Title
WAVELENGTH MODULATION SPECTROSCOPY OF SEMICONDUCTORS

Author
Zucca, Ricardo R.L.

Publication Date
1971

UCRL-20361
c.2

WAVELENGTH MODULATION SPECTROSCOPY OF SEMICONDUCTORS

Ricardo R. L. Zucca
(Ph. D. Thesis)

January 1971

AEC Contract No. W-7405-eng-48

LAWRENCE RADIATION LABORATORY
UNIVERSITY of CALIFORNIA BERKELEY

UCRL-20361
c.2

DISCLAIMER

WAVELENGTH MODULATION SPECTROSCOPY OF SEMICONDUCTORS

Contents

WAVELENGTH MODULATION SPECTROSCOPY OF SEMICONDUCTORS

Ricardo R. L. Zucca

Inorganic Materials Research Division, Lawrence Radiation Laboratory
and Department of Physics, University of California
Berkeley, California

ABSTRACT

A sensitive wavelength modulation spectrometer with a spectral range from 1.75 to 6.0 eV is described. The spectrometer employs a two-beam system with two feedback loops in order to compensate the dispersion of the optical components. The principles of operation, construction details and operational techniques are described.

Wavelength modulation spectra of GaAs, GaSb, InAs, InSb, Ge and Si at 5, 80 and 300°K are presented. The results are compared with electroreflectance and thermoreflectance data. New structures are found in the spectra of all crystals.

With the help of existing band structures of these crystals, all the reflectivity peaks can be consistently assigned to proper critical transitions between the valence and the conduction bands. Values of spin-orbit splittings at several symmetry points can be calculated. Temperature effects on the band spectra are discussed.

I. INTRODUCTION

In recent years the subject of band structures of semiconductors has again attracted much attention. Although the gross features of the band structures of the most common materials were already well known, the availability of high speed computers stimulated attempts of more precise calculations. Calculations from first principles, i.e. starting from free ion potentials have been attempted with some success,[1] but due to the complexity of the problem, they must rely on simple models that hamper the accuracy greatly. The introduction of the empirical pseudopotential method[2] has led to a great improvement of the calculations with a better agreement with the experimental data. Further progress requires improvement of the resolution of experimental spectra.

Experimentally, some of the early information about the band structure in the vicinity of the gap came from cyclotron resonance measurements.[3] But the most powerful tool for obtaining information about the band structures of semiconductors is the measurement of optical transmissivity and reflectivity. The first gives information at energies below and about the band gap and the latter above the gap. A large amount of optical spectroscopical work has already been done[4] in this direction and was the basis for the early band structure calculations. Recently the application of optical derivative spectroscopy has greatly improved the resolution of the optical spectra, and has had a strong impact on the recent advances in band structure calculations.

Many different modulation schemes have been invented for derivative spectroscopy. For measurements of reflectivity spectra of solids, electroreflectance,[5,6] piezoreflectance,[7,8,9] thermoreflectance[10,11]

and wavelength modulation[12,13] methods have been most successful. In all these schemes, except the wavelength modulation method, modulation of the light beam is obtained through application of a direct ac perturbation on the solid, and hence, interpretation of the derivative spectrum depends very much on how the solid responds to the perturbation. Thus, in electroreflectance, we must know how the band structure of the solid changes with an applied electric field.[14] In piezoreflectance and thermoreflectance we must know the variation of the band structure as a function of pressure[15] and temperature[11] respectively. Unfortunately, our knowledge of these properties of solids is generally rather limited. Therefore, the fact that no perturbation on the solid is needed makes the wavelength modulation method most attractive. Since the wavelength modulation spectrum is simply the derivative of the normal spectrum, there is no ambiguity in the interpretation.

However, unlike the other modulation schemes, wavelength modulation requires careful construction of the experimental system in order to eliminate a huge background in the derivative spectrum. This background appears as a result of wavelength modulation on the spectra of various optical components in the system. In particular, because of the many strong spectral lines in the arc source, it is difficult to apply the scheme to the uv region. For this reason wavelength modulation has not been as popular as the other modulation schemes. Work done with wavelength modulation has usually been limited to a narrow region in the visible and near infrared.[12,13,16,17]

In this work we present a wavelength modulation spectrometer which practically eliminates all the background by employing a feedback system.

The sensitivity of the spectrometer is as large as $\Delta R/R \cong 10^{-4}$. This spectrometer has been used to obtain derivative spectra of Si, Ge, GaAs, GaSb, InAs and InSb from 1.75 to 6 eV. In order to reduce lifetime broadening and to resolve fine structures in the spectra, the measurements have been made at liquid nitrogen and liquid helium temperatures. The results of our measurements will be presented and discussed in the following sections. While the gross features of our spectra agree with the results of others,[6,11] new structures and more fine details generally appear in our spectra, particularly in the uv region.

In the following section a complete description of the experimental set-up is given. This section is divided into four parts: part A contains the principles of operation of the spectrometer; part B a description of the spectrometer and operation procedures; part C a description of the Dewar; and part D a description of the samples and their preparation. In Sec. III the wavelength modulation spectra of the six semiconductors at 5, 80 and 300°K are presented. In Sec. IV the theoretical situation is discussed. Starting from the basic assumptions of band structure calculations, the pseudopotential method is described and recently calculated band structures of the six semiconductors are presented and described. The exciton problem is also briefly introduced. In Sec. V the experimental data are compared with the theoretical calculations: the derivative spectra are analyzed and, with the help of the existing band structures, various reflectivity peaks are assigned to proper critical transitions between bands. Emphasis is on the new structures observed. Variation of the derivative spectra with temperature is also discussed and some comments on the contribution of hyperbolic excitons to the spectral structures are given.

II. EXPERIMENTAL DESCRIPTION

A. Principles of Operation of the Wavelength-Modulation Spectrometer.

Let $S(\lambda_o)$ be the signal response of a photomultiplier to a quasi-monochromatic light beam with a spectral distribution function $I(\lambda)g(\lambda-\lambda_o)$, where $g(\lambda-\lambda_o)$ is a normalized, symmetric slit function centered at the wavelength λ_o with an effective bandwidth W. Then, we have

$$S(\lambda_o) = \int_{-\infty}^{\infty} g(\lambda-\lambda_o)\, G(\lambda)\, I(\lambda)\, d\lambda \tag{1}$$

where $\int_{-\infty}^{\infty} g(\lambda-\lambda_o)\, d\lambda \equiv W$, and $G(\lambda)$ is the coefficient for converting the input light signal to the output electrical signal.

If now λ varies with time, such that[18]

$$\lambda = \lambda_o + A\cos\omega t \tag{2}$$

Let $T(\lambda) \equiv I(\lambda)\, G(\lambda)$. Then Eq. (1) becomes

$$S(\lambda_o,t) = \int_{-\infty}^{\infty} g(\lambda'-\lambda_o-A\cos\omega t)\, T(\lambda')d\lambda' \tag{3}$$

In order to understand the effect of the modulation better, let us expand $T(\lambda')$ in a power series about λ_o:

$$S(\lambda_o,t) = \int_{-\infty}^{\infty} g(\lambda'-\lambda_o-A\cos\omega t)\, [T(\lambda_o)+T'(\lambda_o)(\lambda'-\lambda_o) + \tfrac{1}{2}\, T''(\lambda_o)(\lambda'-\lambda_o)^2 + \cdots]d\lambda' \tag{4}$$

Such an expansion converges quickly if A and W are small. We now integrate term by term, taking advantage of the symmetry and normalization

of $g(\lambda)$ and, after some algebra (see details in the Appendix), we finally obtain:

$$S(\lambda_o,t) = [WT(\lambda_o) + \frac{1}{2}(\xi_2 W^2 + \frac{1}{2}A^2)WT''(\lambda_o) + \cdots] +$$

$$+ [WAT'(\lambda_o) + \frac{1}{2}WA(\xi_2 W^2 + \frac{1}{4}A^2)T'''(\lambda_o) + \cdots]\cos\omega t +$$

$$+ [\cdots]\cos 2\omega t + \cdots \qquad (5)$$

where ξ_2 is a constant of order one ($\xi_2 = 1/6$ for the ideal slit function of Fig. 1).

For our wavelength-modulation spectrometer we are only interested in the dc component and the component varying with $\cos\omega t$. If we retain only the leading term in each component, then we have:

$$S(\lambda_o,t) \cong S^{DC}(\lambda_o) + S^{AC}(\lambda_o)\cos\omega t \qquad (6)$$

where:

$$S^{DC}(\lambda_o) \cong WT(\lambda_o) \qquad (7)$$

$$S^{AC}(\lambda_o) \cong WAT'(\lambda_o) \qquad (8)$$

Equations (7) and (8) show that S^{DC} is proportional to the signal intensity per unit wavelength $T(\lambda_o)$ at the wavelength λ_o, while S^{AC} is proportional to the first derivative of the intensity $T(\lambda_o)$ with respect to the wavelength at λ_o. But we must keep in mind that this is an approximation valid only as long as the effective bandwidth W and the modulation amplitude A are kept smaller than the width of the narrowest spectral structures of $T(\lambda_o)$ (see the Appendix for details).

Up to this point we have seen how a wavelength-modulation spectrometer can measure $T(\lambda)$ and its derivative $T'(\lambda)$, but we notice that $T(\lambda)$ is proportional to the product of the spectra of all the optical components in the optical path from the light source to the photomultiplier. Usually the sample spectrum is weak compared with those of other optical components. In order to get a good sample spectrum we must eliminate the background spectrum due to other components. This is often done by a two-beam method. In our system the two beams have identical optical paths except for the fact that the "sample beam" undergoes an extra reflection from the sample.

At the output of the photomultiplier we detect the ac and the dc signals coming from the two beams, both being wavelength-modulated. Then, from the reference beam we have, following Eqs. (7) and (8),

$$S_R^{DC}(\lambda) = WT_o(\lambda) \tag{9}$$

$$S_R^{AC}(\lambda) = WA(dT_o/d\lambda) \tag{10}$$

where $T_o(\lambda)$ is signal intensity per unit wavelength from the reference beam at λ. The signal intensity per unit wavelength from the sample beam is $R(\lambda)T_o(\lambda)$, being $R(\lambda)$ the reflectivity of the sample. Therefore we have from the sample beam:

$$S_S^{DC}(\lambda) = WT_o(\lambda)R(\lambda) \tag{11}$$

$$S_S^{AC}(\lambda) = WAd(T_oR)/d\lambda = WAR(\lambda)(dT_o/d\lambda) + WAT_o(\lambda)(dR/d\lambda) \tag{12}$$

and, replacing Eqs. (9) and (10),

$$S_S^{DC}(\lambda) = S_R^{DC}(\lambda)\ R(\lambda) \tag{13}$$

$$S_S^{AC}(\lambda) = S_R^{AC}(\lambda)\ R(\lambda) + AS_R^{DC}(\lambda)(dR/d\lambda) \tag{14}$$

For transmission measurements R should be replaced by the transmissivity T.

By means of two feedback loops (described with details in Secs. II.B.1 and II.B.2) it is possible to simultaneously make

$$S_R^{DC}(\lambda) = V_c \equiv \text{constant} \tag{15}$$

and

$$S_R^{AC}(\lambda) = 0 \tag{16}$$

Equations (13) and (14) then yield:

$$S_S^{DC}(\lambda) = V_c\ R(\lambda) \tag{17}$$

$$S_S^{AC}(\lambda) = V_c\ A\ dR/d\lambda \tag{18}$$

The system therefore measures $R(\lambda)$ and $dR/d\lambda$ simultaneously.

With another choice of feedback loops we can make

$$S_S^{DC}(\lambda) = V_c \equiv \text{constant} \tag{19}$$

$$S_R^{AC}(\lambda) = 0 \tag{20}$$

instead of (15) and (16). We then find, from Eqs. (13) and (14),

$$S_R^{DC}(\lambda) = V_c/R(\lambda) \tag{21}$$

$$S_S^{AC}(\lambda) = V_c\ A\ dR/Rd\lambda \tag{22}$$

The system measures $1/R$ and $dR/Rd\lambda = d(\ell nR)/d\lambda$ simultaneously.

As a final remark for this section, it is convenient to discuss why we prefer to measure $dR/d\lambda$ (or $dR/Rd\lambda$) instead of simply improving conventional measurements of R. The reason is that while direct measurements of R are quite sufficient for obtaining general shapes of spectra, small fine structures hidden in broad peaks are often of great importance in the optical spectroscopy of solids, as it is true in most fields of optical spectroscopy. These small structures can be most easily recognized by the derivative spectroscopy, a fact well known in magnetic resonance spectroscopy. An example is given in Fig. 2 where the curves resulting from the superposition of two Gaussian peaks at three different separations are compared with their corresponding derivatives. We notice that when the small peak starts hiding behind the other, the derivative still shows the splitting very clearly.

The above example indicates that derivative spectroscopy has better resolution than conventional spectroscopy. It is also interesting to compare sensitivities. We may do this by comparing two ideal experiments, a standard reflectivity measurement and a wavelength modulated reflectivity measurement, both performed with the same basic equipment (light source, photomultiplier, spectrometer, etc.), both employing two-beam systems, and differing only in the use of modulation. We assume both techniques having their best signal-to-noise ratio, which is reached when all the noise introduced by the measuring electronics is negligible with respect to the photomultiplier noise. Under such conditions the amount of noise is the same in both experiments. The signal-to-noise ratio for the wavelength modulation experiment is

$$(S/N)_{WM} \sim \left[\frac{\overline{(dR/d\lambda)^2}\ A^2}{\overline{N^2}}\right]^{1/2} \tag{23}$$

where A is the modulation amplitude and N is the noise.

In order to compare the standard reflectivity measurement with the wavelength modulation measurement we differentiate the reflectivity spectrum; the noise, superimposed on the reflectivity spectrum, is also differentiated. Therefore the S/N ratio is:

$$(S/N)_R \sim \left[\frac{\overline{dR/d\lambda)^2}}{\overline{(dN/d\lambda)^2}}\right]^{1/2} \tag{24}$$

For a comparison of the two S/N ratios we can reasonably estimate $(\overline{N^2})^{1/2} \sim \left[\overline{(dN/d\lambda)^2}\right]^{1/2} \Delta\lambda/4$, $\Delta\lambda$ being the average period of the noise variation. We obtain, after replacing $\overline{(dN/d\lambda)^2}$ in Eq. (24),

$$(S/N)_R \sim \left[\frac{\overline{(dR/d\lambda)^2}\ (\Delta\lambda/4)^2}{\overline{N^2}}\right]^{1/2} \tag{25}$$

If we compare Eqs. (23) and (25), we notice that, as long as we choose $A > \Delta\lambda/4$, we get $(S/N)_{WM} > (S/N)_R$. which means more sensitivity for the wavelength modulation experiment.

Unfortunately we cannot increase the modulation amplitude A at will if we want to avoid distortion of the spectra, because A [see Eq. (12') of the Appendix] cannot be chosen larger than the width of the narrowest spectral peak of the spectrum. Such a width can also be taken as an estimated value of the average noise period, because the electronic system cuts off all wavelengths smaller than the width of the narrowest

spectral peak, for maximum noise elimination. Therefore the restriction on A for distortionless spectra can be written as $A < \Delta\lambda$, and we are limited to play with A in the interval $\Delta\lambda/4 < A < \Delta\lambda$. Consequently $(S/N)_{WM}$ cannot be improved above $4(S/N)_R$, if we are not willing to accept some distortion of the spectra.

There are also practical considerations to bear in mind when discussing the advantages of wavelength modulation. Among them an important one is that it is easy to reach high sensitivity with wavelength modulation, e.g. $\Delta R/R \sim 10^{-4}$ in our spectrometer, because we are handling differential quantities. Reaching the same sensitivities with standard reflectivity is difficult when there is a large background, as in the spectra of solids; matching the sensitivity of our spectrometer, for example, would require a 4-digit accuracy from a conventional spectrometer. These facts, coupled to the better resolution and the improvement of sensitivity strongly support of our choice of wavelength modulation.

B. The Actual Experimental Layout

1. The Optical System

The block diagram of the optical system is shown in Fig. 3. The scheme is typical of a two-beam system. Different parts of the system are described separately in the following.

a. The Lamp. Above 3500 Å, a quartz-iodine filament lamp is used. In the uv, between 2000 and 4500 Å, a 75-Watt xenon arc lamp is employed.

b. The Spectrometer. A 1/2 meter Jarrel-Ash monochromator is used. The slits are adjustable with a maximum opening of 3 mm. The exit slit width is chosen as a compromise between spectral resolution and output

light intensity. The entrance slit is kept at its maximum opening. However an external slit S (Fig. 4) with narrow slitwidth is inserted between the lamp and the entrance slit E, at the image position of the entrance slit E. It is this external slit which does the real function of the entrance slit.

c. The Modulation System. Several devices have been proposed to produce wavelength-modulation: a quartz disc that vibrates or rotates in the optical path inside the spectrometer,[19,20,21] a mirror that oscillates inside the spectrometer,[22] oscillation of the exit slit of the spectrometer[12] or oscillation of an external slit after the spectrometer.[23] They are all based on the same principle that consists of perturbing the light path inside the spectrometer and they differ only in the technical scheme to achieve this purpose. The choice between systems depends only on practical considerations.

We designed our own system, slightly different from the others, such that no modification of a commercial spectrometer is necessary and that the light beam passes through the modulation device before entering the spectrometer. Wavelength-modulation is achieved through the vibration of a mirror M between the entrance slit E and the external slit S (see Fig. 4). The mirror oscillates about a vertical axis, driven by a power oscillator through a piezoelectric bimorphous solid, with a frequency adjustable in a small range about 100 Hz. The image of the external slit S is focused on the plane of the entrance slit E by the concave mirror C. When M oscillates, the position x of this image also oscillates on the plane of E with

$$x = a \cos\omega t \qquad (26)$$

where a is the geometrical amplitude of oscillation; a can be selected by varying the amplitude of the electronic signal that drives the mirror M, but cannot be larger than one half of the maximum opening of the entrance slit (3 mm.).

The effect of this oscillation is the same as if an entrance slit with the size of the image of S were oscillating. The nearly monochromatic light beam which exits from the spectrometer has a modulated wavelength

$$\lambda = \lambda_o + A \cos\omega t \tag{27}$$

Between x and $(\lambda-\lambda_o)$ there is a one to one correspondence given by

$$\lambda-\lambda_o = (d\lambda_o/dx)x \tag{28}$$

where $d\lambda_o/dx$ is the reciprocal linear dispersion of the spectrometer at the entrance slit. Therefore the modulation amplitude A is connected with the geometrical amplitude a by

$$A = (d\lambda_o/dx)a \tag{29}$$

Here, $d\lambda_o/dx$ varies with λ_o. For our spectrometer we derive from the geometry of the instrument:

$$d\lambda_o/dx = 16.9 \left[(1-\lambda_o/16900)^{1/2} + \lambda_o \times 0.083/16900\right] \quad \text{Å/mm} \tag{30}$$

where λ_o is in Angstroms. From 2000 to 7000 Å $d\lambda_o/dx$ varies by 6%. Since a is kept constant during a wavelength scanning, and since S_S^{AC} is proportional to $dR/d\lambda$ (Eq. (18)), or to $dR/R d\lambda$ (Eq. (22)), through the proportionality "constant" A, the output S_S^{AC} must be divided by $d\lambda/dx$

in order to compensate the variation of A with λ. This correction is done numerically when the output is processed (see Sec. II.B.3).

d. The Feedback Diaphragm. A diaphragm is used as a feedback control to make $S_R^{AC} = 0$ as required by Eqs. (16) and (20) for background cancellation. This diaphragm, called D in Fig. 4, is located at a point in the optical path in front of the entrance slit. The opening of the diaphragm is just as large as necessary in order that the diaphragm edges don't touch the oscillating beam when the diaphragm is at a centered position as in Fig. 5a. However, when the diaphragm is shifted from such a centered position as in Fig. 5b for example, part of the beam intensity is cut off by the diaphragm edge. This cutoff is made gradual by giving the diaphragm edges a sawtooth shape and also by having the diaphragm at a point of the optical path where the beam is rather broad.

In order to describe the effect of the diaphragm on the electronic signals, we want to plot the total signal $S_R(\lambda_o,t)$ of the reference beam, a function of time given by Eq. (6), as a function of x, the geometrical position of the oscillating beam. We obtain by inserting Eq. (26) in Eq. (6),

$$S_R(\lambda_o,x) = S_R^{DC}(\lambda_o) + \frac{x}{a} S_R^{AC}(\lambda_o) \tag{31}$$

If the diaphragm is centered, the function $S_R(\lambda_o,x)$ is given by Fig. 6a. If the diaphragm is shifted from the centered position, as in Fig. 5b, $S_R(\lambda_o,x)$ changes to Fig. 6b. As a result of the diaphragm cutting off the light, $S_R(\lambda_o,x)$ decreases more and more as $x \rightarrow a$. Correspondingly, S_R^{AC} also decreases. If the diaphragm is shifted farther, S_R^{AC} becomes

smaller and smaller until we find a position such that $S_R^{AC} = 0$ as shown in Fig. 6c. This is the desired position for background cancellation. It would be still better if we could get $S_R(\lambda_o,x)$ to be perfectly flat like the dashed line of Fig. 6c, but such a perfection could be achieved only with a very careful and difficult design of the edges of the diaphragm and the distribution of light intensity within the beam. This is not necessary because we are only interested in S_R^{AC}, the amplitude of the Fourier component of $S_R(\lambda_o,t)$ that goes like $\cos\omega t$. Lack of linearity of $S_R(\lambda_o,x)$ only introduces higher harmonics of ω in the ac signals that are completely filtered out by the lock-in detectors of the electronic system (see next section).

Finally we want an automatic operation, i.e., the diaphragm must adjust its position automatically to make $S_R^{AC} = 0$. This is easily achieved by negative feedback using S_R^{AC} as an error signal to drive a dc motor which positions the diaphragm. Therefore, whenever S_R^{AC} deviates from zero, the diaphragm moves until S_R^{AC} is brought back to zero.

e. The Beam Chopper. The beam coming out of the spectrometer is split into two by a rotating wheel, which has holes on the 1st and 3rd quadrants. The light beam is then alternatively passing through or reflecting from the wheel. The switching frequency is chosen to be 5 Hz , much smaller than the modulation frequency, in order not to interfere with the detection of the ac signal.

f. The Detector. An EMI 9558Q photomultiplier with a quartz window and an S-20 spectral response is used.

g. The Overall Optical Setup. Aluminum mirrors coated with MgF_2 are used to direct the beams in order to ensure a good reflectivity in the ultraviolet.[24] Quartz lenses have good transmissivity in the uv, but their focal lengths vary appreciably with the wavelength.

As described in Sec. II.A, the sample and the reference beams should be matched such that when the sample reflectivity R is 1, the two beams should yield the same intensity at the photomultiplier. We must therefore do our best to make the two beam paths equivalent. For this reason the equivalent mirrors in the two paths were coated together. When the sample is in the optical dewar, a set of quartz plates is inserted in the reference beam to match the windows of the dewar. Matching is also the reason for our choice of a design that employs the same photomultiplier to detect the reference and the sample beam. Two separate photomultipliers would ease both the optical and electronic design,[18,22] but at the expense of an imperfect matching of the spectral responses.

We show in Fig. 7 the arrangement of the two-beam system in detail. It looks rather complicated because the beams cross several times, but it has several merits: the number of mirrors is minimum, a convenient condition because each reflection implies a loss of light intensity; the angles of incidence of the light beams on the spherical mirrors are rather small (less than 25°), keeping the image distortion low; the angle of incidence on the sample is only 4.5°, a good approximation to normal incidence;[25] finally the angles of incidence of the light beams on the photomultiplier are only 6° simplifying the problem of matching the response for the two beams because at oblique incidence, the response

would depend on the angles very critically. With this arrangement we get a good matching of the sample and reference beams if the optical alignment is done carefully. Certainly we could have found some other arrangement with a simpler geometry of the optical paths, but only at the expense of large angles of incidence on the mirrors (~45°) if the number of mirrors were not to be increased; thus specially designed toroidal mirrors would be necessary, a choice not convenient for a first design, when quick changes for testing are useful.

We tested a scattering chamber in front of the photomultiplier[23] in order to make matching easier by averaging small differences of incidence angles and beam sizes, but the advantages did not justify the loss of light intensity due to backscattering from the chamber. Therefore we chose to focus both light beams directly on the photomultiplier cathode, taking care to get spots of the same size at the same place.

For tests of beam matching we would like to have a good reflectivity standard but, unfortunately, such a material does not exist; nevertheless we can do satisfactory tests by observing the spectrum of an aluminum coated mirror, whose lack of fine structure in the spectral range of our interest is well known.[26] This test is also used to check that there is no structure in the spectra due to condensation on the windows of the dewar after several hours of operation at low temperature.

2. The Electronic System

The block diagram of the electronic system is shown in Fig. 8. The modulated sample and reference beams hit the photomultiplier alternatively. Since the intensities of the two beams are usually different, the output from the photomultiplier appears as a 5 Hz square wave, with ~100 Hz ac modulation superimposed on it (see Fig. 9a). The frequency of the square wave is equal to the frequency of the optical switching (5 Hz).

The photomultiplier output is fed into a high input impedance preamplifier, which shifts the zero level through the addition of a constant voltage V_c and amplifies the shifted input with a constant gain γ. The preamplifier output is shown in Fig. 9b.

After the preamplifier an electronic switch unit, operated synchronously with the optical switching, separates the signals from the sample and the reference beams into two channels. Capacitors are connected at the output terminals of the electronic switch in order to hold each output at a constant level, just before the output terminal is switched off. The two outputs are shown in Figs. 9c and 9d.

After the switching unit each channel has a lock-in detector to measure the modulated ac components S_R^{AC} and S_S^{AC}. The detectors actually measure $\alpha_R \gamma S_R^{AC}$ and $\alpha_S \gamma S_S^{AC}$ where α_R and α_S represent the gains of the lock-in detectors. The lock-in detectors are synchronized with the modulator.

The output from the reference channel $\alpha_R \gamma S_R^{AC}$, amplified by a power amplifier, is used to drive the diaphragm so as to keep $\alpha_R \gamma S_R^{AC} = 0$ and hence $S_R^{AC} = 0$. As described earlier in Sec. II.B.1, this is required for background cancellation (see also Eqs. (16) and (20). A

recorder is used to monitor the signal $\alpha_R \gamma S_R^{AC}$ in order to assure the correct function of this feedback loop.

The output $\alpha_S \gamma S_S^{AC}(\lambda)$ of the sample channel is recorded on the chart recorder. This output is, according to Eq. (18), under the condition of Eq. (15),

$$\text{ac output} \equiv \alpha_S \gamma S_S^{AC}(\lambda) = \alpha_S \gamma \, A \, V_c \, dR/d\lambda = \text{const.} \times dR/d\lambda \qquad (32)$$

With the condition of Eq. (19), the output is (from Eq. (22)),

$$\text{ac output} \equiv \alpha_S \gamma S_S^{AC}(\lambda) = \alpha_S \gamma \, A \, V_c \, dR/Rd\lambda = \text{const.} \times dR/Rd\lambda \qquad (33)$$

In parallel with the lock-in detector, each channel has a dc amplifier and filter to measure the dc components. The outputs are $\gamma\beta_R(S_R^{DC} - V_c)$ and $\gamma\beta_S(S_S^{DC} - V_c)$ where β_R and β_S are the gains of the dc amplifiers. In one mode of operation we would like to keep $S_R^{DC} = V_c =$ constant (Eq. (15)). This is done by using the output $\gamma\beta_R(S_R^{DC} - V_c)$ as an error signal in a negative feedback to control the gain of the photomultiplier. This keeps $\gamma\beta_R(S_R^{DC} - V_c)$ at zero, and hence $S_R^{DC} = V_c$. Correspondingly, the other dc output is:

$$\text{dc output} \equiv \beta_S \gamma(S_S^{DC} - V_c) = \gamma\beta_S(V_c R - V_c) = \gamma\beta_S V_c(R-1) \qquad (34)$$

[(see Eq. (17)]. This output goes to a recorder. The output levels corresponding to R = 0 and R = 1 can easily be calibrated (see Sec. II.B.3).

