

Optical measurements of the minigaps in electron inversion layers on vicinal planes of Si(001)

Avid Kamgar, M. D. Sturge, and D. C. Tsui

Bell Laboratories, Murray Hill, New Jersey 07974

(Received 17 September 1979)

Measurements are reported of the far-infrared absorption of electron inversion layers at Si-SiO₂ interfaces tilted a few degrees from (001). The lowest minigap E_g , as a function of electron density N and tilt angle θ , is found to be $E_g = (7.5 \pm 0.5)N \sin|\theta| + (0.15 \pm 0.02)N^2$ meV, where N is in units of 10^{12} cm⁻². The first term is tentatively interpreted as arising from a term in the surface potential $V_Q e^{i\vec{Q}\cdot\vec{r}}$, where $Q = (4\pi/a) \sin\theta$ in the direction of tilt, and the second as coming from a kinetic energy term of the form $k_x k_y$. The second minigap has also been measured, for $N = 3.1 \times 10^{12}$ cm⁻² and $\theta = 1.9^\circ$, to be 2.75 ± 0.3 meV.

I. INTRODUCTION

Electrons or holes in an inversion or accumulation layer on a semiconductor move in a two-dimensionally (2D) periodic potential whose periodicity, for a stepless unreconstructed interface, is simply the periodicity of the 2D projection of the 3D crystalline potential. If the surface makes a small angle θ with a low index plane whose lowest reciprocal lattice vector is \vec{G} , this potential contains a long-wavelength component with wave vector $Q = G \sin\theta$ in the direction of tilt, all other components having much shorter wavelengths. The effect of this potential is to superimpose a 1D "minizone" structure on the 2D band structure, introducing "minigaps" in the $E(k)$ spectrum at the minizone boundaries. Since Q can (in principle) be arbitrarily small, the "minibands" between minigaps can be arbitrarily narrow in energy.

This scheme has recently been realized in the n -type inversion layer of Si MOSFET's (metal-oxide-semiconductor-field-effect transistors), in which the Si-SiO₂ interface is tilted a few degrees from (001) about [110].¹⁻³ In silicon, there are six valleys in the conduction band, with minima at the points $(\pm k_0, 0, 0)$, $(0, \pm k_0, 0)$, $(0, 0, \pm k_0)$ where $k_0 = 0.85(2\pi/a)$. When the interface lies close to the (001) plane, the $(0, 0, \pm k_0)$ valleys are the lowest, and are the only ones occupied at low electron temperature. When alignment in (001) is exact, the bands associated with these two valleys coincide, but when the interface is tilted, the bands are separated in the 2D \vec{k} space. This lifting of the degeneracy gives rise to minigaps at the crossings of the bands. Figure 1 shows the lowest 4 gaps. The minigaps 1, 2, and 4 are due to the removal of the valley degeneracy by intervalley coupling while intravalley coupling gives rise to minigap 3.

The existence of a minigap can be detected through the anomaly it engenders in the dc mobility when the

Fermi level first touches it, as the electron density N is increased from zero. This anomaly was first observed (for the lowest minigap) by Cole *et al.*,¹ who also observed Shubnikov-de Haas oscillations from the electron pocket just above the minigap. They attributed the minigap to a hypothetical step structure in the Si-SiO₂ interface, but were puzzled by the periodicity deduced from k_c , the value of k_F at the anomaly. Sham *et al.*² proposed the projected surface band-structure model illustrated in Fig. 1, which accounts accurately for the value of k_c . They observed optical transitions across the lowest minigap in a sample with tilt angle $\theta = 8.9^\circ$. Later Tsui *et al.*³ confirmed the model by observing anomalies in dc conductivity corresponding to four different minigaps, in samples with small tilt angles. The model of Sham *et al.* is successful in predicting correctly the values of k_c (and thus of electron density $N_c = k_c^2 \pi$) at which the mobility anomalies occur. They are determined solely by the band structure of silicon and simple geometry.

Volkov and Sandomirski⁴ object to the Sham model on theoretical grounds, and propose an alternative model which predicts a highly nonmonotonic dependence of k_c on θ . This prediction is falsified by the data of Tsui *et al.*³ who find that $k_c \propto \sin\theta$ over a wide range of θ .

