Theory of the Influence of Misfit Dislocations on Interfacial Mobility and Hall Effect

G. F. NEUMARK
Philips Laboratories, Briarcliff Manor, New York 10510
(Received 28 April 1969)

The interfacial mobility and Hall effect have been calculated on the basis of a model of misfit dislocations. It is assumed that the dislocation regions can be represented two-dimensionally as a network of continuous lines containing "trapped" carriers of low mobility, and that they result in interference with current flow between the dislocation-free regions. The mobility and Hall effect are obtained as a function of the area, mobility, and carrier-density ratios in the two regions. Both the conductivity and Hall mobilities show maxima in dependence on the ratio of carrier densities, and the results also show that the Hall effect does not give a good measure of the carrier concentration except over a limited range. The theoretical mobilities are compared with experimental results reported in the literature on the dependence of mobility on gate bias for Si metal-oxide-semiconductor units in inversion operation. For this comparison, some consideration of the additional role of quantum effects was included. It is shown that such effects probably are important mainly at high inversion, with the misfit-dislocation effects dominant at low inversion. Comparison of the present results with the low-inversion experimental data is very satisfactory, especially with the inclusion of some quantum corrections.

I. INTRODUCTION

IT has recently been suggested[1] that so-called misfit dislocations are crucial in determining carrier mobilities in Si–SiO_2 interfacial inversion layers. It is the aim of the present paper to carry out a more exact calculation of mobility reduction resulting from such misfit dislocations, as well as to include an analysis of the Hall mobility and of the Hall coefficient.

The results to be obtained here will be based on a geometrical model of a square grid of continuous dislocation lines (Fig. 1). Reasons for assuming such a model to be typical and also the physical justification for the dislocation mobility reduction effect in general will be discussed in Sec. II. It can be noted that the present model of continuous lines is slightly different from that of Ref. 1, which consisted of long but discontinuous ellipsoidal rods. The present model has the advantage that it is possible to obtain both Hall and conductivity results over the entire range of the model parameters. Moreover, results of the two models are qualitatively comparable over reasonably wide ranges of the parameters.[2]

Following the discussion of the physical situation in Sec. II, the derivation of the equations and representative quantitative results will be given in Sec. III. Subsequently, Sec. IV discusses interrelations of the present theory with surface quantization effects and with experimental data of Fang and Fowler.[3] Also included in Sec. IV is a qualitative analysis of possible refinements and a consideration of the present work in relation to surface mobility in general. Over-all conclusions are presented in Sec. V.

[1] G. F. Neumark, Phys. Rev. Letters **21**, 1252 (1968).
[2] G. F. Neumark, *International Conference on Properties and Use of M.I.S. Structures, Grenoble, 1969*, edited by J. Borel (Centre d'Etudes Nucléaires de Grenoble, Grenoble, 1969), p. 337.
[3] F. F. Fang and A. B. Fowler, Phys. Rev. **169**, 619 (1968).

II. MISFIT-DISLOCATION MODEL

The essential hypotheses of the present treatment are that misfit dislocations are present at the Si–SiO_2 interface, that so-called "surface" states of this system are in fact states of the misfit dislocations, and that in inversion such dislocations are charged and consequently impede the charge carrier flow.

Regarding the main point of the occurrence of misfit dislocations at interfaces, this assumes that lattice mismatch at the interface is accommodated by formation of fairly continuous linear dislocation regions instead of formation of random point defects. Such an effect has long been predicted on energetic grounds,[4] and has more recently been confirmed on a number of semiconductor and other systems.[5] These dislocations would be expected to form a two-dimensional grid of lines or long rods located in the plane of the interface, with the lines of this grid located along appropriate crystallographic directions; moreover, this type of grid structure has been found experimentally.[5] Thus the model of continuous dislocation lines in a square grid is representative of a regular crystallographic structure. It will also be assumed that the electric field is applied parallel to one set of grid lines, with the current flow consequently perpendicular to the other set. A change of geometry is not expected to change the main results, although there will be differences in detail.

Misfit dislocations, like imperfections in general, are expected to introduce electronic states within the forbidden gap. Thus, electronically and spatially, such

[4] F. C. Frank and J. H. van der Merwe, Proc. Roy. Soc. (London) **A198**, 205 (1949); **A198**, 216 (1949). For more recent work, see, for example, J. H. van der Merwe, J. Appl. Phys. **34**, 117 (1963); **34**, 123 (1963); **34**, 3420 (1963); N. H. Fletcher and P. L. Adamson, Phil. Mag. **14**, 99 (1966).
[5] G. O. Krause and E. C. Teague, Appl. Phys. Letters **10**, 251 (1967), and references given there; R. S. Mroczkowski, A. F. Witt, and H. C. Gatos, J. Electrochem. Soc. **115**, 750 (1968); G. B. Stringfellow and P. E. Greene, J. Appl. Phys. **40**, 502 (1969).

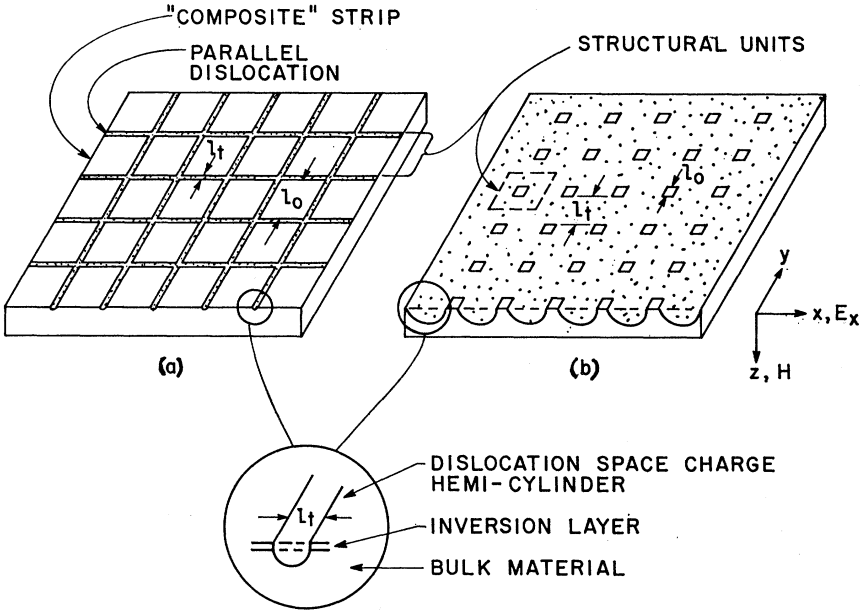

FIG. 1. Geometrical model used for the misfit dislocation. (a) shows the case that the area of the dislocations and associated space-charge cylinders is small compared to the over-all area, (b) shows the case where the dislocation area is large. The respective structural units used in the calculations are also shown. The dislocation space-charge cylinder is assumed to always extend throughout the entire inversion region (insert).

states would act as surface states. In consequence, it seems plausible to assume that the experimentally observed[6–8] surface states at the interface[9] (also referred to as "fast"[7] or "interface"[8] states), are in fact the states of the misfit dislocations. In favor of this model is the good fit[1] which it gives to mobility-surface state data of Arnold and Abowitz.[6] In addition, it provides a logical basis for a number of observed surface state characteristics, as previously discussed in Ref. 1. The surface states are thus not homogeneously distributed (in the surface plane) but are bunched along the dislocation lines, and their average area density is equal to the product of the dislocation area density times the linear state density along the dislocation. A consequence of the resultant spatial continuity is that one expects at least some conduction along the dislocation lines by electrons "trapped" in these states. It should be noted that some conduction across the dislocations (i.e., perpendicular to them) also seems likely, and will be assumed equal to the parallel conduction.