For the other mode of operation we would like to keep S_S^{DC} constant [(Eq. (19)]. Then the output $\gamma\beta_S(S_S^{DC} - V_c)$ should be used instead of $\gamma\beta_R(S_R^{DC} - V_c)$ as the error signal for negative feedback. Consequently the other dc output is:

$$\text{dc output} \equiv \gamma\beta_R(S_R^{DC} - V_c) = \gamma\beta_R(V_c/R - V_c) = \gamma\beta_R\, V_c(1/R-1) \qquad (35)$$

We notice in Eqs. (32) to (35) that all the output signals contain the voltage shift V_c at the preamplifier as a common factor. Therefore a choice of V_c implies a choice of the gain of the whole system. This overall gain is no longer determined by the photomultiplier gain, as it would be in a system without feedback.

In what follows we give a brief description of each electronic component.

a. The Oscillator. The frequency is adjustable between 90 and 100 Hz, with amplitude and frequency stability better than 0.1%. In order to drive the piezoelectric bimorph that moves the mirror to modulate the light beam, the oscillator output is amplified by a commercial audio amplifier.

b. The Photomultiplier and HV Power Supply. The EMI 9558Q photomultiplier tube has the conventional voltage-dividing network for dc measurements with its cathode at high negative voltage.

The high voltage power supply is a commercial unit with a modification such that the output voltage can be controlled by an external feedback. This is done by applying the external error signal to the internal feedback loop for voltage regulation in the power supply.

c. The Preamplifier. This is an operational amplifier with high input and low output impedances. The constant voltage V_c is subtracted from the negative photomultiplier signal (Fig. 8) by applying a negative voltage to the negative imput of the operational amplifier. The constant voltage is taken from a mercury battery through an adjustable voltage divider. The preamplifier is linear (10KHz. bandwidth) and the gain γ

can be adjusted between 1 and 100 by steps of 10.

d. The Electronic Switching Unit. This unit consists of two mercury relays (A and B of Fig. 10) and their control circuits. Fig. 10 shows that relay B performs the switching between sample and reference channel, while relay A disconnects both channels whenever it moves to position 2. The timing of the relays is shown in Fig. 9e. If we compare this diagram with the signals before and after the switching unit, shown in Figs. 9a, 9b and 9c, we realize that relay A determines a short blanking interval encompassing the actual switching between channels. This blanking is necessary because the optical switching of the light beams by the beam chopper is slower than the electronic switching. The blanking interval given by relay A assures that each channel is connected only when the proper signal with its steady value is fed in.

The switching system is triggered by an adjustable lamp-photocell unit mounted in front of the beam chopper to monitor the passage of the edges of the mirrors. The photocell triggers relay A, beginning the blanking interval that has an adjustable length. A delay circuit triggers relay B after relay A, the actual amount of delay being immaterial, as long as it is short, since both channels are disconnected during the blanking interval. The beginning and the length of the blanking period are chosen such that it is as short as possible in order not to waste signal intensity.

The capacitor at the output of each channel (Fig. 10) holds the output signal constant when the channel is disconnected because it can discharge only through the high input resistance of the following dc amplifier (see Fig. 8). When the relays connect the channel, the time

constant is lowered by the low output resistance of the preamplifier and the capacitor does not affect the signal.

e. The dc Amplifiers. They are built with operational amplifiers working as active low pass filters. Gains are adjustable. The dc amplifier whose output is used as an error signal to control the photomultiplier voltage is usually operated at a gain of 200.

f. Other Electronic Components. The lock-in detectors are standard units[27] with a conventional design; they allow for a wide choice of gains (1 to 1000) and time constants (0.01 to 100 sec). The recorders are commercial units built by Hewlett-Packard (680M models).

3. Operation Procedures

The use of feedback controls in our wavelength-modulation spectrometer makes the operation automatic, requiring only a simple surveillance while scanning over broad wavelength ranges. Therefore there is nothing important to say about operational techniques during measurements. On the other hand the criteria for choosing some parameters of the system such as slitwidths, modulation frequency, etc. deserve some consideration.

We start by discussing the choice of modulation amplitude A and slitwidth. For this purpose the effective bandwidth W defined in Sec. II-A can be assumed equal to the slitwidth measured in Angstroms. We notice in Eq. (8) that the ac signals are directly proportional to W and A; therefore both must be chosen as large as possible in order to maximize the signal, and hence the signal to noise ratio. However W and A cannot be increased at will; they must be kept smaller than the width of the narrowest peak we want to observe without distortion

(see Appendix). Since such an upper bound is the same for W and A [Eqs. (11') and (12')] and the ac signals are directly proportional to both of them, the best choice is to keep W and A approximately equal and as large as possible without distortion. A convenient procedure is to start with a tentative choice, rather large, and decrease W and A until a further decrease does not modify the lineshapes of the spectra.

Another restriction on W and A is determined by the maximum opening of the spectrometer entrance slit E of Fig. 4. In order that the oscillating image of S never touches the edges of E, (2A+W) must be smaller than the opening of E. Since this opening is 51 Å in our spectrometer, W and A, if chosen approximately equal, cannot exceed 17 Å. We did not reach this limit in our measurements; however, if necessary, the limit can be easily raised using a diffraction grating with less dispersion. This may be convenient for spectra with very broad peaks.

Another important operational choice is the modulation frequency ω. Although at least 1 KHz would be desirable in order to reduce the 1/f electronic noise, the low mechanical resonance frequency of the oscillating mirror forces us to choose ω at about 100 Hz. This frequency is below resonance of the mirror and ensures a good amplitude stability of the oscillations. An important precaution is that ω must not coincide with a multiple of the beam switching frequency (5 Hz) because switching introduces some noise at the switching frequency and its harmonics. Our choice is 97.5 Hz. The 2.5 Hz difference from the next multiples of 5 Hz is larger than the bandwidth of the phase sensitive detector in a typical operation and the switching noise cannot interfere with the signal.

The choice of time constant of the lock-in detector which detects S_S^{AC} (sample channel in Fig. 8) depends on the scanning rate of the spectrometer and must be kept as large as possible without distorting the spectrum. Data of Sec. III were obtained with a time constant of 10 sec at a scan rate of 50 Å/min. For the other lock-in detector, the one that detects S_R^{AC} and whose output controls the diaphragm for background cancellation, the same choice would be convenient. However, since the motor that positions the diaphragm has a fast response and the gain of the feedback loop is very high, a long time constant would make the system oscillate. Therefore the time constant is kept low (0.1 sec for the spectra of Sec. III) at the expense of some jittering motion of the diaphragm due to noise. This jittering of the diaphragm adds extra noise to the signal S_S^{AC} which is nevertheless eliminated effectively by the other lock-in detector.

Here we conclude our discussion of choice of operational parameters. We only touched those parameters that deserve special considerations associated with the design of the wavelength-modulation spectrometer, omitting those parameters that are chosen following standard rules, i.e., electronic gains, scan rates, etc.

We now describe how we process the raw recorded output of our wavelength-modulation spectrometer in order to have the spectra the way they are presented in Sec. III, suitable for a theoretical discussion. The first step is to determine the scales of the recorded spectra. The dc output that is proportional to (R-1) in one mode of operation [Eq. (34)] and to (1/R-1) in the other mode [Eq. (35)] can be calibrated to read directly R (or 1/R) on the recorder chart: we first block the

light of the sample beam (or reference beam) and the recorder for the DC signal indicates R=0 (or 1/R=0). We then set the recorder input signal at zero, this corresponds to R=1 (or at 1/R=1 in the other mode of operation). The first operation is rigorous because blocking the sample (or reference) light beam is strictly equivalent to make R=0 (or 1/R=0). The second operation, however, is done electronically and is therefore sensitive to errors due to loss of light intensity, such as light scattering by imperfections of the sample surface. Therefore many careful tests would be necessary if we were interested in good absolute measurements of R. Since this is not our goal we prefer to regard our R values as given in arbitrary scales, although we notice that our reflectivity data agree within 5% other reported reflectivity measurements.[28]

The calibration of the ac output [Eqs. (32) and (33)] in units of $dR/d\lambda$ (or $dR/Rd\lambda$) is indirect. It can be done only by comparing the integral of $dR/d\lambda$ (or $dR/Rd\lambda$) with R for some sample whose absolute reflectivity R is previously measured accurately. Again, since our main concern is with detection of small spectral structure and only relative comparison between spectra, we choose to work with an arbitrary scale. Nevertheless we scaled the spectra presented in Sec. III by roughly comparing R and $dR/d\lambda$ for GaAs, such that the ac scales, although arbitrary, reflect the approximate magnitude of the actual scales. Ratios between scales for different measurements are accurate because we took account of changes of electronic gains and modulation amplitude. Scale ratios between different samples are accurate within 5% estimated above for R scales, for the same reasons.

The ac spectra, either $dR/d\lambda$ or $dR/Rd\lambda$, as recorded, need the small correction for changes of spectrometer dispersion, indicated in Sec. II.B.1. The ac output is divided by $d\lambda/dx$ given by Eq. (30). We actually normalize Eq. (30) such that $d\lambda/dx$ is equal to one at the center of the wavelength interval covered by the spectra.

For a theoretical discussion of spectra it is also necessary to convert the wavelength scales of the recorded spectra into energy scales. The relationship between wavelength λ and energy E is

$$\lambda = C/E \tag{36}$$

where C=12398.05 eV Å.

For R or 1/R spectra only the wavelength must be converted to energy scales. For ac spectra, we also want to convert $dR/d\lambda$ (or $dR/Rd\lambda$) into dR/dE (or dR/RdE). Since

$$d/dE = (d\lambda/dE)\, d/d\lambda = -(\lambda^2/C)\, d/d\lambda \tag{37}$$

we multiply the ac spectra by a factor proportional to $-\lambda^2$. Such a multiplication makes the structures at low wavelength (high energy) much less pronounced. Therefore, when we analyze the spectra presented in Sec. III, we must remember that small wiggles in the uv region were large and clear structures, very well distinguished from noise, in the original spectra.

In order to perform all the above operations, we punch the coordinates of the recorded spectra on computer cards, employing a facility currently used for punching coordinates of bubble chamber tracks. A simple computer program performs the dispersion correction

and energy scales conversion, plus a test of integration of the derivative spectra. The output of the computer program is plotted directly by an on-line plotting facility of the computer.

We find it convenient to conclude this section presenting some tests of the wavelength-modulation spectrometer. Here we omit the description of innumerable tests done while debugging the system and we limit ourselves to give an example to show that the system is operating properly.

The essential test is the verification of background cancellation and this test becomes particularly crucial at the wavelength of the xenon spectral lines when the xenon arc lamp is used. If the ac feedback loop were not working properly, the xenon lines would introduce a background two orders of magnitude larger than the signals usually observed. Figure 11a shows our derivative spectrum of GaAs appearing on the recorder chart. The arrows indicate the positions of the xenon lines. It is seen that small residual structure of the xenon lines is still present, but it is within the tolerable limit.

We could ask ourselves whether a slower scanning rate plus a larger integration constant of the lock-in detector, or some improvement of the feedback scheme, would make cancellation of xenon lines still better. The residue we see in the GaAs spectrum of Fig. 11a are due to harmonic distortion because the xenon lines are narrower than the modulation amplitude. In fact let us recall that, in a lock-in detector, when harmonics of the modulation frequency are superimposed on the signal, only even harmonics of the modulation frequency are totally rejected by the synchronous detection scheme.[29] Odd harmonics are only

effectively rejected by the tuned amplifier stage of the lock-in detectors. The rejection, is quite satisfactory for normal operation in our lock-in detectors, only less than 0.1% of third harmonics is not rejected). However, when the oscillating wavelength sweeps through a narrow xenon line, distortion of the ac signals from a sinusoidal shape becomes gigantic, and third harmonics become so huge that the unrejected amount ceases to be negligible. An addition of third harmonics to the normal signal explains the distortion of the output spectrum at the frequency of a xenon line. A way to reduce the incidence of such effect farther would be to reduce the modulation amplitude below the xenon lines width, but this would decrease the signal to noise ratio to an extent untolerable for the kind of broad spectra we are measuring. We prefer to tolerate the small residual wiggles left in Fig. 11a.

Another test of the system is to see whether ac and dc outputs are consistent. Figures 11a and 11b give part of the spectra $dR/d\lambda$ and R, recorded simultaneously, of GaAs at liquid He temperature as they appear on the recorder charts. Numerical integration of Fig. 11a (after dispersion correction) yields the spectrum of R in Fig. 11c. Comparison of Fig. 11b and Fig. 11c shows good agreement between the two. We have also tried to differentiate R of Fig. 11a and then compare the result with $dR/d\lambda$ of Fig. 11b, but the calculated derivative was too noisy even when R was measured with a digital recorder to increase sensitivity. Good agreement could be reached only after smoothing the noise by a least square fitting at each point of R before differentiating, but the choice of how much smoothing was necessary was possible only because $dR/d\lambda$ was already known. This proves the

superiority of our wavelength-modulation spectrometer. Incidentally the advantage of derivative spectroscopy is clearly manifested in Fig. 11 since it would be difficult to recognize the small structure in R without the help of the derivative spectrum $dR/d\lambda$.

C. The Dewar.

The optical dewar is a commercial unit built by Janis Research Corporation. It is a variable temperature dewar: liquid helium stored in a separate reservoir flows into a sample chamber where it is evaporated by a heater. Both the rate of helium flow and the heater current can be varied to adjust the temperature of the sample, helped by an additional heater in the sample holder. The sample temperature can be varied from liquid helium to room temperature arbitrarily. Also samples can be changed by removing the sample holder from the sample chamber while the helium reservoir remains undisturbed.

Temperature is measured with a copper-constantan thermocouple soldered to the sample holder, with its reference junction immersed in liquid nitrogen. We measure the thermocouple e.m.f. with a Model K5 Leeds and Northrup potentiometer. Although a thermocouple is not very sensitive at low temperature, we chose it because it covers the temperature range from liquid He to room temperature and because the thermocouple does not need a previous calibration. We just checked available calibration curves[30] at liquid helium, liquid nitrogen and room temperature. With a carefully constructed thermocouple, temperature measurements can be made with a $\pm 1^{\circ}K$ accuracy. With frequent readings of the temperature and manual adjustments of the heater current while surveying the spectrometer operation, we can easily keep the sample

temperature stable within ±1°K, while scanning a spectrum for about one hour or longer if necessary.

The dewar optical windows are made of Suprasil quartz and keeping them clean is extremely important, because very thin layers of oil from the vacuum pumps, that have no effect in the visible part of the spectrum, produce strong absorption and structure in the uv (2000 to 3000 Å). In order to solve this problem, we use liquid nitrogen traps with the oil diffusion pumps that evacuate the vacuum jacket and the sample chamber of the dewar. Even so, when the dewar is kept in continuous operation for several days, some condensation takes place. An additional heater, very close to the windows, proves very successful in evaporating this film when necessary. While recording the spectra presented in Sec. III, we monitored the spectrum of an aluminum coated mirror very frequently (Sec. II.B.1) in order to be sure that the spectra were free of structure due to oil condensation. The mirror is mounted on the back of the sample holder, so designed that a rotation of 180°, controlled from the outside of the dewar without disturbing the temperature, puts the mirror in the place of the sample.

D. The Samples.

All the samples are wafers of single crystals, commercially available. We chose all of them with (1,1,1) orientation, although this orientation is immaterial for reflectivity measurements, because all our samples have cubic symmetry.[31]

The samples are of either n- or p- type, with a carrier concentration in the range between 10^{13} and 10^{17} cm^{-3}. Values for each sample are given in Table I. We would expect wavelength-modulation

spectra to be independent of carrier concentration to the same extent as reflectivity, because, unlike other techniques such as electroreflectance,[6] the wavelength-modulation scheme has nothing to do with carriers. This is supported by our measurements on Si samples with a carrier concentration of 10^{17} cm^{-3} that give wavelength-modulated spectra identical to those of samples with 10^{13} carrier per cubic centimeter.

Surface preparation is very important, although the penetration depth of light is of the order of several thousands Angstroms[32] for our samples in the wavelength range of our interest. In order to have good reproducibility, it is important to have a surface free of mechanical distortion and chemical contamination. We followed the standard procedure for reflectivity measurements, namely mechanical polishing with finer and finer abrasives ending with 0.01 micron levigated alumina, followed by chemical etching.[33,34] The etchants used and etching times are listed in Table I. After etching, all the samples were rinsed with running distilled water in order to remove the etchant before exposing to the air. The samples were put into the evacuated sample chamber immediately after etching for the measurements.

Although, as a precaution, all measurements were done on freshly etched samples, not all of them exhibited the same aging effects due to surface contamination: Si and InSb in particular showed no change of their spectra after several weeks of exposure to dry air.

Surface contamination effects produce large distortion, mostly a cutoff in the uv, and are much less severe in the visible.

III. EXPERIMENTAL RESULTS

In Figs. 12-17 we present the derivative spectra dR(E)/RdE of GaAs, GaSb, InAs, InSb, Ge and Si respectively at three different temperatures, 5, 80 and 300°K.

The spectra of the six crystals are very much alike, reflecting the similarity of their band structures. Hence, a uniform notation for labeling the structures can be used. Following partially the notation of Cardona et al.,[6] we divide the structures in each spectrum into groups labeled by E_o, E_1, E_o', E_2, E_1', etc., as shown in the figures. Similar groups (having the same label) in different semiconductors are believed to come from transitions in similar general areas of the band structure.

Figure 18 shows the derivative spectrum dR/RdE of InSb at 5°K together with the normal reflectivity spectrum R(E). Group E_o, corresponding to transitions near the direct fundamental gap, has not been covered by this experiment. Groups E_1, E_2 and E_1' correspond, in terms of R(E), to three broad peaks with some splitting, and group E_o' is an intermediate region with only small structure. This pattern is generally the same for all the six crystals.

Several general conclusions can be drawn after a survey of all the spectra and a comparison with available data:

(1) Compared with the derivative spectra obtained from electroreflectance[6] and thermoreflectance[11] measurements, our spectra give the same gross features, but show more fine structures in the E_o', E_2 and E_1' regions, especially at low temperatures. As an example for comparison, we reproduce in Fig. 19 the electroreflectance spectrum of InAs obtained

by Cardona et al.[6] and in Fig. 20 the thermoreflectance spectrum of the same crystal obtained by Matatagui et al.[11]

(2) For most structures in our spectra, they agree well in positions with those obtained from normal reflectivity data at room[35] and liquid nitrogen temperature,[36] electroreflectance data at room temperature and thermoreflectance data at liquid nitrogen temperature.[11]

(3) The temperature dependence of the spectra is striking, although the change is rather gradual. All the main structures shift to higher energy when the temperature is lowered. The shift has a temperature coefficient that does not vary too much for different structures and for different samples. The structures become sharper at low temperature; this effect being particularly strong for the E_1 doublet and the main E_2 structure of the III-IV compounds.

(4) The spectra at liquid helium temperature are new in the sense that no complete spectra of all these materials at temperatures below liquid nitrogen could be found in the literature. However, the 5°K spectra are not very different from the spectra at 80°K except for more shifting and sharpening of structures.

As shown in Figs. 12-17, each group may contain many structures, the E_2 group being the most complicated. These structures presumably come from several reflectivity peaks superimposed on top of one another. Decomposition of a composite line into individual reflectivity peaks is somewhat arbitrary. In our case the decomposition was made with the following general rules:

(1) The low-temperature spectrum of a composite line should be decomposed into a minimum number of individual lines with simple lineshapes.

(2) Recomposition of these individual lines with broadened linewidth should yield the high-temperature spectrum of the composite line.

(3) Similarity in the spectra of different semiconductors should be used as a guide line in the decomposition.

With these rules, the ambiguity in the decomposition of our spectra is suprisingly small, although the positions of some components may not be very accurate and their shapes may be somehow arbitrary. We show in Figs. 21-26 the decomposition of the spectra at 5°K of the six semiconductors. In Table II we list the positions of all the reflectivity peaks (zeros with negative slope in the $\frac{1}{R}\frac{dR}{d\lambda}$ spectrum) obtained from such decomposition. The accuracy of the values, also listed, was mainly due to ambiguity in the decomposition, except for the E_1 structures.

We now proceed to comment on the low-temperature spectrum of each semiconductor separately:

GaAs (Figs. 12, 21): Although the spectrum below 2.7 eV is not shown in Fig. 12, we have explored this region carefully at 5°K. We have not been able to find the small structures at 2.3 and 2.6 eV observed by Greenaway.[36] In the E_1 region our spectrum confirms the absence of the small structures suggested by Lukes et al.[37] The spectrum in the E_o' region looks somehow peculiar, but it can be decomposed into two reflectivity peaks, consistently with the other III-V compounds. Decomposition of the E_2 region is rather arbitrary because part of the E_2 spectrum above 6 eV was cut off by our spectrometer. To be consistent with the other III-V compounds, we decompose the E_2 group into a strong broad peak with three small peaks at higher energy.

The temperature dependence of the spectrum of GaAs was studied more

carefully than that of the other crystals. Spectra at 150 and 225°K were measured in addition to the other three temperatures, but they are not reproduced in Fig. 12 in order not to crowd the figure. They show a gradual transition from the 80°K to the 300°K spectra. The dependence of the position of the E_1 reflectivity peaks and the main E_2 peak on temperature is plotted in Fig. 27.

GaSb (Figs. 13, 22). The structure at 1.9 eV observed by Greenaway[36] is absent in our spectrum. Decomposition of the spectrum gives unambiguously two peaks in the E_o' region (in disagreement with an observation of three peaks in a thermoreflectance[11] spectrum), one strong and three weak peaks in the E_2 region and two peaks in the E_1' region. The shape of the spectrum near 6 eV indicates the presence of additional structures just above 6 eV, belonging to the E_1' group.

InAs (Figs. 14, 23). We cannot identify in our spectrum the peaks at 2.2 and 2.45 eV suggested by Greenaway.[36] The spectrum of InAs is somewhat different from those of other III-V compounds in the sense that the E_1' and the E_2 regions overlap. We can unambiguously decompose the structures in the $E_o' + E_2$ region into six reflectivity peaks. We can assign the two weak ones at lower frequencies to the E_o' group and the rest to the E_2 group. Note that the spectrum of InAs appear to have a very strong temperature dependence.

InSb (Figs. 15, 24). The spectrum of InSb looks very much similar to that of GaSb except for a slightly stronger temperature dependence. Decomposition of this spectrum gives two peaks in the E_o' region, four in the E_2 region, and three in the E_1 region.

Ge (Figs. 16, 25). The spectrum of Ge is still quite similar to those of III-V compounds, but with less structures. Decomposition of the spectrum yields one peak in the E_o' region, one strong and one weak peak in the E_2 region and two peaks in the E_1' region.

Si (Figs. 17, 26). The spectrum of Si is rather different from those of the other materials in the E_1 and E_o' regions, difference that can be attributed to the difference of band structures near the direct gap (see Sec. Iv-C). The discussion of Sec. V will show that we should assign the peak at 3.4 eV to E_o' and the one at 3.45 eV to E_1. The spectrum in the E_2 and E_1' regions is similar to those of the other crystals. We can decompose the spectrum into one strong and one weak peak in the E_2 region and only one in the E_1' region. The temperature dependence of the structures (sharpening and shifting) is the smallest of all the crystals we have investigated.

IV. THEORETICAL BACKGROUND

All the optical properties of a crystal are determined once the dielectric constant is known. For cubic materials in the absence of a magnetic field, the dielectric constant ε is a complex scalar: $\varepsilon = \varepsilon_1 + i\,\varepsilon_2$. The real and imaginary parts are connected by the Kramers-Krönig relations, i.e.

$$\varepsilon_1(\omega) - \varepsilon_o = (2/\pi)\, P \int_o^\infty \omega' \varepsilon_2(\omega')(\omega'^2-\omega^2)^{-1}\, d\omega' \tag{38}$$

$$\varepsilon_2(\omega) = -(2\omega/\pi)\, P \int_o^\infty \varepsilon_1(\omega')(\omega'^2-\omega^2)^{-1}\, d\omega' \tag{39}$$

If one component of ε is known (or can be extrapolated) over a wide range then the other is determined by one of the equations above.

Two other real optical constants, the refractive index n and the extinction coefficient k, are connected to the dielectric constant by the expresssions:

$$\varepsilon_1 = n^2-k^2 \tag{40}$$

$$\varepsilon_2 = 2nk \tag{41}$$

The reflectivity R at normal incidence from air is a function of n and k:

$$R = \frac{(n-1)^2 - k^2}{(n+1)^2 - k^2} \tag{42}$$

A. The One-Electron Approximation

Approximations are often used in the microscopic theory of dielectric constant of a crystal. For the energy range of interest the approximation often used is the one-electron approximation. In such an approximation each valence electron is treated as a single particle that moves in a potential $V(\underline{r})$. However, in order to take the other electrons into account, $V(\underline{r})$ is the sum of the core potentials and a self-consistent Hartree potential of all the other valence electrons. The Hamiltonian, neglecting spin dependence, is

$$\mathcal{H} = p^2/2m + V(\underline{r}) \tag{43}$$

$V(\underline{r})$ has the translational symmetry of the lattice. Therefore application of the well-known Bloch theorem leads to a description of the one-electron states in terms of a band structure.[38]

We use the Born-Oppenheimer approximation. For most of this discussion we assume the temperature to be 0°K, so that no phonons are present. At higher temperatures, the electron-phonon interaction is responsible for increasing the lifetime broadening and producing energy shifts of the states. The latter will be discussed in the next section.

Under the assumptions of the one-electron model, the expression for the imaginary part ε_2 of the dielectric constant as a function of the optical frequency ω in the dipole approximation[39] is

$$\varepsilon_2(\omega) = \frac{4\pi^2 e\hbar}{3m^2\omega^2} \sum_{i,j} \int_{B.Z.} \frac{2}{(2\pi)^3} \left| M_{ij}(\underline{k}) \right|^2 \delta[\omega_{ij}(\underline{k}) - \omega] \, d^3k \tag{44}$$

where subscripts i and j refer to filled and unfilled bands respectively and the integration is performed over the Brillouin zone. $\omega_{ij}(\underline{k})$ is

given by:

$$\omega_{ij}(\underline{k}) = [E_i(\underline{k}) - E_j(\underline{k})]/\hbar \tag{45}$$

and $M_{ij}(\underline{k})$ is:

$$M_{ij}(\underline{k}) = \langle u_{i\underline{k}} | \underline{p} | u_{j\underline{k}} \rangle / \hbar \tag{46}$$

where $u_{i\underline{k}}$ and $u_{j\underline{k}}$ are the periodic parts of the Bloch eigenfunctions of the energies $E_i(\underline{k})$ and $E_j(\underline{k})$.