What still remains in question is the mechanism which gives rise to these optical gaps. There have been a few theoretical attempts on this problem.⁵⁻⁸ Here we present results of optical measurements from which we can determine the magnitude of the two lowest minigaps. We have measured the far-infrared absorption due to transitions across the lowest minigaps, as a function of electron density, in samples with $\theta = 2.9^\circ$, 5.6° , 8.9° , and 11° . We have also observed absorption due to the minigap No. 2 in a sample with $\theta = 1.9^\circ$ for the value of N at which the Fermi level just touches the minigaps. A preliminary

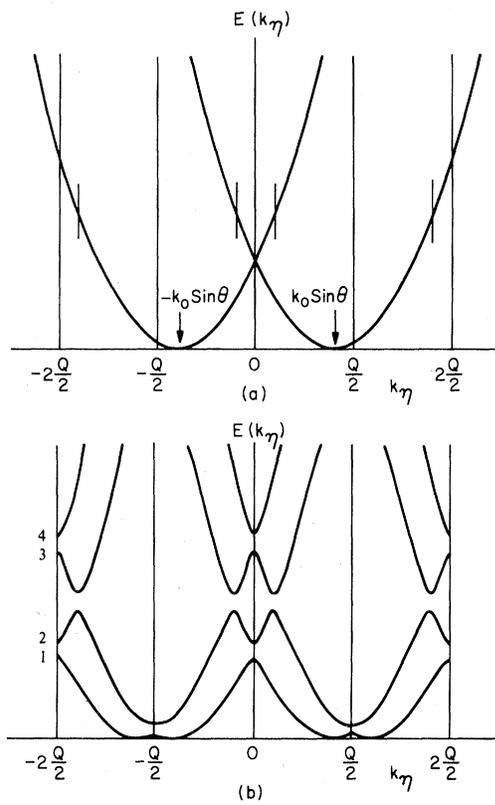


FIG. 1. E vs k_η of inversion-layer electrons on surfaces tilted at θ from (100): (a) in extended-zone scheme in the absence of valley-valley interactions and (b) in the periodic-zone scheme with degeneracies at band crossings removed. k_η is in the direction of tilt.

account has already been given.⁹ Similar results to ours have been recently reported.¹⁰

II. EXPERIMENT

Experiments were done at 4.2 K using Si-MOSFET's fabricated on 24- Ω cm p -type Si wafers, which were oriented using x-rays to better than 0.2° . The gate oxide was thermally grown to ~ 1400 Å in dry oxygen at 1100°C , and subsequently annealed in H_2 at 380°C . The metallization was produced by evaporating a thin Ti layer (~ 50 Å) followed by a $0.7\text{-}\mu\text{m}$ thick layer of Al. The Al layer was subsequently removed from the gate to transmit the far-infrared radiation. These devices have maximum electron mobility varying from ~ 8000 to ~ 20000 $\text{cm}^2/\text{V sec}$ at 4.2 K. The electron density of the inversion layer was determined from $N = (C_{\text{ox}}/e)(V_g - V_t)$, where C_{ox} is the capacitance of the gate oxide, V_g the gate voltage, and V_t the conduction threshold at 78 K.

We used a molecular gas laser pumped with a pulsed CO_2 laser.¹¹ Several lines between 1.36 and

10.45 meV were used. The experiment is made possible through the relatively strong density dependence of the size of the gap. Absorption (α), its derivative ($d\alpha/dV_g$), and photoconductivity were measured as a function of the applied gate voltage V_g in the normal incidence geometry. Absorption due to the carriers in the inversion layer was differentiated from background by chopping the gate voltage. In order to avoid interference effects the Si substrate was either wedged by about 4° , or it was glued to a $4\text{--}6^\circ$ wedge. A Ge bolometer/detector monitored the laser radiation transmitted through the sample. The strongest absorption signal was about 0.5% of the total laser radiation, while the weakest one was roughly two orders of magnitude smaller. In the case of the weaker signal, a signal averager was used.

III. RESULTS

A. Photoconductivity

Using differential absorption techniques, i.e., either modulating or chopping the gate voltage to obtain the absorption derivative or the total absorption, respectively, we have been able to detect transitions across only the lowest two minigaps. However, photoconductivity measurements reveal all four minigaps as shown by the arrows in Fig. 2. This figure shows the photoconductivity as a function of V_g in four different samples. On top of the Drude background, which is independent of tilt, there are anomalies in the photoconductivity quite similar to those in the dc conductivity, although here the signal is much stronger relative to background and has a different line shape. These anomalies are presumably due to a change in the temperature derivative of the mobility, $d\mu/dT$, which occurs when the Fermi level reaches the minigap. This change produces an anomaly in the photoconductive signal, since this arises from the heating of the electrons due to nonresonant (Drude) absorption. There may also be a small contribution from resonant heating due to transitions across the minigap.

