Optical Field Effect in Silicon

B. O. SERAPHIN*

Michelson Laboratory, China Lake, California

(Received 7 July 1965)

Structure in the reflectance is related to interband transitions at critical points and therefore provides a major source of information in the analysis of the band structure. This paper describes a derivative technique which enhances this structure considerably. The density of states at the critical points is modulated through an ac electric field at the reflecting surface and the resulting change in the reflectance is amplified phase-sensitively. At a photon energy of 3.4 eV, three peaks are resolved in silicon within only 0.1 eV. Their dependence upon temperature and orientation of the reflecting surface suggests an assignment to a transition at the Γ point for the peak at 3.34 eV and at the L point for the peak at 3.45 eV, with an excitonic peak in between. Structure in the reflectance response at 4.20 eV is correlated to the X_4-X_1 transition. Since the effect is caused by a modulation of the surface potential, it is possible to draw conclusions about the parameters of the surface from the response of the reflectance change to temperature, dc bias, and modulation frequency. An optical technique is outlined which will complement the electrical field-effect technique.

INTRODUCTION

FURTHER progress in the analysis of the band structure of semiconductors from optical measurements above the fundamental absorption edge will depend upon a refinement of these measurements with respect to resolution in particular. This, however, is difficult because of the large background absorption in this range of photon energies. The fundamental absorption edge itself, as the first true interband transition on the scale of increasing photon energies, is observed clearly against a preceding region of transparency. It cannot be missed by even the crudest optical measurement. Above this threshold, however, any structure in the optical properties of the material must be detected against the background of very strong absorption. This handicap is of particular importance in cases such as silicon where the onset of the first direct transition occurs more than 2 eV above the threshold for indirect transitions. The related structure must be observed on top of an absorption coefficient, which is already greater than 10^5 cm^{-1}. As a result, it required considerable refinement of the experimental technique over more than a decade before the direct threshold in silicon was identified optically beyond any doubt.

A further development seems the more desirable since experiment recently fell behind theoretical capabilities in this field. The original band-structure calculations from first principles produced the critical points in the band structure with an accuracy of approximately 1 eV. Reflectance measurements were capable of resolving structure and checking on the theoretical predictions to within 0.1 eV. The development of interpolation schemes like the pseudopotential method or the **k·p** method, however, which calculate the band structure over the entire Brillouin zone from an experimental knowledge of the energy gaps at a few symmetry points, has changed the picture in favor of theory. Information on these key parameters to better than 0.01 eV could be

handled by this method with advantage, its present limitation being the lack of resolution in the experiment.

Progress could be made if the present "static"methods were replaced by a derivative technique. Any effect which changes the edge of higher interband transitions only, modulating the density-of-states function in the vicinity of critical points, for instance, is inherently superior in sensitivity and resolution. Produced periodically and observed in reflection, for instance, the constant component of the reflectance can be suppressed and the modulated part amplified considerably.

An effect of this type, caused by an electric field and sharply localized at the fundamental absorption edge, was predicted by Franz and Keldysh in 1958.[1] The effect was confirmed experimentally on several materials,[2] and derivative techniques discovered structure related to the assistance of phonons to the indirect transitions at the bottom of the fundamental absorption edge.[3] All of these experiments operated in the absorption mode, measuring the loss of transmittance caused by the shift of the edge in the presence of an electric field. Since all semiconductors are highly opaque above the fundamental edge and since the field-induced changes in the absorption coefficient are too small to be seen in reflection, no attempt was made to see Franz-Keldysh effects on the edges of higher interband transitions. Recent experiments on Ge,[4] Si,[5] and GaAs[6] have

* Present address: U. S. Office of Naval Research, Keysign House, 429 Oxford St., London, W.1, England.

[1] W. Franz, Z. Naturforsch. **13a**, 484 (1958); W. L. Keldysh, Zh. Eksperim. i Teor. Fiz. **34**, 1138 (1958) [English transl.: Soviet Phys.—JETP **7**, 788 (1958)].

[2] T. S. Moss, J. Appl. Phys. **32**, 2136 (1962); R. Williams, Phys. Rev. **126**, 442 (1962).

[3] M. Chester and P. H. Wendland, Phys. Rev. Letters **13**, 193 (1964); A. Frova and P. Handler, Phys. Rev. Letters **14**, 178 (1965).

[4] B. O. Seraphin, *Proceedings of the International Conference on the Physics of Semiconductors, Paris, 1964* (Dunod Cie., Paris, 1964), p. 165; B. O. Seraphin, R. B. Hess, and N. Bottka, J. Appl. Phys. **36**, 2242 (1965).

[5] B. O. Seraphin and N. Bottka, Phys. Rev. Letters **15**, 104 (1965).

[6] B. O. Seraphin and N. Bottka, Bull. Am. Phys. Soc., **10**, 593 (1965); Appl. Phys. Letters **6**, 134 (1965). B. O. Seraphin, Proc. Phys. Soc. (London), (to be published); J. Appl. Phys. (to be published).

shown, however, that the electric fields in the surface potential barrier are strong enough to cause a Franz-Keldysh effect in the surface region, which can be observed as a change in the reflectance caused by a change in the surface potential. This does not invalidate the previous argument that the field-induced change in the absorption coefficient itself is too small to be detected in reflection. If the change is used in the Kramers-Kronig dispersion relation, however, the correlated change in the refractive index is large enough to interpret the experimentally observed field effect of the reflectance for the fundamental edge of germanium.[7] With the experiment operating in the reflection mode, the higher interband transitions are now open to the investigation. Since the electric field changes the density of states effectively in the neighborhood of the critical points only, the effect provides the desired derivative technique for the precise determination of these critical points. A pronounced structure of peaks and dips in the reflectance response is observed on the wavelength scale, which not only measures the energy of the interband edges precisely, but provides clues as to the nature of the critical point at the same time. It was already pointed out early in these investigations that the character of the reflectance response discriminates among the different types of interband transitions, thereby introducing a new parameter into band structure analysis.[8] Phillips' duality theorem confirmed that the field-induced change in the density of states at a saddle-point edge evolves from the correlated change at a threshold edge by a simple symmetry operation.[9] This cleared the way to an interpretation of the optical-field effect in terms of the density of states, and conclusions were drawn in previous papers as to the band structure of silicon.[5,10]

This paper presents an account of the experimental technique employed for the measurements on silicon. It presents results, for material of both conductivity types, which are markedly different in silicon, in contrast to germanium. It further reports measurements on surfaces of different crystalline orientation.