Indirect (phonon assisted) transitions are neglected in Eq. (44). Their contribution, above the fundamental energy gap, is expected to be negligible when compared with stronger direct transitions.[4]

The integral in Eq. (44) can be easily transformed so that it becomes:

$$\varepsilon_2(\omega) = \frac{4\pi^2 e\hbar}{3m^2\omega^2} \sum_{i,j} \int_{\omega_{ij}(\underline{k})=\omega} 2(2\pi)^{-3} |M_{ij}(\underline{k})|^2 \, |\nabla_{\underline{k}} \omega_{ij}(\underline{k})|^{-1} \, ds \tag{47}$$

The integral now extends over the surface (in k-space) defined by $\omega_{ij}(\underline{k}) = \omega$ and ds is a differential element of area of such a surface. The integral of Eq. (47), without the matrix element $|M_{ij}(\underline{k})|^2$, is the joint density of states for bands i and j, which we call $J_{ij}(\omega)$:

$$J_{ij}(\omega) = \int_{\omega_{ij}(\underline{k})=\omega} 2(2\pi)^{-3} |\nabla_{\underline{k}} (\omega_{ij}(\underline{k})|^{-1} \, ds \tag{48}$$

The dependence of $\varepsilon_2(\omega)$ on the band structure, given by Eq. (47), is still very complicated. However, the existence of a gradient in the denominator within the integral suggests that it is important to investigate whether the integrand has singularities at some points in

k-space. Such singularities would certainly give strong contributions to $\varepsilon_2(\omega)$ that might easily overwhelm the effect of variations of $|M_{ij}(\underline{k})|^2$. The points in k-space where singularities in the integrands of (47) and (48) occur, namely where $\nabla_{\underline{k}}\, \omega_{ij}(\underline{k}) = 0$, are called critical points. The singularities, usually called Van Hove singularities, can be classified into four types,[40] M_o, M_1, M_2 and M_3, according to the mathematical behavior of the function $\omega_{ij}(\underline{k})$ near the singularity: M_o designates a point in k-space where $\omega_{ij}(\underline{k})$has a minimum, M_3 corresponds to a maximum of $\omega_{ij}(\underline{k})$, and M_1 and M_2 refer to saddle points. Phillips[41] showed, employing symmetry arguments, that there is a minimum set of critical points for each crystal structure. Critical points usually occur at symmetry points, but there can be exceptions.

The Van Hove singularities create "kinks" in the spectrum of $\varepsilon_2(\omega)$. These "kinks" give rise to structure in the reflectivity spectrum. Therefore an approximate analysis of the interband transitions spectrum of a semiconductor becomes a search for critical points of the band structure. Calculations[42] show that, although the matrix element $|M_{ij}(\underline{k})|^2$ cannot be neglected in a fine analysis, good qualitative results can be obtained in some cases by assuming that $|M_{ij}(\underline{k})|^2$ is a constant. The effect of the matrix element is to increase or decrease the relative strength of some features and it is important only for comparing relative strength of peaks. Eventually it may become important for identifying structures only when it is necessary to judge whether a small peak is strong enough to split from larger peaks nearby.

Interesting conclusions are reached when band structure calculations are complemented by careful analysis of energy contours on different

planes of the Brillouin zone and by separation of the contributions of each pair of bands to $\varepsilon_2(\omega)$. Such analysis shows that, although most critical points are on symmetry lines as expected, some critical points may be located at points in the Brillouin zone without any special symmetry.[43,44] It is also showed that not only critical points are responsible for structure in the optical spectra, but, in some cases, optical transitions over large regions of the Brillouin zone where the bands are not quite parallel (volume effect)[45,46] also yield structure in the optical spectra.

B. Band Structure Calculations

Early calculations by the Wigner-Seitz method are not practical for our purpose. Simple models like the Nearly-Free Electron Model or the Tight-Binding Model are of some interest for metals and insulators respectively, but they are not good approximations for the covalent behavior found in semiconductors of groups IV and III-V. Calculations for such materials meet a computational difficulty: in order to calculate the dielectric constant with Eq. (44), which requires an integration over the Brillouin zone, the energy eigenstates and eigenfunctions must be calculated at a fairly dense mesh of points in the Brillouin zone. The finer the mesh, the more accurate the calculation is. The availability of fast computers has reduced the problem of numerical accuracy to the point that now computational errors are no longer the limiting factor. The accuracy of the calculation is now set by basic assumptions; i.e. how to determine the Hamiltonian, or more precisely the potential.

Several "first principle" methods to calculate band structures have

been developed with some success; a good description of them is given in Ref. (47). These methods, such as Orthogonalized Plane Waves, Augmented Plane Waves and Korringa, Kohn and Rostoker method, differ from one another in calculation procedures, but they all essentially depend on an estimate of $V(\underline{r})$ as a sum of the free-ion potentials known from atomic calculations, and a term which describes the Hartree potential of the other electrons. The assumptions required in order to estimate this potential, plus approximations in the computation, make the accuracy of these methods not better than 0.5 eV for the band energies. This is very satisfactory for "first principles" calculations, but it is too far from the accuracy we need for discussing details of our reflectivity spectra.

For the reason stated above, we turn our attention to semiempirical methods. One possibility has been advanced by Cardona et al.,[48] with $\underline{k}\cdot\underline{p}$ calculations, in which some matrix elements are treated as adjustable parameters. We however prefer for our purposes a much more powerful method: the empirical pseudopotential method. This method was first developed by Phillips and Kleiman[49] as a modification of OPW; Cohen and Heine[50] observed the cancellation properties of pseudopotentials and Austin et al.[51] analysed different kinds of pseudopotentials. The determination of pseudopotentials by fitting spectroscopic energy levels of free ions was also explored with some success.[52] Finally Brust[42] proposed the empirical fitting of pseudopotentials as it is currently done.

In order to understand how the pseudopotential method works, we must first assume that the potential $V(\underline{r})$ can be decomposed into a sum

of "atomic" potentials

$$V(\underline{r}) = \sum_i v(\underline{r}-\underline{R}_i) \tag{49}$$

where $\underline{R}_i$ are the atomic positions in the crystal. Although we are implicitly assuming here to have only one kind of atoms, the following can be easily generalized when different atomic species are present. We must bear in mind that the "atomic" potentials $v(\underline{r}-\underline{R}_i)$ are not just bare core potentials but they also include their share of Hartree potential of valence electrons.

The "atomic" potentials $v(\underline{r}-\underline{R}_i)$ look like deep potential wells centered at the atomic sites $\underline{R}_i$, this form being due to the strong attractive potential of the cores. Such well-like potentials make us think of the valence electrons as moving through an array of scattering centers[53] located at the atomic positions $\underline{R}_i$, with scattering potentials $v(\underline{r}-\underline{R}_i)$. The pseudopotentials we are going to define would be some smaller potentials capable of giving the same scattering or diffraction pattern for the valence electrons as the true scattering potentials $v(\underline{r}-\underline{R}_i)$.

Let us recall from scattering theory,[54] that the scattering amplitude in a direction making an angle θ with the incident beam is

$$f(\theta) = (2ik)^{-1} \sum_{\ell=0}^{\infty} (2\ell+1)(e^{2i\delta_\ell}-1)\, P_e(\cos\theta) \tag{50}$$

where δ_e is the phase shift of the partial wave of angular momentum ℓ. Once $f(\theta)$ is known, the scattering problem is completely determined, because $f(\theta)$ tells us how to add scattered wavefunctions, with the proper phase, in order to construct the diffraction pattern. And we

notice in (50) that $f(\theta)$ is determined once the set of phase shifts δ_ℓ is known.

Without going into details about the actual calculation of phase shifts from the potential, we just recall[54] that δ_ℓ is positive for an attractive potential, and when the potential is deep as in our case, δ_ℓ may be much larger than π. In fact a deep potential well may "pull in" several oscillations of the radial wave function, each one corresponding to a phase change π. However, Eq. (50) shows that $f(\theta)$ does not change if we subtract an integral multiple of π from δ_ℓ leaving δ_ℓ smaller than π. We get such reduction of δ_ℓ if we substitute the original scattering potential by a suitable shallow potential. Physically we are replacing a strongly oscillating wave function with large kinetic energy in a deep negative potential, by a much smoother wave function with less kinetic energy in a shallow negative potential. Neither the energy nor the diffraction pattern of the electrons outside the core wells are affected by the substitution. The new shallow potentials v_i^P, called "atomic" pseudopotentials, must be ℓ-dependent, i.e. non-local, in order to allow us to subtract an integral number of π from each separate δ_ℓ without changing the remainders left by each subtraction. We can write v_i^P as:

$$v_i^P = \sum_\ell v_e(\underline{r}-\underline{R}_i)\, p_\ell \tag{51}$$

where p_ℓ is a projection operator that selects the ℓ component of any wave function. The total pseudopotential is

$$V^P = \sum_i v_i^P \tag{52}$$

We conclude that if we replace the potential $V(\underline{r})$ in the

Hamiltonian (43) by the pseudopotential V^P, and diagonalize the new Hamiltonian, the energy eigenvalues are the true ones although the wave functions, now called pseudo-wave functions, coincide with the true wave functions only outside the atomic cores.

What we gain with the substitution of $V(\underline{r})$ by V^P is that, because V^P is smooth, the pseudo-wave functions don't have strong oscillations at the cores as is the case with the actual wave functions. Therefore expansions of the pseudo-wave functions in a plane-wave series converge quickly. The price paid by this method is the non-locality of the pseudopotential. However, for the semiconductors of our interest, local approximations to the pseudopotential give very satisfactory results.[42,55] Calculations with non-local pseudopotentials[56] show that the non-local terms are very small.

In order to diagonalize the Hamiltonian, all we need from the pseudopotential V^P are the matrix elements of V^P between plane wave states $|\underline{k}\rangle$ and $|\underline{k}+\underline{q}\rangle$ for all the reciprocal lattice vectors $\underline{q}=\underline{G}$ of the crystal. In the local pseudopotential approximation, such matrix elements can be written as:

$$\langle \underline{k}+\underline{q}|V^P|\underline{k}\rangle = S(q)\ v(q) \tag{53}$$

where

$$S(q) = \frac{1}{N} \sum_j \exp(-i\underline{q}\cdot\underline{R}_j) \tag{54}$$

is non-zero only for $\underline{q}=\underline{G}$, and

$$v(q) = \langle \underline{k}+\underline{q}\ |V^P|\ \underline{k}\rangle = \frac{1}{\Omega}\int v^P(z)\ \exp(-i\underline{q}\cdot\underline{r})\ d^3r \tag{55}$$

Ω being the atomic cell volume and N the number of atoms.

S(q) is called "structure factor" because it depends only on the geometry of the lattice. v(q) is called "form factor" and it depends essentially on the properties of the atom, except for the effect of changes of the Hartree potential of the valence electrons when an ion is in different crystals.

To treat v(q) as characteristic of each atom has proved a major success of the pseudopotential method and it has allowed to analyze regular trends in the periodic table[2] and even to do some successful interpolation of pseudopotentials on the periodic table.[55,57] For this purpose, v(q) is considered as a continuous function of q, a typical shape being shown in Fig. 28. For a particular crystal, with a given set of reciprocal lattice vectors $\underline{G}$, the corresponding set of v(G)'s is used.

The empirical pseudopotential method as devised by Brust[42] consists of an empirical fitting of form factors. Since v(q) goes to zero for large q, as shown in Fig. 28, only a few form factors need to be taken into account and many successful calculations have been done assuming only the first three or four form factors to be different from zero. The fitting procedure is by trial and error. The whole band structure must be calculated for each trial by diagonalizing the Hamiltonian at a mesh of points in k-space; E_2 is calculated with Eq. (44) and compared with experimental data. The form factors are then modified and the calculation is repeated. This goes on until a satisfactory fitting of E_2 is found. Although the adjustment of form factors is done with a very limited freedom of choice in order to take systematic trends of the periodic table into account, the agreement with experimental data over a

broad energy range is very good for many crystals.[2]

Spin-orbit splitting was neglected in the early calculations. Later some semi-empirical calculation was done by Cardona et al.[48] in their $\underline{k}\cdot\underline{p}$ band structures. They assume that, given the spin-orbit splittings at k=0, such splittings combine linearly at any point $\underline{k}$ of the Brillouin zone with the same coefficients that mix the $\underline{k}$=0 wave functions. The spin-orbit splitting at the bottom of the conduction band was taken from experimental data.

Recently spin-orbit was introduced in pseudopotential calculations[58] adding, for materials with p-bands as it is our case, only one extra parameter to the calculations. This parameter is usually adjusted in order to get agreement with the experimental spin-orbit splitting of the fundamental gap. Several calculations with spin-orbit effects for some of the materials of interest for us are now available.[46,56]

C. Description of the Band Structure of Six Semiconductors

The six semiconductors of interest for us have cubic symmetry; the III-V compounds GaAs, GaSb, InAs and InSb have the zincblende structure, while Ge and Si have the diamond structure. The Brillouin zone, common to both structures, is the typical Brillouin zone corresponding to a face-centered cubic lattice and it is shown in Fig. 29 in order to illustrate our notation for symmetry points and symmetry lines. Notice that in the band structures of the following figures the portion between X and K is S. This is possible without introducing discontinuities, because Σ and S can be thought of as a unique straight line, as it becomes apparent if we imagine another cell stuck to the side of the cell of Fig. 29. The

prolongation of Σ just happens to be S in the added cell.

Our semiconductors have been the subject of many band structure calculations, and among them we choose to show in Fig. 30 those of Cohen and Bergstresser.[55] Although several more recent calculations are available for most of our semiconductors, and we shall refer to them in our discussion, the pseudopotential calculation without spin-orbit of Cohen and Bergstresser is a systematic work that includes all the semiconductors of our interest, and it is therefore very useful for the observation of systematic trends.

If we compare the band structures of Fig. 30, we immediately notice a remarkable similarity among the III-V compounds. In fact, we can go from one to the other qualitatively, by lowering Γ and reducing the other gaps at X and L accordingly in the sequence GaAs, GaSb, InAs and InSb.

For Ge and Si the existence of inversion symmetry, not present in the III-V compounds, makes the bands more degenerate than those of the III-V compounds. However, the only fundamental change for the bands reproduced in Fig. 30, in which we neglect spin-orbit coupling effects, is that X_1 and X_3 become degenerate. Except for this change the band structure of Ge is very similar to that of GaAs, as expected, since Ge is located between Ga and As in the same row of the periodic table.

The band structure of silicon seems to be very different from the others. However, a careful comparison with Ge shows that most of the difference is simply that $\Gamma_{2'}$ is much higher in Si than in Ge. In Si $\Gamma_{2'}$ is above Γ_{15}, leading to the existence of an indirect gap. This is a trend of group IV, where the height of $\Gamma_{2'}$ above Γ_{25} follows the

sequence C, Ge, Si, Sn, being zero for the last of them, a fact that explains the semimetallic behavior of gray (cubic) Sn.

In order to illustrate which splittings occur when we take spin-orbit coupling into account, we show in Fig. 31 a band structure of GaSb reproduced from a recent calculation.[46] Although we should label the representations in Fig. 31 with the double-group notation, we keep the single-group notation in order to make our discussion of the next section simpler and to be able also to compare Figs. 31 and 30 more easily.

Let us now comment on Fig. 31 in detail, starting from the top valence band (band 4): the point Γ_{15}, six-fold degenerate without spin-orbit interaction, splits into one 4-fold and one 2-fold degenerate states, very much like the splitting of atomic p states into J=3/2 and J=1/2 states. The 4-fold degenerate bands at L_3 split into two doubly degenerate bands; strictly speaking, they split into three bands, one doubly degenerate and two non-degenerate, but the single bands stick together once again because of time reversal symmetry.[59]

The Λ line has the same symmetry as L, but the time reversal argument is no longer applicable (it is valid for a point on a zone face only[59]). Therefore Λ_3 is allowed to split into three bands, one doubly degenerate and two non-degenerate. However, with both ends at L and Γ degenerate, the separation of the single bands along Λ is probably small, and indeed calculations find no appreciable splitting.[46,56] Spin-orbit effects also split the 4-fold degenerate bands at Δ_5 and X_5 into two doubly degenerate bands.

The first conduction band (band 5) is doubly degenerate at L, Λ, Γ, Δ and X when spin-orbit interaction is neglected, and spin-orbit cannot

split this degeneracy.

The next conduction band (band 6) has the same symmetry of the upper valence band (band 4) at Γ, Δ and L, and the splittings are therefore the same. Along Δ the representation is Δ_1 which goes into X_3 at X; the band is doubly degenerate and it is not split by spin-orbit interaction.

Along the Σ line all the bands, either Σ_1 or Σ_2, are doubly degenerate when spin-orbit coupling is neglected. When spin-orbit interaction is introduced, symmetry no longer requires such degeneracy, but calculations show very small splittings.[46]

A rough estimate based on the degeneracy of the bands[60] predicts a spin-orbit splitting of the upper valence band at L_3 of order 2/3 the spin-orbit splitting at Γ_{15} (fundamental gap). More accurate calculations agree with this rule.[46] Calculations predict still smaller splittings for Δ_5 (band 4)[61] and for L_3 (band 6).[46]

For the remaining III-V compounds, spin-orbit splittings are expected to be in some degree proportional to the splitting of the fundamental gap. Such splitting increases in the sequence GaAs, InAs, GaSb, InSb. For InSb, the spin-orbit splitting is larger than the fundamental gap.

The spin-orbit splittings of Ge are smaller than those of GaAs.[48] The picture of the split bands still resembles that of the III-V compounds (Fig. 31) with few changes: the already described degeneracy of X_1 and X_3 is not split by spin-orbit interaction; the 4-fold degenerate X_4 remains unsplit when spin-orbit coupling is introduced, and therefore the valence band splits into two doubly degenerate bands along Δ between $\Gamma_{25'}$ and X_4, as in the III-V compounds, but both ends remain degenerate.

Finally another minor difference is that the doubly degenerate bands along Σ are not split by spin-orbit interaction.[59]

The spin-orbit splitting of the Si band structure is similar to that of Ge, but the splittings are quite small because the atomic spin-orbit coupling constant of Si is very small.[62]

D. Excitons

Excitons play an important role in the optical spectra of solids and, although they have been the subject of extensive investigations,[63] many questions still remain open. Excitons are collective excitations of the electrons for which two pictures are currently given: Frenkel[64] excitons are excited states of individual atoms or molecules only perturbed by neighbors; although localized in space, the excitation may propagate through the crystal transferring energy from one part to another. This description applies well to molecular crystals. Wannier's description,[65] more suitable for semiconductors, where the atoms interact strongly, is that of an electron wave packet and a hole wave packet moving rather freely through the crystal, bound to each other by the Coulomb interaction.

A crude but physically clear description of Wannier excitons can be given by treating an electron at the bottom of the conduction band as a quasiparticle with effective mass m_e, and a hole at the top of the valence band as a quasiparticle with effective mass m_h. The attractive Coulomb energy of the pair is $V = -e^2/\varepsilon_o r$, where r is the relative electron-hole coordinate and ε_o, the static dielectric constant, represents the screening of the other electrons. Such a system has a set of

hydrogenic "bound" states below the bottom of the conduction band. Strictly speaking, these are not bound states but rather the first excited states of the crystal. However, we call the hydrogenic states "bound" states because they are below the first excited one-electron state, their energy being lowered by an attractive interaction.

Such a crude description holds because the static dielectric constant is large, of the order of 10 for many semiconductors, and the effective mass of the electron-hole system is of the order of 1/10 the true electron mass. Since the radius of the exciton orbit is proportional to ϵ_o and inversely proportional to the effective mass, the radius is of the order of 100 times the Bohr radius, much larger than the lattice constant. Therefore the electron-hole pair extends over many core sites and the effective mass approximation works acceptably.

An exciton as described above is usually called a parabolic exciton, because the electron and the hole are about a parabolic (M_o) critical point of the band structure; in fact, the fundamental gap is always an absolute minimum of the energy difference between bands (see Section IV-A). However, at any point in k-space where the energy difference between two bands is extremal, the bands are parallel at least along one direction, and the group velocity of the electron and hole are the same, making a resonant motion possible. Phillips[4] pointed out this possibility and called such resonances hyperbolic or saddle point excitons, when they occur at hyperbolic (M_1 or M_2) critical points. Hyperbolic excitons are metastable because they are degenerate with continuum states into which they decay.

Theoretical calculations are difficult at saddle points because

the effective mass equation is not separable, although an adiabatic approximation can be done.[66] Another approach is to truncate the Coulomb interaction between electron and hole to extend to a finite number of neighboring cells.[67,68] The calculation by Toyazawa et al.[68] with a simple δ-function interaction predicts for the exciton part of $\varepsilon_2(\omega)$ at a critical point M_i, a shape similar to the contribution to $\varepsilon_2(\omega)$ from one-electron states at M_{i+1} (assuming $M_o=M_4$).

Excitons, both of parabolic and saddle point type, have been identified in solid rare gases[69] and alkali-halides.[70] In semiconductors, which have large dielectric constants (~10 compared with 5 for alkali-halides and 2.5 for solid rare gases) and a smaller conduction band mass, the exciton binding energy is smaller and the exciton structure can hardly split from the one-electron structure. Nevertheless, parabolic excitons were unmistakingly observed,[71] and there is evidence to support the existence of hyperbolic excitons,[13,16] although their precise role in the optical spectra of semiconductors is still not clear.

V. COMPARISON WITH THEORY AND EXPERIMENT

There are two ways to compare theoretical calculations with reflectivity measurements. One is to compare the dielectric constant obtained from the experimental data with the theoretical values, and the other is to compare the measured reflectivity with the one calculated from the theoretical dielectric constant.

The first method is based on the relations:

$$R = |r|^2 \tag{56}$$

$$r \equiv |r|e^{i\theta} = \frac{n-ik-1}{n-ik+1} \tag{57}$$

If R is measured over a sufficiently broad frequency range, then θ can be obtained from a Kramers-Krönig relation between the real and imaginary parts of the complex function $\ln r \equiv \ln|r| + i\theta$:[72]

$$\theta(\omega) = (2\pi/\omega) \int_0^\infty \ln|r(\omega')|(\omega'^2-\omega^2)^{-1}\, d\omega' \tag{58}$$

Once $|r|$ and θ are known, it is straightforward to obtain n and k from Eq. (57). From n and k one finds $\varepsilon_2(\omega)$ from Eq. (41). The experimental curve for $\varepsilon_2(\omega)$ is then compared with theoretically calculated $\varepsilon_2(\omega)$.

The trouble with the above procedure is that experimental measurements seldom cover a wide enough frequency range and the Kramers-Krönig transform of Eq. (58) must rely on extrapolations at the ends of the spectrum. For this reason it is more appropriate to follow the second procedure, which compares the measured reflectivity spectrum directly with the theoretical reflectivity spectrum

We should remark that for the III-V and IV semiconductors in the

energy region we are interested, R and ε_2 are very much alike.[73] They don't have strictly the same shape, but peaks of R usually correspond to peaks of ε_2 with only small shifts of their energies, and the strengths are also proportional. The main difference is that ε_2 goes to zero below the gap, while R decreases without going to zero.

Since our discussion is rather qualitative, we shall assume that the structure of $R(\omega)$ always follows that of $\varepsilon_2(\omega)$ closely. In particular we expect that spin-orbit splittings and temperature shifts of the ε_2 peaks can be measured directly from the reflectivity spectrum.

A. Discussion of Each Spectral Region

We start now discussing each spectral region in detail, taking advantage of the similarities between our semiconductors and the systematic trends described in Section IV-C. For this discussion we refer to Figs. 12 to 17 with the original spectra, and Figs. 21 to 26 with our decomposition of the E_0', E_2 and E_1' regions into individual structures.

We summarize in Table II our assignments of all the observed reflectivity peaks to the corresponding interband transitions, and we summarize in Table III the spin-orbit splittings at various symmetry points deduced from our spectra. In the tables and in the discussion, bands (without spin-orbit coupling) are numbered when some confusion may arise because two bands have the same representation. Also, in order to compare quickly with the III-V compounds, we label the Ge and Si transitions with the representation of the zincblende structure; we add between parenthesis the diamond structure notation when it is different.

The following discussion of each spectral region refers to spectra at 5°K.

1. The E_1 Region

Except for Si, all our spectra show a sharp doublet. It is generally agreed that this doublet correspond to $\Lambda_3 \rightarrow \Lambda_1$ transitions with the Λ_3 level spin-orbit split.[2,6] As shown in Table III, the observed splittings of the doublet agree with the 2/3 rule (see Sec. IV-C).

The non-existence of additional reflectivity peaks for $L_3 \rightarrow L_1$ transitions has been a puzzling fact.[2,36,45] That we have seen no additional structure in the E_1 region indicates that either the $L_3 \rightarrow L_1$ transitions are too weak or they are hidden in the strong $\Lambda_3 \rightarrow \Lambda_1$ structures.

For Si, raising of Γ_1 (Γ_2, in the diamond structure notation) above Γ_{15} (Fig. 30), makes the $\Lambda_3 \rightarrow \Lambda_1$ transitions and the $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions partially degenerate in energy. As shown in Figs. 17 and 26, there are two overlapping reflectivity peaks at 3.40 and 3.45 eV. Assuming that similar transitions in different crystals would yield reflectivity peaks of similar strength, we should assign the 3.45 eV peak to the $\Lambda_3 \rightarrow \Lambda_1$ transitions, and the 3.40 eV peak to the $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions, partially degenerate in energy. We already pointed out in Sec. IV-C, that spin-orbit coupling is particularly small for Si. The splitting is only of 0.04 eV at Γ_{15} ($\Gamma_{25'}$)[6,62] and even smaller splittings are expected along Λ_3 and Δ_5. The resolution of our spectrum is clearly not sufficient to show the spin-orbit splitting of either $\Lambda_3 \rightarrow \Lambda_1$ or $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions. We therefore rule out the possibility that the two peaks at 3.40 and 3.45 eV could correspond to the spin-orbit doublet of the $\Lambda_3 \rightarrow \Lambda_1$ transitions.

2. The E_o' Region

We can identify a doublet in the E_o' group for the four III-V compounds and a single peak for the group IV elements. These E_o' peaks were originally assigned to 4→5 transitions at or near Γ,[4,6] but the low joint density of states near Γ rules out such an assignment.[45,46] The peaks are more likely due to $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions away from Γ. Our results agree with the prediction, mentioned in Section IV-C, that the spin-orbit splitting along Δ_5 should be smaller than the splitting at L_3 (band 4). The observed E_o' doublet for the III-V compounds has indeed a splitting (see Table III) smaller than that of the E_1 doublet and proportional to such a splitting within 5%. In Ge, the degeneracy of X_5 (X_4) (see Sec. IV-C) makes the splitting of Δ_5 (band 4) small,[74] and we would not expect to resolve the spin-orbit doublet in the Ge spectrum. We therefore assign the single E_o' peak of Ge to $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions. The same is true for Si, which has an even smaller spin-orbit coupling.[62]

3. The E_2 Region

Our spectra for all the six crystals seem rather complicated in this region. They generally show more structures than either electroreflectance and thermoreflectance spectra. However we can always decompose the E_2 group quite unambiguously into a broad strong reflectivity peak and several small peaks at higher energies. Pseudo-potential calculations[2,43,45,46] indicate that this broad peak should be due to $\Sigma_2 \rightarrow \Sigma_1$ (4→5) transitions over a large region in the Brillouin zone. The small spin-orbit splitting of the III-V compounds along Σ would be difficult to resolve. From the band structures of Figs. 30 and 31, one might expect to observe a reflectivity peak corresponding to $X_5 \rightarrow X_1$ transitions at an energy between

$\Sigma_2 \rightarrow \Sigma_1$ (4→5) and $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions. We cannot recognize any such structure in all our spectra. This suggests that either $X_5 \rightarrow X_1$ transitions are too weak, or they are hidden in the broad $\Sigma_2 \rightarrow \Sigma_1$ (4→5) peak.

