Self-Consistent Results for *n*-Type Si Inversion Layers

Frank Stern

IBM Watson Research Center, Yorktown Heights, New York 10598 (Received 13 January 1972)

Self-consistent results for energy levels, populations, and charge distributions are given for n-type inversion layers on p-type silicon. Quantum effects are taken into account in the effective-mass approximation, and the envelope wave function is assumed to vanish at the surface. Approximate analytic results are given for some special cases. Numerical results are given for representative surface orientations, bulk acceptor concentrations, inversion-layer electron concentrations, and temperatures.

I. INTRODUCTION

External electric fields or charges can change the properties of semiconductor surfaces significantly, and the consequences of such changes have been systematically explored for many years.^{1,2} More recently there has been considerable work on the quantum effects that arise when band bending confines the carriers to a narrow surface channel. Some of these effects have been described elsewhere.^{3,4} In this paper we give the results of numerical self-consistent calculations for the energy levels and subband populations of *n*-type inversion layers on *p*-type Si. We use the notation of Stern and Howard⁵; readers are referred to that paper for an introduction to *n*-type inversion layers.

The first numerical self-consistent calculations for inversion layers were made by Howard.⁶ Selfconsistent results have also been obtained by Maeda⁷ and by Stern.⁸ Self-consistent solutions for accumulation layers were discussed by Duke,⁹ and have been studied in detail by Appelbaum and Baraff.^{10,11}

Our calculation is based on three major approximations. We assume (i) that the effective-mass approximation is valid, so that we can neglect the periodic potential and use the effective masses and the dielectric constant of the perfect crystal; and (ii) that the envelope wave function vanishes at the surface. Neither approximation is likely to be valid at high surface electric fields, for which the calculated wave functions can extend 10 Å or less into the semiconductor. The third major approximation (iii) is that we can neglect surface states and can replace the effect of any charges in the oxide or insulator adjacent to the semiconductor by an equivalent electric field. This approximation is reasonably well justified for the interface between silicon and silicon oxide when it is appropriately treated.¹² Estimates of the errors that result from use of the effectivemass approximation have been made by Howard.¹³

Section II gives the equations to be solved, and Sec. III gives some approximate results and compares them with the corresponding self-consistent results. In Sec. IV we briefly describe the numerical calculation, and in Sec. V we give numerical results for representative surface orientations, bulk-carrier concentrations, temperatures, and inversion-layer carrier concentrations. Section VI discusses the relation of the calculated results to experiment and includes an estimate of the changes in the calculated energies that result from relaxation of the condition that the envelope wave function vanish at the surface. Two appendixes describe depletion-layer edge effects and give integrals of powers of z times the square of the Airy function.

II. EQUATIONS

The band bending at a semiconductor surface can be characterized by an electrostatic potential $\phi(z)$. In the effective-mass approximation, the electronic wave function for the *i*th subband is the product of the Bloch function at the bottom of the conduction band and an envelope function⁵

$$\psi_i(x, y, z) = \zeta_i(z) e^{i\theta z} e^{ik_1 x + ik_2 y}, \qquad (1)$$

where k_1 and k_2 are measured relative to the band edge, θ depends on k_1 and k_2 ,⁵ and $\zeta_i(z)$ is the solution of

$$\frac{d^2\zeta_i}{dz^2} + \frac{2m_3}{\hbar^2} \left[E_i + e\phi(z) \right] \zeta_i(z) = 0.$$
⁽²⁾

We are interested in the bound solutions of (2), so we require that $\zeta_i(\infty) = 0$. In addition, we require that $\zeta_i(z)$ vanish at the surface, where z = 0. This should be a good approximation for the Si-SiO₂ interface, for which the potential barrier for electrons is approximately 3 eV,¹⁴ until we reach electric fields strong enough to confine the electrons within 1 nm or less.¹⁵ An estimate of the error arising from this approximation is given in Sec. VI.

Each eigenvalue E_i found from the solution of (2) is the bottom of a continuum of levels called

5

a subband, with energy levels given by

$$E_{i}(\vec{\mathbf{k}}) = E_{i} + \hbar^{2} k_{1}^{2} / 2m_{1} + \hbar^{2} k_{2}^{2} / 2m_{2}, \qquad (3)$$

where m_1 and m_2 are the principal effective masses for motion parallel to the surface, which can be obtained in a straightforward way from the bulk masses.⁵ Values for these quantities and for the mass m_3 for motion perpendicular to the surface, used in (2), are given in Table I for the surface orientations used in this paper.

There can be as many as three values of m_3 for a given surface orientation because the conduction band of Si has six valleys along the $\langle 100 \rangle$ directions of the Brillouin zone. In the effective-mass approximation, the valleys are degenerate in pairs, but the degeneracy is found to be lifted in magnetoconductance experiments at high magnetic fields.¹⁶ The origin of that splitting has not yet been explained.