Since the reflectance change in this method is caused by a change in the electronic situation inside the semiconductor surface, it must, in turn, be possible to draw conclusions as to the nature of this electronic situation from the particular type of response. The information on which the present model of the semiconductor surface is based was derived mainly from electrical measurements. This imposed restrictions which permitted a wealth of measurements to be made on germanium and silicon, but produced only a few papers on GaAs and

almost nothing on the rest of the semiconductors. The marginal magnitude of surface effects explains this situation, if high resistivity and the tendency to form rectifying contacts are taken into account. It is pointed out in this paper how some features of the field effect of the reflectance, observed in particular at low temperatures, indicate the possibility of determining the parameters of the surface strictly optically. Such an optical technique would alleviate some of the problems encountered in making field-effect measurements on most of the semiconductors.

EXPERIMENTAL PROCEDURE

Most of the measurements were taken on 100-Ω cm p-type silicon, with the reflection taking place at the (111) face. After it was noticed that, in contrast to germanium, the conductivity type is of importance, the investigation was extended to n-type material and reflection was studied from (110) and (100) faces as well. In all cases the largest effect was found for material in the range from approximately 10 to 100 Ω cm. This is plausible since the effect depends upon formation and control of a space-charge layer by means of an external electric field, which is easiest for a background carrier density neither too high nor too low.

The samples were cut into wafers of 10$\times$10$\times$2-mm size and polished, using $\frac{1}{4}$-μ diamond abrasive as the last step. The reflecting face was etched, applying either "white etch" for 5 sec or CP 4-A for 15 sec. The first etch produced a strong n-type layer on the surface, while the second etch placed the surface on the n-side, too, but closer to the flat-band position. The final point of operation was adjusted by dc bias at low temperatures. This permits the superimposed ac modulating field to operate around different values of surface potential, swinging the bands, and therefore the internal electric field, around different center positions.

The importance of mounting the sample properly was pointed out previously.[4] The configuration of the parallel-plate condenser, which applies the modulating field perpendicular to the reflecting surface by means of a transparent field electrode, represents an optical interferometer at the same time. Even cementing the components of the sandwich rigidly together does not prevent the Coulomb forces from changing the sample-electrode separation. Unwanted reflectance variations are produced, which appear at the signal frequency if dc bias is applied at the same time. Besides these "interference fringes," improper mounting leads to "background" in the regions between peaks. Such background has no physical significance, depends upon the rigidity of the mounting, and inverts phase with the direction of the bias, thereby indicating its character as a mechanical vibration. It may be impossible to prevent vibration completely. But the problem can nevertheless be solved by matching the system optically, so that the strongest reflection comes from the

[7] B. O. Seraphin and N. Bottka, Phys. Rev. 139, A560 (1965).
[8] B. O. Seraphin and R. B. Hess, Phys. Rev. Letters 14, 138 (1965).
[9] J. C. Phillips, Proceedings of the International School of Physics "Enrico Fermi," Varenna Lectures, 1965 (Nuovo Cimento, Suppl., to be published).
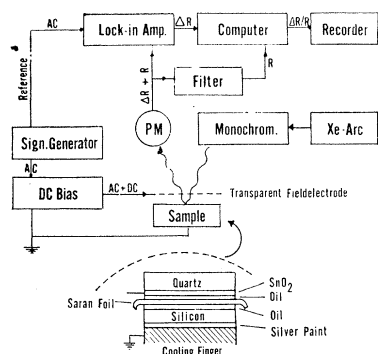[10] J. C. Phillips and B. O. Seraphin, Phys. Rev. Letters 15, 107 (1965).

Fig. 1. Block diagram of the experimental setup for the measurement of the optical-field effect in silicon. The detail shows a cross section (not to scale) of the sample sandwich.

sample surface. The mixture of different lens cements used previously proved unsatisfactory at low temperatures, since the hard compound tended to crack upon cooling. A mixture of Canada Balsam and diffusion-pump oil is now being used which stays viscous down to 80°K, matches refractive indices sufficiently, and is transparent to about 4.7 eV. The field-electrode consists of SnO_2 deposited on fused quartz. The SnO_2 cutoff limits the uv performance of the system seriously, but no better uv-transparent conductor could be found. Electrode and sample are separated by a sheet of Saran wrap, which is sufficiently transparent in the uv, in contrast to the previously used Mylar.

The whole sandwich, shown in Fig. 1, is clamped rigidly against a copper cooling finger, making electrical contact to the lapped back side of the silicon by means of silver paint.