The top three spectra show the lowest gap in 11.0° , 8.9° , and 5.6° samples, while the lowest spectrum shows the three higher gaps in a 1.9° sample.

The relative amplitude of this signal can be qualitatively understood by considering the size of the minigap and the electron mobility μ where the signal occurs. The 8.9° sample shows a stronger signal compared with the 5.6° sample because of the larger gap and therefore a stronger departure from a free-electron behavior. However, the 11° sample shows a smaller signal because it occurs at a gate voltage where $d\mu/dT$ and hence the photoconductivity signal is anyway practically zero. In the 1.9° sample the anomaly due to the lowest minigap, which is expected

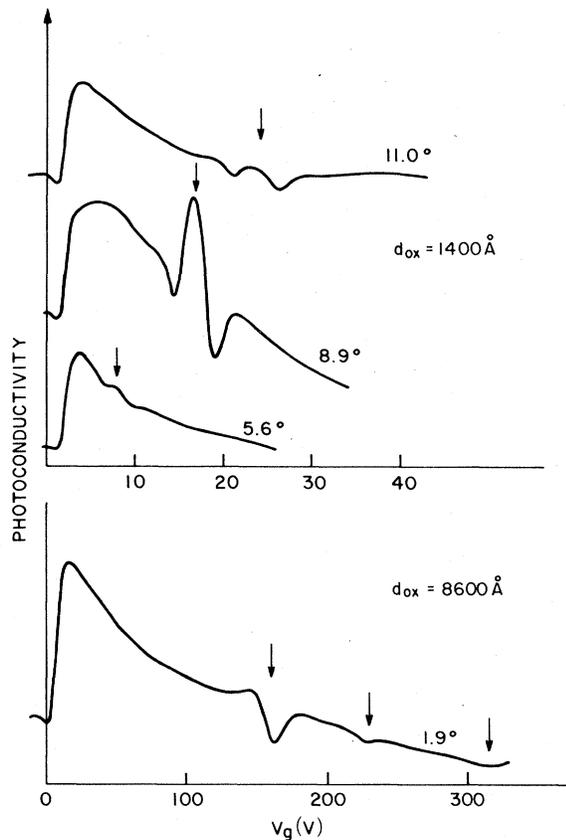


FIG. 2. Photoconductivity as a function of V_g in four different samples. The 10.45-meV laser line is used to obtain all of these traces. However the signal is quite insensitive to the laser energy.

at around $N = 1.7 \times 10^{11} \text{ cm}^{-2}$, is not strong enough to be seen, but the three higher gaps show up quite distinctly.

In the following we will divide the discussion on the absorption measurements into two sections concerning, respectively, the lowest two gaps.

B. Absorption due to the lowest minigap

The first optical data on the lowest minigap were obtained using a Fourier spectrometer.² In this experiment the size of the minigap was shown to change from 2.5 to 5.5 meV as the electron density N varied between 2.3 and $4.7 \times 10^{12} \text{ cm}^{-2}$, for $\theta = 8.9^\circ$. We have extended this measurement to the range 1.2 to $6 \times 10^{12} \text{ cm}^{-2}$ in N and to a wide range of θ . Figure 3 shows an example of the absorption measurements for $\theta = 8.9^\circ$ sample. The absorption data are obtained by sweeping the gate voltage, therefore the electron density, at fixed laser frequency ω . In Fig. 3 this frequency is chosen so that at $N = N_c$, $\hbar\omega$ is very

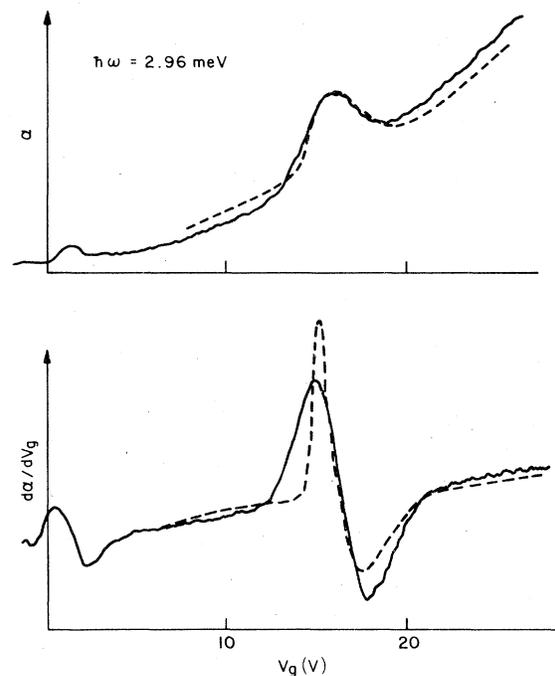


FIG. 3. Absorption α and absorption derivative $d\alpha/dV_g$ as a function of V_g in the $\theta = 8.9^\circ$ sample. At this laser energy $E_g(N_c) = \hbar\omega$. Dashed lines are the calculated line shapes, with vertical scale chosen to give the best fit.

