Localized states of p-type inversion layers in semiconductor field-effect transistors

N. Garcia* and L. M. Falicov†
Department of Physics, University of California, Berkeley, California 94720
(Received 20 August 1974)

A variational calculation is presented for the electronic configuration of a p-type surface accumulation layer of a metal-oxide-semiconductor device. The self-consistent equations show that the criterion for the penetration of a second (light-mass) subband of carriers, after a first (heavy-mass) subband is present depends crucially on the oxide-layer thickness and the transverse mass of the heavy carriers. Agreement with reported experimental results is excellent.

Several experimental papers in recent months[1-6] have reported interesting results on the electronic properties of inversion layers in semiconductor surfaces. These layers are obtained experimentally by applying suitable gate biases in metal-oxide-semiconductor (MOS) devices. The experiments are performed in both p-type and n-type layers in various types of silicon and on various crystallographic surfaces. Theoretical calculations of properties of semiconductor inversion layers are available in the literature.[7-13] Although the general physical principles involved in the experiments are well understood, several points remain to be clarified and some discrepancies to be resolved.

In particular, for p-type inversion layers on the (110) surface in n-type silicon, von Klitzing *et al.*[1] in their study with a 1200-Å oxide-layer MOS sample report the existence, up to 23-V gate bias, of only one bound level (subband) with a transverse cyclotron mass $m_c = 0.33-0.47m_0$. For gate biases larger than 23 V a second subband appears with a different cyclotron mass.

On the other hand, Lakhani *et al.*,[2] who performed similar experiments on (100) and (111) Si surfaces with MOS devices of 1400- and 1900-Å-thick oxide layers, find that up to gate voltages of about 36 V only one hole subband appears. The cyclotron (transverse) mass of these samples varies strongly with gate voltage between 0.6 and approximately $1m_0$.

The variation of the mass with carrier concentration has been presumed to be either a two-dimensional many-body effect[1,2] or a consequence of the nonparabolicity[2] of the top of the valence band in Si. We shall show elsewhere[14] that ordinary band theory predicts anisotropy and nonparabolicity of the bands, which are sufficient to explain adequately the observed experimental results.

In the present paper, we want to study in a self-consistent way the *appearance and properties of a second subband*. Our investigation shows that the gate voltage at which a second subband begins to be populated depends crucially on the oxide-layer thickness and on the density-of-states (transverse) effective mass. Our results are compatible with both sets of experimental data.[1,2]

Our calculation consists essentially of a minimization of the grand potential at $T = 0$ in the semiconductor. The grand potential rather than the energy must be taken because the gate-voltage connection between the semiconductor and the metal in the MOS device allows for electronic charge transfer between them. Schrödinger and Poisson's equations should be solved simultaneously and self-consistently, subject to the conditions of fixed electrostatic-potential difference, as set by the experimental conditions.

We assume the following.

(a) The device extends indefinitely and uniformly on the x and y directions.

(b) The oxide (SiO_2), with a dielectric constant $\epsilon_0 = 4.55$, extends between $z = -\delta$ and $z = 0$. The thickness δ is our *first variable parameter*.

(c) The semiconductor extends to the right of the oxide between $z = 0$ and $z = L$. The thickness L of the semiconductor is such that $L \gg \delta$; i.e., the semiconductor may in principle be thought to extend indefinitely and constitute a semi-infinite bulk sample.

(d) The semiconductor (Si) is intrinsically n type, has a dielectric constant $\epsilon_s = 11.7$, and has its Fermi level at an energy $C = 1.01$ eV above the top of the valence band.

(e) The top of the valence band is taken to consist of two bands, degenerate at $k = 0$. These two bands have a longitudinal mass[1,15] along the z axis (normal to the surface) equal to $m_{LH} = 0.52m_0$ and $m_{LL} = 0.17m_0$.