Using the electric field in the surface potential barrier is certainly not the best way of producing the effect. Most of the potential drops across the dielectric, and the maximum field strength is limited to about 10^5 V/cm. From this maximum value at the interface the field drops toward the bulk in a rather complicated function, derived from Poisson's equation. This function must be evaluated with respect to the penetration depth of the light, which varies drastically in certain regions of the spectrum. Suggestions for improvements are obvious: Surface-barrier structures were tried which by means of thin gold films on either side are able to deplete the silicon wafer completely, producing a uniform field strength in the order of 10^5 V/cm.[11] Reflectance and absorption of gold are so high, however, that even with thin films very little modulated light returns from the Au-Si interface. Experiments are now being made with electrode materials of smaller reflectance and absorption. A second obvious suggestion—use of shallow *pn*-junctions—must be considered questionable because of the dependence of the effect upon the conduction type of the material.

The actual instrumentation is shown as a block diagram in Fig. 1: The light of a Xe high-pressure arc lamp passes through a Perkin-Elmer 99 monochromator

and is reflected from the sample under an 8-deg angle of incidence.[12] The surface potential at the reflecting surface is modulated by the transparent field electrode driven by the ac output of a signal generator superimposed onto the variable dc level of a battery pack. Dc and ac voltages of up to 600 V produce field strengths in the order of up to 2.5×10^5 V/cm in the dielectric. A photomultiplier tube monitors the reflected light. Modulation impressed onto the reflected beam is filtered out by a lock-in amplifier which beats the phototube output against the modulation reference taken from the signal generator. The dc output of the amplifier, proportional to the modulation depth and signifying by its sign the phase relation between modulation voltage and reflectance response, is connected to the numerator input of a desk computer. The filtered output of the phototube, representing by its dc value the total reflected intensity, provides the denominator input, and the quotient $\Delta R/R$ is displayed on a recorder which receives wavelength markers from the monochromator.

In all of the following figures, the $\Delta R/R$ trace going positive indicates that the reflectance increases during the positive half-wave of the modulating voltage. The trace going negative signifies a reduction of the reflectance caused by the same positive half wave.

The optical part of the instrumentation gives a spectral resolution of at least 0.003 eV in all parts of the spectrum considered here. Noise limited the detectability toward the uv end of the range to $\Delta R/R\geq3\times10^{-6}$, the performance improving over this figure toward the visible end.

RESULTS AND INTERPRETATION

A. Dependence upon Temperature

No response of the reflectance to the electric field is observed for photon energies of up to 3.3 eV, although measurements in transmission have shown that the indirect transition threshold at 1.1 eV is affected by an electric field.[3] Changes in the absorption coefficient of less than 1 cm^{-1}, as they are typically observed as phonon-assisted structure in the transparent region at the bottom of the first absorption edge, are much too small to be seen in reflection. The more sensitive transmission measurement, however, fails beyond this first edge.

Above 3.3 eV the reflectance responds to the electric field in a pronounced structure, which was not resolved by previous optical or photoemission studies. Two peaks appear at room temperature, located with opposite sign at 3.34 and 3.45 eV. The slight shoulder in between them indicates the presence of a third peak, which is resolved at low temperatures (Fig. 2) and eventually overtakes the negative neighbor in size.

[11] I am indebted to Professor M. Chester, University of California at Los Angeles, for supplying these structures.

[12] Dr. H. E. Bennett of this laboratory designed the optical part of the instrumentation.

Before discussing this peak group any further, it must be established that the observed structure really represents independent components. The phenomenon being basically a dispersive effect, one expects satellite structure to the left and to the right of a main peak. These satellite peaks are usually present and were studied in detail for the fundamental edge of germanium.[4] Since they are derivatives of the main peak, their response to temperature, surface potential, or modulating field follows a pattern, which was interpreted qualitatively on the basis of the Kramers-Kronig dispersion relation.[7] This pattern should be the same for any material for which the model of the surface potential barrier holds, as is the case for silicon. None of the properties of the three peaks which are described in the following fits the satellite pattern, and there is no evidence that any one of these peaks is a satellite of one of the others.

The upper part of Fig. 3 shows the temperature dependence of the wavelength position of the three peaks. The cross-hatched areas enclose the positions as measured on about 50 p-type runs for different dc bias. It will be shown in the following section how a position shifts toward the upper edge marked $+$ with positive dc bias and toward the lower edge marked $-$ with negative dc bias.

The temperature gradient of the wavelength position places the peaks in two distinctly different classes: Peak I moves only slowly with temperature, the crosshatched area described by $(-1.35\pm0.10)\times10^{-4}$ eV/deg with respect to center line and boundaries on either side. The two upper peaks II and III move more than twice as fast with temperature, their coefficients being $(-3.25\pm0.45)\times10^{-4}$ and $(-3.40\pm0.45)\times10^{-4}$ eV/deg. These values are approximately representative for the two groups of transitions band structure analysis describes as "insensitive" and "sensitive" to small changes of the crystal potential.[13] It is interesting to note that previous reflectance studies, not able to resolve any structure within the one peak observed in this region, assigned a temperature coefficient of -2.7×10^{-4} eV/deg to this peak, which is the arithmetical mean of the coefficients of the three resolved peaks.[14]

The temperature dependence of the peak size is shown in the lower part of Fig. 3. Although qualitative features of this diagram are always reproduced, the exact size of the peaks can vary widely according to surface conditions. The run shown in Fig. 3 represents a strong inversion layer on a p-type sample.