A further assumption of the present model is that the dislocations are charged in inversion. This type of behavior follows if the states in the upper half of the gap are of acceptor type, and hence negatively charged when occupied by electrons. Correspondingly, the states in the lower half of the gap would be of donor type. Experimental evidence for such an energy-level structure of the surface states has been presented by Gray and Brown[7] and by Whelan.[8]

Charged dislocations result in mobility reduction, since consequent repulsion of free charge gives rise to a poorly conducting space-charge cylinder (or strictly speaking, a hemicylinder for surface geometry) around the dislocations. This cylindrical region in turn causes a displacement of the current lines. A mechanism of this type has long been postulated for bulk dislocations.[10] However, in the present instance, mobility reduction can be much more effective than in bulk material. This follows if the space-charge cylinder extends down in the z direction (insert of Fig. 1) far enough to fill the entire depth of the inversion region (the "channel"). Under this condition the flow of the "free" carriers is completely blocked, i.e., conduction can only take place via current flow through the dislocations; there can be no carrier flow around the dislocations, as there would be in the bulk. Moreover, it appears very likely that the space-charge cylinder does extend through the entire channel depth. Theoretical results[11] indicate that for the usual parameter ranges of metal-oxide-semiconductor (MOS) systems the channel depth for the (100) surface is $\lesssim 60$ Å; substitution of the appropriate effective mass from Stern and Howard[11] Table I into their Eqs. (22) and (42) gives a corresponding channel depth of $\lesssim 240$ Å for the (111) surface. Experimentally, space-charge cylinders around dislocations in Si have

[6] E. Arnold and G. Abowitz, Appl. Phys. Letters 9, 344 (1966); E. Arnold, Trans. IEEE **ED-15**, 1003 (1968).

[7] P. V. Gray and D. M. Brown, Appl. Phys. Letters **8**, 31 (1966); D. M. Brown and P. V. Gray, J. Electrochem. Soc. **115**, 760 (1968).

[8] M. V. Whelan, Philips Res. Rept. **22**, 289 (1967).

[9] It is known that additional charge, variously referred to as "built-in" (Ref. 6) "fixed" (Ref. 7), or "oxide" (Ref. 8) charge, is induced by work-function differences and by charge located within the oxide, and also influences various properties of the MOS system. We assume that mobility does *not* depend on this type of charge, which will therefore be ignored in the present paper.

[10] W. T. Read, Jr., Phil. Mag. **46**, 111 (1955). For a more recent review and extension, see R. Broudy, Advan. Phys. **12**, 135 (1963).

[11] F. Stern and W. E. Howard, Phys. Rev. **163**, 816 (1967); F. Stern, Phys. Rev. Letters **21**, 1687 (1968).

been shown to extend over the order of microns.[12] It can be noted that for the extension of the dislocations through the entire channel depth, the mobility and Hall coefficient calculations become a two-dimensional problem.

Based on the above over-all view of the role of misfit dislocations, the mobilities and the Hall coefficient will be derived in Sec. III as a function of three parameters: (1) the density of the dislocations, (2) the ratio of the mobility in the dislocations to that in the "bulk," and (3) the corresponding carrier ratio. For comparison to experimental results it would also be desirable to know the dependence on gate bias. This would require additional assumptions but it is obvious in any case that the dependence on gate bias will be qualitatively similar to that on the carrier ratio, since in accordance with the Fermi distribution the ratio of free to trapped charge will be a monotonically increasing function of the degree of inversion, and thus will depend correspondingly on gate bias. In the present paper, the dependence on the carrier ratio will therefore be used as a qualitative measure of the variation with gate bias; a more detailed analysis of this aspect is planned for a subsequent paper.

III. THEORY

A. Model and Definitions

In the present treatment the dislocation space-charge cylinders are represented by a two-dimensional regular square array of lines of width l_t, with islands of "bulk" material of width l_0 in between (Fig. 1). Two subcases, with a different mathematical approximation for each, will be considered: (1) The "dislocation density," i.e., the fractional area of the dislocation lines and their associated space-charge cylinders, is small [Fig. 1(a)]; and (2) the dislocation density is large [Fig. 1(b)]. Defining

$$\gamma \equiv l_t/l_0, \tag{1}$$

$$\epsilon = \text{relative area of bulk material} = (1+\gamma)^{-2}, \tag{2}$$

the first case corresponds to $\gamma \ll 1 (\frac{1}{2} \lesssim \epsilon < 1)$, the second to $\gamma > 1 (\epsilon \ll 1)$.

The basic methods of calculation for the two cases are the following: For a low dislocation density ($\gamma \ll 1$), we use a "strip" structural unit, [Fig. 1(a)] with this unit consisting of a (continuous) "parallel" dislocation, i.e., a dislocation parallel to the applied electric field (the x direction), plus a contiguous "composite" strip consisting of the islands of bulk material plus the "perpendicular" dislocations (those in the y direction); the conductivity and Hall effect are obtained by adding the contributions of the parallel dislocation and of the composite strip. For the case of high dislocation densities, it is more appropriate to use as structural unit a square of side (l_0+l_t), consisting of one (small) bulk island and a half-width of dislocation on each side [Fig. 1(b)]. Replacing the squares by circles of equal area and assuming no interaction between units, one can use the approach outlined by Juretschke *et al.*[13] for nonconducting inclusions. This involves solution of the Laplace equation subject to the boundary conditions of continuity of potential and normal current between the bulk and the dislocations, and of zero average y current. The conductivity and Hall effect are then obtained by averaging over the unit cell. Equations for the small-γ case will be given in terms of γ, whereas use of ϵ is more convenient for the large-γ case; for a given physical variable, equations in the former case are labeled "a," and those in the latter "b." Since it will be shown in the following sections that for intermediate values of $\gamma (\gamma \sim 1)$ both approaches yield similar results, it follows that the basic results for mobility and Hall coefficient are independent of the mathematical approximation.