The valleys which present the highest mass for motion perpendicular to the surface have the lowest kinetic energy and the lowest energy levels. We label the subbands arising from these valleys with indices 0, 1, 2,... The {111} surfaces have only one ladder of subbands, since all the valleys have the same orientation with respect to the surface. The {100} and {110} surfaces have a second ladder of subbands, which we label 0', 1', 2', For a general surface orientation, those subbands which arise from the smallest value of m_3 are labelled 0'', 1'', 2'', ... in this notation

labelled 0", 1", 2", ... in this notation. The potential $\phi(z)$ which appears in (2) is the solution of Poisson's equation¹⁵

$$\frac{d^2\phi}{dz^2} = -\left[\rho_{dep1}(z) - e\sum_i N_i \zeta_i^2(z)\right] / \kappa_{sc} \epsilon_0, \tag{4}$$

TABLE I. Parameters used in the calculation.^a

Surface		(100)		(110)		(111)
Valleys		Lower	Higher	Lower	Higher	All
Degeneracy	n.,	2	4	4	2	6
Normal mass	m_3	0.916	0.190	0.315	0.190	0.258
Longitudinal	Ū					
masses	m_1	0.190	0.190	0.190	0.190	0.190
	m_2	0.190	0.916	0.553	0,916	0.674
Conductivity	-					
mass ^b	me	0.190	0.315	0.283	0.315	0.296
Density-of-						
states mass						
per valley ^b	m_d	0.190	0.417	0.324	0.417	0.358
Dielectric constant		Kac		11.7		
Permittivity		$\epsilon_{sc} = \kappa_{s}$	e€0 :	1.04×10^{-10}	F/m	
Fermi level in	the bu	lk, for N	$-N_{\rm D} = 10^2$	¹ m ⁻³ . See	Eq.(9).	
<i>T</i> (K)		0	77	3	00	
$F = F (\alpha V)$		1 19	1 1		89	

^aAll effective masses are in units of the free-electron mass; they are based on the conduction-band masses $m_i = 0.190m$ and $m_i = 0.916m$ given by J. C. Hensel, H. Hasegawa, and M. Nakayama [Phys. Rev. <u>138</u>, A225 (1965)]. See also, Table I of Ref. 5.

 ${}^{\mathbf{b}}m_{s} = m_{1} m_{2} / (m_{1} + m_{2}); m_{d} = (m_{1} m_{2})^{1/2}.$

where κ_{sc} is the dielectric constant of the semiconductor, N_i is the carrier concentration in the *i*th subband, given by ⁵

$$N_{i} = (n_{vi} m_{di} KT / \pi \hbar^{2}) F_{0} [(E_{F} - E_{i}) / KT], \qquad (5)$$

 $F_0(x) = \ln(1 + e^x)$, n_{vi} and m_{di} are the valley degeneracy and the density-of-states effective mass per valley, given in Table I, and E_F is the Fermi energy. $\rho_{depl}(z)$ is the charge density of the depletion layer, which we take to be

$$\rho_{dep1}(z) = -e(N_A - N_D), \quad 0 \le z \le z_d$$

$$\rho_{dep1}(z) = 0, \qquad z \ge z_d \tag{6}$$

where z_d is the depletion-layer thickness, given by

$$z_{d} = [2\kappa_{\rm sc} \epsilon_{0} \phi_{d} / e(N_{A} - N_{D})]^{1/2}.$$
(7)

Here ϕ_d is the effective band bending from the bulk to the surface, apart from the contribution of the inversion layer itself; its value is given below.

The boundary conditions for Eq. (4) are that $d\phi/dz$ vanish for large z and that its value at the surface be $-F_s$, where

$$F_{s} = e(N_{inv} + N_{depl}) / \kappa_{sc} \epsilon_{0}; \qquad (8)$$

 $N_{depl} = z_d(N_A - N_D)$ is the number of charges per unit area in the depletion layer and $N_{inv} = \sum N_i$ is the total number of charges per unit area in the inversion layer. Note that the electric field just outside the semiconductor is larger than the field just inside by a factor κ_{sc}/κ_{ins} , where κ_{ins} is the dielectric constant of the insulator. We have specified $d\phi/dz$ at the boundaries, so ϕ itself is uncertain by a constant. We sometimes choose ϕ to vanish in the bulk, but the energy levels obtained from the solution of (2) assume that ϕ vanishes at the surface.

The assumption that the depletion charge is constant for a distance z_d from the surface and then goes abruptly to zero fails in the transition region from depletion to bulk, in which the field decays to zero exponentially with a characteristic distance given by the bulk screening length.¹⁷ In Appendix A, we discuss the correction to the depletion-layer charge, and therefore to $z_d = N_{depl}/(N_A - N_D)$, which arises from this effect. When the correction to ϕ_d is taken to be -KT/e, the band bending which enters in (7) is

$$\phi_d = \left[\left(E_c - E_F \right)_b + E_F - KT \right] / e + \phi_{sb} - e N_{inv} z_{av} / \kappa_{sc} \epsilon_0,$$
(9)

where the first term is the energy difference between the bottom of the conduction band in the bulk and the Fermi level, E_F is the Fermi energy relative to the nominal conduction-band edge at the surface, and z_{av} is the average penetration of the inversion-layer charge from the surface. The substrate bias ϕ_{sb} is the voltage difference between the source and drain electrodes, which are diffused into the surface to make electrical contact to the inversion layer, ¹² and the bulk; it is normally equal to zero. The last term in Eq. (9) is the potential drop across the inversion layer. Figure 1 shows some of the quantities which enter in Eq. (9). Values of $E_c - E_F$ in the bulk are given in Table I for several values of Tfor $N_A - N_D = 10^{21} \text{ m}^{-3}$.