nearly equal to the size of the gap. The top trace represents the total absorption, and the derivative of absorption is shown in the lower trace. The upper trace shows a general rise in the absorption (at given laser frequency) of an electron gas with increasing density and scattering rate. The structure at very low V_g is present even for $Q = 0$, and has nothing to do with minigap. It is due to the abrupt change in the Drude conductivity near threshold. Like the broad Drude background, it is most pronounced at low $\hbar\omega$ (see Fig. 4). The extra increase in the absorption at $V \approx 16$ volts is due to the electron transition across the minigap.

Superimposed on the experimental traces (full line), the dashed lines show the results of a calculation based on that of Ref. 2. The absorption $\alpha(\omega)$ (which is proportional to the real part of the conductivity) is calculated for a 2D electron gas moving in the 1D potential $V = \frac{1}{2} E_g \cos(2\pi\eta/L)$, where η is measured in the direction of tilt.

It is easily shown² that this model gives the same frequency dependence of α in the vicinity of a minigap as the model of Fig. 1, so long as the minigaps are well separated. The absolute value of α differs by a factor of 2 but this is not measured in our experiments.

We assume a functional form for $E_g(N)$, discussed below, and calculated $\alpha(N)$ at fixed laser energy $\hbar\omega$.

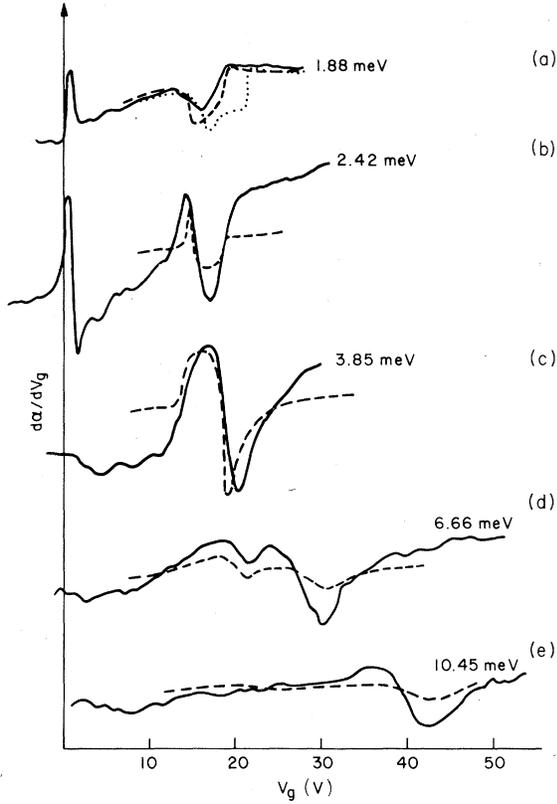


FIG. 4. Absorption derivative $d\alpha/dV_g$ as a function of the gate voltage in the $\theta = 8.9^\circ$ sample at various laser energies. The dashed and dotted lines are the result of calculations as discussed in the text. The calculated curves are all on the same vertical scale. The data curve on arbitrary scales. This also applies to Fig. 5.

The vertical scale for the theoretical curves is chosen to give the best fit to the data obtained at 3.85 meV. The electron scattering time τ is determined as a function of N from the dc mobility; the periodicity L is

$$L = 2\pi(Q - k_0 \sin\theta)^{-1}, \quad (1)$$

where $Q = (4\pi/a) \sin\theta$ and $k_0 = 0.85(2\pi/a)$.

While the theory of Ref. 2 is quantitative, and gives reasonable agreement with experiment, it does neglect any dependence E_g might have on \bar{k} , at fixed N . We will show later that such dependence is in fact present, and probably limits the accuracy of the fit. We see that the experimental curves are somewhat broader than the theoretical; this can, at least in part, reasonably be attributed to the k dependence of E_g . Figure 3, which compares the fit to the direct absorption and to the derivative, demonstrates that this deviation is not serious.