The various required parameters include the carrier density (per unit area), conductivity mobility, Hall mobility, Hall coefficient (per unit area), and conductivity (per unit area); within the dislocation regions these are denoted by n_t, μ_t, μ_t^H, R_t, and σ_t, respectively, and by n_0, μ_0, μ_0^H, R_0, and σ_0 for the bulk regions. The measured quantities, which are the appropriate averages over the whole sample, are denoted by $\bar{n}$, $\bar{\mu}$, $\bar{\mu}^H$, $\bar{R}$, and $\bar{\sigma}$, with the additional use of $\bar{J}$ for average current density and $\bar{E}$ for average field. Other quantities of interest are the Hall coefficient ($\equiv R_c$) and the conductivity ($\equiv \sigma_c$) for the composite strip. In addition, we will use the subsidiary definition

$$\beta \equiv \mu_t/\mu_0 \tag{3}$$

and will for the sake of simplicity use the same symbol for the ratio of Hall mobilities,

$$\beta = \mu_t^H/\mu_0^H. \tag{4}$$

It should also be noted that for results of physical interest one expects $\beta \ll 1$.

The average carrier concentration, i.e., the number of free plus trapped carriers per unit area, can immediately be obtained as

$$\bar{n} = [n_0 + \gamma(2+\gamma)n_t]/(1+\gamma)^2 \tag{5a}$$

or

$$\bar{n} = \epsilon n_0 + (1-\epsilon)n_t. \tag{5b}$$

B. Conductivity and Conductivity Effective Mobility

For the case of small γ, the average conductivity, to a good approximation, is given by the sum over that in a parallel dislocation (i.e., parallel to the current) plus that in the composite strip of mixed bulk material and perpendicular dislocations [Fig. 1(a)]

$$\bar{\sigma} = (l_0+l_t)^{-1}(l_0\sigma_c + l_t\sigma_t). \tag{6a}$$

[12] H. F. Mataré and C. W. Laasko, Appl. Phys. Letters **13**, 216 (1968); J. Appl. Phys. **40**, 476 (1969).

[13] H. J. Juretschke, R. Landauer, and J. A. Swanson, J. Appl. Phys. **27**, 838 (1956).

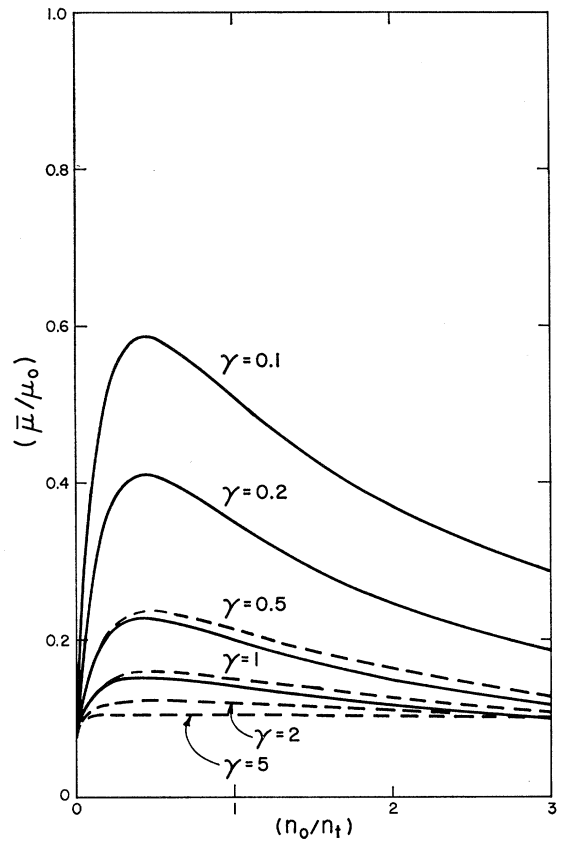

FIG. 2. Normalized conductivity mobility as a function of the carrier ratio n_0/n_t for various values of the parameter $\gamma=l_t/l_0$ and for a representative value of $\beta=\mu_t/\mu_0$ of 0.1. The solid lines are calculated based on the theory appropriate for a low density of dislocations [Eq. (9a)]. The dashed lines are for the high dislocation case [Eq. (9b)]. The two cases are compared at intermediate dislocation densities ($\gamma=0.5$, 1).

Moreover, it is obvious that

$$\sigma_c=(l_0+l_t)\sigma_0\sigma_t/(l_0\sigma_t+l_t\sigma_0). \tag{7}$$

The normalized average or effective mobility $(\bar{\mu}/\mu_0)$ is given by

$$\bar{\mu}/\mu_0=\bar{\sigma}n_0/\sigma_0\bar{n}, \tag{8}$$

and use of Eqs. (1), (3), (5a), (6a), and (7) yields

$$\frac{\bar{\mu}}{\mu_0}=\left[\frac{1}{\beta(n_t/n_0)+\gamma}+\frac{\gamma}{1+\gamma}\right]\left[\frac{(1+\gamma)^2\beta}{(n_0/n_t)+\gamma(2+\gamma)}\right]. \tag{9a}$$

The corresponding conductivity for large γ is obtained, analogously to Ref. 13, as

$$\bar{\sigma}=\bar{J}_x/\bar{E}_x, \tag{10}$$

$$\bar{\sigma}=\sigma_t\left\{1-\epsilon\left[\frac{1-(\sigma_0/\sigma_t)}{1+(\sigma_0/\sigma_t)}\right]\right\}\left\{1+\epsilon\left[\frac{1-(\sigma_0/\sigma_t)}{1+(\sigma_0/\sigma_t)}\right]\right\}^{-1}, \tag{6b}$$

and the average mobility then follows by use of Eqs. (3), (5b), and (8) as

$$\frac{\bar{\mu}}{\mu_0}=\beta\left\{1-\epsilon\left[\frac{\beta-(n_0/n_t)}{\beta+(n_0/n_t)}\right]\right\}\left\{1+\epsilon\left[\frac{\beta-(n_0/n_t)}{\beta+(n_0/n_t)}\right]\right\}^{-1}\times\left[1-\epsilon\left(1-\frac{n_0}{n_t}\right)\right]^{-1}. \tag{9b}$$

The results obtained from Eqs. (9a) and (9b) are shown in Fig. 2 for a representative value of $\beta=0.1$; results from Eq. (9a) are given by the solid lines, and from Eq. (9b) by the dashed. It can be seen that in the intermediate range ($\gamma=0.5$ and $\gamma=1$) the two equations give quite similar curves.