The small E_2 peaks have higher energies than the $\Sigma_2 \rightarrow \Sigma_1$ transitions. As seen from Figs. 30 and 31, they should correspond to transitions between the valence band (band 4) and the second conduction band (band 6). There is some ambiguity in decomposing the small E_2 structures, but we can unambiguously identify one peak in Ge and Si, three in InAs, InSb and GaSb, and probably also three in GaAs. We then recognize that for all the III-V compounds, the spacing between two of the three peaks agree quite well with the spin-orbit splitting of the Δ_5 level (see Table II). We therefore assign the doublet to 4→6 transitions along Δ (close to x).* The same transition should give rise to only one reflectivity peak in Ge and Si, since the spin-orbit splitting along Δ_5 for these two elements is small (see Sec. IV-C). Accordingly, the small E_2 peak of Ge and Si should correspond to $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions. The remaining small E_2 peak of the III-V compounds is assigned to 4→6 transitions around Σ as suggested by pseudopotential calculations.[75] Such a peak did not show up in the spectra of Ge and Si presumably because of the slight difference in their band structures.

4. The E_1' Region

Our spectra, limited by the uv cutoff of the spectrometer, cover only part of the E_1' region in Si, Ge, GaSb and InSb and none in GaAs and InAs. The E_1' peaks are normally assigned to 4→6 transitions along Λ,

---

* This assignment is also suggested by pseudopotential calculation for GaSb, which shows two spin-orbit split peaks coming from 4→6 transitions along Δ in the neighborhood of X. (Ref. 46).

close to L.[6,35] Both Λ_3 levels are spin-orbit split, but the splitting of Λ_3 (band 6) is expected to be considerably smaller than that of Λ_3 (band 4).[46,56] We should therefore expect to see two doublets separated by the spin-orbit splitting of Λ_3 (band 4).

In Si, the spin-orbit coupling is small,[62] and we have observed only one unresolved reflectivity peak, as expected. In Ge we can identify two peaks with a separation somewhat larger than the splitting of the $\Lambda_3 \rightarrow \Lambda_1$ doublet. This suggests that the splitting of Λ_3 (band 6) in Ge is very small, and the observed $\Lambda_3 \rightarrow \Lambda_3$ transitions are closer to L than the $\Lambda_3 \rightarrow \Lambda_1$ transitions. In InSb, where the spin-orbit coupling is larger, we have actually observed three peaks with the fourth one being cut off by our spectrometer. The spin-orbit splitting of Λ_3 (band 4) derived from them is again somewhat larger than the splitting of the $\Lambda_3 \rightarrow \Lambda_1$ doublet. The separation of the two overlapping peaks gives the spin-orbit splitting of Λ_3 (band 6) near L (see Table III). In GaSb, we can observe only one doublet with a small splitting corresponding to the splitting of Λ_3 (band 6). The other doublet at higher energy should be outside the range of our spectrometer.

B. The Temperature Dependence

We can obtain from our spectra at various temperatures some information about the temperature dependence of the band structure. Figures 12 to 17 show that all the major reflectivity peaks shift to lower energies at higher temperatures. The temperature shift of each peak is rather small and gradual.

We present in Table IV the observed temperature coefficients of

the E_1 and major E_2 peaks for all the six crystals. Note that Si has a smaller temperature dependence than the other crystals, presumably because it has a higher Debye temperature.

We have studied the temperature effect on GaAs in more detail and Fig. 27 shows the temperature shifts of the E_1 doublet and the major E_2 peak for this compound. The three curves behave similarly. The observed temperature shifts fit well with the exponential dependence:

$$\Delta E(T) \equiv E(T)-E(5^\circ K) = A \exp(-B/T) \tag{59}$$

with, for example, A=0.28 eV, B=320°K for the E_2 peak. The order of magnitude of B is that of the Debye temperature, indicating a close connection between the energy shifts and the phonon spectrum.

The principal factors governing the temperature-dependence of the reflectivity spectrum are the thermal expansion of the crystal and the thermal vibration of the nuclei (Debye-Waller effect). Since an expanded lattice reduces the average potential seen by the valence electrons, the energy splittings between bands are generally smaller at higher temperatures, and the energies of the reflectivity peaks decrease accordingly. The thermal vibrations of the nuclei reduce the core scattering potential seen by the electrons by the Debye-Waller factor e^{-W}.[76] This additional reduction of the cores potential also makes the band gaps smaller and decreases the energies of the reflectivity peaks.

We introduce both factors, thermal expansion of the crystal and Debye-Waller factor, in a pseudopotential calculation for GaAs.[77] For this purpose, we start with the choice of a suitable set of pseudopotential form factors and spin-orbit parameter (at 0°K), such that we

obtain a good agreement between the calculated reflectivity and the 5°K spectrum of GaAs. In this theoretical band structure, we identify the critical points responsible for the E_1 doublet, which corresponds to a $\Lambda_3 \rightarrow \Lambda_1$ spin-orbit split transition, and the critical point responsible for the main E_2 peak, which corresponds to a $\Sigma_2 \rightarrow \Sigma_1$ (4→5) transition, as predicted in our discussion of Sec. V-A. Thereafter, we analyze quantitatively the temperature variation of the band gaps at the three critical points.

We introduce the thermal expansion of the lattice in the calculation by varying the lattice constant. The temperature dependence of the lattice constant is obtained from the thermal expansion function for GaAs.[78] The lattice constants used at 5, 80, 150, 225 and 300°K are 5.640, 5.640, 5.641, 5.643 and 5.645 Å respectively. Since both the volume of the unit cell and the values of the reciprocal lattice vectors change slightly with variations in the lattice constant, we scale the pseudopotential form factors, the criterion being that the actual atomic potentials remain unchanged.

The Debye-Waller factor e^{-W} is calculated from the experimental phonon spectrum and, since this calculation is greatly simplified for a monoatomic crystal, we use the phonon spectrum of Ge instead of GaAs. This is a reasonable approximation because the GaAs phonon spectrum is nearly identical to that of Ge, and the average density of GaAs is the same as that of Ge to within 0.5%. Accordingly, we use an expression given by Blackman[79] for a monoatomic crystal:

$$W = \frac{hG^2}{8\pi^2 m} \frac{\int \rho(\nu)\nu^{-1}\left[1/2 + \frac{1}{e^x - 1}\right] d\nu}{\int \rho(\nu) d\nu} \tag{60}$$

where $x=h\nu/kT$, $\rho(\nu)$ is the density of phonon modes for Ge, G is a reciprocal lattice vector and m is the mass of the nucleus. The values of (W/G^2) we obtain by using Eq. (60) at 5, 80, 150, 225 and 300 °K are 0.0010, 0.0015, 0.0024, 0.0034 and 0.0044 respectively.

The Debye-Waller and lattice expansion effects are incorporated in the pseudopotential calculation of the band structure to give the temperature shifts of the band gaps at the three critical points identified above as responsible for the E_1 doublet and the main E_2 peak. The temperature shift is also calculated at Γ (the fundamental gap) as a test, and it is compared with the experimental temperature dependence obtained by Oswald,[80] finding a good agreement.

The theoretical temperature dependence of the E_1 and E_2 peaks is shown, with dashed lines, superimposed on the experimental curves in Fig. 27. The comparison between theory and experiment for the E_2 peak is good, and for the E_1 doublet, the comparison is excellent.

Separate calculation of the Debye-Waller and lattice expansion effects shows that the major part of the energy shift is caused by the Debye-Waller effect, with only a small fraction caused by lattice expansion.

The temperature shifts of the reflectivity peaks of GaSb, InAs, InSb and Ge are similar to those of GaAs. Si also behaves similarly but with smaller shifts presumably because of its high Debye temperature. Therefore we expect that similar theoretical calculations of the temperature shifts of peaks for these crystals would also yield good results.

A survey of the spectra of Figs. 12 to 17 also suggests that there

must be a close connection between broadening and the phonon spectrum. In all the spectra, the peaks become sharper when the temperature is decreased, presumably due to reduction of lifetime broadening, but the magnitude of the effect is not the same for all the crystals. We observe that sharpening of peaks at low temperature increases in the sequence Si, GaSb, Ge, GaAs, InAs, InSb. This correlates almost exactly with the sequence of decreasing Debye temperatures, which is Si, Ge, GaAs, GaSb, InAs, InSb; only GaSb occupies different places in the two sequences. This correlation indicates that further study of the connection between the broadening parameter of the reflectivity peaks and the phonon spectrum may yield interesting results for these crystals.

C. Excitons

Strong experimental evidence that the E_1 doublet of GaAs has a large exciton contribution is presented by Rowe et al.[16] with a measurement of wavelength-modulated reflectivity of a sample under uniaxial stress, at 77°K. They find as the only explanation for a polarization dependent splitting of the E_1 peaks, the existence of a hyperbolic exciton.

Additional evidence on the existence of excitons in the E_1 spectral region is presented by Shaklee et al.[13] with a wavelength modulation reflectivity measurement on InSb at 77°K. They show, by analyzing the lineshapes of the E_1 peaks, that such lineshapes are a mixture of those corresponding to M_1 and M_2 critical points. Since it is generally accepted that the E_1 peaks correspond to M_1 critical points,[2] the M_2-like portion of the lines is assigned to hyperbolic excitons, following

the prediction of Toyozawa et al.[68] on the exciton lineshapes at different critical points (see Sec. IV-D).

Our spectra show, for the E_1 low temperature peaks of all six semiconductors, lineshapes very similar to those reported by Shaklee et al. for InSb, therefore allowing us to extend their conclusions on the existence of hyperbolic excitons at E_1 to all the six crystals. Furthermore, we notice that the positive parts of the dR/RdE lines are approximately the M_1-like parts, while the negative parts are approximately the M_2-like parts of the E_1 structure. The negative M_2-like parts, which we assign to hyperbolic excitons, are the ones that sharpen very drastically at low temperature in agreement with the generally accepted idea that lifetime broadening for exciton peaks is more sensitive to temperature than for one-electron states peaks. These arguments on the existence of hyperbolic excitons at M_1 are also strengthened by band structure calculations, since the theoretical reflectivity spectra, which are calculated neglecting excitons, predict peaks at E_1 much weaker than the experimental ones, without the sharp M_2-like negative portions.[46]

As a concluding remark, we notice in other spectral regions, particularly in E_2, that some peaks sharpen at low temperature more than others. Although the complexity of the E_2 region rules out a simple lineshape analysis, this fact is an indication that there might be significant exciton contributions at some critical points responsible for the E_2 structure. Such possibility requires further investigation.

VI. CONCLUSIONS

Two goals have been achieved in this work: one is the construction of a very sensitive wavelength modulation spectrometer, the other is the measurement of the derivative reflectivity spectra of six semiconductors, yielding new information about their band structures.

We have shown how to build a sensitive wavelength modulation spectrometer to measure dR/dE (or dR/RdE) over a wide spectral range which extends to the near ultraviolet, employing mostly commercially available components. Modulation is introduced without basic modification of a commercial spectrometer. A two-beam system with electronic-optical feedback loops eliminates the noisy background effectively, even when an arc lamp is used.

The spectrometer was used to obtain the spectra of six semiconductors at several temperatures. Our derivative spectra show clear improvement of resolution over other techniques. In particular our low temperature spectra give more clearly defined reflectivity peaks than either electroreflectance or thermoreflectance spectra. With available information about the band structure, the spin-orbit splittings and similarities among the semiconductors, we can consistently assign all the observed reflectivity peaks to proper critical transitions between bands. Values of spin-orbit splittings at various symmetry points can then be deduced. Results agree well with theoretical calculations.

Our measurements at various temperatures also yield valuable information about the temperature dependence of the band structure. All the reflectivity peaks shift to higher energies at lower temperature. A theoretical calculation of some shifts by the pseudopotential method

yield good agreement with the experimental values. The structures in the spectra generally become much more pronounced at lower temperatures. Sharpening of the E_1 peak at low temperature is particularly striking and can be explained in terms of the hyperbolic excitons associated with Λ. Whether the exciton effect is also important in the other transitions remains to be investigated.

To help us make sure that our assignment of reflectivity peaks is correct, measurements on samples under uniaxial stress should be performed. That a stress can be exerted on the sample without much complication is another advantage of the wavelength modulation scheme. The pressure dependence of the reflectivity spectrum should also yield valuable information about hyperbolic excitons associated at various symmetry points.[81]

With simple modifications, it is possible to adapt the wavelength modulation spectrometer for transmission measurements and also to extend the spectral range to the near infrared. This would allow us to extend the investigation of this work to the band gap of the III-V compounds, and probably gain new information on parabolic excitons. Wavelength modulation can also be applied to other semiconductors, metals,[82] and materials of biological interest.

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Prof. Y. R. Shen for suggesting this problem and for his encouragement and guidance through all this work. I also thank Prof. M. L. Cohen, Dr. J. P. Walter, R. Cahn and Dr. C. Y. Fong for helpful discussions on the interpretation of the experimental data, and Prof. M. P. Klein for helping the initial design of the experimental setup. P. Saltz helped with the electronic design and adjustment.

I would also like to thank Prof. B. Loeb for his kind support and advice in his role of student advisor.

I am grateful to the Pan American Union and the Consejo Nacional de Investigaciones Cientificas y Tecnicas of Argentina for the financial support. This work was done under the auspices of the Atomic Energy Commission.

APPENDIX

WAVELENGTH-MODULATION OF A NEARLY MONOCHROMATIC LIGHT BEAM

We have shown in Eq. (4) that the signal intensity of a wavelength-modulated light beam as a function of its central wavelength λ_o and time t can be written, after a Taylor expansion, as:

$$S(\lambda_o,t) = \int_{-\infty}^{\infty} g(\lambda'-\lambda_o-A\cos\omega t)[T(\lambda_o)+T'(\lambda_o)(\lambda'-\lambda_o) + \frac{1}{2}T''(\lambda_o)(\lambda'-\lambda_o)^2 + \ldots]\,d\lambda' \tag{1'}$$

If we change the variable of integration replacing λ' by λ, such that

$$\lambda = \lambda'-\lambda_o-A\cos\omega t \tag{2'}$$

Eq. (1') becomes:

$$S(\lambda_o,t) = \int_{-\infty}^{\infty} g(\lambda)[T(\lambda_o)+T'(\lambda_o)(\lambda+A\cos\omega t) + \frac{1}{2}T''(\lambda_o)(\lambda+A\cos\omega t)^2+\ldots]d\lambda \tag{3'}$$

Integrating term by term,

$$S(\lambda_o,t) = T(\lambda_o)\int_{-\infty}^{\infty} g(\lambda)d\lambda+T'(\lambda_o)\int_{-\infty}^{\infty} g(\lambda)(\lambda+A\cos\omega t)\,d\lambda + \frac{1}{2}T''(\lambda_o)\int_{-\infty}^{\infty} g(\lambda)(\lambda+A\cos\omega t)^2 + \ldots \tag{4'}$$

If we now expand each parenthesis of the type $(\lambda+A\cos\omega t)^n$ and integrate term by term again, we find integrals of the type

$$\int_{-\infty}^{\infty} g(\lambda)\,\lambda^{\alpha}\,d\lambda \tag{5'}$$

where α is a positive integer. Since $g(\lambda)$ is a symmetric function, only those integrals with even α are different from zero. Also, because $g(\lambda)$ is normalized to W, we can write, for $\alpha=2$,

$$\int_{-\infty}^{\infty} g(\lambda)\, \lambda^2\, d\lambda = \xi_2 W^3 \tag{6'}$$

where ξ_2 is adimensional ($\xi_2=1/6$ for the ideal $g(\lambda)$ of Fig. 1), and we can write similar expresions for $\alpha=4$, etc.

Finally Eq. (4') becomes:

$$S(\lambda_o,t) = T(\lambda_o)W+T'(\lambda_o)WA\cos\omega t + \frac{1}{2} T''(\lambda_o)W(\xi_2 W^2+A^2\cos\omega t) + \frac{1}{6} T'''(\lambda_o)WA(3\xi_2 W^2\cos\omega t+A^2\cos^3\omega t) + \ldots \tag{7'}$$

We now expand powers of $\cos\omega t$ into Fourier components $\cos2\omega t$, $\cos3\omega t$, etc. using well known trigonometric identities, and collect terms; we finally obtain:

$$S(\lambda_o,t) = [T(\lambda_o)W + \frac{1}{2} T''(\lambda_o)(\xi_2 W^2+ \frac{1}{2} A^2)W + \ldots] + [T'(\lambda_o)WA+ \frac{1}{2} T'''(\lambda_o)WA(\xi_2 W^2+ \frac{1}{4} A^2) + \ldots] \cos\omega t + [\ldots] \cos2\omega t + \ldots \tag{8'}$$

This is the expression we sought, in which we are interested only in the first two terms, that are the ones we measure. We now want to retain only the leading terms in each of the first two brackets. We can do this if

$$\frac{1}{2} T''(\lambda_o)(\xi W^2 + \frac{1}{2} A^2) \ll T(\lambda_o) \tag{9'}$$

and

$$\frac{1}{2} T'''(\lambda_o)(\xi_2 W^2 + \frac{1}{4} A^2) << T(\lambda_o) \qquad (10')$$

For a spectral line with a characteristic width Δ, we can estimate orders of magnitude assuming that $T'(\lambda_o) \approx T(\lambda_o)/\Delta$, $T''(\lambda_o) \approx T(\lambda_o)/\Delta^2$ and $T'''(\lambda_o) \approx T(\lambda_o)/\Delta^3$. Therefore conditions (9') and (10') are satisfied if

$$W^2 << \Delta^2 \qquad (11')$$

and

$$A^2 << \Delta^2 \qquad (12')$$

Both requirements can be shown to hold exactly for Gaussian and Lorentzian lineshapes.

REFERENCES

1. F. Herman, R. L. Kortum, C. D. Kuglin, J. P. VanDyke and S. Skillman in Methods in Computational Physics, edited by B. Alder, S. Fernbach and M. Rotenberg (Academic Press, New York, 1968) Vol. 8, and references therein.

2. M. L. Cohen and V. Heine, The Fitting of Pseudopotentials to Experimental Data and Their Subsequent Applications (to be published in Solid State Physics, edited by F. Seitz, D. Turnbull and H. Ehrenreich, Academic Press Inc., New York); and references therein.

3. D. Dresselhaus, A. F. Kip and C. Kittel, Phys. Rev. 98, 368 (1955).

4. J. C. Phillips, The Fundamental Optical Spectra of Solids, in Solid State Physics, edited by F. Seitz, D. Turnbull and H. Ehrenreich, Academic Press Inc., New York, 1966, Vol. 18, p. 55; and references therein.

5. B. O. Seraphin and N. Bottka, Phys. Rev. 145, 628 (1966).

6. M. Cardona, K. Shaklee, and F. H. Pollak, Phys. Rev. 154, 696 (1967).

7. G. W. Gobeli and E. O. Kane, Phys. Rev. Letters 15, 142 (1965).

8. W. E. Engeler, H. Fritzsche, M. Garfinkel and J. J. Tiemann, Phys. Rev. Letters, 14, 1605 (1965).

9. M. Garfinkel, J. J. Tiemann and W. E. Engeler, Phys. Rev. 148, 659 (1968).

10. B. Batz, Solid State Commun. 4, 241 (1966).

11. M. Matatagui, S. G. Thompson and M. Cardona, Phys. Rev. 176, 950 (1968).

12. I. Balslev, Phys. Rev. 143, 676 (1966).

13. K. L. Shaklee, J. E. Rowe and M. Cardona, Phys. Rev. 174, 828 (1968).

14. D. E. Aspnes, P. Handler and D. F. Blossey, Phys. Rev. 166, 921 (1968).

15. E. O. Kane, Phys. Rev. 178, 1368 (1969).

16. J. E. Rowe, F. H. Pollak and M. Cardona, Phys. Rev. Letters, 22, 933 (1969).

17. J. E. Rowe, M. Cardona and K. L. Shaklee, Solid State Commun. 7, 441 (1969).

18. G. Bonfiglioli and P. Brovetto, Applied Optics, 3, 1417 (1964).

19. K. G. McWilliam, Journal of Scient. Instr. 36, 51 (1959).

20. A. Gilgore, P. J. Stoller and A. Fowler, Rev. Scient. Instr., 38, 1535 (1967).

21. A. Perregaux and G. Ascarelli, Appl. Optics, 7, 2031 (1968).

22. G. Bonfiglioli, P. Brovetto, G. Busca, S. Levialdi, G. Palmieri and E. Wanke, Appl. Optics, 6, 447 (1967).

23. C. S. French and A. B. Church, Annual Report of Carnegie Institution of Washington, p. 162, (1954-55).

24. L. R. Canfield, G. Hass and J. E. Waylonis, Appl. Optics, 5, 45 (1966).

25. V. A. Kizel, Soviet Physics Uspekhi, 10, 485 (1968) (English translation); 92, 479 (1967) (original version).

26. W. R. Hunter, Journ. Opt. Soc. of America, 54, 208 (1964).

27. The lock-in detectors are units built for general purpose by the Electronic Shop of the Physics Department of the University of California, Berkeley.

28. H. R. Philipp and H. Ehrenreich, Phys. Rev., 129, 1550 (1963).

29. T. Coor, Technical Note T-198A, Princeton Applied Research, Princeton, N. J.

30. P. L. Richards, Physics Department, University of California, Berkeley, private communication, September 1968.

31. M. Born and W. Wolf, Principles of Optics, Pergamon Press, London, 1959, p. 675.

32. H. R. Philipp and E. A. Taft, Phys. Rev., 113, 1002 (1959).

33. H. C. Gatos and M. C. Lavine, in Progress in Semiconductors, edited by A. F. Gibson and R. E. Burgess, Temple Press, London (1965), Vol. 9, p. 1.

34. R. G. Rhodes, Imperfections and Active Centres in Semiconductors, Pergamon Press, London, 1964, Chapter 9.

35. M. Cardona in Proceedings of the 7th International Conference on the Physics of Semiconductors, Dunod, Paris, 1964, p. 181.

36. D. L. Greenaway, Phys. Rev. Letters 9, 97 (1962).

37. F. Lukes and E. Schmidt in Proceedings of the International Conference on the Physics of Semiconductors, Exeter, 1962, p. 389.

38. C. Kittel, in Quantum Theory of Solids, J. Wiley & Sons, New York, 1966, Ch. 9.

39. M. H. Cohen, Phil. Mag. 3, 762 (1958).

40. L. Van Hove, Phys. Rev. 89, 1189 (1953).

41. J. C. Phillips, Phys. Rev. 104, 1263 (1956).

42. D. Brust, Phys. Rev. 114, A1337 (1964).

43. E. O. Kane, Phys. Rev. 146, 558 (1966).

44. L. R. Saravia and D. Brust, Phys. Rev. 171, 916 (1968).

45. J. P. Walter and M. L. Cohen, Phys. Rev. 183, 763 (1969).

46. R. Cahn and M. L. Cohen, Phys. Rev. B1, 2569 (1970).

47. J. Callaway, Energy Band Theory, Academic Press, New York, 1964.

48. M. Cardona and F. H. Pollak, Phys. Rev. 142, 530 (1966).

49. J. C. Phillips and L. Kleinman, Phys. Rev. 116, 287 (1959).

50. M. H. Cohen and V. Heine, Phys. Rev. 122, 1821 (1961).

51. B. J. Austin, V. Heine and L. J. Sham, Phys. Rev. 127, 276 (1962).

52. I. V. Abarenkov and V. Heine, Phil. Mag. 12, 529 (1965).

53. J. M. Ziman, Advances in Physics, 13, 89 (1964).

54. L. I. Schiff, in Quantum Mechanics, McGraw Hill, New York, 1955, Ch. V.

55. M. L. Cohen and T. Bergstresser, Phys. Rev. 141, 789 (1966).

56. H. I. Zhang and J. Callaway, Phys. Rev. 181, 1163 (1969).

57. L. M. Falicov and P. J. Lin, Phys. Rev. 141, 562 (1966).

58. G. Weisz, Phys. Rev. 149, 504 (1966).

59. R. H. Parmenter, Phys. Rev. 100, 573 (1955).

60. M. Cardona in Semiconductors and Semimetals, edited by R. K. Willardson and A. C. Beer, Academic Press, New York (1967), v.3, p. 125.

61. F. H. Pollak, C. W. Higginbotham and M. Cardona in, Proceedings of the International Conference on the Physics of Semiconductors, Kyoto, 1966, p. 20.

62. F. Herman, C. D. Kuglin, K. F. Cuff and R. L. Kortum, Phys. Rev. Letters 11, 541 (1963).

63. R. S. Knox in Solid State Physics, edited by F. Seitz, D. Turnbull and H. Ehrenreich, Academic Press, New York, 1963, Suppl. No. 5.

64. J. Frenkel, Phys. Rev. 37, 17 and 1276 (1931).

65. G. H. Wannier, Phys. Rev. 52, 191 (1937).

66. B. Velicky and J. Sak, Phys. Stat. Sol. 16, 147 (1966).

67. J. Hermanson, Phys. Rev. 166, 893 (1968).

68. Y. Toyozawa, M. Inoue, T. Inui, M. Okazaki and H. Hanamura, in Proceedings of the International Conference on the Physics of Semiconductors, Kyoto, 1966, p. 133.

69. G. Baldini, Phys. Rev. 128, 1562 (1962).

70. J. E. Eby, K. J. Teegarden and D. B. Dutton, Phys. Rev. 116, 1099 (1959).

71. G. G. MacFarlane, T. P. McLean, J. E. Quarrington and V. Roberts, Phys. Chem. Solids 8, 388 (1959).

72. F. C. Jahoda, Phys. Rev. 107, 1261 (1957).

73. H. R. Philipp and H. Ehrenreich, Phys. Rev. 129, 1550 (1963).

74. L. R. Saravia and D. Brust, Phys. Rev. 176, 915 (1968).

75. J. P. Walter, Private Communication, August 1969.

76. S. C. Yu, Ph. D. Thesis, Harvard (1964).

77. J. P. Walter, R. R. L. Zucca, M. L. Cohen and Y. R. Shen, Phys. Rev. Letters 24, 102 (1970).

78. S. I. Novikova, Soviet Physics, Solid State, 3, 129 (1961).

79. M. Blackman, in Handbuch der Physik, VI, Sec. 1, p. 325 (1962).

80. F. Oswald, Z. Naturforsh. 10a, 927 (1959).

81. E. O. Kane, Phys. Rev. 178, 1368 (1969).

82. C. Y. Fong, M. L. Cohen, R. R. L. Zucca, J.Stokes and Y. R. Shen, Phys. Rev. Letters 25, 1486 (1970).

83. G. Koster, in Solid State Physics, edited by F. Seitz and D. Turnbull, Academic Press, New York, 1957.

Table I. Sample properties and etching information.

Material	Type	Carrier Concentration (cm^{-3})	Etching Compound (parts per volume)	Etching Time (sec)
GaAs	n	10^{16}-10^{17}	$3HNO_3$+1HF+$4H_2O$	120
GaSb	p	1.5-3.5×10^{17}	1Br_2+15CH_3COOH	30
InAs	n	2-5×10^{16}	1Br_2+15CH_3OH	120
InSb	n	1-5×10^{14}	1Br_2+10CH_3COOH	30
Ge	p	5×10^{15}-5×10^{16}	$2HNO_3$+1HF+$4CH_3COOH$	90
Si	p	0.5-1×10^{13}	$6HNO_3$+3HF+$3CH_3COOH$	30

Table II. Energies (eV) of observed reflectivity peaks in the six crystals. Assignment of various peaks to particular interband transitions is also given.