5

III. APPROXIMATE RESULTS

A. Triangular Potential

Because of the way in which the Schrödinger equation (2) and Poisson's equation (4) are coupled it is necessary in general to solve these equations numerically to obtain self-consistent results. We can, however, find approximate results for some simple limiting cases.

The simplest approximation is to replace the potential $\phi(z)$ in (2) by $-F_s z$ for $z \ge 0$ and by an infinite barrier for z < 0. This is sometimes called the triangular-potential approximation. It leads to the Airy equation with solutions¹⁸

$$\zeta_i(z) = \operatorname{Ai}\left((2m_3 eF_s/\hbar^2)^{1/3} \left[z - (E_i/eF_s)\right]\right),$$
 (10a)

$$E_i \approx (\hbar^2/2m_3)^{1/3} \left[\frac{3}{2} \pi eF_s(i+\frac{3}{4})\right]^{2/3}.$$
 (10b)

The eigenvalues E_i are asymptotic values for large i, but they are amazingly close even for the ground state i = 0. The exact eigenvalues for the three lowest states have $i + \frac{3}{4}$ in (10b) replaced by 0.7587, 1.7540, and 2.7525, respectively.¹⁸ Another property of these solutions is that the average separation of carriers in the *i*th subband from the surface, defined by

$$z_i = \int z \zeta_i^2(z) dz / \int \zeta_i^2(z) dz, \qquad (11)$$

is $2E_i/3eF_s$, and that the average of z^2 is $\frac{6}{5}z_i^2$. See Appendix B for the relevant integrals, including the normalization of (10a).

B. Variational Results for Electric Quantum Limit

The triangular-potential approximation which we have just described is a reasonable approximation when there is little or no charge in the inversion layer, but fails when the charge density per unit area in the inversion layer is comparable to or exceeds that in the depletion layer. When only one subband is occupied, i.e., in the electric quantum limit, a variational approach gives a good estimate for the energy of the lowest subband. Fang and Howard¹⁹ used the trial eigenfunction

$$\zeta_0(z) = \left(\frac{1}{2}b^3\right)^{1/2} z \, e^{-bz/2},\tag{12}$$

with a single undetermined parameter b. The energy of the lowest state is found after a straightforward calculation to be

$$E_0 = \hbar^2 b^2 / 8 m_3 + (3 e^2 / \kappa_{sc} \epsilon_0 b) \\ \times \left[N_{dep1} + \frac{11}{16} N_{inv} - (2/b) (N_A - N_D) \right]$$
(13)

when the potential ϕ vanishes at z = 0. The term involving N_{inv} gives the interaction of the inversion-layer charge density with itself; the correct choice of b minimizes the total energy of the system, in which the coefficient $\frac{11}{16}$ is replaced by $\frac{11}{32}$. The last term in (13) arises from the curvature of the potential of the depletion charge density; it will be small in Si, for which the inversion layer is generally much thinner than the depletion layer. We can therefore treat this term as a small perturbation.

The average distance the charge density associated with the variational wave function (12) penetrates into the semiconductor is found from (11) to be $z_0 = 3/b$. We therefore find the following variational results:

$$b = b_0 + \delta b,$$

$$b_0 = (12m_3 e^2 N^* / \kappa_{sc} \epsilon_0 \hbar^2)^{1/3},$$

$$\delta b = -4N_A / 3N^*;$$

$$z_0 = z_{00} + \delta z_0,$$

(14)

$$z_{00} = (9\kappa_{sc}\epsilon_0 \hbar^2 / 4m_3 e^2 N^*)^{1/3}, \qquad (15)$$

$$\delta z_0 = 4N_A z_{00}^2 / 9N^*;$$

$$E_{0} = E_{00} + \delta E_{0},$$

$$E_{00} = (\frac{3}{2})^{5/3} \left(e^{2} \hbar / m_{3}^{1/2} \kappa_{sc} \epsilon_{0} \right)^{2/3} \left(N_{dep1} + \frac{55}{96} N_{inv} \right) / N^{*1/3},$$

$$\delta E_{0} = - \left(2N_{A} e^{2} z_{00}^{2} / 3 \kappa_{sc} \epsilon_{0} N^{*} \right) \left(N_{dep1} + \frac{11}{96} N_{inv} \right).$$
(16)

Here $N^* = N_{dep1} + \frac{11}{32} N_{inv}$ and N_A is the net acceptor concentration. In the limit $N_{inv} = 0$, for which the



FIG. 1. Potential energy $V(z) = -e\phi(z)$ at 0 K for electrons near a (111) surface of p-type silicon with 10^{21} net acceptor ions per m³ and with 10¹⁷ inversion-layer electrons per m². The Fermi energy E_F and the energy E_0 of the bottom of the lowest subband are shown. The dotdash curve is the potential energy with the inversionlayer contribution removed. The potential V(z) was used to calculate the results in Table III.

TABLE II. Comparison of variational (var) and self-consistent (sc) values of energy E_0 and spatial extent z_0 for the lowest subband, as a function of N_{inv} , the number of inversion-layer electrons per unit area, and N_{depl} , the number of depletion-layer charges per unit area. The results are for a (100) surface at 0 K, with 10^{21} m^{-3} net acceptors, and with the parameters taken from Table I. See Eqs. (15) and (16). All energies are in meV and all distances in nm.