It can also be seen from Fig. 2 that the mobility goes through a maximum as a function of n_0/n_t. By differentiation of Eqs. (9a) and (9b), one obtains, for $\beta\ll1$

$$(n_0/n_t)_{\mu=\max}\approx(2\beta)^{1/2}, \qquad \gamma\ll1 \tag{11a}$$

$$\left(\frac{\bar{\mu}}{\mu_0}\right)_{\max}\approx\frac{(1+\gamma)^2}{[1+\gamma(2/\beta)^{1/2}]^2}, \qquad \gamma\ll1 \tag{12a}$$

$$(n_0/n_t)_{\mu=\max}\approx\{[4\beta+(\beta-\epsilon^2)^2]^{1/2}-(\beta-\epsilon^2)\}, \qquad \epsilon\ll1 \tag{11b}$$

and, for $\epsilon^2\ll\beta$ (in addition to $\beta\ll1$)

$$(\bar{\mu}/\mu_0)_{\max}\approx\beta[1+\epsilon(3-4\beta^{1/2})], \quad \epsilon\ll1. \tag{12b}$$

It is also of interest to consider limiting values of mobility away from the maximum, for n_0/n_t either large or small. In the low-γ case: for a high free-carrier concentration, such that the conditions

$$n_0/n_t\gg\beta/\gamma, \tag{13a}$$

$$n_0/n_t\gg2\gamma \tag{14a}$$

are satisfied, Eq. (9a) gives

$$\bar{\mu}/\mu_0\approx(\beta/\gamma)(1+\gamma)^2(n_t/n_0). \tag{15a}$$

Thus the mobility in this case keeps decreasing with increasing free-carrier concentration; this follows since the over-all conductivity becomes limited by the relatively poorly conducting dislocation regions, while more and more carriers are going into the "useless" bulk regions. The corresponding situation for a low free-carrier concentration requires the inverse inequality (14a) as well as

$$n_0/n_t\ll\gamma\beta. \tag{16a}$$

The resultant mobility is

$$\bar{\mu}/\mu_0\approx\beta/2. \tag{17a}$$

In this case the conduction is entirely along the parallel dislocations, with the mobility ratio (β) modified by the fraction of (trapped) carriers located in the parallel dislocations (since carriers in the perpendicular dislocations do not contribute). For the high-γ case, quite

analogous results obtain. For high n_0/n_t, satisfying the inequalities

$$n_0/n_t \gg \beta, \tag{13b}$$

$$n_0/n_t \gg 1/\epsilon; \tag{14b}$$

the mobility given in Eq. (9b) reduces to

$$\bar{\mu}/\mu_0 \approx (\beta/\epsilon)(1+2\epsilon)(n_t/n_0). \tag{15b}$$

For low n_0/n_t, satisfying the inverse inequality (13b) and

$$n_0/n_t \ll 1 \tag{16b}$$

the corresponding result is

$$\bar{\mu}/\mu_0 \approx \beta/(1+\epsilon). \tag{17b}$$

As to further limiting mobility values, it can readily be determined that Eqs. (9a) and (9b) reduce properly to $\bar{\mu}/\mu_0=1$ for the case of no dislocations ($\gamma=0$) and $\bar{\mu}/\mu_0=\beta$ for no bulk material ($\epsilon=0$). The proper value of $\bar{\mu}/\mu_0=1$ is also obtained under the conditions $\beta=1$, $n_0/n_t=1$. It is also apparent that no conduction is obtained for $\beta=0$, and correspondingly Eqs. (9a) and (9b) give $\bar{\mu}/\mu_0=0$ in this case.

In summary—as regards the variation of the conductivity mobility with the ratio of free to trapped carriers: (1) There is a maximum at a value of n_0/n_t which for $\beta \ll 1$ and $\epsilon^2 \ll \beta$ is in the range of $(2\beta)^{1/2}$ to $2(\beta)^{1/2}$; (2) the mobility becomes constant at low values of the carrier ratio and decreases continuously at high values.

C. Hall Mobility

In the usual Hall measurement on MOS structures a magnetic field is applied in the z direction (Fig. 1), perpendicular to the surface, and a Hall voltage is measured in the y direction under the condition $\bar{J}_y=0$. (It can be noted that in the present inhomogeneous case the quantity J_y is not zero everywhere, but the condition $\bar{J}_y=0$ nevertheless applies as a "boundary" condition.)

For the low-γ case, the over-all unit (dislocation plus composite strip) will display a Hall voltage $\bar{E}_y(l_0+l_t)$. For an applied voltage V in the x direction the standard Hall-effect equations give

$$\bar{E}_y = \bar{R}H\bar{\sigma}(V/L), \tag{18}$$

where L is the length in the x direction. It is also apparent that $\bar{E}_y(l_0+l_t)$ will be the sum of the Hall voltages in the strips of parallel dislocations plus those in the composite strip. Consequently,

$$\bar{E}_y(l_0+l_t) = R_t H\sigma_t(V/L)l_t + R_c H\sigma_c(V/L)l_0. \tag{19}$$

Equating Eqs. (18) and (19) thus gives

$$\bar{R}\bar{\sigma} = (l_0+l_t)^{-1}(l_0 R_c\sigma_c + l_t R_t\sigma_t). \tag{20}$$

The normalized average mobility $(\bar{\mu}^H/\mu_0{}^H)$ is obtained by use of the standard equation

$$\mu^H = R\sigma_c \tag{21}$$

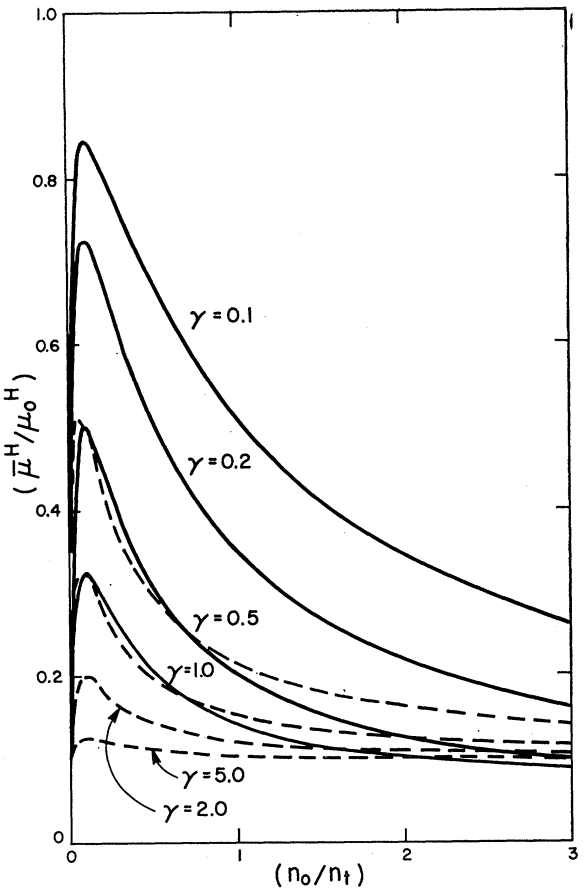

FIG. 3. Same as Fig. 2, but for the normalized Hall mobility [Eqs. (23)].

as

$$\bar{\mu}^H/\mu_0{}^H = (l_0+l_t)^{-1}[l_0\sigma_c + l_t\sigma_t(R_t/R_c)](R_c/R_0\sigma_0). \tag{22}$$

For strips which are wide in the y direction (small γ) the quantity R_c can be easily derived, as shown in the Appendix. Use of Eq. (A12), and of Eqs. (1), (4), and (7) gives

$$\frac{\bar{\mu}^H}{\mu_0{}^H} = \left[\frac{1}{\beta(n_t/n_0)+\gamma} + \frac{\gamma}{1+\gamma}\left(\frac{(n_0/n_t)+\gamma\beta}{1+\gamma\beta}\right)\right] \times \left(\frac{(1+\gamma\beta)\beta}{(n_0/n_t)+\gamma\beta}\right). \tag{23}$$