The dependence of the size upon temperature places the peaks again in two distinctly different groups, with the members of the group being different, however. Peaks I and III slowly decrease with increasing temperature, still being present at 300°K. Peak II de-

FIG. 2. The field-induced change in reflectance $\Delta R/R$ as a function of photon energy for four different temperatures. The recorder trace going positive indicates an increase in reflectance caused by the positive half wave of the modulating field.
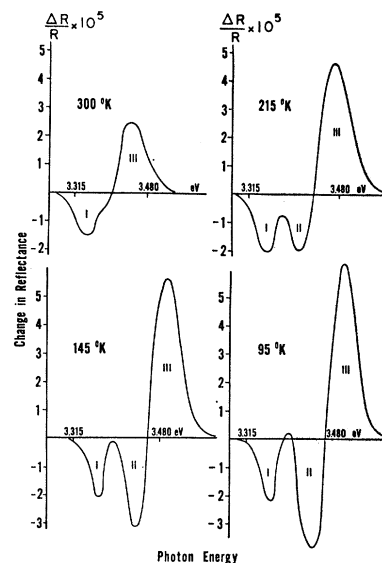
creases much faster, however, and all but disappears toward room temperature.

Notice that the temperature dependence alone—placing II and III in one group with respect to position and I and III with respect to size—provides information as to the assignment of these peaks. A previous paper suggested from arguments of this type that peak I should be assigned to the direct threshold at the Γ point of the Brillouin zone, while peak III represents a transition, presumably at the L point, preceded by an exciton indicated by peak II.[10]

Outside the region 3.3–3.6 eV and up to 4.7 eV, only one broad peak of 0.2-eV width is observed, centered at 4.20 eV. It probably correlates to the structure seen at 4.3 eV in photoemission[15] and in reflectance[16] and usually
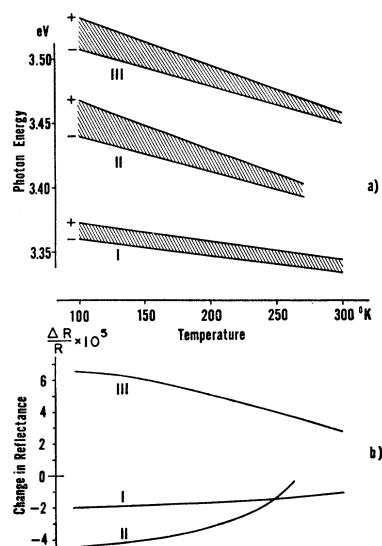

FIG. 3. (a) The wavelength position of the three peaks as a function of temperature. The $+$ or $-$ on the edges of the crosshatched areas indicate in which direction a dc bias of this sign shifts the peak. (b) Peak heights as a function of temperature for a strong inversion layer in the reflecting surface of the p-type sample.

[13] F. Herman and S. Skillman, *Proceedings of the International Conference on Semiconductor Physics, Prague, 1960* (Academic Press Inc., New York, 1961). p. 20.

[14] M. Cardona, J. Appl. Phys. **32**, 2151 (1961).

[15] W. E. Spicer and R. E. Simon, Phys. Rev. Letters **9**, 385 (1962); G. W. Gobeli and F. G. Allen, Phys. Rev. **137**, A245 (1965).

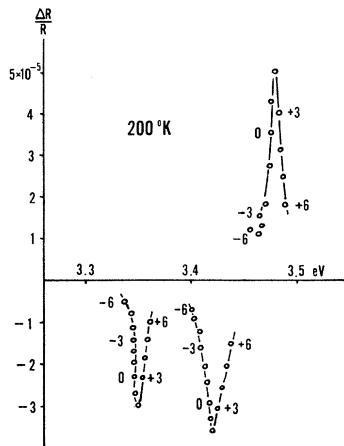[16] H. R. Phillipp and E. A. Taft, Phys. Rev. **120**, 37 (1960).

FIG. 4. The dependence of peak position and size upon surface potential, as adjusted by different dc bias (volts divided by 100 indicated along the traces) at 200°K for an inversion layer on a p-type sample.

assigned to the X_4-X_1 transition.[17] The temperature coefficient of this 4.20-eV peak is in the -1×10^{-4}-eV/deg class, its broadness making a more precise determination difficult.

The following sections will show how the response of the peak structure to surface potential, modulating field, crystal orientation, and conductivity type provides even more specific information as to the character of the transitions to which these peaks correlate.

B. Dependence upon Surface Potential

After adjusting the point of operation roughly as to a strong or moderate inversion layer on the p-type samples by choice of the chemical etch, the final adjustment was accomplished by dc bias at low temperatures. A range of dc bias between -600 and $+600$ V moves the surface from a point close to the flat-band position to approximately 0.30 eV up the n branch.

Figure 4 demonstrates how such a move affects wavelength position and size of the individual members of of the peak group. Every point represents one peak for one particular dc bias, and the diagram, although similar in appearance, must not be confused with the actual recorder traces of Fig. 2. A -600-V dc bias places the surface at a point close to the flat-band position. The modulating ac field causes a small swing of the bands only, resulting in a small optical response. The small heights of the potential barrier cause the peaks to appear at a lower photon energy than for a higher average field, in accordance with the analysis for germanium. The decrease in negative bias permits the buildup of the potential barrier, shifting the peak position toward greater photon energies. The modulating field causes a wider band swing, producing a larger optical response. Positive bias further supports the buildup of the barrier, shifting all peaks continuously in the same direction. The optical response, however, decreases again. Large positive bias brings the Fermi level so close to the con-

duction band that the necessary change in charge during one cycle can be produced by a small swing of the bands only. Notice that the variations are more pronounced for peaks II and III, confirming the greater stability of peak I against changes of the basic parameters.

The full range of Fig. 4 is seen only if the initial point of operation is selected properly by the choice of the etch. Starting out with a strong n-type layer, for instance, only the right-hand half of the "peaks" in Fig. 4 may be seen within the range of feasible dc bias.

The dependence of peak position and size provides a simple technique of measuring type and strength of a surface layer by strictly optical means. Without attaching contacts to the sample or making any electrical measurements, the response of the peak position to a change in dc bias will indicate the type of the surface layer. The variation in peak size will roughly indicate the approximate distance from the flat-band position. In the following sections will be demonstrated how bulk conductivity type and crystal orientation can be evaluated from diagrams of the type shown in Fig. 4.