A series of curves representing absorption derivative for a full spectrum of laser energies is shown in

Fig. 4. The vertical scale of each experimental trace is arbitrary. In trace (a), the laser energy is less than the gap for $N = N_c$, i.e., $\hbar\omega < E_g(N_c)$. Nevertheless, we see structure at $N = N_c$, which arises from the changes in the density of states which occur when the Fermi level is swept across the minigap.

Jumping to traces (d) and (e), for which $\hbar\omega \gg E_g(N_c)$, we see structure which moves to greater N with increasing $\hbar\omega$. This arises from transitions across the minigap, and the minimum in $d\alpha/dV_g$ corresponds approximately to the value of N at which $\hbar\omega = E_g(N)$. (To get a more exact value, a quantitative fit must be made, as we shall see.) In trace (d) there is still a vestige of the structure at $N = N_c$, which has disappeared in (e), the highest $\hbar\omega$ for which any structure was observable. Traces (b) and (c) are for $\hbar\omega \sim E_g(N_c)$. The two structures are superimposed, and give their maximum combined effect when $\hbar\omega = E_g(N_c)$.

Figure 4(a) also illustrates the sensitivity of our data to L , the periodicity. The dashed curve was calculated using $L = 117 \text{ \AA}$, the value given by Eq. (1) for $\theta = 8.9^\circ$; the dotted curve was calculated for $L = 109 \text{ \AA}$, the value obtained from dc conductivity measurements. A clear distinction in favor of the larger value can be made; the use of a more quantitative theory in analyzing these data (as compared with dc conductivity) has made this distinction possible.

In Figs. 5 and 6 we show the spectra obtained in the 5.6° and 2.9° samples. The 5.6° traces are quite similar to the 8.9° sample for high values $\hbar\omega$. In the

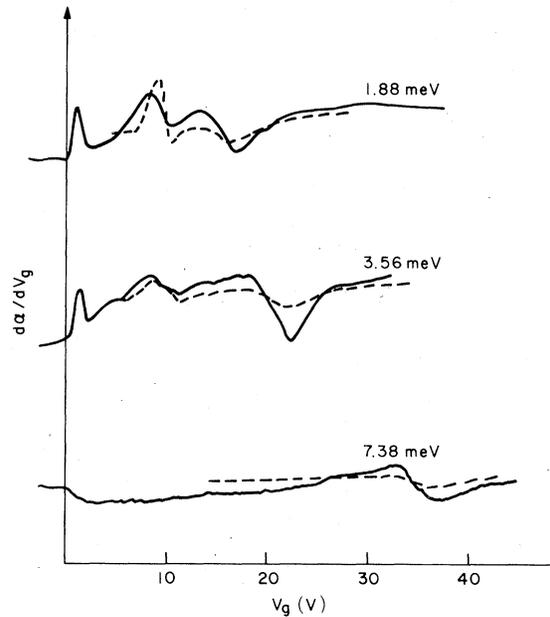


FIG. 5. Absorption derivative as a function of V_g in the $\theta = 5.6^\circ$ sample at three different laser energies, together with the calculated line shapes, shown as dashed lines.

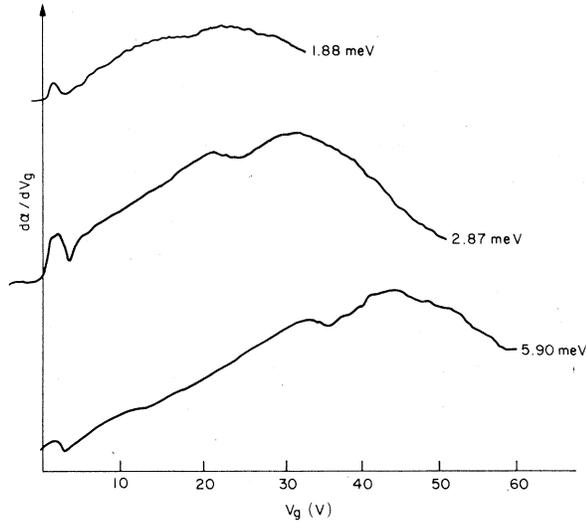


FIG. 6. Absorption derivative as a function of V_g in the $\theta = 2.9^\circ$ sample.

2.9° sample the signal is noticeably weaker relative to the background, which is due to the Drude absorption.