$N_{inv} (m^{-2})$	$N_{depl}(m^{-2})$	<i>E</i> ₀₀	$\delta \boldsymbol{E}_0$	E_0 (var)	E_0 (sc)	z ₀₀	δz0	<i>z</i> ₀ (var)	z ₀ (sc)
0	1.214×10^{15}	19.91	-0.052	19.86	18.76	7.072	0.018	7.09	6.69
$1 imes 10^{14}$	$1.213 imes 10^{15}$	20.65	-0.050	20.60	19.53	7.008	0.018	7.03	6.63
2×10^{14}	$1.213 imes 10^{15}$	21.31	-0.048	21.34	20.28	6.945	0.017	6.96	6.56
5×10^{14}	$1.213 imes 10^{15}$	23.54	-0.043	23.50	22.48	6.768	0.015	6.78	6.38
1×10^{15}	$1.212 imes 10^{15}$	26.96	-0.037	26.92	25.96	6.511	0.012	6.52	6.13
2×10^{15}	1.211×10^{15}	33.31	-0.029	33.28	32.35	6.093	0.009	6.10	5.72
5×10^{15}	$1.210 imes 10^{15}$	49.83	-0.017	49.81	48.81	5.273	0.004	5.28	4.95
1×10^{16}	1.207×10^{15}	72.75	-0.011	72.74	71.50	4.522	0.002	4.52	4.26
$2 imes 10^{16}$	1.204×10^{15}	110.4	-0.006	110.4	108.7	3.760	0.001	3.76	3.56
5×10^{16}	1.198×10^{15}	197.9	-0.003	197.9	195.1	2.858	0.000	2.86	2.72
1×10 ¹⁷	1.192×10^{15}	311.2	-0.002	311.2	306.8	2.294	0.000	2.29	2.19

potential approaches the triangular potential, the variational values of both E_{00} and z_0 are 5.9% larger than the corresponding exact results. The trial function (12) has too slow a decay at large values of z; it gives an average of z^2 equal to $\frac{4}{3}z_0^2$, compared to the value $\frac{4}{5}z_0^2$ found for the Airy function.

The variational results are compared with numerical self-consistent calculations in Table II for several values of N_{inv}/N_{depl} , and the ratios of the variational and self-consistent values of E_0 and of z_0 are plotted in Fig. 2. Because of the scaling property discussed in Sec. III D, the results in Fig. 2 are valid for all electric-quantum-limit cases provided the correction terms in (15) and (16) are sufficiently small.

C. Excited States in Electric Quantum Limit

Approximate energy levels for the excited states can be obtained in the electric quantum limit by treating the inversion-layer potential and the cur-



FIG. 2. Ratios of variational to self-consistent values of the energy E_0 and of the average penetration z_0 of the lowest subband as functions of the ratio of the inversionlayer charge density to the depletion-layer charge density in the electric quantum limit, when only the lowest subband is occupied. Note the change of scale at $N_{inv}/N_{depl} = 5$.

vature of the depletion potential as perturbations. The inversion-layer charge density is assumed to be a sheet located a distance z_0 from the surface. If we neglect the oscillations in the Airy function¹⁸ and assume that the probability density for the excited states varies as $(a_i - z)^{-1/2}$,²⁰ where $a_i = E_{i,d}/eF_{dep1}$ and $F_{dep1} = eN_{dep1}/\kappa_{sc}\epsilon_0$, then we find that the energy levels for the excited states are approximately given by

$$E_{i} = E_{i,d} - e^{2} F_{dep1} F_{inv} z_{0}^{2} / 4E_{i,d}$$
$$- 4E_{i,d}^{2} / 15eF_{dep1} z_{d} + eF_{inv} z_{0}, \qquad (17)$$

where $F_{inv} = eN_{inv}/\kappa_{sc}\epsilon_0$ and $E_{i,d}$ is the energy obtained from (10b) by using the depletion-layer field F_{depl} as the surface field. The second term in (17) is the approximate lowering of the excited-state energy by the inversion-layer potential well. The third term gives the approximate contribution of the depletion-potential curvature to the energy levels; it is much more important for the excited states than is the corresponding term for the lowest subband, given in (16), because of the larger spatial extent of the excited states. The last term in (17) gives the inversion-layer contribution to the potential energy at the surface.

Equation (17) applies to the electric quantum limit, but it may still have some validity when more than one subband is occupied, provided z_0 is replaced by

$$z_{av} = \sum_{i} N_{i} z_{i} / N_{inv} , \qquad (18)$$

the average distance of all the inversion-layer electrons from the surface.

The quantity z_0 in (17) can be found by combining the variational expression (15) and the correction factor in Fig. 2. The magnitudes of the terms in (17) and of the energy obtained from a numerical self-consistent calculation are given in Table III for excited states of a sample inversion layer.

D. Scaling

In the electric quantum limit, Eqs. (2) and (4) can be written in dimensionless form, with $(\kappa_{sc}\epsilon_0\hbar^2/2m_3e^{2}N_{depl})^{1/3}$ as the unit of distance and $(\hbar^2/2m_3)^{1/3} \times (N_{depl}e^2/\kappa_{sc}\epsilon_0)^{2/3}$ as the unit of energy, as first pointed out by Howard.⁶ If the inversion-layer thickness is much smaller than the depletion-layer thickness, then the only parameter characterizing the dimensionless equations is N_{inv}/N_{depl} . Thus solutions for other semiconductors can be obtained from those given in this paper for Si in the electric quantum limit by appropriate rescaling There does not appear to be a simple scaling relation when more than one subband is occupied.