	E_1		E_0'		E_2				E_1'		
	$\Lambda_3 \rightarrow \Lambda_1$		$\Delta_5 \rightarrow \Delta_1$ (4→5)		$\Sigma_2 \rightarrow \Sigma_1$ (4→5)	$\Sigma_2 \rightarrow \Sigma_1$ (4→6)	$\Delta_5 \rightarrow \Delta_1$ (4→5)		$\Lambda_3 \rightarrow \Lambda_3$		
GaAs	3.017	3.245	4.44	4.60	5.11	5.64	5.91	6.07(?)			
GaSb	2.154	2.596	3.35	3.69	4.35	4.55	4.75	5.07	5.51	5.65	
InAs	2.612	2.879	4.39	4.58	4.74	4.85	5.33	5.52			
InSb	1.983	2.478	3.39	3.78	4.23	4.75	4.56	4.92	5.33	5.50	5.96
Ge	2.222	2.420	3.20		4.49		5.01		5.58		5.88
Si	3.45		3.40		4.44		4.60		5.53		
Estimated Uncertainty	±0.004		±0.008		±0.01	±0.03	±0.03		±0.03		

Table III. Spin orbit splittings (in eV) at various points in the band structures of the six crystals obtained from analysis of our reflectivity spectra. The values of spin-orbit splittings at Γ_{15} are obtained from Ref. 6.

S-O Splitting at / Material	Γ_{15}	Λ_3(band 4)		Λ_3(band 6)	Δ_5	
		from $\Lambda_3 \rightarrow \Lambda_1$ transitions	from $\Lambda_3 \rightarrow \Lambda_3$ transitions	from $\Lambda_3 \rightarrow \Lambda_3$ transitions	from $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions	from $\Delta_5 \rightarrow \Delta_1$ (4→5) transitions
GaAs	0.34	0.22			0.16	0.16 (estimate)
GaSb	0.80	0.45		0.14	0.34	0.32
InAs	0.43	0.27			0.19	0.19
InSb	0.82	0.50	0.63	0.17	0.39	0.36
Ge	0.29	0.20	0.30			
Si	0.04					

Table IV. Average temperature shifts in unit of 10^{-4} eV/°K between 80 and 300 °K for the E_1 doublet and the major E_2 peak in the six crystals. Accuracy is $\pm 0.4 \times 10^{-4}$ eV/°K.

	GaAs	GaSb	InAs	InSb	Ge	Si
E_1	-5.3	-4.5	-5.0	-4.4	-4.2	-2.2
E_2	-3.6	-4.1	-5.6	-3.6	-2.4	-2.2

FIGURE CAPTIONS

Fig. 1. Ideal shape of the slit function $g(\lambda)$ for a monochromator with entrance and exit slit width equal to W.

Fig. 2. Superposition of two Gaussian peaks, $y_1 = \exp(-x^2/2)$ and $y_2 = 0.25 \exp[-(x-d)^2/0.5]$, for three different separations d. y_1 and y_2 are shown with dotted lines. The upper curves show $y=y_1+y_2$. The lower curves show dy/dx.

Fig. 3. Block diagram of the optical system.

Fig. 4. Optical parts for modulation and background compensation. The light beam is shown for the central (solid lines) and extreme positions (dashed lines) of the mirror oscillation.

Fig. 5. Light beam converging on the entrance slit of the spectrometer with diaphragm D at a) central position; b) off-central position.

Fig. 6. $S_R(\lambda_o,x)$ for three different positions of the diaphragm: a) centered; b) off-center; c) diaphragm off-centered for background cancellation.

Fig. 7. Optical parts after the spectrometer for the two-beam system. Mirrors C_1 and C_2 are a matched pair. Mirror C_3 and the mirrors of the beam chopper are the other matched pair.

Fig. 8. Block diagram of the electronic system.

Fig. 9. Signal waveforms at different points: a) photomultiplier output; b) preamplifier output; c) reference channel after switch, d) sample channel after switch; e) operational diagram of the electronic switch unit. The shift V_C at the preamplifier is for the operational mode in which R and $dR/d\lambda$ are measured. Sizes of ac signals are exaggerated. Signals are negative because they

come from a photomultiplier.

Fig. 10. Diagram of the Electronic Switch Unit.

Fig. 11. Test of GaAs spectrum at liquid helium temperature. a) ac spectrum ($dR/d\lambda$); b) dc spectrum (R) recorded simultaneously with a); c) R calculated by integration of $dR/d\lambda$. The arrows in a) point to the residues of xenon lines.

Fig. 12. Logarithmic derivative of the reflectivity spectrum of GaAs at 5, 80 and 300 °K.

Fig. 13. Logarithmic derivative of the reflectivity spectrum of GaSb at 5, 80 and 300 °K.

Fig. 14. Logarithmic derivative of the reflectivity spectrum of InAs at 5, 80 and 300 °k.

Fig. 15. Logarithmic derivative of the reflectivity spectrum of InSb at 5, 80 and 300 °K.

Fig. 16. Logarithmic derivative of the reflectivity spectrum of Ge at 5, 80 and 300 °K.

Fig. 17. Logarithmic derivative of the reflectivity spectrum of Si at 5, 80 and 300 °K.

Fig. 18. Reflectivity spectrum and logarithmic derivative reflectivity spectrum of InSb at 5°K in the range between 1.75 and 6 eV.

Fig. 19. Electroreflectance spectrum of InAs at room temperature (reproduced from Ref. 6).

Fig. 20. Thermoreflectance spectrum of InAs at 77°K (reproduced from Ref. 11).

Fig. 21. Decomposition of the derivative spectrum of GaAs at 5°K into many components.

Fig. 22. Decomposition of the derivative spectrum of GaSb at 5°K into many components.

Fig. 23. Decomposition of the derivative spectrum of InAs at 5°K into many components.

Fig. 24. Decomposition of the derivative spectrum of InSb at 5°K into many components.

Fig. 25. Decomposition of the derivative spectrum of Ge at 5°K into many components.

Fig. 26. Decomposition of the derivative spectrum of Si at 5°K into many components.

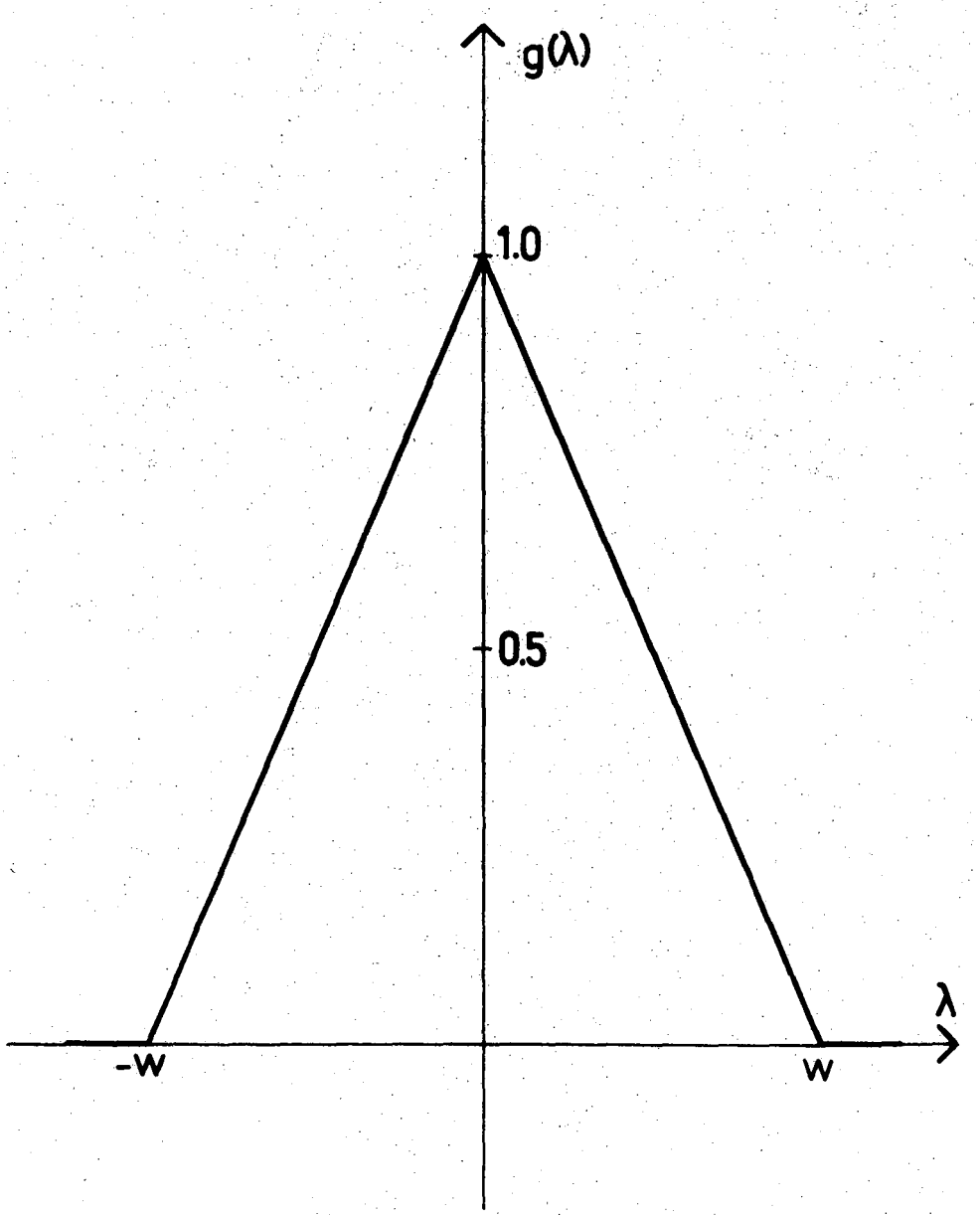
Fig. 27. Temperature shifts of the E_1 doublet and the major E_2 reflectivity peak of GaAs.

Fig. 28. Typical shape of v(q).

Fig. 29. Brillouin zone of a face centered cubic lattice, showing symmetry points and symmetry lines (from Ref. 83).

Fig. 30. Band structure of GaAs, GaSb, InAs, InSb, Ge and Si (from Ref. 55); spin-orbit coupling is neglected.

Fig. 31. Band structure of GaSb (from Ref. 46); spin-orbit coupling is considered; the single group notation is used for the representation labels.



XBL 7011-6888

Fig. 1.

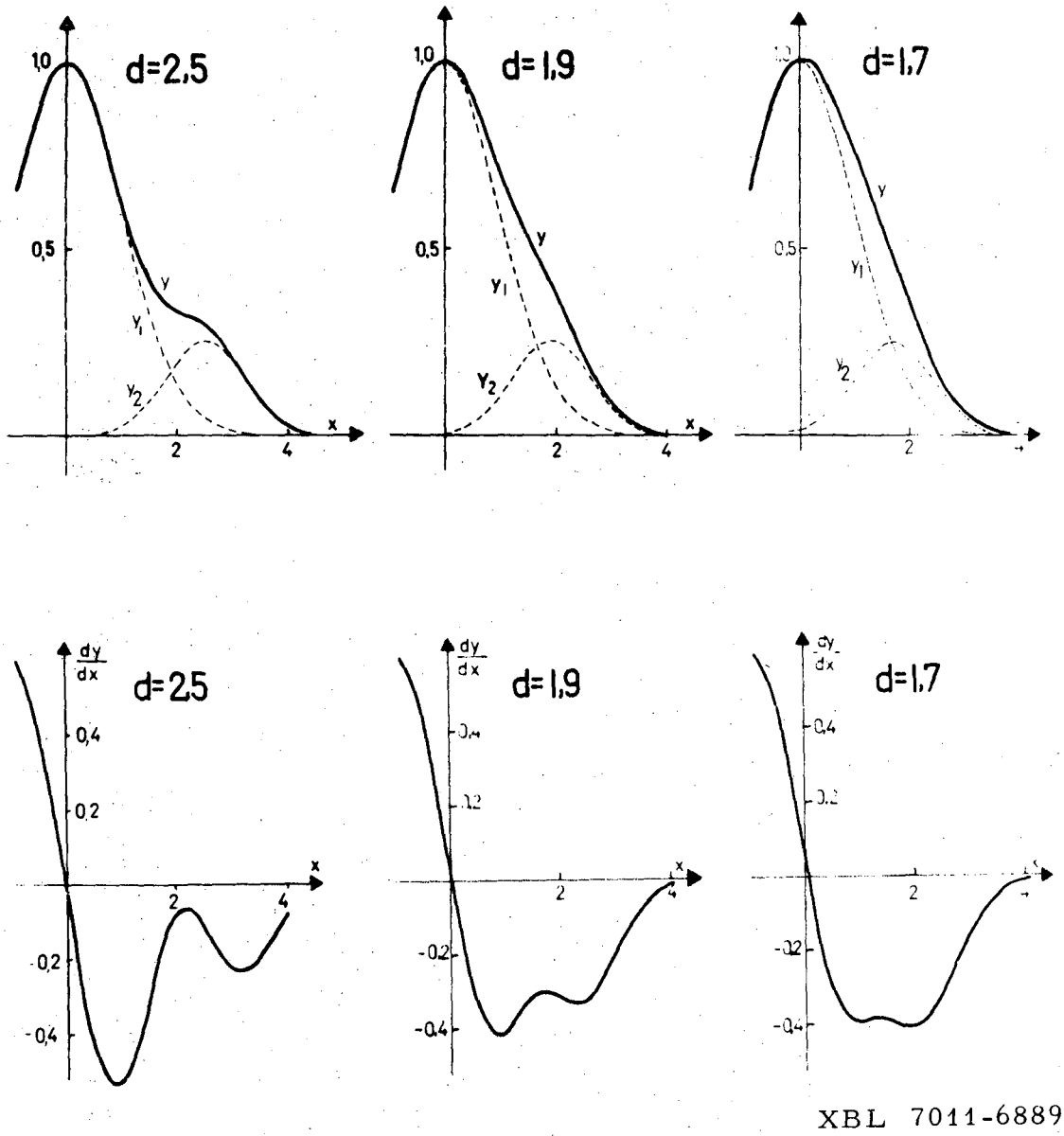
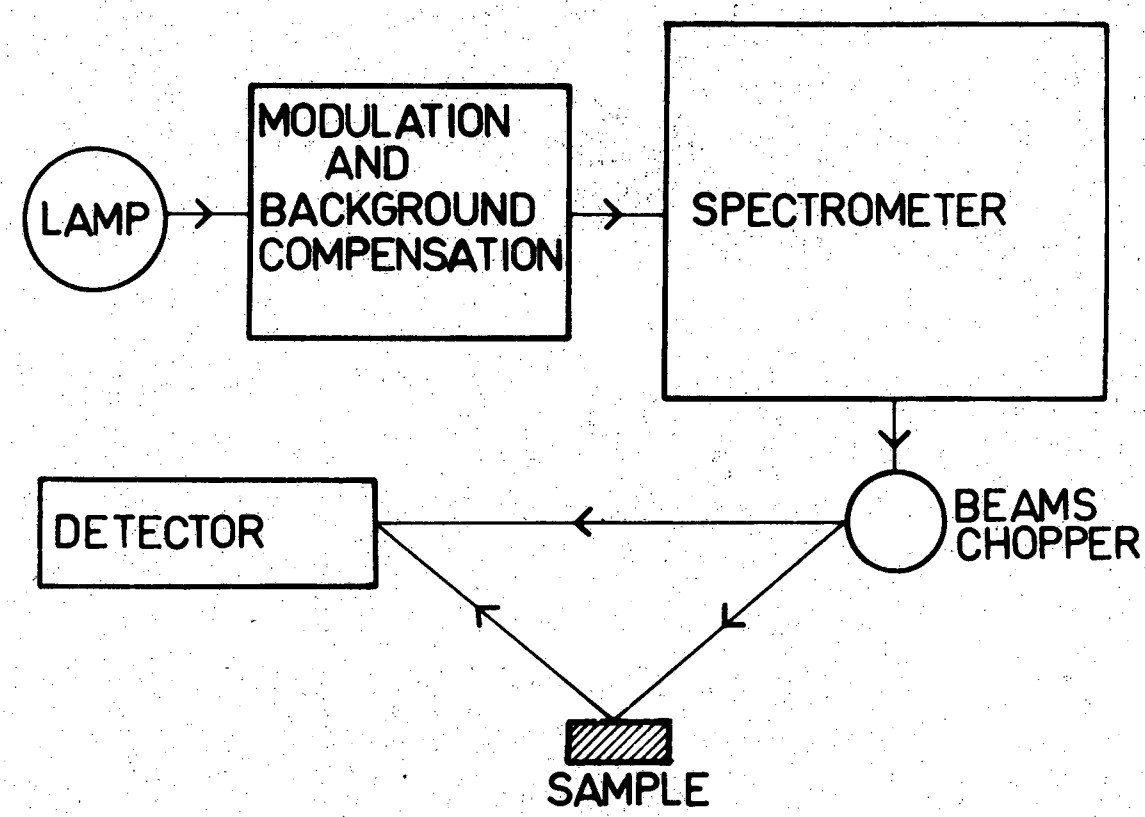
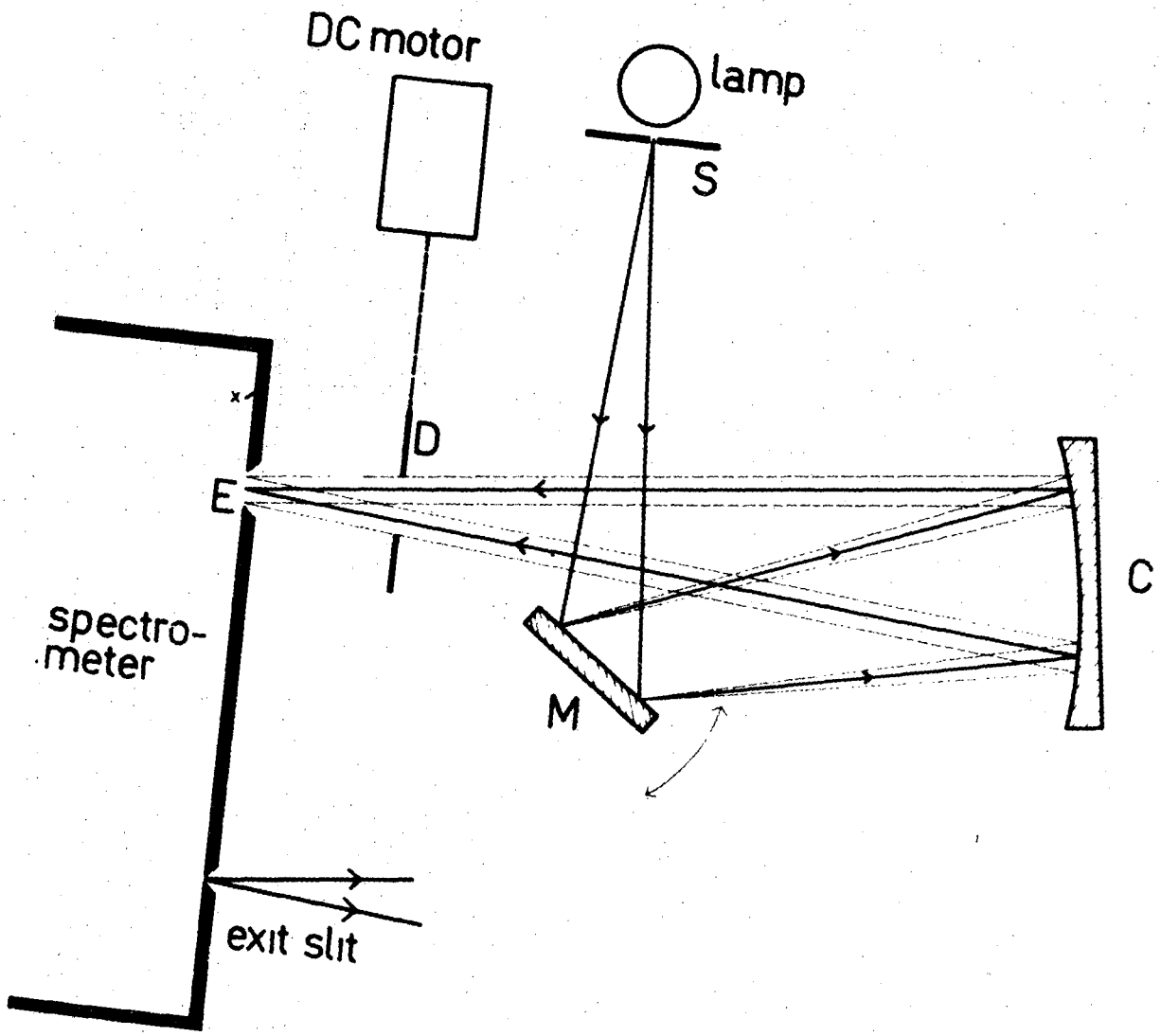


Fig. 2.



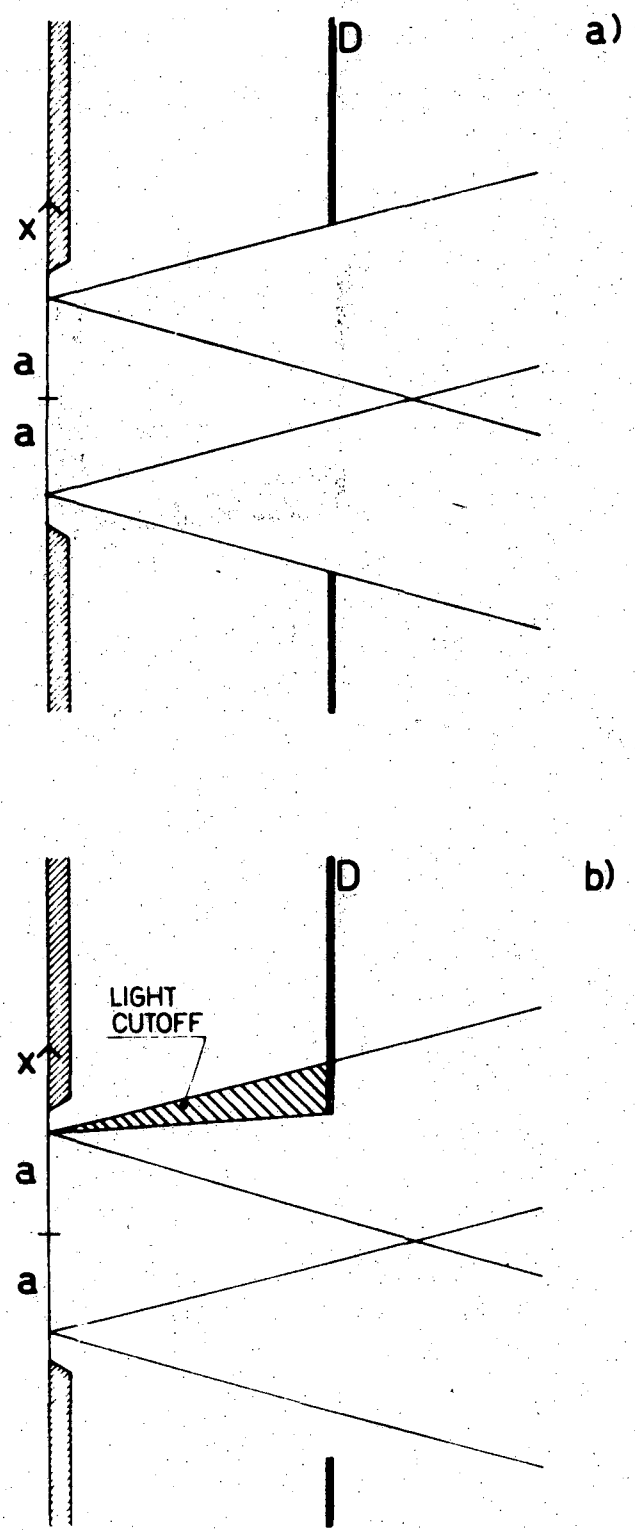
XBL 7011-6890

Fig. 3.



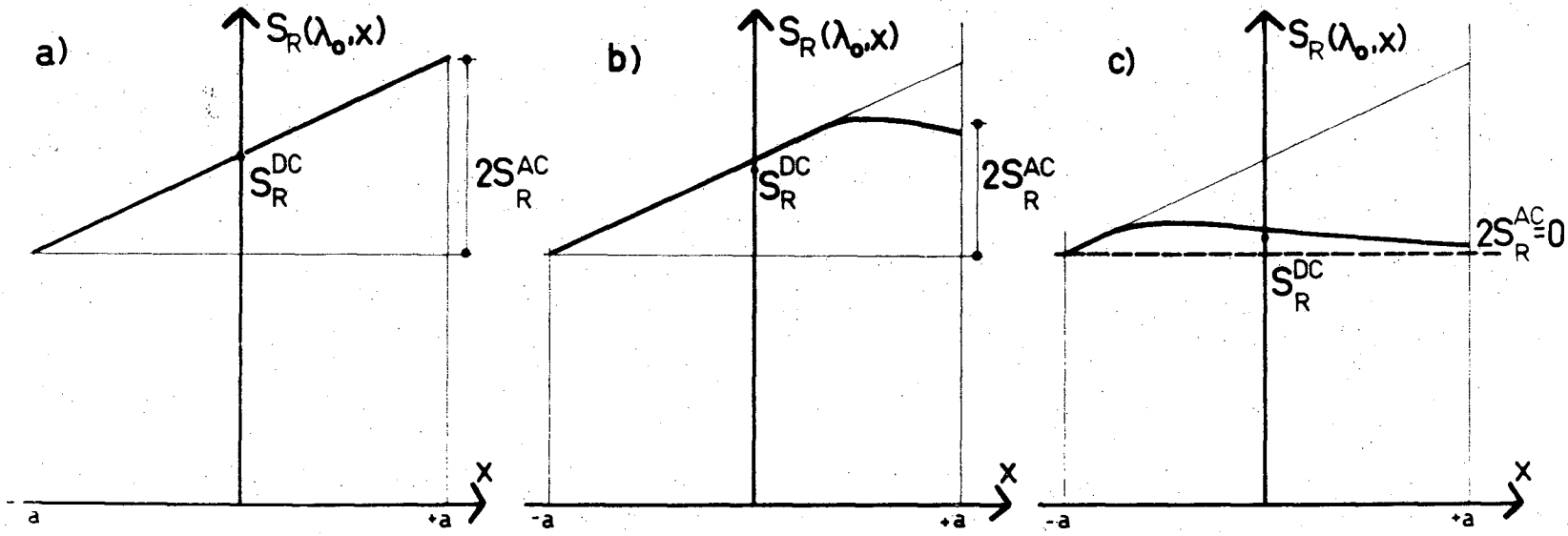
XBL 7011-6891

Fig. 4.