In order to calculate the conductivity as a function of N , we had to assume a functional form for the variation of the minigap with electron density. We tried two different functions, suggested by the Taylor expansion for small N :

$$E_g = c + dN, \quad (2)$$

and

$$E_g = aN + bN^2, \quad (3)$$

choosing constants a and b , or c and d , to give the best fit to the data. Table I shows these fitted values for various samples. Both functions give quite satisfactory fit to the data, but we will discuss only the second function, since not only does it have some theoretical basis (to be discussed in the next section), but the parameters obtained are physically more acceptable (the negative value of c makes no physical

TABLE I. Parameters in Eqs. (2) and (3).

θ (degrees)	a	b	c	d
2.9	0.37	0.15	-2	1.5
5.6	0.75	0.15	-1.8	1.8
8.9	1.2	0.15	-1.7	2.3
11.0	1.55	0.15		

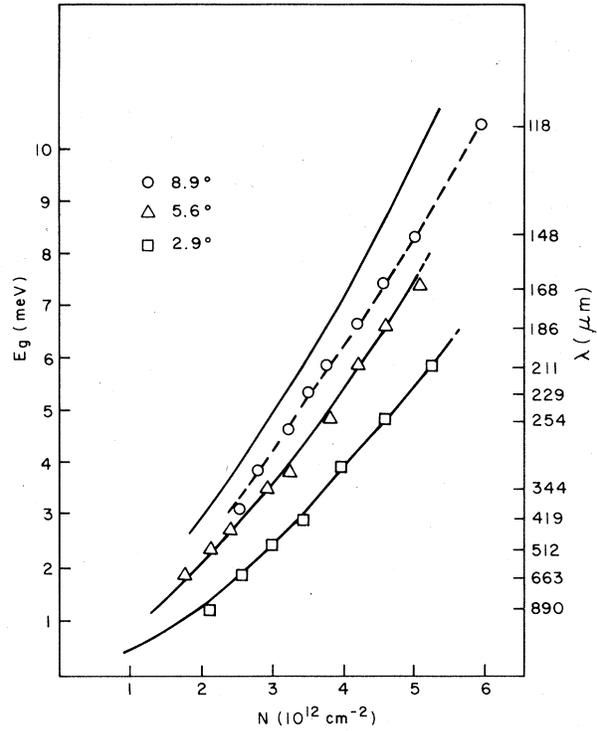


FIG. 7. Solid lines represent the size of the gap as function of electron density as obtained from a theoretical fit to the data. The points show the position of the minima in the experimental traces. The dashed line indicates the position of minima in the calculations. In case of the 5.6° and 2.9° samples the dashed and the solid lines coincide.

sense, since it implies $E_g < 0$ at small N).

Figure 7 summarizes our results for the lowest minigap. The points give the positions of the minima in $d\alpha/dV_g$ obtained experimentally. The full lines represent Eq. (3) with the parameters of Table I. The dashed line passes through the positions of the minima in $d\alpha/dV_g$ calculated from Eq. (3), for $\theta = 8.9^\circ$. The fit to the data is seen to be good. Comparison with the full curve for this θ illustrates the importance of making the fit to the full theoretical curve, rather than relying on some arbitrarily chosen point (such as the minimum) to give the value of N where $E_g(N) = \hbar\omega$. Nevertheless, for $\theta = 5.9^\circ$ and 2.9° , the dashed curve would be close to the full curve and is not shown.

In Fig. 8 we show that the fitting parameter a is approximately proportional to $\sin\theta$, leading to the following empirical relation:

$$E_g = (7.5 \pm 0.5)N \sin|\theta| + (0.15 \pm 0.02)N^2, \quad (4)$$

where N is in units of 10^{12} cm^{-2} , and E_g in meV.

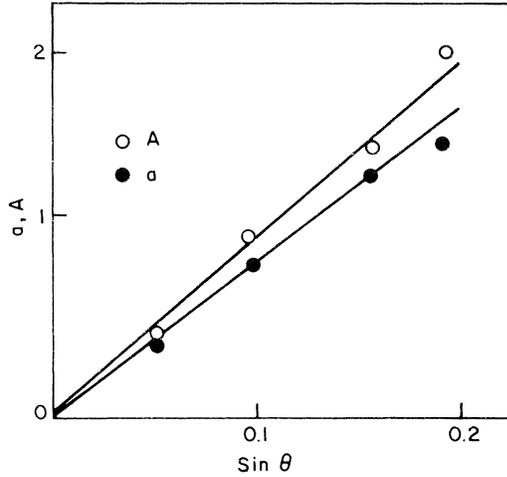


FIG. 8. Fitting parameter a , and $A = a + 2bN_c$, as a function of $\sin\theta$.