Alternatively, Eq. (23) can be put into a form to show its symmetry with respect to σ_t/σ_0 around the point $\sigma_t=\sigma_0$, as follows:

$$\frac{\bar{\mu}^H}{\mu_0{}^H} = \frac{1+\gamma\beta}{[1+\gamma(\sigma_t/\sigma_0)][1+\gamma(\sigma_0/\sigma_t)]} + \frac{\gamma\beta}{1+\gamma}. \tag{23a}$$

For the case of large γ, using the relation

$$\bar{\mu}^H(H/c) = \bar{E}_y/\bar{E}_x \tag{24}$$

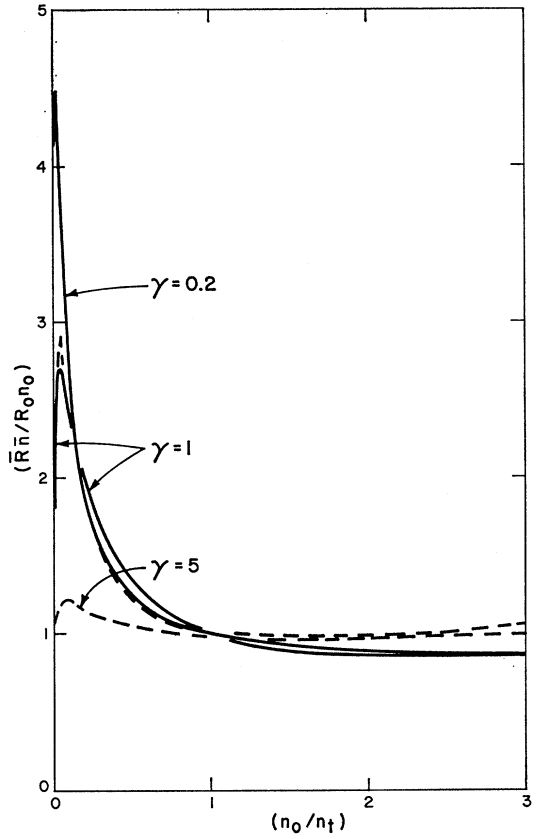

FIG. 4. Same as Fig. 2, but for the normalized product $(\bar{R}\bar{n}/R_0n_0)$ [Eq. (32)].

one obtains, again analogously to Juretschke *et al.*,[13]

$$\frac{\bar{\mu}^H}{\mu_0{}^H}=\beta\left[1-\epsilon\left(1-\frac{2(\beta+1)(n_0/n_t)}{[\beta+(n_0/n_t)]^2}\right)\right]\times\left[1-\epsilon\left(1-\frac{2(n_0/n_t)}{\beta[\beta+(n_0/n_t)]}\right)\right]\Big/\left\{\left[1+\epsilon\left(1-\frac{2(n_0/n_t)}{[\beta+(n_0/n_t)]}\right)\right]\times\left[1-\epsilon\left(1-\frac{2(\beta+1)(n_0/n_t)^2}{\beta[\beta+(n_0/n_t)]^2}\right)\right]\right\}. \quad (23b)$$

Values of $\bar{\mu}^H/\mu_0{}^H$ are shown in Fig. 3 (for $\beta=0.1$). The over-all dependence for both high and low dislocation density models on n_0/n_t at intermediate values of γ is basically similar.

It can be seen from Fig. 3 that the normalized Hall mobility also displays a maximum with n_0/n_t. For the low-γ case it is obvious from Eq. (23a) that this maximum is at

$$(\sigma_0/\sigma_t)_{\mu=\max}=1, \quad (25)$$

$$(n_0/n_t)_{\mu=\max}=\beta. \quad (26)$$

The resultant maximum mobility is then obtained from Eq. (23b) as

$$(\bar{\mu}^H/\mu_0{}^H)_{\max}\approx(1+2\beta\gamma)/(1+\gamma)^2, \quad \gamma\ll1. \quad (27a)$$

For the high-γ case differentiation of Eq. (23b) for $\epsilon\ll(\beta/2)$ yields the same conditions (25) and (26), with a resultant mobility

$$(\bar{\mu}^H/\mu_0{}^H)_{\max}\approx\beta+\epsilon(1-\beta), \quad \epsilon\ll1. \quad (27b)$$

Next we examine the behavior for a high free-carrier concentration. In the low-γ case this limit is obtained when the carrier ratio satisfies the inequalities (13a) and

$$(n_0/n_t)\gg(1/\gamma)^2(1+\gamma\beta). \quad (28a)$$

The resultant expression for the mobility is

$$\bar{\mu}^H/\mu_0{}^H\approx\gamma\beta/(1+\gamma). \quad (29a)$$

In the case of high n_t/n_0 satisfying the conditions (16a) and

$$(n_0/n_t)\ll(\gamma\beta)^2/(1+\gamma\beta), \quad (30a)$$

the resultant expression for the mobility is that already given in Eq. (29a). Thus, as expected from the symmetry shown by Eq. (23a), one obtains the same Hall mobility at very high and at very low free-carrier concentration. Moreover, this value is independent of n_0/n_t.

For large γ, the free-carrier concentration can be considered large under the conditions (13b) and

$$n_0/n_t\gg2(\beta+1), \quad (28b)$$

which for $\epsilon\ll(\beta/2)$, leads from Eq. (23b) to

$$\bar{\mu}^H/\mu_0{}^H\approx\beta(1-2\epsilon). \quad (29b)$$

Small free-carrier concentration on the other hand requires

$$n_0/n_t\ll(\tfrac{1}{2})\beta^2, \quad \beta\ll1 \quad (30b)$$

$$n_0/n_t\ll(\beta/2), \quad \beta\gg1 \quad (31b)$$

with the resultant mobility again identical to that obtained for large free-carrier concentration [Eq. (29b)].

As to other limiting behavior of the Hall mobility, for $\gamma=0$ and $\epsilon=0$ one obtains the proper limits of $\bar{\mu}^H/\mu_0{}^H=1$ and β, respectively. Similarly, the proper result of $\bar{\mu}^H/\mu_0{}^H=1$ is obtained for $\beta=1$, $n_0/n_t=1$, and that of $\bar{\mu}^H/\mu_0{}^H=0$ for $\beta=0$.

In summary: For low dislocation densities, the Hall mobility goes through a maximum at a value of $n_0/n_t=\beta$ and reaches the same constant value at both high and low free-carrier concentrations. For high dislocation densities one obtains the same results, but subject to the additional condition $\epsilon\ll\beta/2$.