The relation between the wavelength position of a peak and the surface potential can be used to monitor the character of a surface as a function of temperature. It was observed that the wavelength position for an unbiased surface did not proceed on a straight line upon cooling. A peak located close to the $+$ edge of Fig. 3 moves typically closer to the $-$ edge, if cooled to about 200°K, then turns around and approaches the $+$ edge again below this temperature. If the surface was kept under strong positive bias during the cooling, the wavelength position moved parallel to the $+$ edge.

The unbiased surface apparently decreases the strength of the inversion layer upon cooling, transferring the point of operation closer to the flatband position into the direction of majority-carrier accumulation. Below approximately 200°K, the strength of the n-type layer recovers. If, on the other hand, the surface is kept strongly n-type by applying positive bias, this transfer does not take place. This observation could be confirmed on a sample, on which larger negative bias produced p-type accumulation, passing through the flat-band position, in which all three peaks invert sign. Figure 5 plots the height of peak III as a function of dc bias for
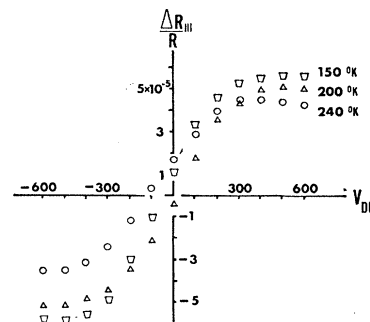
[17] J. C. Phillips, J. Phys. Chem. Solids **12**, 208 (1960); H. Ehrenreich, H. R. Phillipp, and J. C. Phillips, Phys. Rev. Letters **8**, 59 (1962).

FIG. 5. Size and sign of peak III as a function of dc bias for three different temperatures. The range of dc bias swings this particular p-type sample through the flat-band position over to the accumulation side.

different temperatures. Notice that the phase inversion, marking approximately the flat-band position, occurs for different values of the dc bias at different temperatures: Starting at 240°K with an unbiased surface which is n-type, a slight p-type layer is observed without bias at 200°K, which converts back to n-type at 150°K.

The kinetics of the silicon surface, especially at low temperatures, are not well enough known to interpret these shifts unambiguously. A common feature of surfaces, however, is the increase of the relaxation time into the fast states toward lower temperatures. If there is not enough time during one cycle of the 700-cps modulation frequency for the fast states to communicate with the conduction band, the type of the surface layer will shift toward the majority carrier side, as observed. The shift back toward stronger inversion must go unexplained. Yunovich and Talat[18] report, however, that the relaxation time passes through a maximum and falls to smaller values again beyond approximately 200°K, which would agree with our interpretation. Measurements at lower frequencies indicate that the shift can be decreased considerably between 700 and 70 cps, indicating the order of magnitude of the relaxation time.

If these observations are calibrated in terms of surface potential and modulation frequency, a convenient technique can be developed for the measurement of the dynamic parameters of a surface, as relaxation time, capture cross section, etc. The kinetics of a semiconductor surface could be monitored at all temperatures, using a simple reflection from an etched surface.

C. Dependence upon Modulation Frequency

It was pointed out in the previous section that wavelength position of a peak and the response of its size to a change in dc bias are representative for the point of operation on the spectrum of different surface conditions, opening the way for an optical investigation of type and strength of the surface space-charge layer as a function of temperature. As in measurements of the high-frequency field effect,[19] the ratio of the relaxation parameters to the length of time available during one cycle of the modulating field determines the type of response in a characteristic manner.

This observation of a frequency dependence was extended by measuring peak position and size as function of dc bias not only at the usual 700-cps modulation frequency, but also at 70 cps. For inversion as well as for accumulation layers, the low-frequency runs are similar to a high-frequency run at a much lower amplitude of the modulation. A 70-cps run of 600-V modulation amplitude, for instance, is similar to a 700-cps run with only 200-V modulation amplitude. Adjusting the amplitude for the high-frequency run until the same

over-all appearance of the peak group is produced provides a criterion for the comparison. The strong dependence of the excitonic peak II upon surface potential, its appearance or disappearance when the band swing hits upon its threshold potential, gives in particular a suitable marker for the measurement.

For the interpretation of this effect the sequence of different processes is recalled by which the surface adjusts to a change in the electric field.[19] Due to their high density, the majority carriers respond in a very short time, providing the necessary charge, on which the new field lines terminate, by either flowing in or out of the surface region. The bands are bent differently now, leaving fast states either filled above or empty below the Fermi level. Only after a much longer time τ, typically between fractions of a millisecond to fractions of a second, depending upon temperature, is equilibrium reestablished by communication between the fast states and the nearest band.

It is important for the interpretation that the direction into which the first process bends the bands be opposite to the direction of the second process in the sense that the communication between the fast states and the band tends to restore partially the configuration before the field was changed. If the surface potential is modulated at a frequency above $1/\tau$, the effective band swing will be large and correlated to the motion of the majority carriers only. If there is sufficient time during one cycle, however, to reduce the band swing by permitting the fast states to communicate with the adjacent band, the response will be equivalent to a smaller modulation amplitude at the higher frequency.

The drastic reduction in effective band swing between 700 and 70 cps indicates that the relaxation time τ is in the order of a few tens of a millisecond at 200°K. Closer sampling of the frequency range would make the determination more precise, providing an optical technique for the measurement of this parameter.

It cannot be decided at this time whether, in analogy to the electrical measurement, a "pulsed optical-field-effect" technique will be feasible. The necessary wide-band detection may present noise problems which could prevent the optical determination of surface state location and density otherwise possible with such a technique.