IV. DESCRIPTION OF CALCULATION

The self-consistent calculation starts with an initial estimate for the potential $\phi(z)$ and then solves Eqs. (2) and (4) successively until the output potential from (4) agrees with the input potential in (2) to within specified limits. The general method used to achieve self-consistency has been described elsewhere.²¹ Here we give only a brief description of the calculation.

The trial potential is obtained by assuming that all the carriers are in the lowest subband, for which we use the variational approximation described above. The trial eigenvalues for the initial integration of (2) are obtained by an approximation similar to (17). The trial potential generally leads to good convergence except at high temperatures, for which the assumption that all carriers are in the lowest sub-

TABLE III. Energies of excited states for a (111) surface at 0 K, given in meV. Columns 2-4 correspond to the first three terms on the right-hand side in Eq. (17). The inversion-layer contribution to the surface potential energy is $eF_{inv}z_0 = 338.5$ meV. The fifth column is the approximate energy calculated from (17), while the last column gives the energy from a numerical self-consistent calculation. We use $N_{depl} = 1.2 \times 10^{15}$ m⁻² (corresponding to $N_A - N_D = 10^{21}$ m⁻³) and $N_{inv} = 10^{17}$ m⁻²; the constants are taken from Table I. The numerical calculation gives $z_0 = 2.189$ nm and $E_0 = 306.8$ meV; the Fermi level is at 317.9 meV.

i	E _{i,d}	T_2	T_3	E (approx)	$E_i(sc)$
1	32.43	-10.53	-0.13	360.3	361.9
2	43.83	-7.79	-0.23	374.3	375.4
3	53.90	-6.34	-0.35	385.7	386.5
4	63.10	-5.41	-0.48	395.7	396.3
5	71.67	-4.76	-0.62	404.8	405.2
6	79.75	-4.28	-0.77	413.2	413.5
7	87.45	-3.91	-0.93	421.1	421.4
8	94.82	-3.60	-1.09	428.6	428.8
9	101.91	-3.35	-1.26	435.8	436.0
10	108.76	-3.14	-1.44	442.7	442.8
15	140.30	-2.43	-2.39	474.0	474.0
20	168.61	-2.03	-3.45	501.6	501.6

band is particularly bad. At high temperatures we start with small values of N_{inv} and then gradually go to larger values, taking the results of each case to construct the starting potential for the next.

The convergence criteria are that successive eigenvalues in the solution of the Schrödinger equation (2) with a given potential agree to within 10^{-6} eV and that the potentials in successive rounds of the iteration differ by no more than the larger of KT/2000, $(E_F - E_0)/10^4$, or 10^{-6} eV. This does not mean that the solutions are correct within these limits. We have not made a systematic study of the accuracy of our results, but it is thought to be considerably better than 1% in most cases. Greater accuracy is not needed for comparison with experimental results because physical parameters like the acceptor concentration, the position of the Fermi level in the bulk, and the dielectric constant are not known to greater accuracy.

The present program handles up to 25 subbands numerically, and a crude scaling approximation for the contribution of higher subbands to the potential is used if required. This is adequate at 300 K for (100) *n*-type inversion layers on Si if $N_A > 10^{21}$ m⁻³. Fewer subbands are needed for the other high-symmetry surfaces because they have a largerenergy splitting between the lowest subband and the first excited subband. At lower temperatures the number of subbands required drops sharply. The present version of the program automatically assigns the grid interval used in the numerical integration, and fixes the number of grid points needed for each subband. Up to 1201 grid points can be used, but will normally not be required except for cases with very many subbands.

The computer program has evolved over a period of years. It contains 13 subroutines and about 2000 FORTRAN statements. Execution of the program requires 280K bytes of storage, much of it for the wave functions themselves. A run with 17 subbands and 748 grid points, which took six iterations to converge, required 5.2 sec of central-processingunit time on an IBM System/360 Model 91 data processing system.

V. SELF-CONSISTENT RESULTS

In this section we give some representative results of our self-consistent calculations to illustrate properties and trends in the solutions. A set of results complete enough to encompass the entire range of temperatures and doping levels used by experimenters would be too lengthy to be published here. We hope nevertheless that this sampling will be useful.

Figure 1 shows the potential energy at 0 K for a (111) surface with $N_A - N_D = 10^{21}$ m⁻³ and with $N_{inv} = 10^{17}$ m⁻², the same values as for Table III. This is a case in which N_{inv} is much larger than N_{depl}

= 1.2×10^{15} m⁻², so that the electric field at the surface is much larger than the field in the depletion layer.

As N_{inv} increases from zero at low temperatures, the separation $E_F - E_0$ of the Fermi level from the bottom of the lowest subband increases faster than the splitting $E_1 - E_0$ between the next subband and the lowest subband. But after the Fermi level crosses the bottom of the first excited subband, the rate of change of $E_F - E_0$ decreases and the rate of change of $E_1 - E_0$ increases; the Fermi level seems to be pinned near the bottom of the second subband.⁶ This effect is shown in Fig. 3 for a (100) surface at 0 K for a sample with $N_A - N_D = 10^{20}$ m⁻³.

