Effects of higher sub-band occupation in (100) Si inversion layers*

W. E. Howard and F. F. Fang
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 6 October 1975)

At low temperatures and for moderate surface charge densities, electrons in inversion layers on (100) p-type Si are two-dimensional in character; i.e., they occupy only one electric quantum level, or sub-band. For electron densities greater than about 6×10^{12} per cm^2, depending upon substrate doping, we have observed deviations from two-dimensional behavior in oscillatory magnetoconductance (Shubnikov–de Haas) measurements, which are interpreted as evidence of the onset of occupation of a higher sub-band. The effects of such occupation of higher sub-bands have been calculated for the four different resistivities used in the experiments. The observed change in effective density of states [(10–20)%] and its dependence upon substrate resistivity are in good agreement with theory. The experimental values for the threshold charge densities for higher sub-band occupation are higher than the theoretical values by a factor of about 1.5.

I. INTRODUCTION

It is well known that the properties of electrons in inversion layers on Si are influenced by the quantization of the motion of the electrons in the direction perpendicular to the surface.[1] (This is a consequence of the strong electric field associated with the inversion layer, which forms a narrow potential well at the surface.) In particular, it has been demonstrated that at low temperature the inversion-layer electrons on p-type Si exhibit a density of states which is independent of energy.[2] Such behavior is characteristic of a two-dimensional electron gas and therefore indicates that the system is in the electric quantum limit, with all the electrons in a single sub-band corresponding to the lowest eigenstate for motion normal to the surface.

If the situation is viewed in the effective-mass approximation, the sixfold degeneracy of the conduction ellipsoids in Si is in general lifted by the process of confinement in the potential well,[2] since for a given crystal orientation of the surface the effective masses perpendicular to the surface may be different for the various pairs of ellipsoids. Indeed, for the case of (100) surfaces on Si, which have been most often used for quantization experiments, the lowest sub-band corresponds to the twofold set of ellipsoids with major axis normal to the surface. Since the observed degeneracy and transverse effective mass for the lowest sub-band are consistent with the expectations of the effective-mass analysis,[3] there has been considerable interest in comparing actual sub-band separations with effective-mass predictions.

We report here experimental evidence of the occupation of higher sub-bands even at low temperatures, for surface-electron densities greater than $(5–8)\times10^{12}$, the threshold for occupation being dependent upon the resistivity of the bulk silicon. In each case the experimental value for the electron density at threshold is approximately 50% higher than expected from effective-mass-type calculations, although the variation with substrate resistivity is as expected. In addition, the manner in which the occupation of the higher sub-band manifests itself provides quantitative confirmation of one of the key features of the self-consistent calculations of sub-band energies. Recently, Tsui and Kaminsky[4] reported a similar observation of higher sub-band occupation, also on a (100) Si surface, but for a single substrate resistivity of approximately 25 Ω cm.

II. EXPERIMENTS

The experiments were carried out on circular metal-oxide-semiconductor field effect transistors, similar to those described in earlier works,[3] with channel lengths of ~10 μm and channel widths of ~500 μm. The SiO_2 gate insulation layers varied in thickness from 1000 to 5500 Å for different samples. The experimental technique employed was the measurement of Schubnikov–de Haas-type oscillations of surface transconductance as a function of electric field perpendicular to the surface, for constant magnetic field also perpendicular to the surface. Essentially this is the same technique that was used to demonstrate that at low charge densities only one sub-band is occupied; what we have done is focus on the deviations from single-sub-band behavior at high electron densities. The transconductance g_m, defined as the first derivative of source-drain current I_{SD} with respect to gate voltage V_G, was recorded as a function of V_G, since the increase of gate voltage beyond a threshold value is proportional to the induced electron density in the channel. A fixed value of magnetic field of 90 kOe was used and sample temperatures were nominally between 1.2 and 1.4 °K. Figure 1 shows the results for a sample with boron substrate doping level of 1.6×10^{15}/cm^3 (~ 8.5 Ω cm at room temperature), having a gate oxide thickness