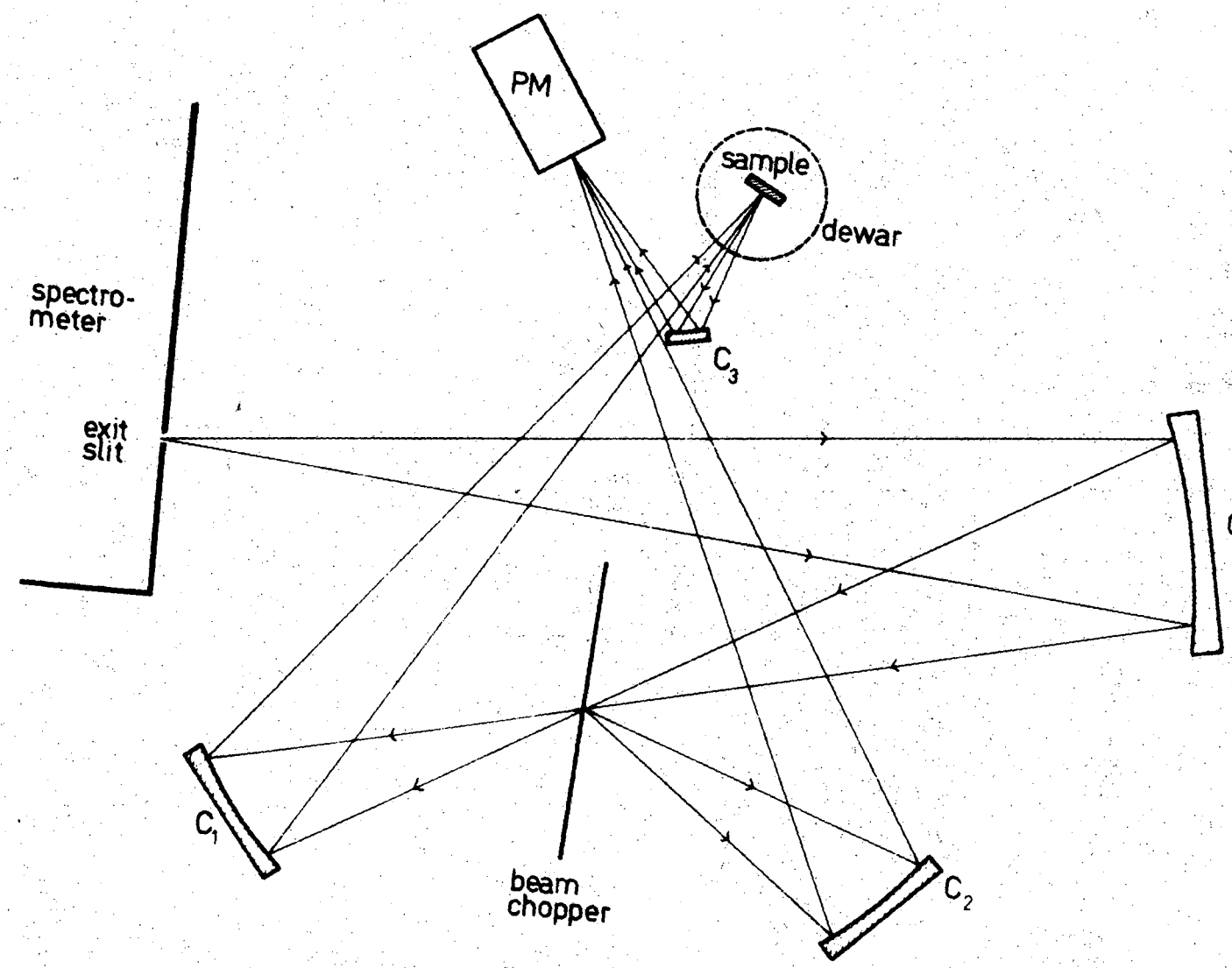
XBL 7011-6892

Fig. 5.



XBL 7011-6893

Fig. 6.



XBL 7011-6894

Fig. 7.

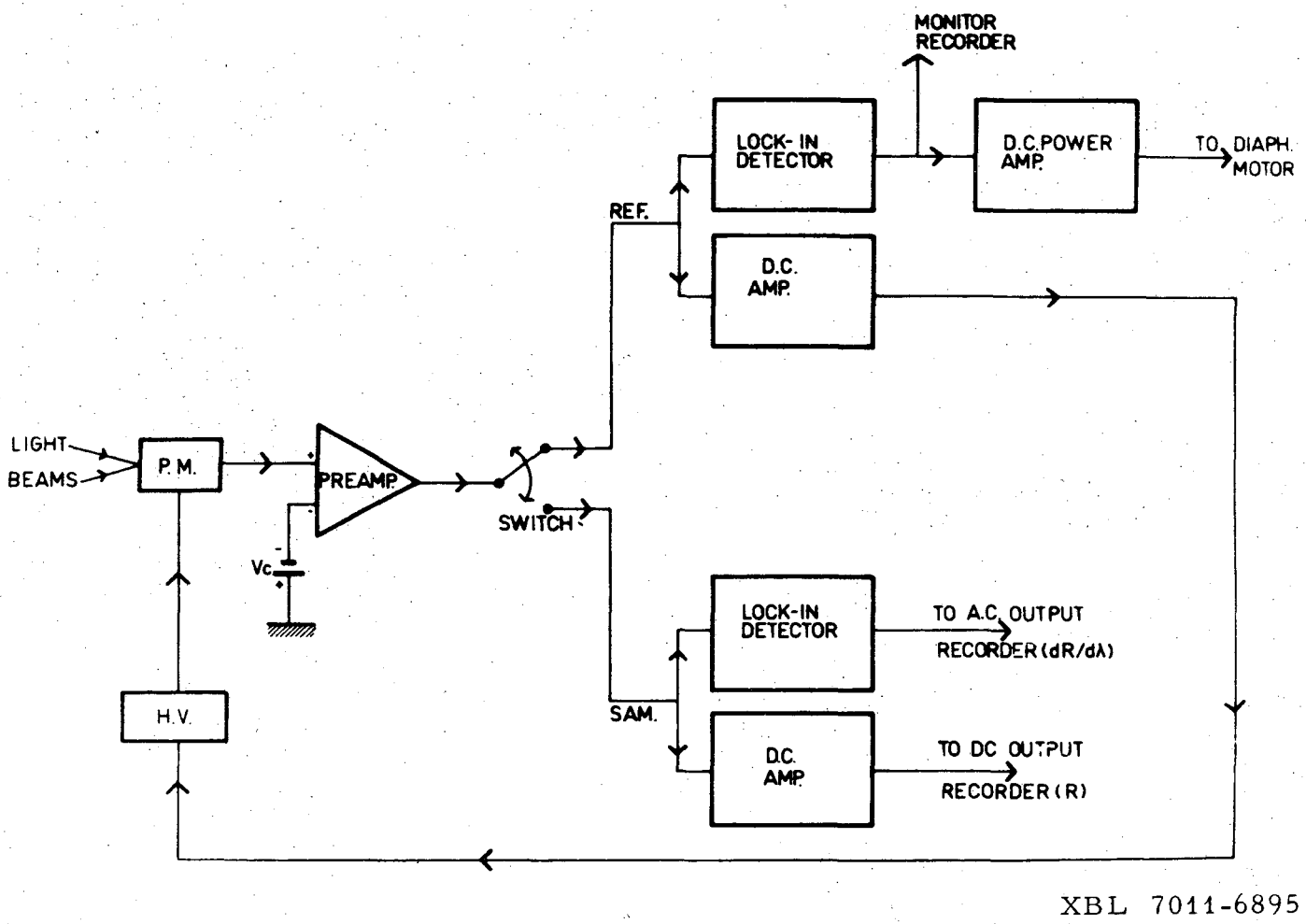
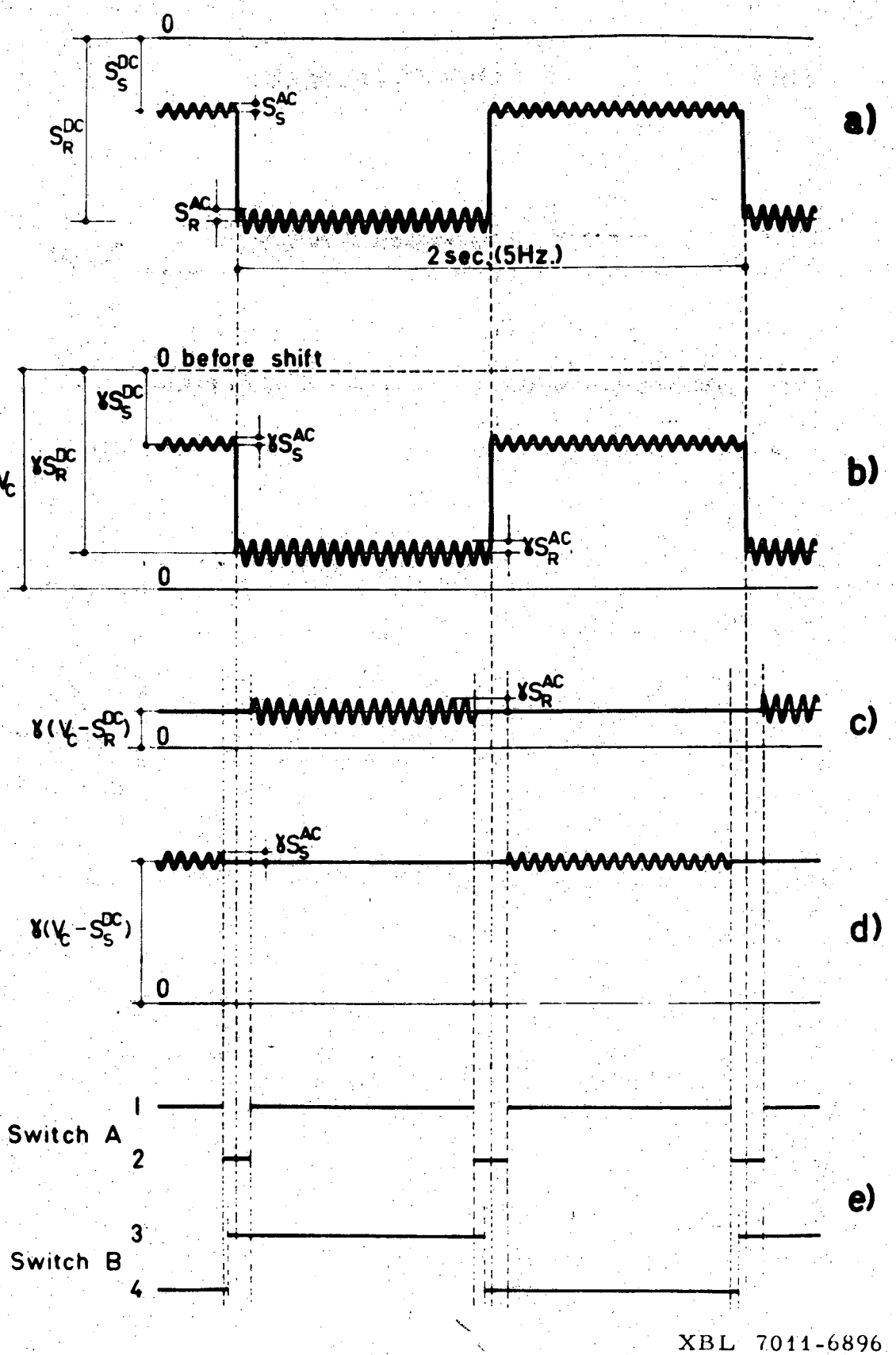
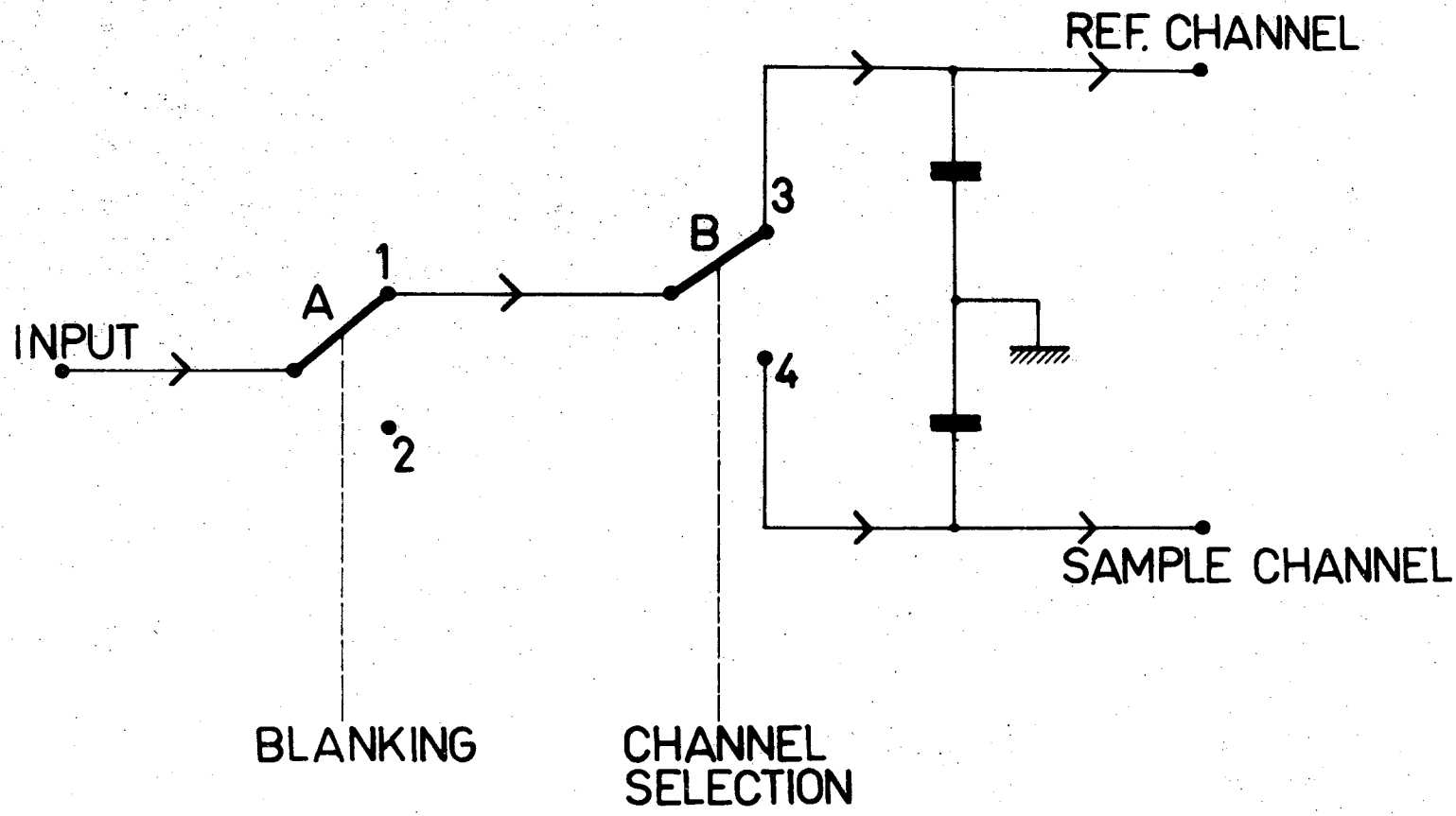

MONITOR
RECORDER
LOCK- IN
DETECTOR
D.C.POWER
AMP.
TO DIAPH.
MOTOR
REF.
D.C.
AMP.
LIGHT
BEAMS
P.M.
PREAMP.
SWITCH
Vc
LOCK-IN
DETECTOR
TO A.C. OUTPUT
RECORDER (dR/dλ)
H.V.
SAM.
D.C.
AMP.
TO DC OUTPUT
RECORDER (R)
XBL 7011-6895

Fig. 8.



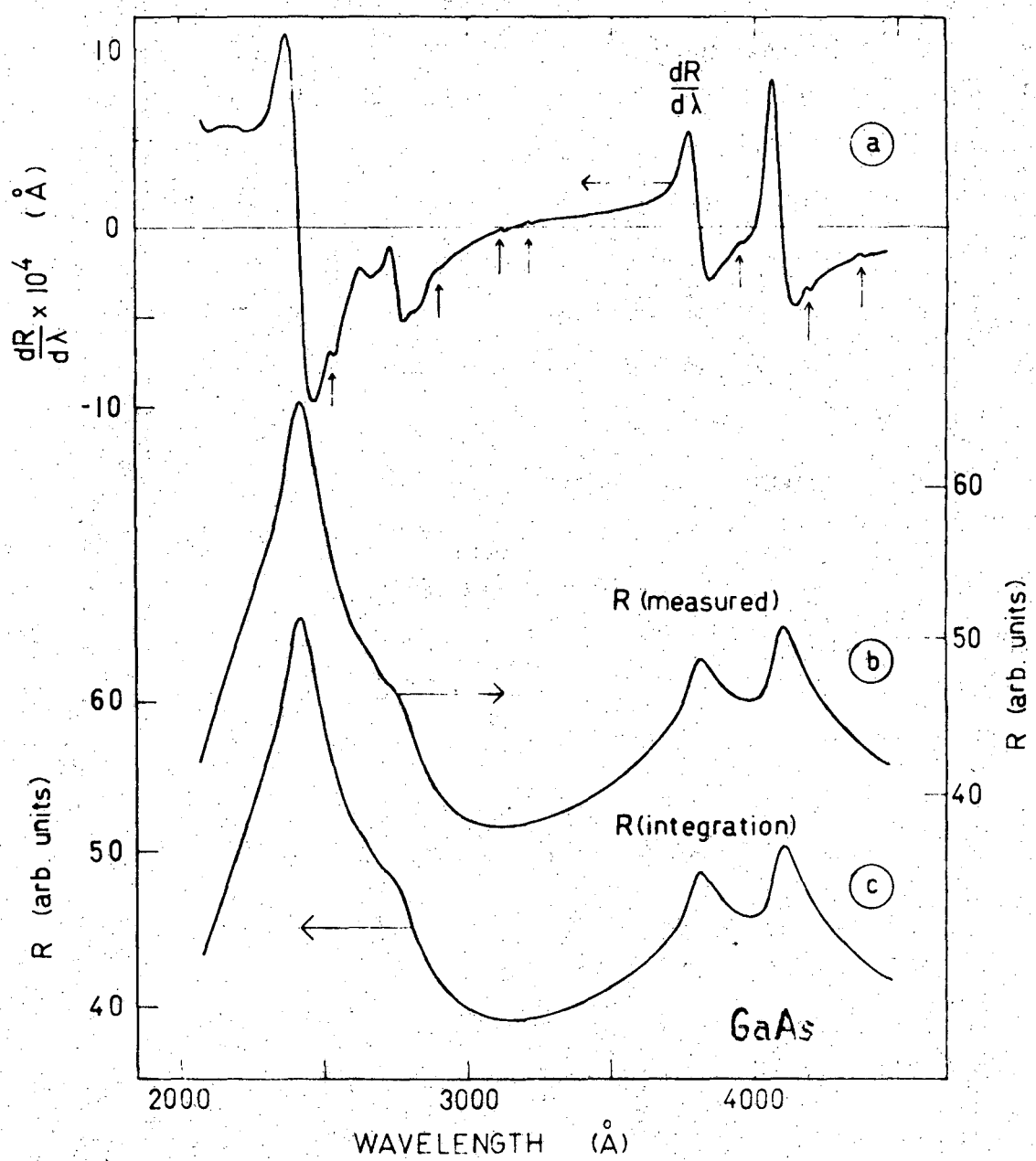
XBL 7011-6896

Fig. 9.



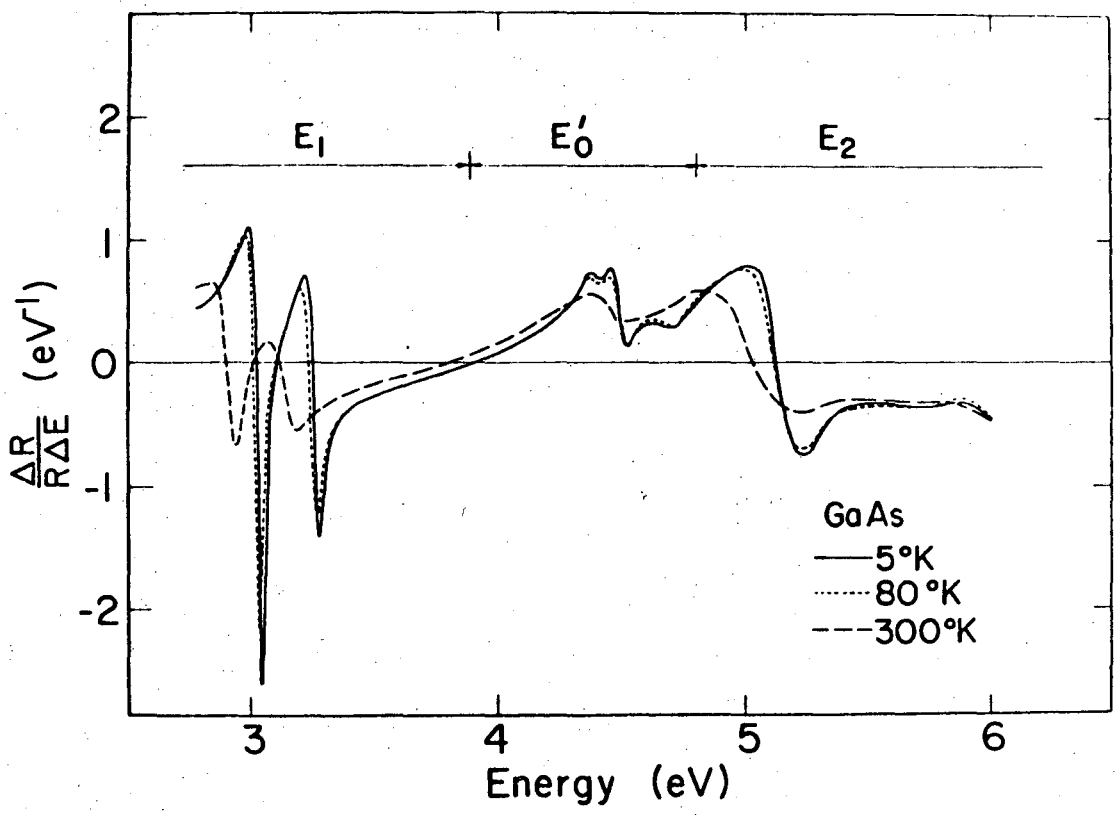
XBL 7011-6897

Fig. 10.



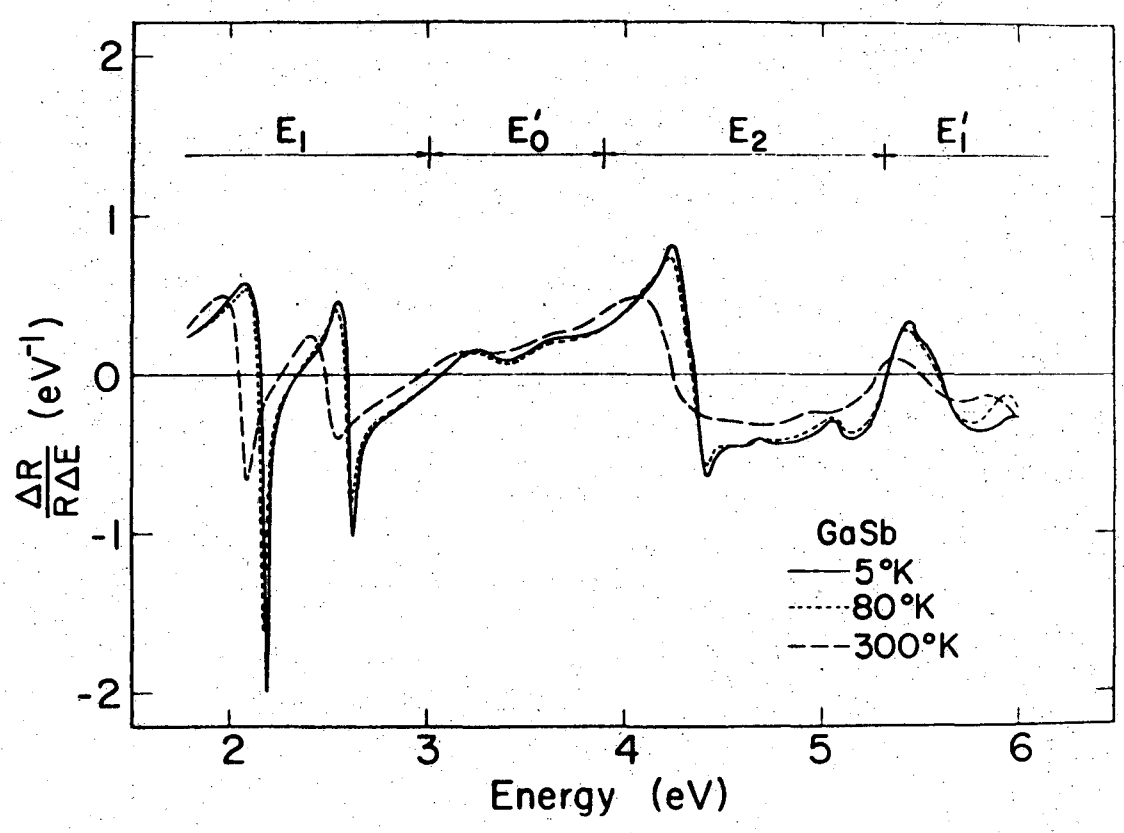
XBL 7011-6898

Fig. 11.



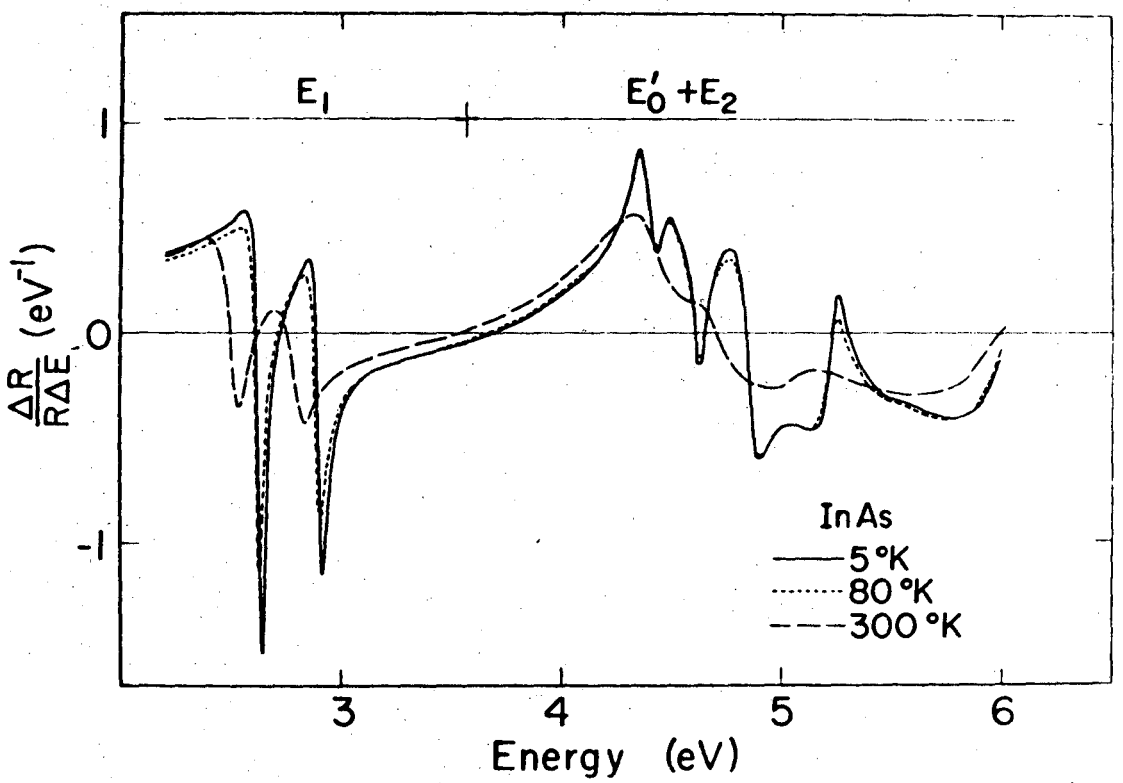
XBL 7011-6899

Fig. 12.