The results of the self-consistent calculations are not easily characterized by a single parameter, but one that is quite important is z_{av} , the average penetration of the inversion-layer charge density from the surface. This quantity is shown in Fig. 4 for $N_A - N_D = 10^{21}$ m⁻³ for the (100) and (111) surfaces as a function of the total number of charges in the space-charge layer, i.e., $N_{inv} + N_{depl}$, at 2, 77, and 300 K. Near threshold, i.e., when the inversion layer is first populated, z_{av} is almost the same for all three surfaces at 300 K, showing that quantum effects are relatively unimportant there. A calculation on the continuum model confirms that the results at 300 K are very close to those obtained here; they differ substantially at lower temperatures.⁴

As the temperature is increased, inversion-layer electrons will be excited to higher-lying subbands, with larger average values of z, and the inversionlayer contribution to the ptoential well increases. The temperature dependence of z_0 and z_{av} and of the



FIG. 3. Energy-level splittings and Fermi energy at 0 K for a (100) surface inversion layer on *p*-type silicon with 10^{20} net acceptor ions per m³, showing the change in slope when the Fermi level passes into the first excited subband. Note that the bottoms of subbands 1 and 0' cross where $N_{\rm inv}/N_{\rm depl} = 5.3$. All energies are measured from the bottom of the lowest subband.



FIG. 4. Average penetration of inversion-layer electrons into the semiconductor as a function of total density of charges in the inversion and depletion layers for (a) (100) and (b) (111) surfaces on p-type Si with 10^{21} net acceptor ions per m³. The values for a (110) surface are very close to those for a (111) surface.

energy-level splittings is shown for a (100) surface in Fig. 5 for $N_A - N_D = 10^{21} \text{ m}^{-3}$ and $N_{inv} = 10^{16} \text{ m}^{-2}$.

A measure of the importance of quantum effects is the distribution of carriers among the subbands. Figure 6 shows the fraction of electrons in the lowest subband for several temperatures for $N_A - N_D = 10^{21} \text{ m}^{-3}$ as a function of $N_{inv} + N_{devl}$ for the three high-symmetry surface orientations. Also shown is the fraction of the electrons in all the subbands in the ladder associated with the highest value of m_3 , the mass for motion perpendicular to the surface. When quantum effects are unimportant, this ratio should be the degeneracy for this ladder, as given in Table I, divided by 6, the total number of valleys. We see in Fig. 6 that the fraction of electrons in the lowest ladder of subbands does indeed approach the expected value at 300 K when $N_{inv} + N_{depl}$ is small, but departs from it both for low temperatures and for large values of $N_{inv} + N_{depl}$. Thus, quantum effects must be considered even at room temperature if accurate re-



FIG. 5. Temperature dependence of (a) z_0 and z_{av} , the average penetration of electrons in the lowest subband and in all subbands, respectively, and (b) the energies of the lowest seven subbands, for a (100) surface with $N_A - N_D = 10^{21} \text{ m}^{-3}$ and $N_{iav} = 10^{16} \text{ m}^{-2}$. Note that levels 0' and 1 cross at about 35 K. Note also that N_{dep1} decreases from $1.21 \times 10^{15} \text{ m}^{-2}$ at 0 K to $0.96 \times 10^{15} \text{ m}^{-2}$ at 300 K. If the surface field were held constant at the value for 0 K, splittings at 300 K would be about 1% larger than the values shown here for a fixed value of N_{iav} .

sults are to be obtained for high surface-charge densities.

VI. DISCUSSION

Sections $\Pi I-V$ give representative results of selfconsistent calculations for *n*-type inversion layers on Si. In this section we comment on the validity of some of the approximations that we use.

The assumption that the envelope wave function vanishes at the surface simplifies the calculation considerably, but a more realistic boundary condition specifies the ratio of the wave function to its derivative at the surface, a value which depends on the nature and height of the barrier there. The energy change connected with this more general boundary condition is easily calculated for the triangular potential, whose eigenfunctions and eigenvalues are given in (10), by using tabulated values of the Airy function and its derivative.¹⁸ The dependence of the eigenvalue on the boundary condition is shown in Fig. 7.

For a numerical estimate of the effect of more general boundary conditions, let us take for $\psi(d\psi/dz)^{-1}$ the value 0.2 nm appropriate for a free electron in a 1-eV barrier, a smaller barrier than the one present at the Si-SiO₂ interface.¹⁴ If the field just inside the semiconductor is 3×10^8 V/m, a relatively high field, and $m_3 = 0.916m$, then the scaling distance a^* for the abscissa of Fig. 7 is



FIG. 6. Fraction of carriers in the lowest subband, and in all the subbands associated with the valleys having the largest value of m_3 , the mass for motion perpendicular to the surface, for the (a) (100), (b) (110), and (c) (111) surfaces. The fraction in the lowest valleys should be $\frac{1}{3}$ and $\frac{2}{3}$, respectively, for the (100) and (110) surfaces when quantum effects are negligible. It is identically unity for the (111) surface, for which all valleys have the same value of m_3 .

4897

FIG. 7. Effect of boundary conditions on energy levels for the triangular potential. The abscissa is the scaled ratio of the envelope wave function at the surface to its derivative there. The scaling distance a^* is $(\hbar^2/2m_3eF_s)^{1/3}$, where F_s is the electric field, and the scaling energy E^* is eF_sa^* . The energies for $d\psi/dz=0$. are marked at the right-hand side.