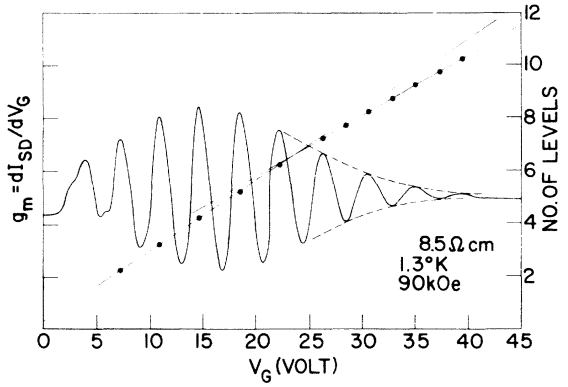

FIG. 1. Oscillatory transconductance (first derivative of source-drain current I_{SD} with respect to gate voltage V_0) as a function of gate voltage for an n-type inversion layer on an 8.5-Ω cm (100) p-type Si substrate, with $H = 90$ kOe perpendicular to the surface. The dots are the Landau level indices plotted against the voltages of the transconductance extrema. The slope of the line through the dots is proportional to the inverse of the density of states of the inversion layer electrons.

of approximately 1000 Å. Since the density of states for the lowest sub-band in the absence of a magnetic field is constant, so is the number of states per Landau level. Superimposed on this figure is a plot of the number of Landau levels occupied versus gate voltage. The points are taken from the extrema of the oscillations, taking into account the envelope as a reference. Each extremum represents a gate voltage (or total surface electron density) at which a given portion of a Landau level has been filled. Thus a plot of Landau index versus corresponding gate voltage gives a curve the slope of which is a measure of the number of electrons required to fill one Landau level. The inverse of this slope is consequently a measure of the effective density of states of the sub-band. As can be seen in Fig. 1, there is clearly a break in the curve of levels versus voltage, indicating a departure from a constant density of states. We regard this break in density of states an evidence of the onset of a population of higher sub-bands in the inversion layer. The corresponding carrier density Q_B/q at which this onset occurs is about $5.8\times10^{12}/\text{cm}^2$, based on the voltage and the oxide thickness, as determined by the period of the oscillations. The fractional change in slopes at the breakpoint, which we denote by $(S_1 - S_2)/S_2$, in this case is equal to 0.12 ± 0.05; this represents the fractional change in effective density of states for the lowest sub-band.

Figure 2 shows the corresponding results for a sample having a substrate resistivity of 1.9 Ω cm. In this case the break occurs at a density $7.6\times10^{12}/\text{cm}^2$ and the fractional slope change is 0.18 ± 0.05. We have also examined samples with substrate resistivities of 5.4 and 100 Ω cm; the results are summarized in Table I. In the case of the 100-Ω cm sample, the slope change was too small to be detected in comparison to our experimental uncertainty.

TABLE I. The electron densities Q_B/q beyond which the effective inverse densities of states change and the corresponding fractional changes, for four different substrate resistivities. Both theoretical and experimental values are listed for comparison.

ρ (Ω cm)	Q_B/q (cm^{-2})		$(S_1 - S_2)/S_2$	
	Theor.	Expt.	Theor.	Expt.
1.9	5.5×10^{12}	7.6×10^{12}	0.20	0.18 ± 0.05
5.4	4.3×10^{12}	6.7×10^{12}	0.16	0.19 ± 0.05
8.5	3.9×10^{12}	5.8×10^{12}	0.15	0.12 ± 0.05
100	2.3×10^{12}	...	0.06	...

III. DISCUSSION

The fact that the effective density of states for the lowest lying sub-band changes by such a relatively small amount [i.e., (10–20%)] after the onset of higher sub-bands may at first seem surprising, since the actual densities of states for the next two higher sub-bands are equal to, and about 4.5 times, the density of the lowest sub-band, for the excited twofold band and ground state of the fourfold band, respectively. This is, however, a direct consequence of the strong self-consistent component of the inversion layer potential, wherein the sub-band energies themselves are greatly dependent upon the distribution of the induced electron population among the various sub-bands. Indeed, the experimental results can be seen as confirming the theoretical expectations in this regard.