D. Dependence upon Crystalline Orientation

Several n-type samples were cut so that the reflecting surface was coplanar with the (110) or (100) face. No qualitative difference to (111) surfaces was found. The size of peaks II and III, however, seemed smaller, measured with respect to the size of peak I which did not change noticeably. Since the absolute peak height is the least reproducible parameter, due to the strong dependence upon the surface potential, the absolute value of the ratio of the height of peak III to the height of peak I is plotted in Fig. 6 as a

[18] A. E. Yunovich and G. Kh. Talat, Fiz. Tverd. Tela 6, 2369 (1964) [English transl.: Soviet Phys.—Solid State 6, 1880 (1965)].
[19] H. C. Montgomery, Phys. Rev. 106, 441 (1957); C. G. B. Garrett, Phys. Rev. 107, 478 (1957).
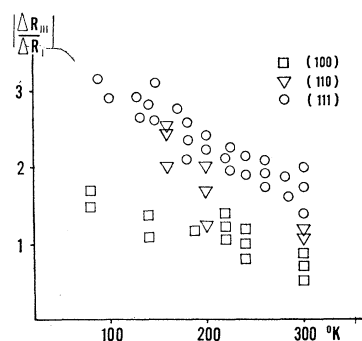
Fig. 6. The absolute value of the ratio of the height of peak III and peak I as a function of temperature for three different orientations of the reflecting surface.

function of temperature for a number of measurements on surfaces of different crystalline orientation. It is evident from this diagram that the ratio is greatest for the [111] direction, decreases slightly for the [110] direction, and is on the average smaller by a factor of two for the [100] direction. The temperature gradient of the ratio decreases in the same proportion.

This dependence of the height of peak III upon the crystalline orientation provides support for the assignment of this peak to a saddle-point edge, which has a principal axis oriented in k space in a certain direction, so that the observed dependence upon the orientation of the reflecting surface can result. The independence of peak I, on the other hand, must relate this peak to an "isotropic" transition, suggesting an assignment to the center of the Brillouin zone at Γ. The theory of the effect as expressed in Phillips duality theorem[9] predicts a field shift of the edge if the electric field is parallel to the principal axis of the saddle-point surface and no systematic shift for the field perpendicular to this principal axis. In cubic crystals, an average must be taken over parallel and perpendicular orientations. Degeneracy of the saddle point, however, together with an anisotropy induced by the electric field in the previously cubic symmetry, can still produce an optical field effect which is different for electric fields along different crystal axes.

E. Optical Field Effect on n-Type Samples

Eight n-type samples, of a resistivity comparable to the p-type samples, were cut so that the reflectance from surfaces of the three major crystalline orientations could be investigated. The field-induced change in the reflectance was measured in the range of photon energies between 3.3 and 3.6 eV as a function of temperature and surface potential, and the following major differences with respect to similar runs on p-type samples were observed:

a. Peaks II and III invert their phase with respect to peak I, which stays unchanged. Beginning at the low-energy end, the sequence of signs in an n-type sample is therefore negative-positive-negative, while in p-type material the sequence reads negative-negative-positive.

b. Peaks II and III shift toward smaller photon energies when the average field strength in the potential barrier of the n-type surface is increased by a positive dc bias. This is opposite to the shift observed in p-type material. Peak I shifts in the same direction regardless of the conductivity type of the material.

c. Peak II is more dominant in n-type material and does not disappear toward room temperature. Its rapid growth on the scale of different surface potential influences position and size of the neighbor peak III to an extent which makes it difficult to distinguish between the true position and the minimum which is observed as a result of the superposition.

d. Although otherwise unaffected by the type of the material, peak I is observed in n-type material at photon energies typically 10–15 meV larger than in p-type material.

The surfaces were treated in exactly the same manner as described for the p-type samples. Both etches, white etch as well as CP 4-A, produce n-type surfaces regardless of the doping of the bulk. On an n-type sample, however, the point of operation is located further away from the flat-band position than on a p-type sample. This explains why in Fig. 7, which shows the differences mentioned under a–d in a diagram similar to Fig. 4, peaks II and III increase in size for negative bias. The surface is so strongly n-type, and the Fermi level so close to the conduction band, that the potential barrier must be decreased by the negative bias to produce considerable band swing. Application of an etch which supports the formation of p-type surfaces, as soaking in dichromate,[20] produces the other half of the "peaks" shown in Fig. 4, where an increase in positive bias increases the peak size. The different transitions seem to respond to the surface field in a different manner, however: Although the full spectrum of peak I is seen in Fig. 7 between -600 and $+600$ V_{dc}, the same range
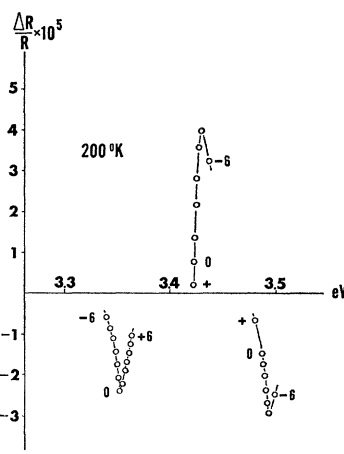

Fig. 7. The dependence of peak position and size upon surface potential, as adjusted by different dc bias (volts divided by 100 indicated along the traces). N-type sample with strong accumulation layer. Notice the inverted sign of peaks II and III and the shift of the position into the opposite direction, while peak I is unchanged in both respects.

[20] T. M. Buck and F. S. McKim, J. Electrochem. Soc. 105, 709 (1958).

of dc bias produces only the "declining half" for peaks II and III.