0.52 nm and the scaling factor E^* for the energy scale is 0.155 eV. For a value of 0.38 on the abscissa of Fig. 7 the energy levels are reduced by less than 15% from the values found with a wave function which vanishes at the surface, and the energy-level splittings are increased by less than 2%. We conclude that within the framework of the effective-mass approximation the errors made in taking the wave function to vanish at the surface are relatively small.

The effective-mass approximation itself will begin to fail when the calculated energies are comparable to the energy gaps in the bulk and when the spatial extent of the wave function for the lowest subband is comparable to atomic dimensions. The resulting errors have been estimated by Howard¹³; they become significant for wave functions whose average penetration z is of order 2 nm or less. Exchange and correlation effects in the inversion layer have been ignored in our work, an approximation which requires further examination.

At the highest surface fields it may be necessary to undertake a more fundamental calculation of the effects of interfaces on wave functions and eigenvalues. This problem is somewhat like the problem of low-energy electron diffraction, because both involve the matching of waves across an interface between two media. It may be possible to include within the framework of a simple model both the surface subbands associated with band bending and the conventional surface states which arise from the presence of the interface itself.

The most promising experimental method to determine inversion-layer subband splittings for comparison with the calculations appears to be to look for structure in the photoconductivity ex-

citation spectrum. This method was first used by Katayama *et al.*²² for InSb, and has been used for Si by Wheeler and Ralston.²³ The latter authors used a laser with a fixed photon energy, and varied the surface electric field to change the subband splittings. We have calculated the energy levels expected from the parameters they report, but find considerably smaller subband splittings than the 44 meV of their laser if we use their subband assignments. Alternative assignments also did not fit. It is probably too early to draw conclusions from this comparison for a single sample, but more extensive measurements should allow a comparison with self-consistent calculations to be made. Experimental information on subband splittings in accumulation layers is provided by the tunneling measurements of Tsui.²⁴

ACKNOWLEDGMENTS

The author is indebted to A. B. Fowler and W. E. Howard for many valuable discussions, and to M. I. Nathan for comments on the manuscript.

APPENDIX A: DEPLETION-LAYER-EDGE EFFECTS

We have assumed in Eq. (7) that the depletionlayer charge density is constant for $0 < z < z_d$, and then changes abruptly to zero. The actual charge density decays smoothly, with a characteristic length given by the bulk screening length. Near the surface, this effect can be accounted for by adding a constant to the band bending used to calculate z_d in Eq. (7). We must now think of z_d not as the abrupt termination of the depletion layer, but as a measure of its charge density, defined by $z_d = N_{depl} / (N_A - N_D)$. We give the required correction term in this Appendix, and also give values for the screening length. The derivation is straightforward, and has been omitted.¹⁷

If the bulk is characterized by acceptor and donor concentrations N_A and N_D , respectively, and by hole concentration p, if all the minority impurities are ionized, and if there is a negligible minority-carrier concentration, then we find that the bulk-screening length L is given by²⁵

$$L^{2} = \kappa_{sc} \epsilon_{0} K T N_{A} / e^{2} [N_{A} p + (N_{D} + p)(N_{A} - N_{D} - p)] .$$
(A1)

At high temperatures, where $p \sim N_A - N_B$, we get the familiar result

$$L^2 \approx \kappa_{\rm sc} \epsilon_0 KT / e^2 (N_A - N_D) \quad , \tag{A2}$$

while at low temperatures, where $p \ll N_A - N_B$,

$$L^2 \approx \kappa_{sc} \epsilon_0 K T N_A / e^2 N_D (N_A - N_D) . \tag{A3}$$

The surface-charge density $\kappa_{sc} \epsilon_0 (d\phi/dz)$ can be found as a function of ϕ by multiplying both sides of Poisson's equation by $\phi' \equiv d\phi/dz$ and integrating. We find

$$\phi'^{2} = (2/\kappa_{sc}\epsilon_{0})((N_{A} - N_{D})e\phi - KT\{p + N_{A}\ln[N_{A}/(N_{D} + p)]\}).$$
(A4)

The term in KT is the correction term which takes the edge effects into account. At high temperatures, ¹⁷

$$\phi'^2 \approx (2/\kappa_{sc}\epsilon_0)[(N_A - N_D)(e\phi - KT)]$$
, (A5)

while at low temperatures,

$$\phi^{\prime 2} \approx (2/\kappa_{sc}\epsilon_0) \left[(N_A - N_D) e \phi - N_A K T \ln(N_A/N_D) \right].$$
(A6)

At very low temperatures it is necessary to consider the width of the acceptor level. If we characterize it by a Gaussian distribution whose full energy width at half-height is Δ , we find the following approximate results when $KT \ll \Delta$:

$$L^{2} \approx 0.4 \kappa_{\rm sc} \epsilon_{0} \Delta / e^{2} N_{A}^{0.17} N_{D}^{0.83} , \qquad (A7)$$

$$\phi'^2 \approx (2/\kappa_{sc}\epsilon_0) [(N_A - N_D)e\phi$$

- 0.45 $\Delta N_A^{1.17} N_D^{-0.17}]$. (A8)

The numerical coefficients and exponents are approximately correct for $N_D/N_A = 0.01$.