In order to compare our experiments with theory, we have carried out the calculation of self-consis-

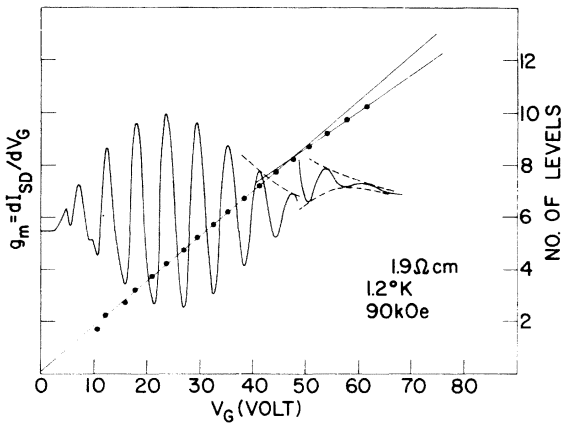

FIG. 2. Oscillatory transconductance as a function of gate voltage for a 1.9-Ω cm sample (see caption Fig. 1).

tent energies for the lowest three sub-bands in a (100) Si surface as a function of induced electron density, for the four substrate resitivities indicated in the table. The calculations were done in the limit of 0 °K.

The approach used is similar to that described by Stern[5] and is, for the special case treated here, simply an extension of that used in an earlier paper[2] on (100) Si surfaces. The effective-mass equation for the envelope function in the presence of a surface potential energy $V(z)$ is given by

$$\tfrac{1}{2}\hbar^2(\underline{m}^{-1}\cdot\nabla)\nabla\psi+[\mathcal{E}-V(z)]\psi=0, \tag{1}$$

where $\underline{m}$ is the effective-mass tensor for an ellipsoidal energy surface. In the case of a (100) surface, the effective-mass equations for the two inequivalent sets of two and four ellipsoids are both separable and one has the one-dimensional equations

$$\frac{\hbar^2}{2m_d}\frac{d^2\psi_{d,i}}{dz^2}+[\mathcal{E}_{d,i}-V(z)]\psi_{d,i}=0, \tag{2}$$

where d indicates the degeneracy for the equivalent ellipsoids (two or four) and $\mathcal{E}_{d,i}$ is the energy of bottom or the ith sub-band of the sequence. In this notation m_2 is equal to the longitudinal mass and m_4 the transverse mass for Si. $\psi_d(0)$ and $\psi_d(\infty)$ are both assumed to be equal to zero. The potential energy is given by the Poisson equation

$$\frac{d^2V}{dz^2}=-4\pi q\epsilon^{-1}Q(f_{2,0}|\psi_{2,0}|^2+f_{2,1}|\psi_{2,1}|^2+f_{4,0}|\psi_{4,0}|^2), \tag{3}$$

where q is the electronic charge, Q is the induced free-electron density, ϵ is the dielectric constant for Si, and the f's are the fractional populations of the three lowest-lying sub-bands being considered. The depletion layer charge density, which is ignored here, is taken into account through the boundary condition

$$\left.\frac{dV}{dz}\right|_{z\to\infty}=4\pi q\epsilon^{-1}Q_{sc}, \tag{4}$$

where Q_{sc} is the total ionized acceptor space charge in the depletion layer. (This is justified by the narrow extent of the inversion layer wave functions relative to the widths of the depletion region.) With the introduction of the dimensionless quantities $E=\mathcal{E}/[(\hbar^2/2m)^{1/3}(4\pi qQsc/\epsilon)^{2/3}]$ and $Z=z/(\hbar^2\epsilon/8m_0qQ_{sc})^{1/3}$, Eqs. (2)–(4) are transformed to become

$$\frac{d^2\psi_{d,i}}{dZ^2}+\frac{m_d}{m_0}[E_{d,i}-V(Z)]\psi_{d,i}=0,$$

$$\frac{d^2V}{dZ^2}=\alpha(f_{2,0}|\psi_{2,0}|^2+f_{2,1}|\psi_{2,1}|^2+f_{4,0}|\psi_{4,0}|^2), \tag{5}$$