D. Hall Coefficient and Ratio of Hall to Conductivity Mobility

The product $\bar{R}\bar{n}$, or, with elimination of numerical factors, the quantity $\bar{R}\bar{n}/R_0n_0$, determine the accuracy

with which the Hall coefficient gives the average carrier concentration. By Eq. (21), and use of $\sigma=e\mu n$, the latter can be shown to be equal to the ratio of the normalized Hall mobility to the normalized conductivity mobility,

$$\frac{\bar{R}\bar{n}}{R_0 n_0}=\frac{(\bar{\mu}^H/\mu_0{}^H)}{(\bar{\mu}/\mu_0)}. \tag{32}$$

A plot is shown in Fig. 4, again for $\beta=0.1$. It can be seen that, as for the individual mobilities, there is a maximum. The numerical results also show two additional features: (1) One obtains $\bar{R}\bar{n}/R_0n_0=1$ at $n_0/n_t=1$; (2) there is a shallow minimum at $n_0/n_t\gtrsim 1.4$. The former relation can readily be confirmed analytically by equating Eqs. (9) and (23). As to the latter, a numerical check showed that for $\beta=0.1$ and $\beta=0.01$ (and $\gamma=0.1$–5, as in Figs. 2 and 3) the minimum does not go below a value of 0.8, i.e., $\bar{R}\bar{n}/R_0n_0$ does not decrease much below unity.

Since the maximum in $\bar{R}\bar{n}/R_0n_0$ appears rather pronounced in the low-γ case, additional information for this case appeared desirable. Differentiation of Eq. (32) with use of Eqs. (9a) and (23a) gave a fifth-order equation, and a general analytical solution did not appear feasible; however, an approximate evaluation for $\beta\ll 1$, $\gamma\ll 1$, and $n_0/n_t<1$ gave the location and height of the maximum as follows:

$$(n_0/n_t)_{\max}\approx\beta\gamma, \tag{33}$$

$$(\bar{R}\bar{n}/R_0n_0)_{\max}\approx 1/2\beta. \tag{34}$$

It follows that the height of the maximum can be quite appreciable (as is also apparent from Fig. 4), so that Hall data in this region does not give a measure of the average carrier concentration.

Evaluation for the case of a high concentration of free carriers by substitution of Eqs. (15) and (29) into Eq. (32) yields

$$(\bar{R}\bar{n}/R_0n_0)\approx[\gamma^2/(1+\gamma)^3](n_0/n_t), \quad \gamma\ll 1 \tag{35a}$$

$$(\bar{R}\bar{n}/R_0n_0)\approx\epsilon(1-4\epsilon)(n_0/n_t), \quad \epsilon\ll\beta/2 \tag{35b}$$

and shows that there is a continuous increase with n_0/n_t. Analysis of Eq. (35b) with use of Eq. (5b) in the limit of large n_0/n_t shows that

$$\bar{R}\approx(1-4\epsilon)R_t, \tag{36}$$

i.e., the Hall coefficient is approximately equal to that of the dislocations, with the high free-carrier concentration in the (small) bulk regions not making itself felt except for a minor correction dependent only on the volume of the bulk.

As to the situation for high n_t/n_0, use of Eqs. (17) and (29) gives

$$(\bar{R}\bar{n}/R_0n_0)\approx 1-(1+\gamma)^{-2}, \quad \gamma\ll 1 \tag{37a}$$

$$(\bar{R}\bar{n}/R_0n_0)\approx 1-\epsilon, \quad \epsilon\ll 1. \tag{37b}$$

In this case, for a small bulk volume ($\epsilon\ll 1$) and small bulk carrier concentration Eq. (37b) shows the that average Hall constant does properly give the average carrier concentration, determined largely by n_t.

In summary: The normalized quantity $\bar{R}\bar{n}/R_0n_0$ is equal to the ratio of the normalized Hall mobility to the normalized conductivity mobility. With increasing n_0/n_t this ratio first displays a maximum, with an amplitude which can be considerably in excess of unity. It then decreases to unity at $n_0/n_t=1$ and subsequently goes through a shallow minimum; numerical evaluation indicates that the minimum value does not decrease below 0.8 (for $\beta=0.1$, 0.01, $\gamma=0.1$–5). For higher values of n_0/n_t, there is a continuous increase. Thus Hall data will give an approximate measure of the average carrier concentration only over limited ranges of n_0/n_t; in fact for large dislocation densities, the Hall coefficient over a good part of the range gives the approximate concentration of trapped carriers rather than the average concentration.

IV. DISCUSSION

A. Experimental Results Used for Comparison

Extensive data on Hall and conductivity mobilities as a function of gate bias have been reported by Fang and Fowler.[3] The main results apply to both (100) and (111) surfaces and can be summarized as follows: There is a maximum in both mobilities very close to the pinch-off region; this maximum is present at room temperature, and persists to at least 7.5°K in some samples. There is a second maximum at higher inversion in the lower temperature range; this maximum is just apparent in one sample at 145°K, and becomes more pronounced at lower temperatures. The mobilities at both maxima are generally lower than bulk, although bulk mobilities are approached in some cases in the first maximum.

B. Quantum Effects

Recent work has shown[14] that at least under certain conditions the carrier bands in Si inversion layers form subbands, each corresponding to a quantized level for motion perpendicular to the surface. In view of various treatments[11,15,16] emphasizing the influence of this effect on carrier mobilities, we would like to discuss the likely role of quantization in relation to the misfit-dislocation theory.

Conditions favorable for the occurence of quantization are a low temperature, high inversion, and/or a (111) surface. It was already pointed out in Ref. 1 that

[14] A. B. Fowler, F. F. Fang, W. E. Howard, and P. J. Stiles, Phys. Rev. Letters **16**, 901 (1966); J. Phys. Soc. Japan Suppl. **21**, 331 (1966).

[15] B. Tavger, Phys. Status Solidi **22**, 31 (1967); V. Ya. Demikhovskii and B. A. Tavger, Fiz. Tverd. Tela **6**, 960 (1964) [English transl.: Soviet Phys.—Solid State **6**, 743 (1964)].

[16] C. B. Duke, Phys. Rev. **168**, 816 (1968); M. E. Alferieff and C. B. Duke, *ibid.* **168**, 832 (1968).

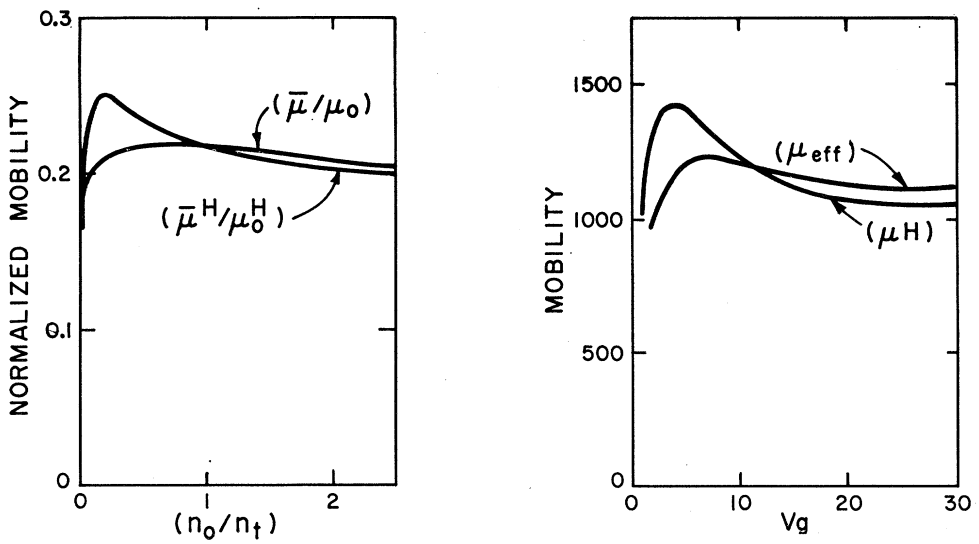

FIG. 5. Left-hand side: Theoretical results for the normalized Hall and conductivity mobilities for $\beta=0.2$, $\gamma=3$, as a function of n_0/n_t. Right-hand side: Experimental 77°K mobility results of Fang and Fowler (Ref. 3), as a function of gate bias. Implicitly, the theoretical results assume a bulk mobility of ≈ 5600 cm^2/V sec.

