academic, New York, 1973), p. 22] we can consider that in samples grown with a great excess of Te (which is the case of A and B samples), the substitution of Sb by Te is achieved at the expenses of the lattice defects. That is the reason why we have not considered in the A and B sample cases, transitions initiating at the acceptor states. Otherwise the energy correction introduced (of the order of the acceptor binding energy) should increase the disagreement between the theoretical and experimental curved [Fig. 7(a)] over the entire temperature range.

[21]H. Mathieu, J. Camassel, and D. Auvergne (unpublished).

[22]D. Auvergne, J. Camassel, H. Mathieu, and M. Cardona, Phys. Rev. B 9, 5168 (1974).