We have used the high-temperature result (A5) in Eq. (9) and in our calculations over the entire temperature range. The difference between (A5) and the more exact results (A4) or (A8) is negligible

¹A Many, Y. Goldstein, and N. B. Grover, Semiconductor Surfaces (North-Holland, Amsterdam, 1965).

²D. R. Frankl, Electrical Properties of Semiconductor Surfaces (Pergamon, New York, 1967).

³R. F. Greene, in Solid State Surface Science, edited by M. Green (Marcel Dekker, New York, 1969), Vol. 1, p. 87.

⁴F. Stern, in Proceedings of the Tenth International Conference on the Physics of Semiconductors, Cambridge, Massachusetts, 1970 (U. S. Atomic Energy Commission, Oak Ridge, Tenn., 1970), p. 451.

⁵F. Stern and W. E. Howard, Phys. Rev. <u>163</u>, 816 (1967).

⁷H. Maeda (unpublished).

⁸F. Stern, Phys. Rev. Letters <u>21</u>, 1687 (1968).

⁹C. B. Duke, Phys. Rev. <u>159</u>, 632 (1967).

 10 J. A. Appelbaum and G. A. Baraff, Phys. Rev. B <u>4</u>, 1235 (1971); 4, 1246 (1971).

¹¹G. A. Baraff and J. A. Appelbaum, J. Vac. Sci.

Technol. (to be published); Phys. Rev. B 5, 475 (1972). The author is indebted to these authors for preprints.

¹²See, for example, A. B. Fowler, F. F. Fang, and F. Hochberg, IBM J. Res. Develop. 8, 427 (1964).

¹³W. E. Howard (unpublished).

¹⁴R. Williams, Phys. Rev. <u>140</u>, A569 (1965).

¹⁵We use mks units in this paper. To convert to cgs units, replace ϵ_0 by $1/4\pi$. Note that 1 nm = 10 Å.

for the cases we consider, but might be significant in other calculations.

APPENDIX B: INTEGRALS OF AIRY FUNCTIONS

In connection with the normalization of the eigenfunction (10a) of the triangular-potential well and with the calculation of averages of powers of z, we note the following results:

$$\int_{z}^{\infty} \operatorname{Ai}^{2}(z') dz' = -z \operatorname{Ai}^{2}(z) + \operatorname{Ai}^{\prime 2}(z), \qquad (B1)$$

$$\int_{z}^{\infty} z' \operatorname{Ai}^{2}(z') dz' = \frac{1}{2} \int_{z}^{z} \operatorname{Ai}^{2}(z) + z \operatorname{Ai}^{\prime 2}(z), \qquad (B1)$$

$$\int_{z}^{\infty} z'^{2} \operatorname{Ai}^{2}(z') dz' = \frac{1}{5} [-z^{3} \operatorname{Ai}^{2}(z) + z^{2} \operatorname{Ai}'^{2}(z)], \quad (B2)$$

$$-2z \operatorname{Ai}(z) \operatorname{Ai}'(z) + \operatorname{Ai}^{2}(z)] . \quad (B3)$$

These results are easily verified if we use Ai''(z) $=z \operatorname{Ai}(z)$, where Ai' and Ai'' denote first and second derivatives, respectively, of Ai with respect to z. For the boundary condition used in our work, the lower limit of the integrals is $(\hbar^2/2m_3eF_s)^{1/3}$ times the (i+1)st zero of the Airy function, ¹⁸ for which the results take a much simpler form. To obtain the results quoted in Sec. III A, note that the z axis here has its origin at the turning point where the energy is equal to the potential energy, while in the body of the paper the origin is at the surface. The difference is $a_i = E_i / eF_s$.

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

¹⁷The high-temperature case has been treated by J. A. Lewis, J. McKenna, and E. Wasserstrom [J. Appl. Phys. <u>41</u>, 4182 (1970)]. See also Refs. 1 and 2.

¹⁸See, for example, in Handbook of Mathematical Functions, edited by M. Abramowitz and I. A. Stegun (U. S. GPO, Washington, D. C., 1964), Chap. 10.

¹⁹F. F. Fang and W. E. Howard, Phys. Rev. Letters

<u>16</u>, 797 (1966). ²⁰This is equivalent to the classical result that the probability density is inversely proportional to the velocity. See, for example, L. I. Schiff, Quantum Mechanics (McGraw-Hill, New York, 1955), 2nd ed., pp. 65-66.

²¹F. Stern, J. Compt. Phys. <u>6</u>, 56 (1970).

²²Y. Katayama, N. Kotera, and K. F. Komatsubara, in Proceedings of the Tenth International Conference on the Physics of Semiconductors, Cambridge, Massachusetts, 1970 (U. S. Atomic Energy Commission, Oak Ridge, Tenn., 1970), p. 464.

²³R. G. Wheeler and R. W. Ralston, Phys. Rev. Letters 27, 925 (1971).

²⁴D. C. Tsui, Phys. Rev. B <u>4</u>, 4438 (1971).

²⁵Corresponding results for n-type material are given by H. Brooks [in Advances in Electronics and Electron Physics, edited by L. Marton (Academic, New York, 1955), Vol. VII, p. 158].

⁶W. E. Howard (unpublished).