$$\left.\frac{dV}{dZ}\right|_{Z\to\infty}=1,$$

where $\alpha=Q/Q_{sc}$. The fractional populations are determined from the 0 °K condition that

$$\left(\frac{dN}{dE}\right)_{2,0}E_F+\left(\frac{dN}{dE}\right)_{2,1}(E_F-E_{2,1})$$

$$+\left(\frac{dN}{dE}\right)_{4,0}(E_F-E_{4,0})=\frac{Q}{q}, \tag{6}$$

where, of course, only positive quantities are considered. Since the density of states $(dN/dE)_{4,0}=4.5\,(dN/dE)_{2,0}$ and $(dN/dE)_{2,1}=(dN/dE)_{2,0}$, one can define $E_{F0}=Q_{sc}/q(dN/dE)_{2,0}$ such that (6) becomes

$$E_F+(E_F-E_{2,1})+4.5\,(E_F-E_{4,0})=\alpha E_{F0}. \tag{7}$$

Equations (5) and (2) are solved iteratively as a function of α for various values of E_{F0} corresponding to different substrate resitivities. In Fig. 3 we show some of the results for three different resistivities. In the upper left-hand side portion of the figure are plotted the energies of the first excited sub-band for the two valley set and the lowest-lying four-valley sub-band, relative to the bottom of the lowest subband. In such an E-vs-α plot the energy variations for only the lowest sub-band occupied are independent of substrate resistivity and the curves are therefore superimposed. The dashed curves represent the Fermi energies in the three cases. Only a portion of the 100-Ω cm results have been plotted to show the trend with doping. In the lower right-hand side portion of the figure we have plotted the fractional population in the upper sub-bands as a function of α for the three cases.

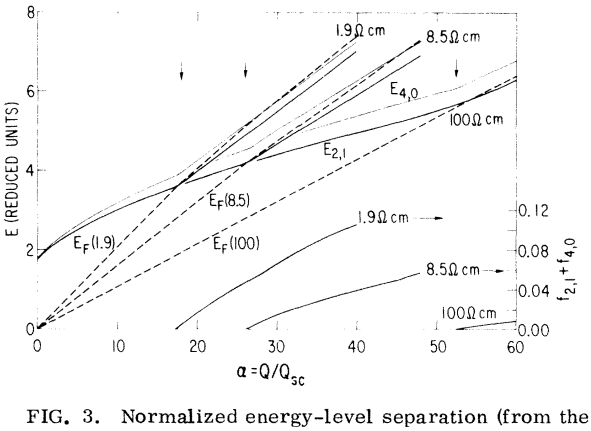

FIG. 3. Normalized energy-level separation (from the lowest sub-band) for the two lowest-lying upper sub-bands as a function of the normalized electron density, for 1.9-, 8.5-, and 100-Ω cm substrate resistivities. The Fermi energy is also shown for each substrate doping level. The curves in the lower right-hand side portion of the figure represent the calculated fractional populations of the upper sub-bands for each case of substrate doping.

One notices in each case that above the intersection of the Fermi energy with the next higher sub-band ($E_{2,1}$) the population of that sub-band takes place at a rather slow rate because the sub-band itself is moving away at an increased rate with added electron density. This is a result of the change in shape of the channel potential as the higher sub-bands, with their less confined wave functions, are occupied and contribute to the potential. It is interesting that in each case, as the second sub-band is being populated, the variation of the Fermi level is still linear with electron density, but the slope has been reduced by (10–20%), depending upon doping. This theoretical slope change is compared with the experimental results in the table. Note in particular the results for 100-Ω cm substrate. Since in this case the expected fractional slope change is only 0.06, it is not surprising that no change in period showed up in early measurements (made mostly with high resitivity samples) and that none could be unequivocally seen in the present work for 100-Ω cm substrates. Actually, we feel that we have seen the expected small change for 100-Ω cm samples, but our experimental uncertainty is such that we have omitted these results from the table.