the misfit-dislocation model gave better quantitative agreement to the Arnold-Abowitz[6] data with an appropriate quantization correction for the (111) surface. *However*, quantum effects would be expected[3] to be minor or nonexistent for the (100) surface at room temperature in general, and especially at low inversion, i.e., near pinch-off. And yet there is a maximum in mobility also under these conditions.[3] We consequently propose that the mobility maximum close to pinch-off is *primarily* due to the misfit-dislocation effect, with some quantization corrections where required [i.e., for the (111) surface in general, and for the (100) surface at lower temperatures]. It seems likely that at higher inversion, such as in the region of the second maximum, the relative importance of these two factors reverses, with quantization effects dominant and corrections due to the misfit dislocations. In this connection it can be noted that the Duke[16] quantization analysis does predict maxima in mobility with degree of inversion, but also predicts these maxima to occur at higher inversion with increasing temperature; such behavior is clearly shown by the peak observed by Fang and Fowler[3] at high inversion (their Figs. 16 and 17). However, the temperature-bias trend[3] of the peak close to pinch-off is the *inverse* of the above, again showing that quantization is not likely to be of major importance for this peak.

As just discussed, the mobility peak close to pinch-off is assumed to be primarily due to the misfit-dislocation effect, but with quantum corrections. For these corrections, we will assume validity of the Tavger[15] analysis, which involves solution of the Boltzmann equation for the two-dimensional quantum case.[17] The relevant equations, considering μ_0 to be determined primarily by acoustic lattice scattering, have been summarized in Ref. 1 [Eqs. (6)].

C. Comparison with Experimental Results

This section will compare the theoretical Hall and conductivity mobility variation with n_0/n_t with the experimental results of Fang and Fowler[3] for the low gate-bias maximum. As previously discussed (Sec. II), the theoretical variation with respect to n_0/n_t is expected to be analogous to the experimental gate bias variation.

The features shown by Figs. 2–4 which are of the most relevance for comparison with the experimental results are the following: (1) The maximum in the Hall mobility occurs at a lower value of n_0/n_t than the maximum in conductivity mobility. This result can be confirmed analytically, for $\beta \ll 1$, by a comparison of Eqs. (11) and (26), and therefore holds over the entire validity range of these equations. (2) The maximum Hall mobility is higher than the maximum conductivity mobility. Analytical confirmation is in this instance provided over the same validity range by Eqs. (12) and (27). (3) Subsequent to the region of the maxima (i.e., at higher values of n_0/n_t) the two mobility values approach each other. Equality is reached at $n_0/n_t=1$, and the Hall mobility is subsequently lower over at least some range of n_0/n_t.

[17] It appears of interest to note that the Tavger theory predicts a mobility decreasing with L, where L is the inversion layer thickness. A similar dependence is also predicted by the diffuse surface scattering theory (without quantization) over certain ranges, but the Tavger theory predicts this dependence for a specular surface. [The diffuse surface scattering theory is summarized for instance by A. Many, Y. Goldstein, and N. B. Grover, *Semiconductor Surfaces* (North-Holland Publishing Co., Amsterdam, 1965).]

In view of the above results it is of interest that the data given in Fig. 4 of Fang and Fowler show the same behavior: (1) The maximum in the Hall mobility is at a lower gate bias. (2) The maximum Hall mobility is higher than the maximum conductivity mobility. (3) The two mobilities become equal at a value of gate bias higher than that of the maxima, and the Hall mobility is subsequently lower. A direct comparison of the features we have just discussed is shown in Fig. 5. The similarity, within the limitations imposed by comparing a dependence on n_0/n_t with one on gate bias, is very striking.

In connection with Fig. 5 it should still be noted that values of μ_0 for both the Hall and the conductivity mobilities were also required for the comparison. We used a value of ≈ 5600 cm^2/V sec for both. This value was derived by applying the Tavger[15] quantum corrections to the 77°K Hall mobility value of ≈ 8000 cm^2/V sec given by Morin and Maita[18] for a sample of approximately comparable doping. These corrections give a mobility reduction of ≈ 0.5–0.65 for full quantization of a (100) surface for the parameter ranges of interest; quantization of the (100) surface at 77°K and low inversion is probably only partial, so that a reduction factor of ≈ 0.7 and a corresponding 5600-cm^2/V sec mobility value appear reasonable. As to values of the conductivity mobility, this is probably very closely equal to the Hall mobility: For bulk Si, again of approximately comparable doping, Long and Myers[19] obtain a Hall factor ($=\mu_0^H/\mu_0$) very close to unity (their Fig. 5, sample SM3); moreover, Tavger[15] predicts a Hall factor of unity for total quantization, so a Hall factor of unity appears expected over the entire range of quantization.[20]

D. Refinements on the Model

We still wish to mention several nonquantum refinements which, based on physical reasoning, may be causing perturbations on the model. These are mentioned primarily for completeness; in view of the good match to the data presented in the preceding Sec. IV C, there is no present evidence that any of them are important.

(1) There may be variations in geometry. For instance, misfit dislocations would be expected to form a hexagonal, and not a square, array on (111) Si surfaces; this would likely cause some quantitative but not qualitative changes. Or, the dislocations may be discontinuous (but still long); results for this case are discussed in Ref. 2.

(2) There is no obvious physical reason why the (parallel) mobility along the dislocations in the field direction should be equal to the (perpendicular) one across the dislocations. For this case, in the limit that either conduction in the parallel dislocation or conduction in the composite strip dominates, the results would reduce to the present ones; curves with different β's on the low and high n_0/n_t sides would result in some cases.

(3) The quantity γ may well depend on screening, and thus be a function of n_0/n_t. If such a dependence obtains, the parallel conduction would probably be independent of γ, since it is expected to be along the dislocation core; thus the effective n_t would vary. Predictions on resultant variations in perpendicular conduction are more difficult (is the mechanism tunneling, space-charge injection, etc.?); a subsidiary dependence on γ can certainly not be excluded.

(4) The value of β for the Hall mobility [Eq. (4)] may not be equal to that for the conductivity mobility; deviations in this regard would, however, be unlikely to exceed 20% at most temperatures (ionized impurity scattering may cause a somewhat larger error at lower temperatures).