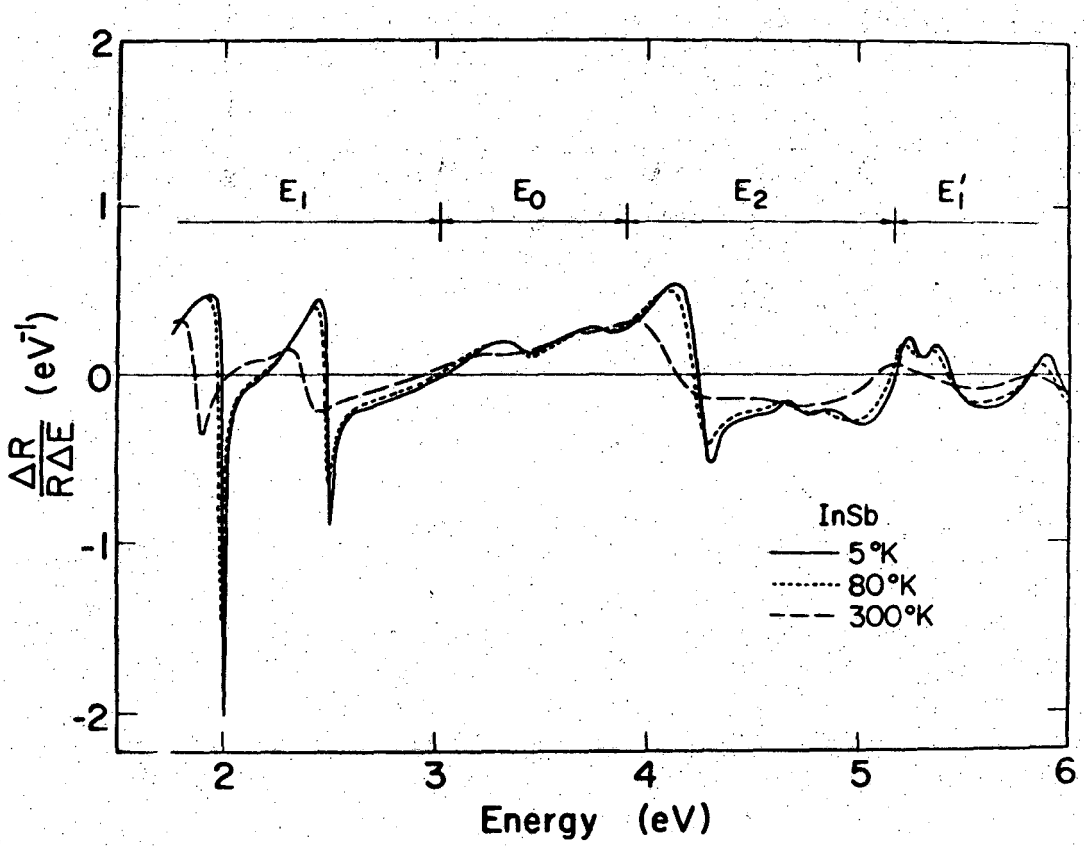
XBL 7011-6900

Fig. 13.



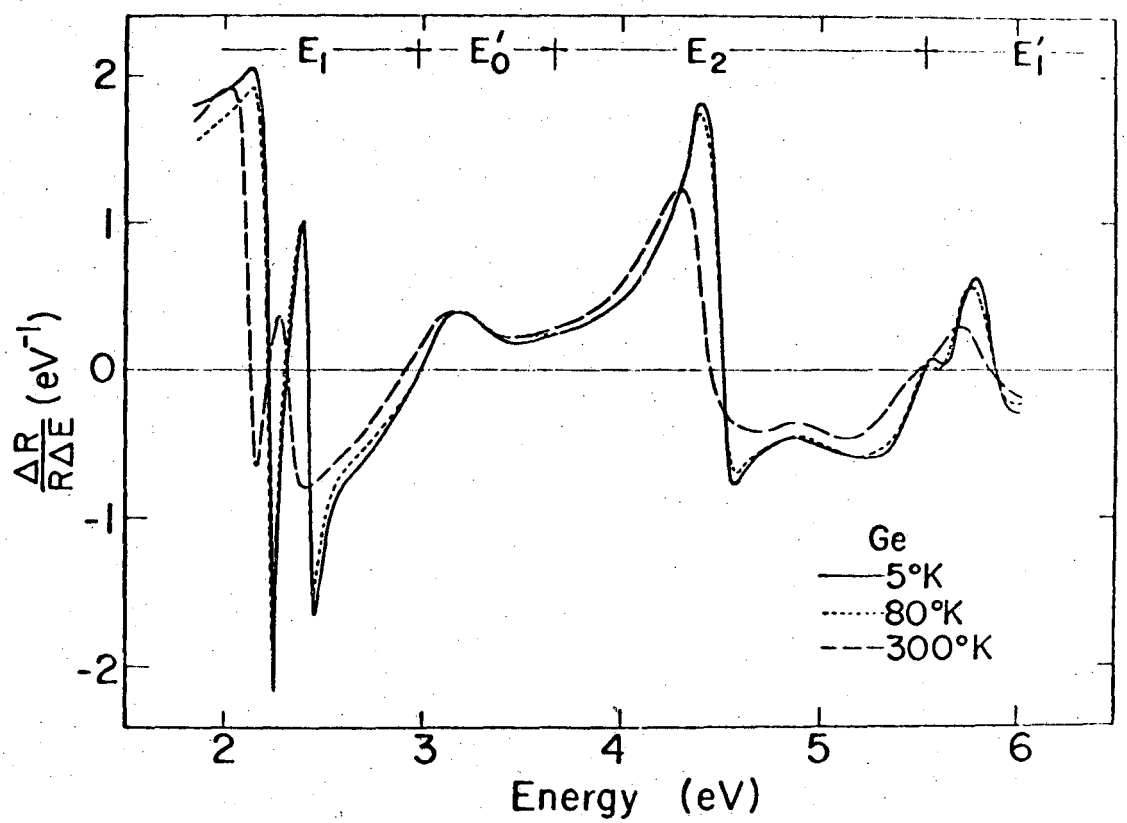
XBL 7011-6901

Fig. 14.



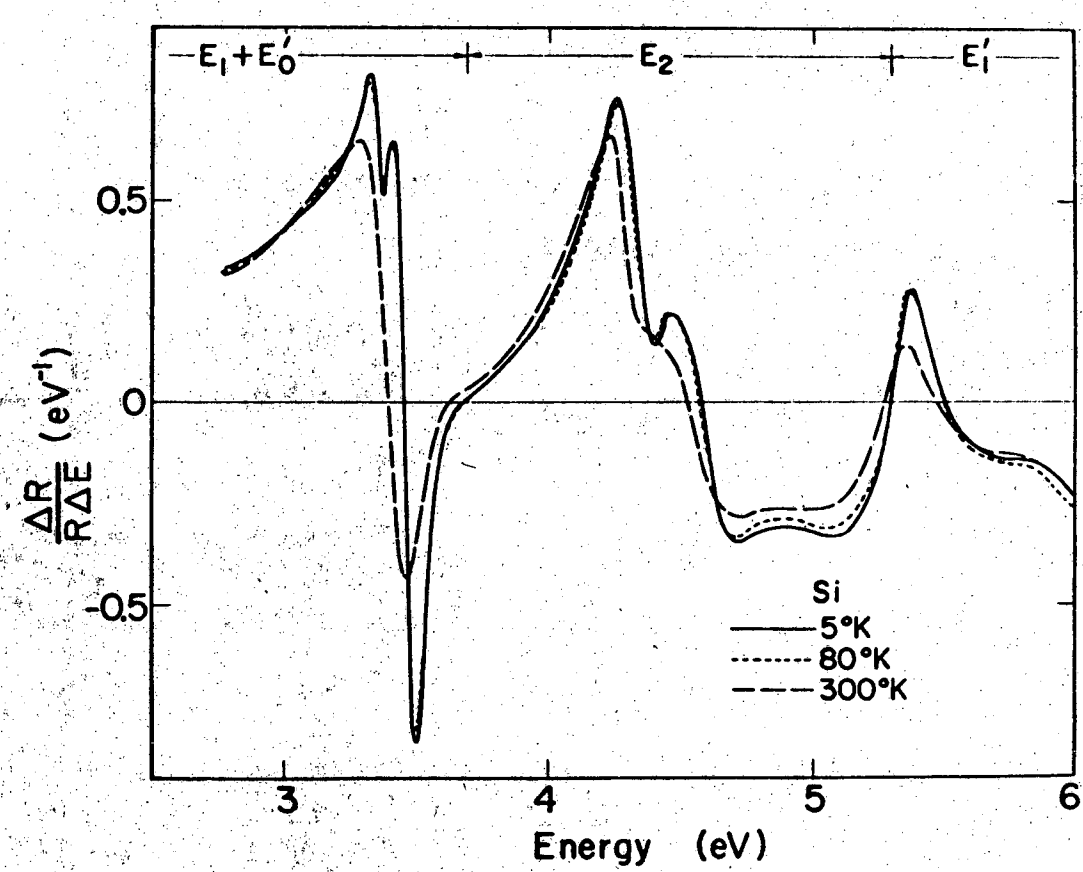
XBL 7011-6902

Fig. 15.



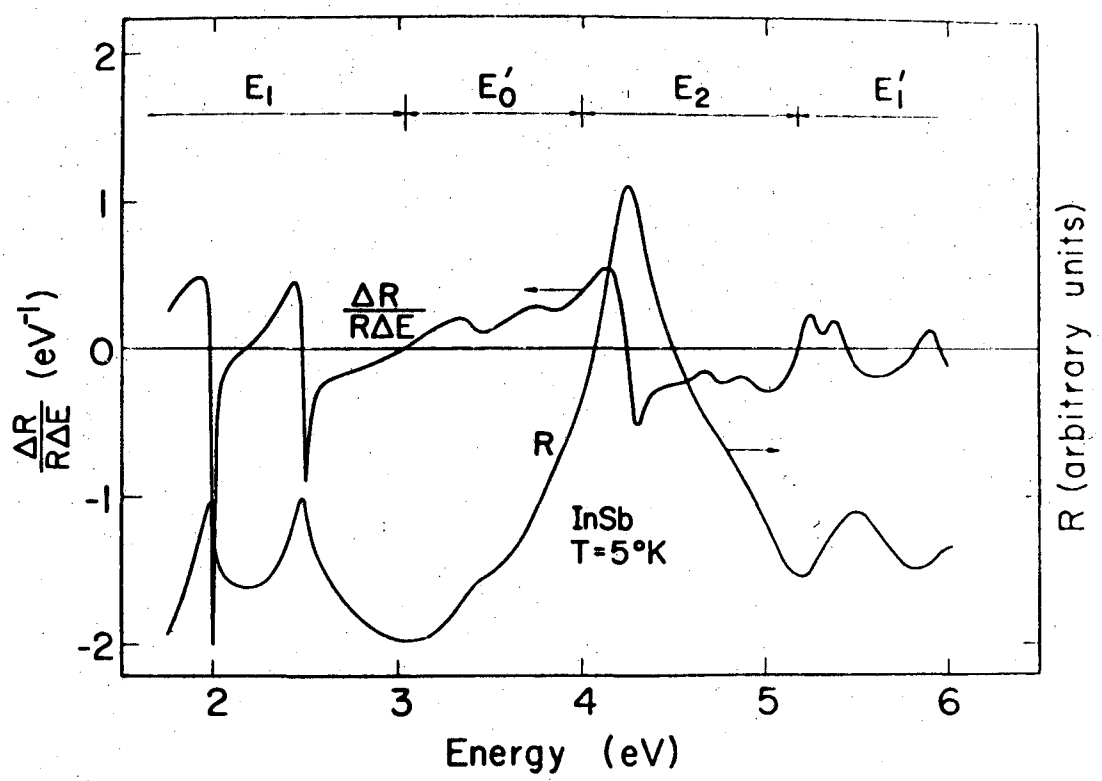
XBL 7011-6903

Fig. 16.



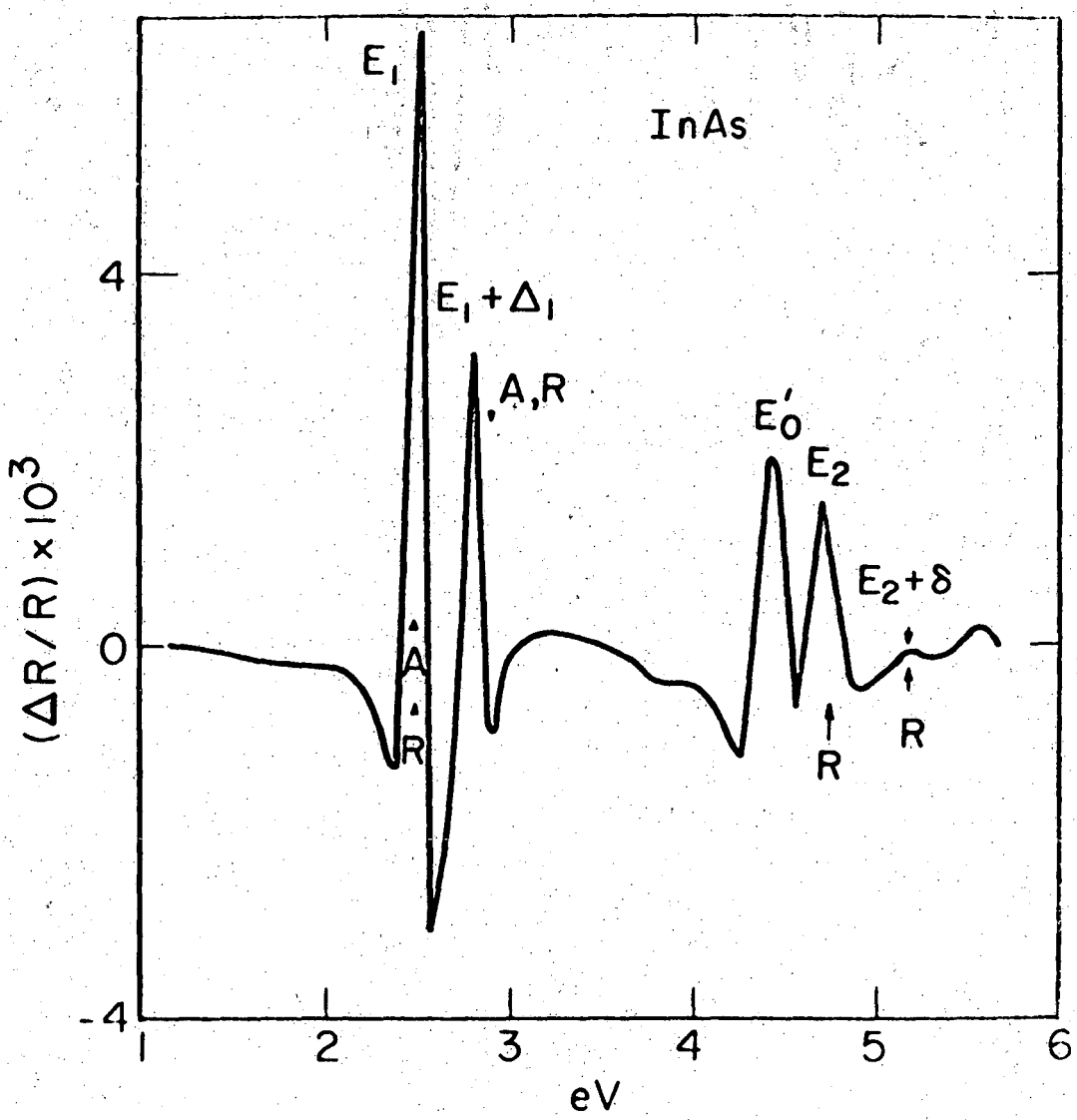
XBL 7011-6904

Fig. 17.



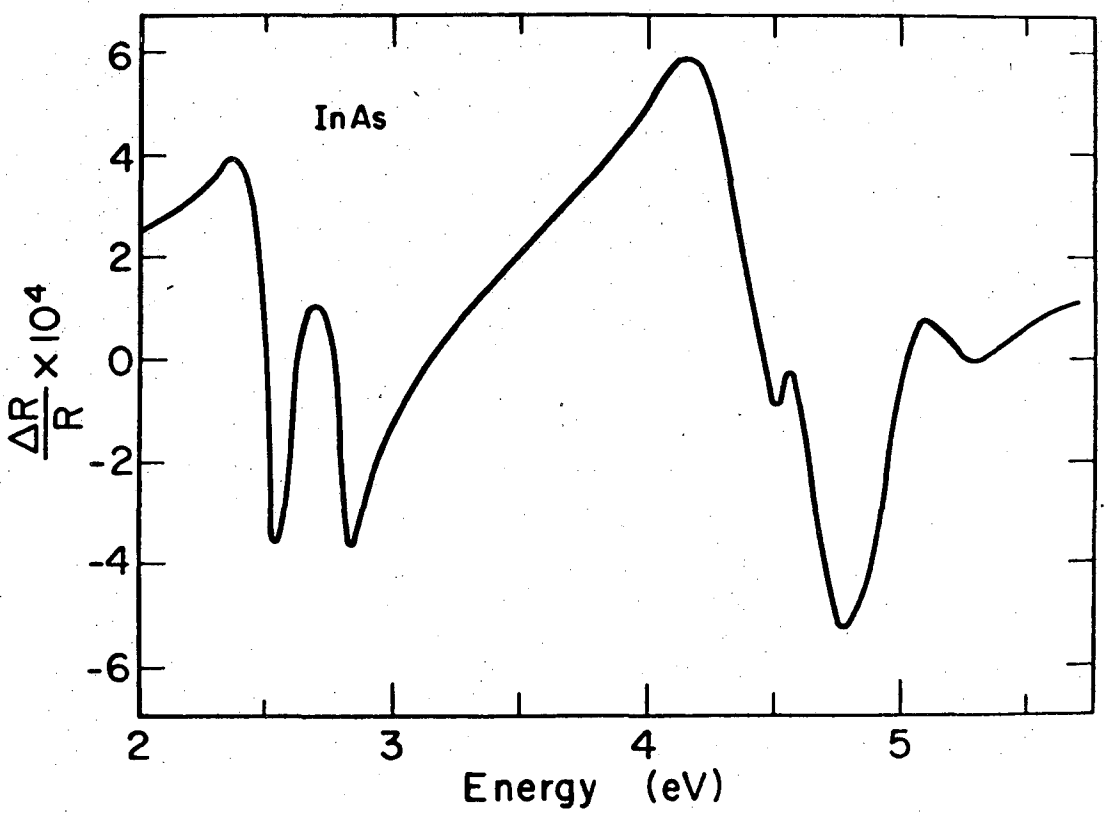
XBL 7011-6905

Fig. 18.



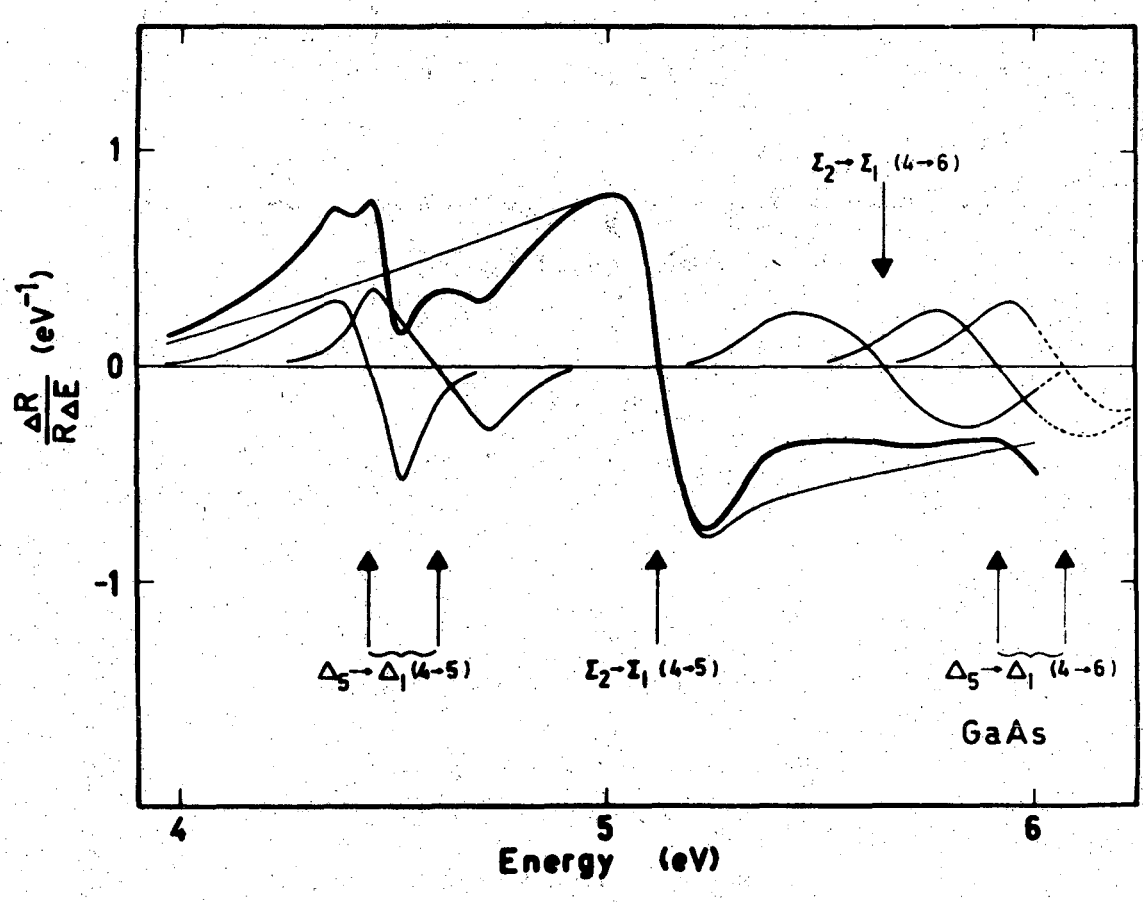
XBL 7011-6906

Fig. 19.



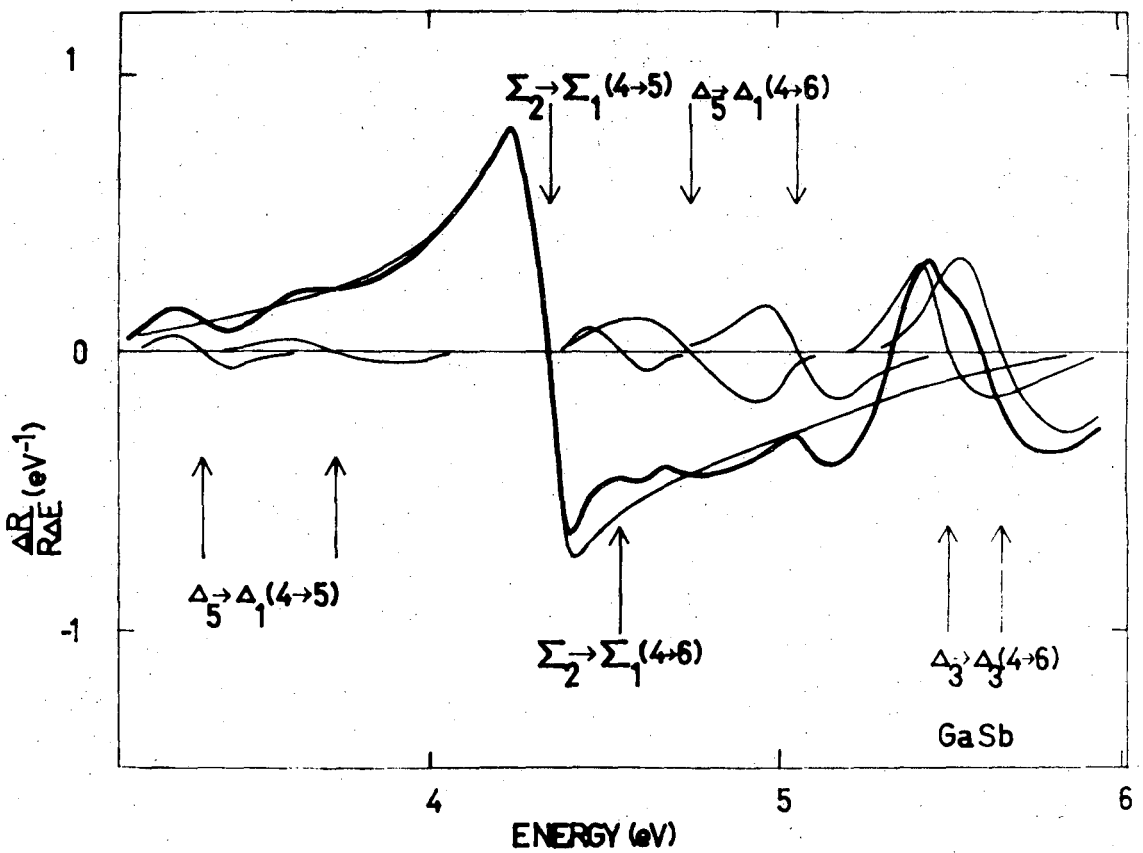
XBL 7011-6907

Fig. 20.



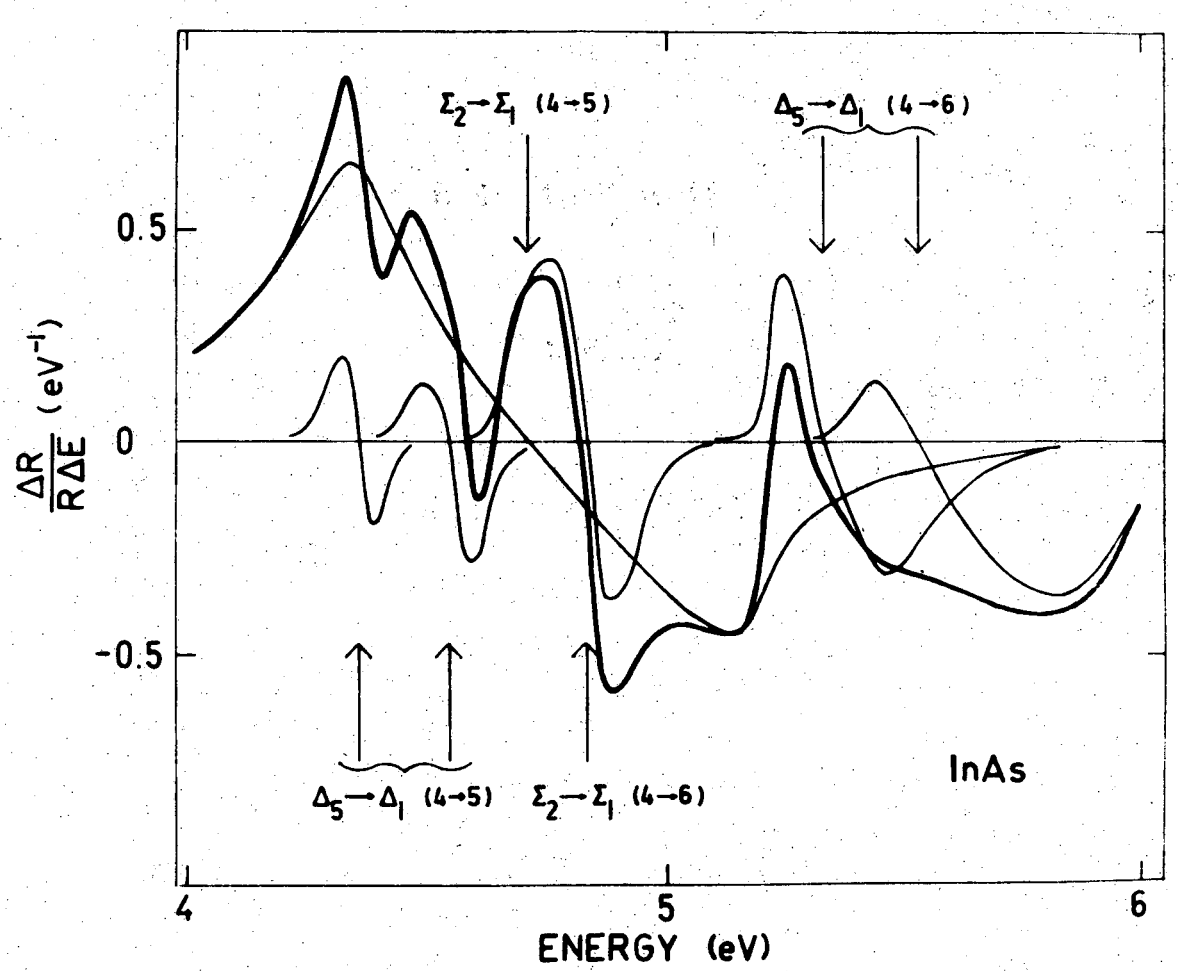
XBL 7011-6908

Fig. 21.