These minor differences can easily be interpreted on the basis of surface effects. Of greater consequence and most probably not a surface effect are the phase inversal of peaks II and III and the reversal of their shift with electric field. While the phase reversal for the whole peak group is observed upon passage through the flat-band position and can be explained as a surface effect, inversions of phase and shift for just one part of the group are beyond an easy explanation based on direction or strength of the potential barrier.

CONCLUSIONS

The preceding sections suggest that a large amount of information can be obtained from the optical field effect, in respect to the analysis of the band structure as well as to the physics of the semiconductor surface. Because of the derivative nature of the effect, the results are of high contrast and show a pronounced structure which can be read out with sufficient precision.

It is much less obvious, however, how these results must be interpreted in order to draw conclusions from the characteristic pattern in which peaks and dips respond to electric field, temperature, frequency, crystal orientation, and doping. Several features—most prominent among them the phase reversal of peaks II and III upon going from p-type to n-type material described in the foregoing section—demonstrate that more measurements on silicon as well as on other materials will be necessary before a full understanding of the effect will evolve. Alternative methods, changing the density of states at the critical points periodically by either changing a magnetic field, the pressure,[21] or the temperature, will contribute substantially to such an understanding.

Even at this early stage, however, a preliminary three-step procedure can be laid out along which the final analysis will proceed.

The basis from which this analysis starts is provided by the relation between a periodic variation of a modulation parameter (electric or magnetic field, pressure, or temperature) and the change in the density-of-state function which this variation induces in the neighborhood of critical points. Since the density-of-state function is proportional to the imaginary part ϵ_2 of the complex dielectric constant, the reflectance response by which the periodic change of ϵ_2 is read out will show a definite and significant relation of its phase to the driving force. It is essential, for instance, whether an increase in the modulation parameter increases the absorption for certain photon energies or reduces it. Accordingly, the reflectance response will be positive or negative, expressing this fundamental phase relation by the sign of the experimentally observed structure.

It will be demonstrated for the case of an exciton at a parabolic edge how general conclusions with respect to the assignment of structure can be drawn from this phase relation. The effect of an electric field on the absorption in front of the fundamental edge, which is necessarily of the parabolic type, is well known as the Franz-Keldysh effect and can be described in very simple terms as a smearing out of the band edges into the forbidden gap. It follows that for photon energies slightly smaller than the energy gap, the larger electric field induces a larger absorption, and a phase relation in accordance with this "large field–large absorption" relation will be observed. If an exciton is preceding such an edge, the relation between the field and the absorption due to this exciton will be of opposite sign: For small electric fields, the exciton is well separated from the band edge and the absorption will be significant. If the field-stretched band edge engulfs the exciton level, however, it decays into the background more easily and the absorption is reduced. The relation is therefore "large field-small absorption" and the exciton will be observed as structure of a sign opposite to its parent transition. This is in agreement with the observation in silicon: Peak II always shows the opposite sign of peak III, regardless of the conduction-type of the material.

This example of a parabolic exciton demonstrates how the up-down or down-up sequence of two peaks separated by the appropriate energy distance helps in the identification of such structure. The relation between the electric field and the induced absorption proves of value in the identification of the type of transitions in general, however. Criteria for certain features of the experiment are established which enable the analysis to discriminate among parabolic and saddle-point edges. Again, an example will demonstrate this most efficiently.

The way in which an electric field changes the density of states around a critical point of the parabolic type was extensively studied in the Franz-Keldysh theory. Since the transmission effect could be observed in front of the fundamental absorption edge only, the change of density of states in the presence of an electric field was never investigated at a saddle-point edge. Stimulated by the observation of the higher interband transitions in the field effect of the reflectance, Phillips studied the changes at these saddle-point edges with the result that the field-induced change of ϵ_2 at such an edge can be derived from the similar change at a parabolic edge by a simple reflection at the origin, as shown in Fig. 8.[9,10] The transformation properties of this "duality theorem" suggest that the experiment, if performed at a saddle-point edge, will show features which may be termed "anti-Franz–Keldysh effect." It was shown in detail, for instance, that the normal Franz-Keldysh mechanism not only explains the observed reflectance response at the fundamental edge of germanium,[7] but also correctly interprets the shift of the components of the peak group with the electric field. Using the transformation

[21] W. E. Engeler, H. Fritzsche, M. Garfinkel, and J. J. Tiemann, Phys. Rev. Letters **14**, 1069 (1965); G. W. Gobeli and E. O. Kane, *ibid.* **15**, 142 (1965).
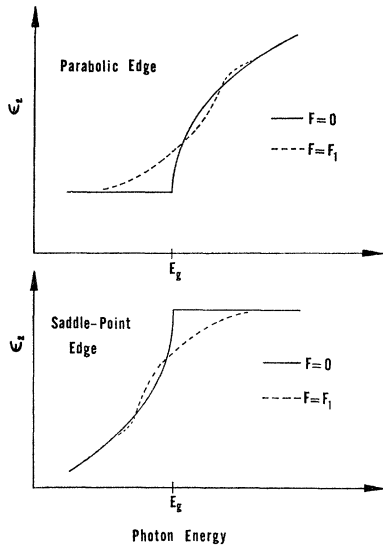
FIG. 8. The imaginary part ϵ_2 of the complex dielectric constant with and without an electric field at (a) a parabolic edge and (b) a saddle-point edge (after J. C. Phillips, Ref. 9).

shown in Fig. 8, it must be expected that a peak correlated with a saddle-point edge shifts into the opposite direction with the electric field. This is actually the case for GaAs,[6] where two major peak groups shift in opposite directions with increasing electric field, suggesting the assignment of these two groups to transitions of different character.