(5) The ratio β may be a function of gate bias, for instance because of changes in degree of quantization or screening.

E. Over-All Considerations Regarding Surface Mobilities

The results of the present paper raise a speculation regarding surface mobilities in general. As discussed, for instance by Greene,[21] many surfaces show a mobility behavior attributed to diffuse surface scattering (which gives results qualitatively similar to the decrease from the maximum of the present theory), whereas other surfaces are thought to scatter specularly. As previously noted[1,21] agreement with this theory is, however, not always satisfactory, even using the diffusivity as an adjustable parameter. As an alternative, Greene and O'Donnell[22] proposed scattering by localized surface charges. Although this approach cannot explain all MOS results,[1] it did receive experimental support for the case that charge is specifically placed on the surface.[23] Considering this in addition to the present results, there thus appears to be a good possibility that surface mobility reduction (of the nonquantum type) is in fact generally caused by surface inhomogeneities rather than by diffuse surface scattering.

V. CONCLUSIONS

It has been shown that the misfit-dislocation model gives a very satisfactory agreement with experimental

[18] F. J. Morin and J. P. Maita, Phys. Rev. **96**, 28 (1954).
[19] D. Long and J. Myers, Phys. Rev. **115**, 1107 (1959).
[20] It can also be shown that even for slight deviations of the free-electron Hall factor from unity a good match can still be obtained, since there would be a proportional change in β [i.e., the assumption of Eq. (4) would not hold]; the Hall factor for the trapped electrons should be unity (as follows for degenerate statistics, which presumably apply for the filled dislocations), leading to $\beta_{\text{conductivity}}/\beta_{\text{Hall}}=\mu_0^H/\mu_0$.
[21] R. F. Greene, Surface Sci. **2**, 101 (1964).
[22] R. F. Greene and R. W. O'Donnell, Phys. Rev. **147**, 599 (1966).
[23] T. I. Kamins and N. C. MacDonald, Phys. Rev. **167**, 754 (1968).

maxima in Hall and conductivity mobility of MOS structures in the pinch-off region. This, together with the earlier[1] fit to data of Arnold and Abowitz[6] as well as earlier[1] arguments based on surface-state behavior, gives strong and probably overwhelming support to the model. One obvious corollary to this conclusion is that experimental work specifically demonstrating the existence of misfit dislocations at the Si–SiO_2 interface, and work correlating the density of surface states with the density of such misfit dislocations would be highly desireable. A second line of promising work, following the discussion of Sec. IV E, would be to investigate the relation between surface inhomogeneities and surface mobilities in general.

Other aspects of the present work have clarified the probable role of quantum effects on MOS inversion regime mobilities. Although presumably dominant at high inversion, such effects appear to play a minor role in the mobility variations seen at low gate bias; in the low bias range approximate quantum corrections, taken independent of gate bias, result in very satisfactory quantitative agreement between the dislocation model and the data.

A further interesting conclusion is that the Hall coefficient can yield reliable information on carrier concentration only over a limited range of the ratio of free to trapped carriers.

ACKNOWLEDGMENTS

The author is indebted to Dr. E. Arnold, Dr. S. Kurtz, and Dr. F. K. duPré for helpful discussions, to Miss T. Gendron for programming the calculations, and to Professor H. Juretschke for an unpublished report on his calculations for nonconducting inclusions.

APPENDIX

Under the usual experimental conditions of a magnetic field **H** in the z direction and no over-all y current, the Hall coefficient R can in general be expressed as[24]

$$RH=\rho_{yx}=-\sigma_{yx}/(\sigma_{xx}\sigma_{yy}-\sigma_{xy}\sigma_{yx}), \tag{A1}$$

with the further simplification at low magnetic fields that

$$\sigma_{xy},\sigma_{yx}\ll\sigma_{xx},\sigma_{yy}, \tag{A2}$$

$$RH\approx-\sigma_{yx}/\sigma_{xx}\sigma_{yy}. \tag{A3}$$

[24] See, for example, A. C. Beer, *Galvanomagnetic Effects in Semiconductors* (Academic Press Inc., New York, 1963).

In the present case we are interested in the coefficient R_c [see Eq. (19)] of a strip of height l_0 in the direction of the Hall field (y), and consisting of islands of (bulk) material of width l_0 and intervening layers of dislocations of width l_t in the direction of the current (x) [see Fig. 1(a)]. For a high strip (small γ) this can be carried out approximately by use of equations given by Herring[25] [Eqs. (53) and (54)] for an infinitely high strip. With the notation σ_c to correspond to Herring's σ_{eff}, these equations are

$$(\sigma_c)_{yx}=\langle 1/\sigma_{xx}\rangle^{-1}\langle\sigma_{yx}/\sigma_{xx}\rangle, \tag{A4}$$

$$(\sigma_c)_{xx}=\langle 1/\sigma_{xx}\rangle^{-1}, \tag{A5}$$

$$(\sigma_c)_{yy}=\langle\sigma_{yy}\rangle+\langle 1/\sigma_{xx}\rangle^{-1}\langle\sigma_{yx}/\sigma_{xx}\rangle\langle\sigma_{xy}/\sigma_{xx}\rangle - \langle\sigma_{yx}\sigma_{xy}/\sigma_{xx}\rangle, \tag{A6a}$$

where the angular brackets are used to denote spatial averages. Use of the low magnetic field approximation, inequalities (A2), leads to

$$(\sigma_c)_{yy}\approx\langle\sigma_{yy}\rangle. \tag{A6b}$$

Substitution of Eqs. (A4)–(A6) into Eq. (A3) gives

$$-R_cH=\langle\sigma_{yx}/\sigma_{xx}\rangle\langle\sigma_{yy}\rangle^{-1}. \tag{A7}$$

For the further evaluation of Eq. (A7), use of Eq. (A3) gives

$$-\langle\sigma_{yx}/\sigma_{xx}\rangle=\langle RH\sigma_{yy}\rangle, \tag{A8}$$

where R is now the microscopic Hall coefficient within each region, given by the standard equation

$$R=1/nec. \tag{A9}$$

Consequently,

$$-\langle\sigma_{yx}/\sigma_{xx}\rangle=(H/c)\langle\mu\rangle =(H/c)(l_0\mu_0+l_t\mu_t)(l_0+l_t)^{-1} =(H/c)(\mu_0+\gamma\mu_t)(1+\gamma)^{-1}. \tag{A10}$$

Evaluation of $\langle\sigma_{yy}\rangle$ is obvious, and gives

$$\langle\sigma_{yy}\rangle=(\sigma_0+\gamma\sigma_t)(1+\gamma)^{-1}. \tag{A11}$$

By substitution of Eqs. (A10) and (A11) into Eq. (A7) one obtains the desired expression for R_c,

$$R_cc=(\mu_0+\gamma\mu_t)(\sigma_0+\gamma\sigma_t)^{-1}. \tag{A12}$$

[25] C. Herring, J. Appl. Phys. **31**, 1939 (1960).