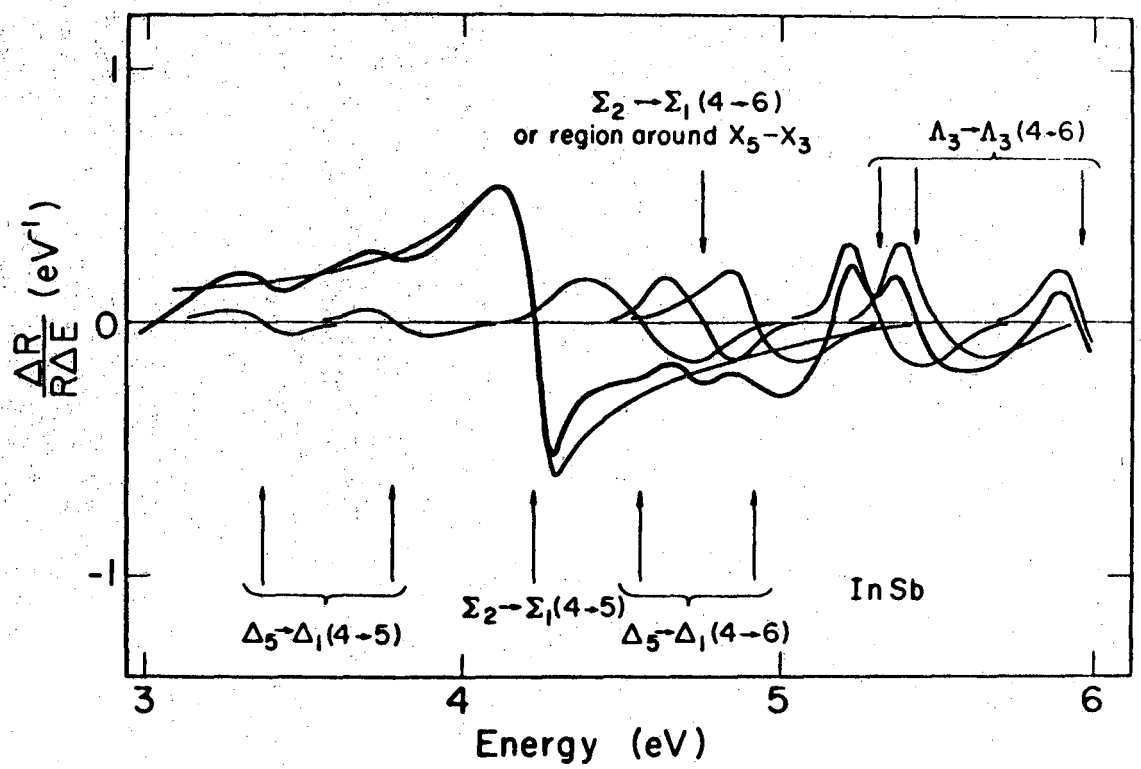
XBL 7011-6909

Fig. 22.



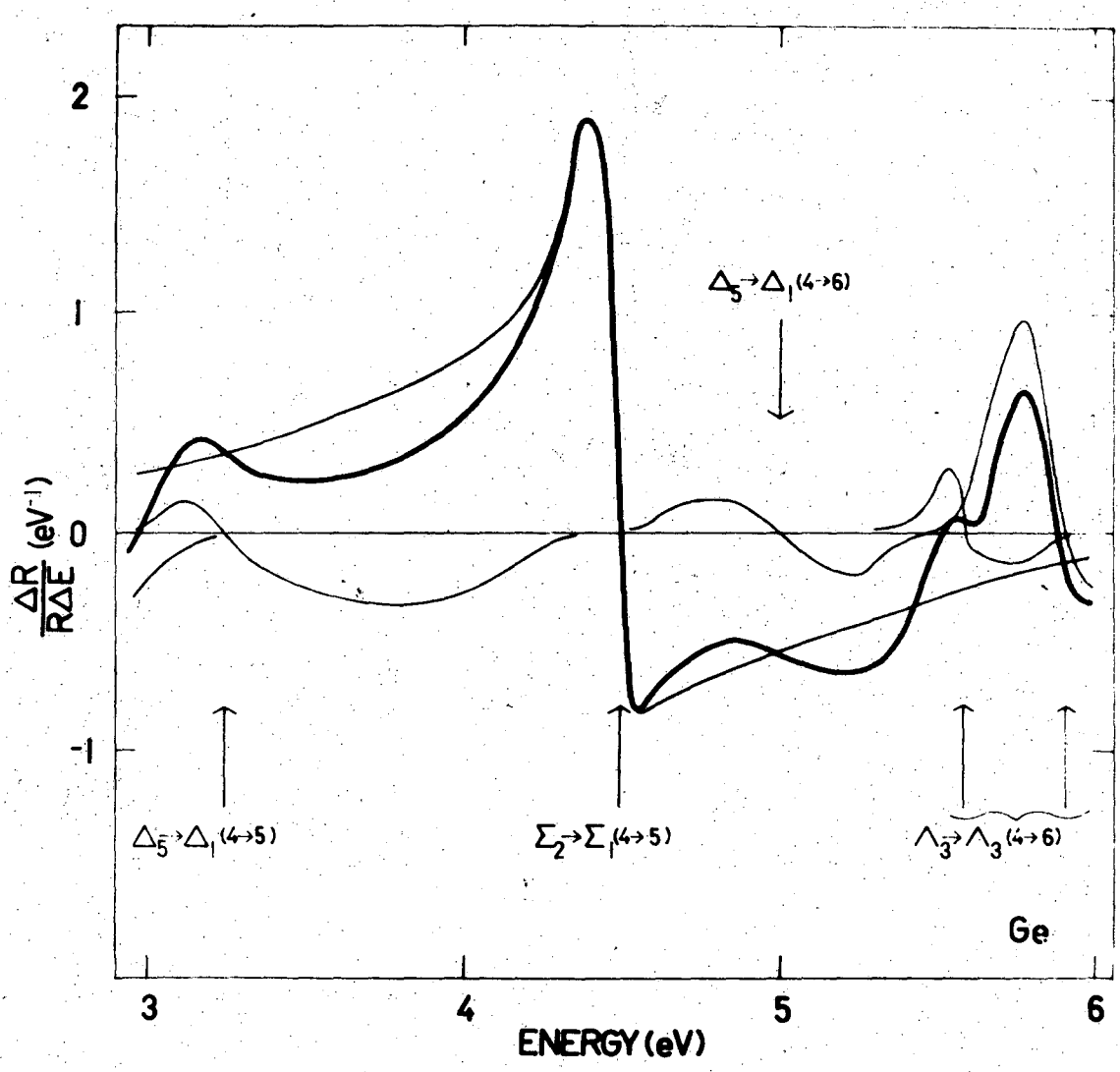
XBL 7011-6910

Fig. 23.



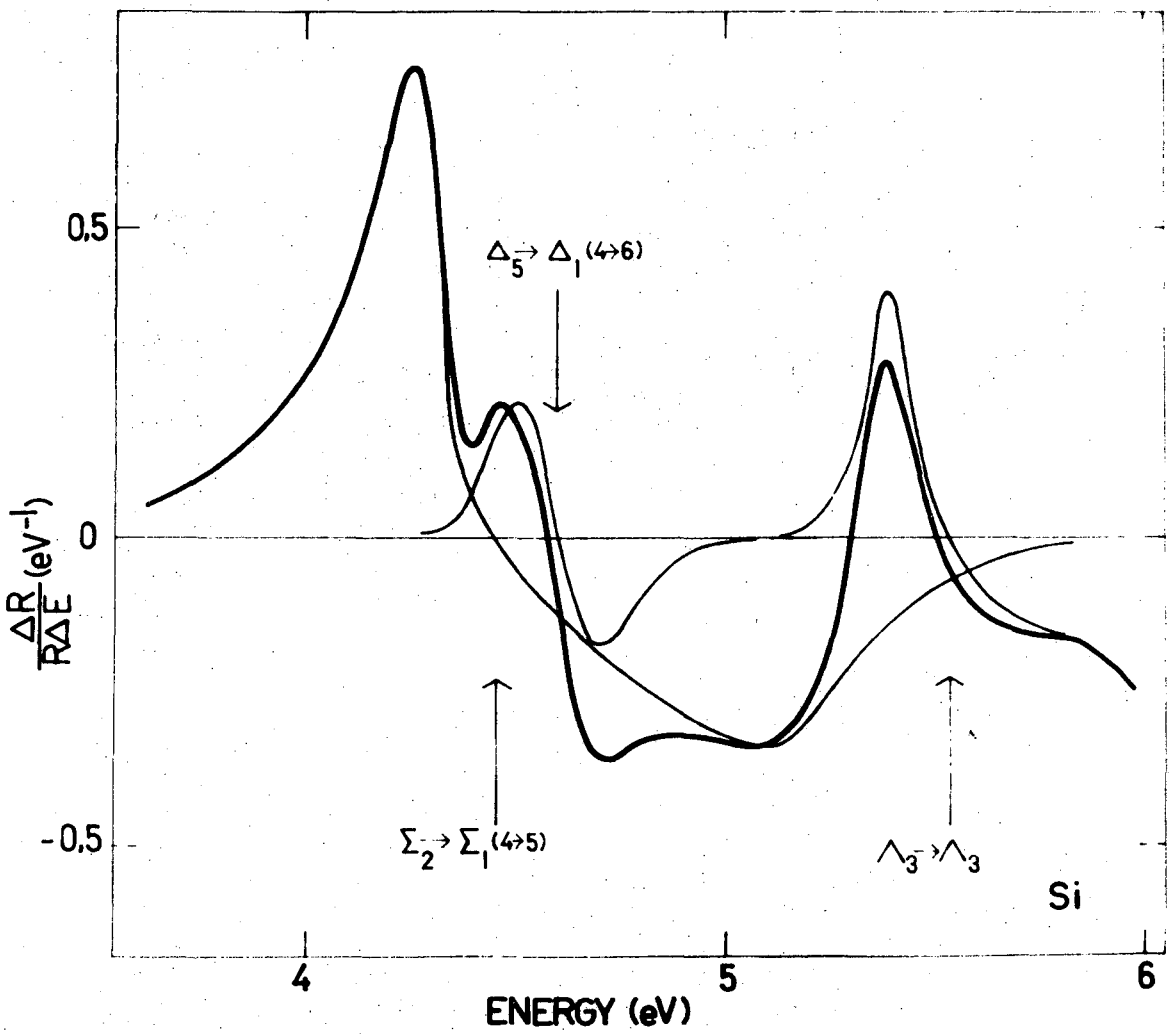
XBL 7011-6911

Fig. 24.



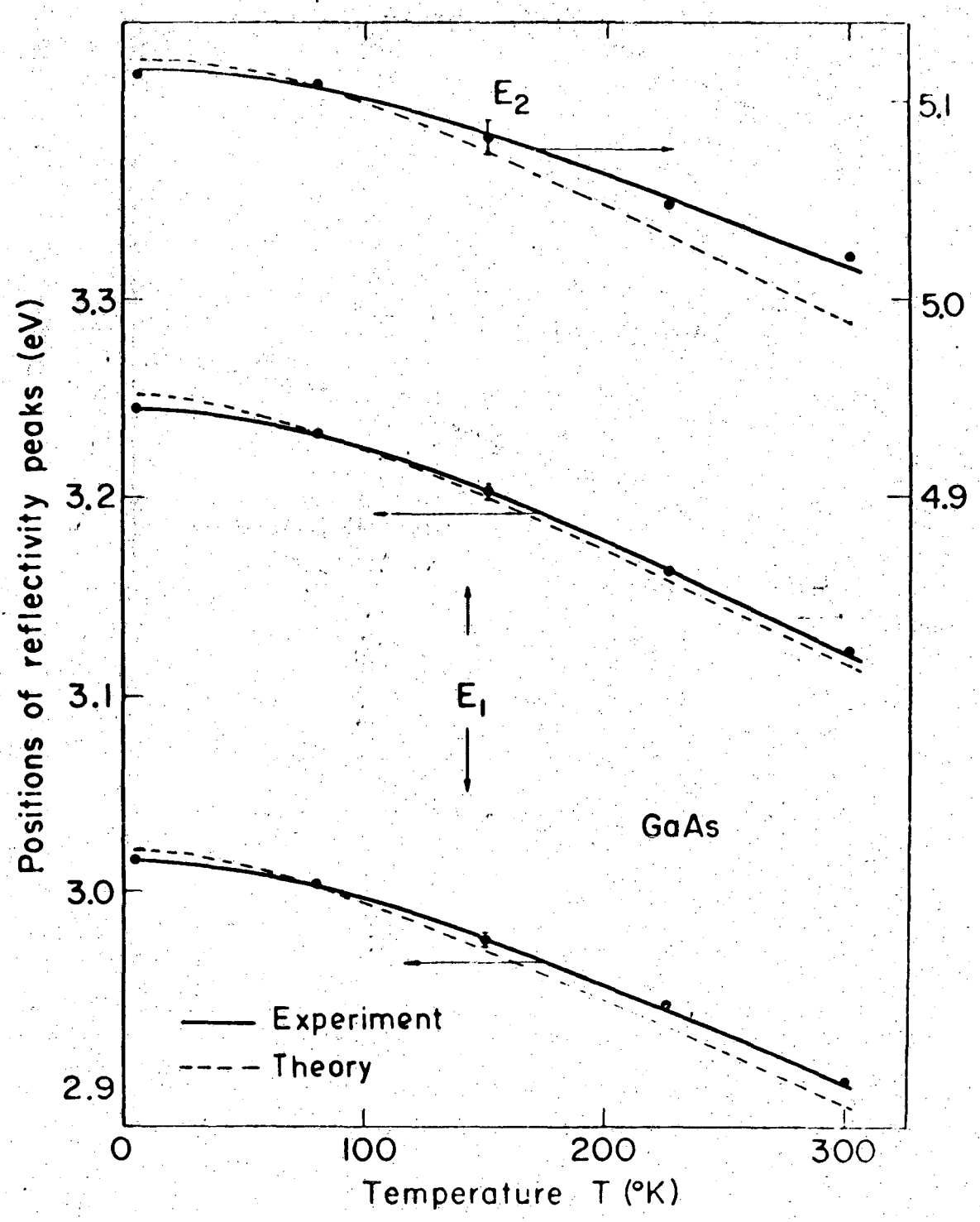
XBL 7011-6912

Fig. 25.



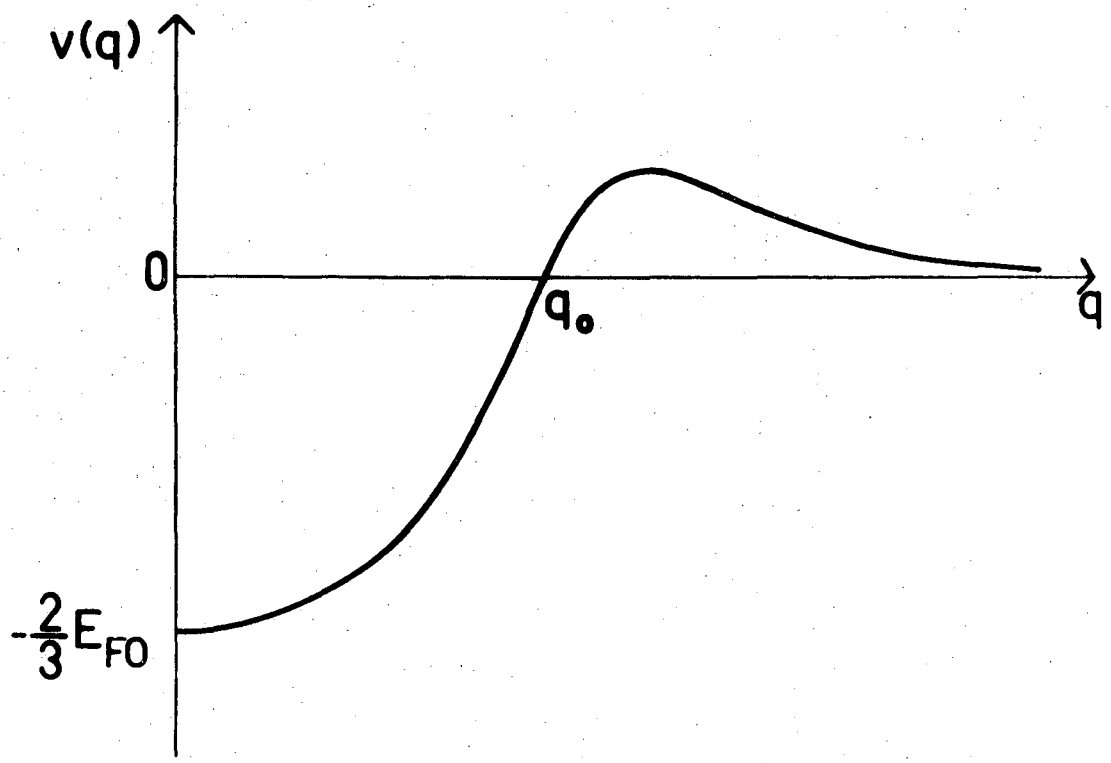
XBL 7011-6913

Fig. 26.



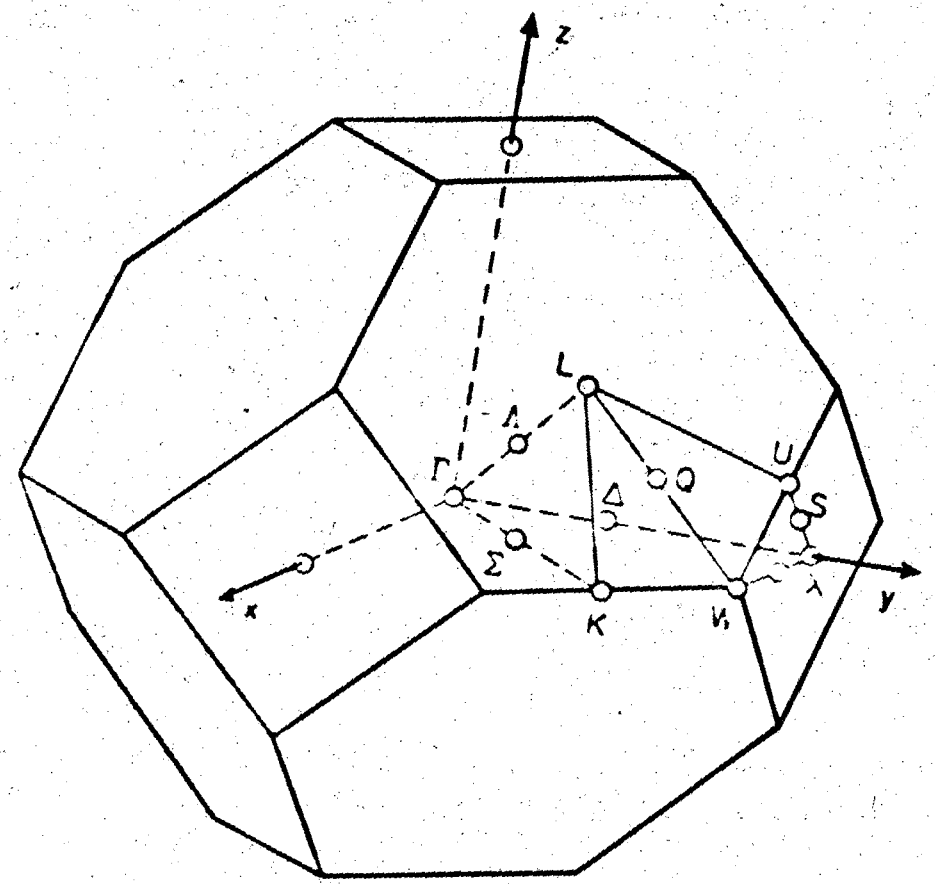
XBL 7011-6914

Fig. 27.



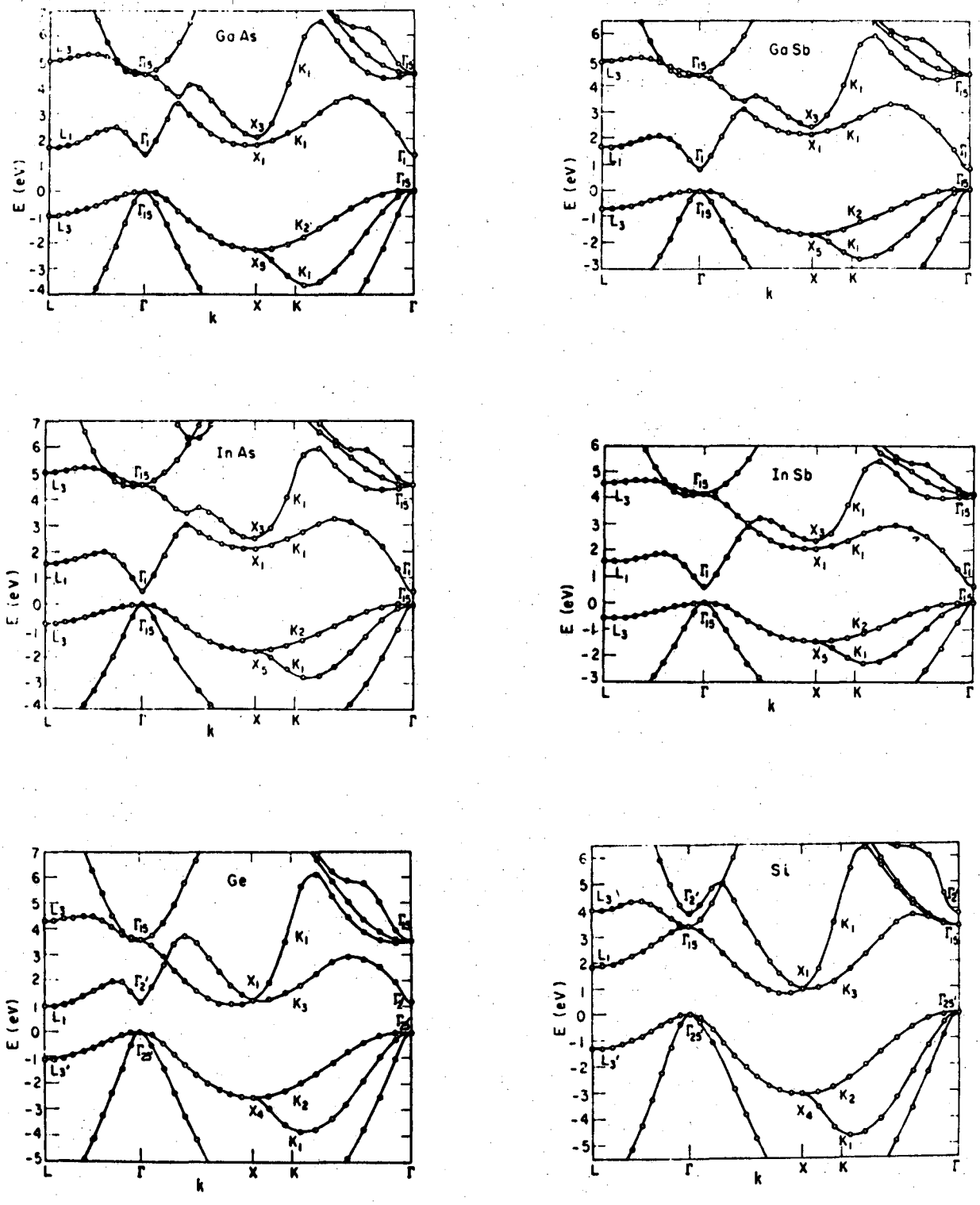
XBL 7011-6915

Fig. 28.



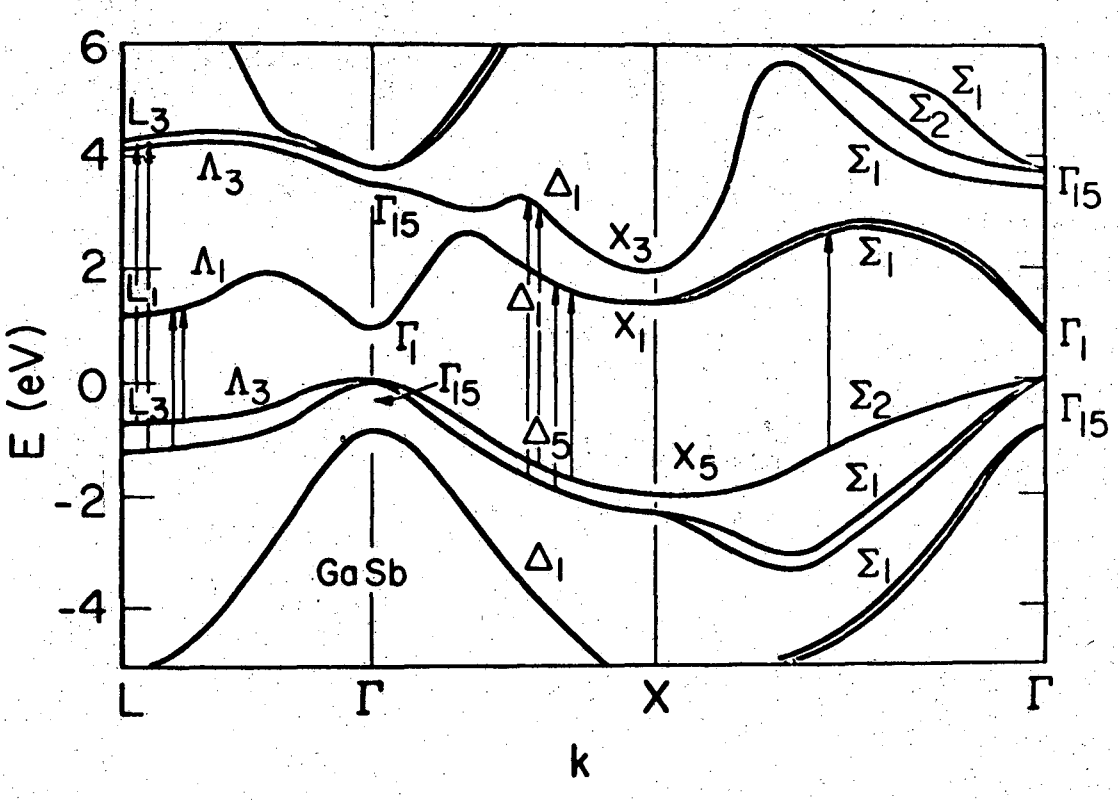
XBL 7011-6916

Fig. 29.



XBL 7011-6917

Fig. 30.



XBL 7011-6918

Fig. 31.

TECHNICAL INFORMATION DIVISION
LAWRENCE RADIATION LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720