Up to this point, only contributions of the change in the density of states to the imaginary part ϵ_2 of the dielectric constant were considered. Since real and imaginary parts are connected by the dispersion relation, however, any change $\Delta\epsilon_2$ of the imaginary part will induce a related change $\Delta\epsilon_1$ in the real part. This contribution will in many cases be the decisive one. It was shown previously for the fundamental edge of germanium that agreement with the optical field effect in this spectral region can be obtained only by computing $\Delta\epsilon_1$ from the Kramers-Kronig relation applied to the field-induced $\Delta\epsilon_2$.[7]

This second step of the procedure now leaves the analysis with an estimate as to what extent the change in the modulation parameter affects the real as well as the imaginary part of the dielectric constant. Proceeding from such a knowledge of $\Delta\epsilon_1$ and $\Delta\epsilon_2$ as a function of the modulation parameter, either in scalar or tensorial form, the analysis must now evaluate in a third and final step to what extent these two functions contribute to the observed change in the reflectance. Expressing the relative change in reflectance as

$$\Delta R/R = \alpha(\epsilon_1, \epsilon_2)\Delta\epsilon_1 + \beta(\epsilon_1, \epsilon_2)\Delta\epsilon_2, \qquad (1)$$

the coefficients α and β are found to be functions of the photon energy, whose sign and relative magnitude determine the result of the analysis in the different parts of the spectrum. Figure 9 shows the coefficients α and β as a function of photon energy for the case of silicon, using the experimental values of Philipp and Taft[16]

for ϵ_1 and ϵ_2. There are regions in which $\Delta\epsilon_1$ dominates the reflectance response (0 to 3 eV), others in which $\Delta\epsilon_2$ dominates (3.5 to 4.0 eV), and regions in which both contribute comparable amounts. The analysis must discuss diagrams of the type of Fig. 9 for every material before assigning structure to transitions. The reflectance response of silicon at 4.20 eV can serve as an example for this third step of the analysis: As mentioned previously in this study, a negative dip is observed at 4.20 eV of approximately the line shape and the sign of peak I. It is seen from Fig. 9 that β is small in this region and that the response is dominated by the negative α. If this peak were of parabolic origin, a structure should be expected opposite in sign of peak I for which α also dominates, with positive sign, however. It is concluded, therefore, that the observed similarity of the structure results from a double inversion of sign, the first resulting from the negative α and the second from the saddle-point origin of the related transition, which inverts the sign due to the transformation properties of the duality theorem. This conclusion is in agreement with previous work which assigned structure in reflectance and photoemission at this photon energy to the transition X_4-X_1 of the saddle-point type.[15-17]

Attention is called to the crossover of the coefficients approximately between peak I and peak II. Since the contributions to the response in this spectral region are distributed among ϵ_1 and ϵ_2 in a manner which changes drastically even within this small range of photon energies, the final analysis must depend upon numerical calculations rather than qualitative viewpoints. Calculations of this type, including lifetime broadening, are presently being carried out.[22]

This concludes the preliminary sketch of a procedure which the final analysis will have to follow regardless of the modulation parameter which is periodically varied in the intended change of the density-of-state function. It may ultimately be possible, with all the functional relations of the type of the duality theorem worked out, to obtain a system of integrodifferential equations, which for an observed structure of peaks and dips, has one and only one set of transitions as a solution. Such a
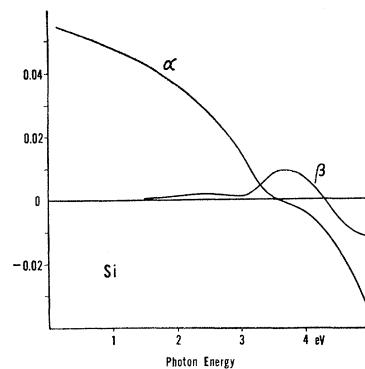

FIG. 9. The coefficients α and β in the equation $\Delta R/R = \alpha\Delta\epsilon_1 + \beta\Delta\epsilon_2$, as calculated from the experimental values of Ref. 16.

[22] B. O. Seraphin and N. Bottka, Phys. Rev. (to be published).

broad view on band-structure analysis is a long way off, however. At present, the variety of the pattern of experimental observations seems to be larger than the theoretical imagination. It should be pointed out, for instance, that none of the previously suggested steps of the analysis will explain why peaks II and III in silicon invert sign and direction of field-shift upon going from p-type material to n-type material. Any interpretation of what appears from the present, limited knowledge of the effect as an anomaly must assume an absorption mechanism, which discriminates among n- and p-type materials in the sense that the absorption is raised by the presence of the electric field in n-type material and lowered in p-type material. An ideal candidate for such an explanation would be a near-degeneracy of a parabolic and a saddle-point edge.

More information must be gathered on this effect, however, before any interpretation can leave the range of speculation. Experiments over a wider range of surface conditions, covering the full spectrum from strong p-inversion layers to strong n-accumulation layers on n-type material and measuring the surface potential simultaneously, must rule out the possibility that the peaks invert their phase separately for different surface potentials. While such measurements were made for p-type samples with the result that size and position are determined by the surface potential for all peaks in a similar manner, the situation seems to be more complicated for n-type samples, with the peaks displaying more individuality in their dependence upon the surface potential.

Note added in proof. M. L. Cohen (private communication) points out that the edges of valence and conduction band in silicon run very close to parallel between Γ and L. This may change the spectrum of critical points along this line, if differences of the "electrical deformation potential" on either end of it are assumed. Since the chemical etch produces an n-type surface on p-type as well as on n-type silicon, the average potential barrier is different in the two materials. This difference could cause the observation of two parabolic edges in p-type material, but a sequence of a parabolic and a saddle-point edge in n-type material.

ACKNOWLEDGMENTS