The agreement between theory and experiment, in regard to both magnitude of the fractional change of effective density of states and its dependence on doping, represents a quantitative confirmation of one important aspect of the Hartree self-consistent treatment of the inversion layer. Less gratifying is the agreement between observed and predicted values for the charge densities Q_B at which the density of states curves break, i.e., beyond which upper sub-bands begin to become populated. Clearly, the experimental values are consistently and significantly higher, specifically by a factor of about 1.5. This can only be interpreted as signalling a shortcoming in the theoretical treatment of the problem. There are several possible corrections to the theory which would improve the situation, but we have no basis as yet for determining which of these is most significant. Within the effective mass approximation, nonparabolicity effects would be expected to lower further the lowest sub-band. However, in view of the results of Smith and Stiles,[6] showing the transverse mass decreasing with increasing electron density, this is unlikely. Also, deviations from effective-mass theory, due to the cell perodic factor in the potential energy becoming less than one, can lower the ground-state sub-band preferentially. This qualitative conclusion is based on computations of the triangular well case wherein the cell function integral $\int u_k u_{k'} d\vec{r}$ was set equal to zero for $|k-k'|>k_c$, an arbitrary parameter. Finally, the inclusion of exchange and correlation effects may possibly account for the larger observed values of Q_B. Vinter[7] has recently calculated both exchange and correlation effects for the case in point here of (100) Si inversion layers. The magnitudes of the corrections he obtains are too large to explain our results. It appears that further refinements of the model as well as more experimental data are needed to resolve the quantitative discrepancies.

In addition to the discrepancy between our experimental and theoretical results, there is also a significant quantitative difference between our results and those of Tsui and Kaminsky, mentioned earlier. Whereas our results imply, by interpolation, a threshold density of approximately $4.7\times10^{12}/\text{cm}^2$ for occupation of the next higher sub-band in 25-Ω cm Si, they observed a value of $7.4\times10^{12}/\text{cm}^2$. Also, our theoretical expectation for the change in apparent density of states at the break for such a resistivity would be less than 12%; they observed a change of approximately 20%. Although our theoretical results predict that the next higher sub-band to be occupied would be the excited states of the two-valley set, Tsui and Kaminsky claim on the basis of strain experiments that the lowest four valley sub-band is occupied next. Since the two sub-bands are close in energy, as Fig. 3 indicates, it would not be surprising if some built-in strain caused an interchange of their relative position in energy. Moreover, it can be shown theoretically that such an interchange would not be expected to affect appreciably the change in effective density of states at the break; thus the change in slope does not provide a means of identifying the higher sub-band.

IV. SUMMARY

We have observed changes in the effective density of states for the lowest lying sub-band in the Si (100) inversion layer which is consistent with the onset of occupation of the next higher sub-band. The magnitude of the change in apparent density of states and its dependence upon substrate doping is in good agreement with the predictions of a simple Hartree self-consistent calculation. The charge densities for which the onsets of higher sub-band occupation occur are consistently higher by a factor of about 1.5 than predicted by the simple theory.

ACKNOWLEDGMENTS

We are pleased to acknowledge helpful discussions with A. B. Fowler, F. Stern, and P. J. Stiles.

*Some of the results in this paper were presented at the American Physical Society Meeting in Berkeley, Calif. in 1968.

[1]See, for example, F. Stern and W. E. Howard, Phys.

Rev. 163, 816 (1967).
[2]F. F. Fang and W. E. Howard, Phys. Rev. Lett. 16, 797 (1966).
[3]A. B. Fowler, F. F. Fang, W. E. Howard, and P. J. Stiles, Phys. Rev. Lett. 16, 901 (1966); and J. Phys. Soc. Jpn. Suppl. 21, 331 (1966).
[4]D. C. Tsui and G. Kaminsky, Phys. Rev. Lett. 35, 1468 (1975).
[5]F. Stern, Phys. Rev. B 5, 4891 (1972).
[6]J. L. Smith and P. J. Stiles, Phys. Rev. Lett. 29, 102 (1972).
[7]B. Vinter, Phys. Rev. Lett. 35, 598 (1975).