C. Absorption due to the second minigap

We have searched for anomalies in the absorption derivative due to the higher minigaps in 1.9 and 2.9° samples. In the 2.9° sample the second gap has $N_c = 7.5 \times 10^{12} \text{ cm}^{-2}$, corresponding to a V_g only just within the limit for the dielectric breakdown of the oxide. We were not able to detect the absorption across this minigap, possibly because of the high electron scattering rate at this density.

In the 1.9° sample all three higher gaps were observed in the photoconductivity (Fig. 2); however, only the absorption due to the second gap was detected (Fig. 9) and then only in the region $N \sim N_c$,

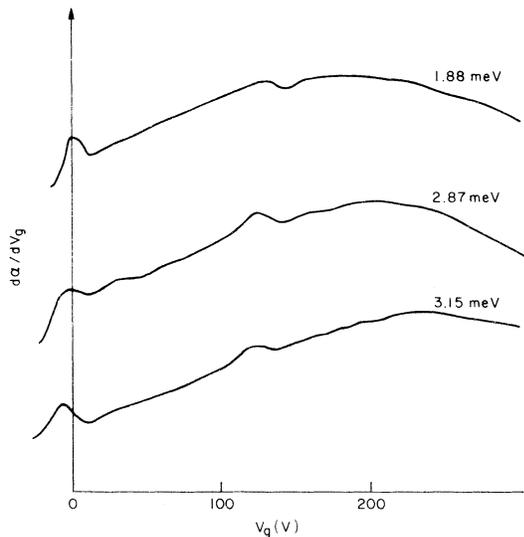


FIG. 9. Absorption derivative as a function of gate voltage in the $\theta = 1.9^\circ$ sample.

where the two structures overlap and reinforce each other. Thus a determination of E_g as a function of N was not possible. We attempted to obtain $E_g(N_c)$ in two ways. The first was to find the value of $\hbar\omega$ for which the absolute magnitude of the anomaly in $d\alpha/dV_g$ is a maximum. While comparison of absolute magnitudes at different $\hbar\omega$ is difficult, the maximum in the intensity is sufficiently pronounced that an unambiguous value of $E_g = 2.75 \pm 0.3 \text{ meV}$, at $N = N_c = 3.1 \times 10^{12} \text{ cm}^{-2}$ can be obtained. In the second method we assume that E_g is proportional to N and make a computer fit, as in the previous section. The fitted value of $E_g(N_c)$ agrees with our first result. The signal is between one to two orders of magnitude smaller than the comparable signal due to the lowest minigap. Furthermore, the position of the absorption peak does not vary with N between 2.4 and 3.15 meV. These two facts suggest that the density dependence of E_g for the second minigap may be weaker than for the first.

IV. DISCUSSION

The positions of minigaps in \bar{k} space depend only on the band structure and can be determined geometrically. Calculation of the size of a minigap, on the other hand, presents formidable difficulties. While some very sophisticated calculations have been made,⁵⁻⁸ they are based on the unphysical model of an infinitely sharp, mathematically flat interface, taking no account of its atomic structure. It is, therefore, unlikely that they can yield quantitatively correct results.

There are, however, some general observations which can be made. First, the symmetries of the conduction bands near k_0 are Δ_1 and Δ'_2 , respectively. It follows that the intervalley interaction (responsible for the lowest two minigaps) must have Δ'_2 (i.e., xy) symmetry. It might have the form of a static potential, such as e_{xy} strain, or of a "kinetic energy term" proportional to $k_x k_y$. Second, the potentials responsible for the different minigaps have different translational properties; by comparison of Figs. 1(a) and 1(b) it becomes obvious that the lowest minigap arises from a potential of wave vector \bar{Q} , while the second comes from a uniform (zero wave vector) potential. Finally, in all the calculations the matrix element is proportional to the normal electric field at the interface; i.e., to N if the small contribution from the depletion charge is neglected. Thus the simplest general form for the variation of the minigap with N and \bar{k} is

$$E = N(A + 2Bk_x k_y) \quad (5)$$

Since, in our samples, the tilt is about ξ , parallel to $\langle 110 \rangle$, it is convenient to rewrite this:

$$E_g = N[A + B(k_\xi^2 - k_\eta^2)] \quad (6)$$

Our fitting procedure ignored any \bar{k} dependence of E_g ; hence Eq. (5) must be appropriately averaged over \bar{k} space. The predominant contribution to the absorption, and still more to its derivative, comes from regions of \bar{k} space close to the two points where the Fermi circle cuts the minizone boundary. At these points (which only exist for $k_F > k_c = \frac{1}{2}Q - k_0 \sin\theta$), $k_\eta = k_c$, and $k_\xi^2 = k_F^2 - k_c^2$, so that

$$E_g = N[A + B(\frac{1}{2}k_F^2 - k_c^2)] \\ = N[A - BN_c/\pi + BN/2\pi] \quad (7)$$

where $N = k_F^2/\pi$ and $N_c = k_c^2/\pi$. Equation (7) has the same form as the empirical relation (3), with $B = 2\pi b$ and $A = a + 2bN_c$. A is plotted in Fig. 8; like a , it varies roughly as $\sin\theta$, while B is independent of θ within experimental accuracy. Theory suggests that B should vary as θ^2 for small θ ^{7,8}; this is quite inconsistent with our data.

We now discuss briefly the possible origin of the dominant A term. Consider the mathematical plane interface as a zeroth order approximation, and impose the discrete atomic structure of the crystal on it. It is reasonable to neglect the effects of reconstruction since this will produce only short-wavelength terms in the potential. The predominant extra potential of long wavelength has the form $V \sim V_Q e^{iQ\eta}$, which has the correct symmetry to produce the lowest minigap. V_Q can arise either from a macroscopic strain, or from the atomic structure itself. Note that if \bar{Q} is parallel to a cube axis, the contribution to V_Q from the atomic structure no longer has

the required xy symmetry, and the A term vanishes. This is consistent with experiment.¹⁰ Experimentally it appears that $V_Q \propto \sin|\theta|$, a reasonable dependence. Insufficient data on the second minigap are available to distinguish between the contributions of a static potential and of kinetic energy effects, although the possibly weaker dependence of E_g on N suggests that the former is dominant. If this is the case, the minigap must be due to strain, since there can be no intrinsic potential of xy symmetry with zero wave vector.

We conclude that the lowest minigap varies with electron density in the expected manner. However, a quantitative understanding of its size and angular dependence requires a theory which takes the atomic nature of the interface into account. The magnitude of the second minigap is reasonable, but requires further experimental study.

ACKNOWLEDGMENTS

We are most grateful to M. Nakayama, Frank Stern, and T. Ando for stimulating theoretical discussions, and for communicating unpublished work. Our thanks are also due to G. Kaminsky for help with the sample preparation, to S. J. Allen, Jr., for many helpful discussions and for the use of his computer program to calculate $\alpha(\omega)$. This program was ably adapted to our purposes by K. F. Rodgers, now deceased, who also provided expert technical assistance in the early stages of this work. We are also grateful to J. R. Brews, G. A. Baraff, and R. Roedel for helpful comments on the manuscript.

¹T. Cole, A. A. Lakhani, and P. J. Stiles, Phys. Rev. Lett. **38**, 722 (1977).

²L. J. Sham, S. J. Allen, Jr., A. Kamgar, and D. C. Tsui, Phys. Rev. Lett. **40**, 472 (1978); P. J. Stiles, T. Cole, and A. A. Lakhani, J. Vac. Sci. Technol. **14**, 909 (1977).

³D. C. Tsui, M. D. Sturge, A. Kamgar, and S. J. Allen, Jr., Phys. Rev. Lett. **40**, 1667 (1978).

⁴V. A. Volkov and V. B. Sandomirski, Pis'ma Zh. Eksp. Teor. Fiz. **27**, 668 (1978) [JETP Lett. **27**, 651].

⁵M. Nakayama and L. J. Sham, Solid State Commun. **28**, 393 (1978).

⁶L. J. Sham, in *Proceedings of the 14th International Conference*

on Physics of Semiconductors, Edinburgh, 1978, edited by B. H. L. Wilson (IOP, London, 1978) p. 1339.

⁷F. J. Ohkawa, J. Phys. Soc. Jpn. **45**, 1427 (1978).

⁸M. Nakayama (unpublished).

⁹D. C. Tsui, A. Kamgar, and M. D. Sturge, in *Proceedings of the 14th International Conference on Physics Semiconductors, Edinburgh, 1978*, edited by B. L. H. Wilson (IOP, London, 1978), p. 1335.

¹⁰W. Sesselmann and J. P. Kotthaus, Solid State Commun. **31**, 193 (1979).

¹¹T. Y. Chang, IEEE Trans. Microwave Theory Tech. **22**, 983 (1974).