(f) If the zero of energy is taken at the Fermi energy, the band dispersion relations close to the top of the valence band are taken to be

$$E_H(\vec{k}) = -C - \frac{\hbar^2}{2m_{LH}} k_z^2 - \frac{\hbar^2}{2m_{TH}} (k_x^2 + k_y^2) , \tag{1}$$

$$E_L(\vec{k}) = -C - \frac{\hbar^2}{2m_{LL}} k_z^2 - \frac{\hbar^2}{2m_{TL}} (k_x^2 + k_y^2) . \tag{2}$$

The transverse masses of the heavy (m_{TH}) and light (m_{TL}) holes enter into our calculation essentially

through a density of states. We take them as parameters; in fact only m_{TH} is needed, and we take it as our *second variable parameter*.

(g) In the absence of a charge layer, the electrostatic potential Φ_0 is given by

$$\Phi_0(z) = \begin{cases} V\dfrac{\epsilon_s z - \epsilon_0 L}{\epsilon_s \delta + \epsilon_0 L}, & -\delta \le z \le 0 \\ V\dfrac{\epsilon_0 (z - L)}{\epsilon_s \delta + \epsilon_0 L}, & 0 \le z \le L \,. \end{cases}$$

The parameter V (our *third variable parameter*) is the value of the applied potential at the metal (with negative sign). The potential is zero at $z = L$, deep into the bulk of the semiconductor. In the neighborhood of the O-S interface, in the absence of trapped charge, Φ is a constant in the limit $L \to \infty$, equal to $(-V)$.

It should be emphasized that the potential Φ_0 given above corresponds to that in the absence of any mobile charge. When charge is trapped in the interface layer the electrostatic potential Φ has a completely different form which is determined self-consistently by solving Poisson's equation. It satisfies, however, the same boundary conditions: zero value deep into the semiconductor and value $(-V)$ at $z = -\delta$.

(h) The hole carriers can only move in the semiconductor ($0 \le z \le L$), and they feel a potential W which is due to the Si lattice and the applied (outside) potential. If a bound state with energy greater than zero—the Fermi level—appears, such state will be emptied, i.e., occupied by a hole. The grand potential

$$G = E_T - N\epsilon_F \tag{3}$$

is therefore the quantity to minimize. With the convention $\epsilon_F = 0$, $G = E_T$, but it should be remembered that there is no conservation of number of carriers. The conserved quantity is the chemical potential or Fermi energy. We should like to remark that minimization of the grand potential G is completely equivalent to minimization of the total energy E_T, with the constraint that the trapped charge in the semiconductor is equal in absolute value to the charge in the condensor plate at $z = -\delta$.

(i) The total energy E_T in the Hartree (*effective mass*) approximation is given by

$$E_T = E_T^0 - \sum_{n\nu} \langle \Psi_{n\nu}(\vec{r}) \left| E_\nu(-i\vec{\nabla}) - |e| \Phi_0(z) \right| \Psi_{n\nu}(\vec{r}) \rangle$$

$$- \frac{1}{2} \sum_{nn'\nu\nu'}{}' \langle \Psi_{n\nu}(\vec{r}_1)\Psi_{n'\nu'}(\vec{r}_2) \left| \frac{e^2}{|\vec{r}_1 - \vec{r}_2|} \right| \cdot \times \Psi_{n\nu}(\vec{r}_1)\Psi_{n'\nu'}(\vec{r}_2) \rangle \,, \tag{4}$$

where $\Psi_{n\nu}$ is a one-electron wave function characterized by the n quantum number corresponding to the ν carrier ($\nu = H$ or L). The operator $E_\nu(-i\vec{\nabla})$ is the band-structure expression (1) or (2), in which $k \to (-i\vec{\nabla})$. The summations in (4) are over those states which are occupied by carriers, i.e., those $(n\nu)$ states which contribute to lower E_T. This is equivalent to saying that the summation is over those $(n\nu)$ states such that their one-electron Hartree energy is positive, greater than the Fermi energy. Standard minimization of (4) with respect to $\Psi_{n\nu}$ yields a Hartree-Poisson system of

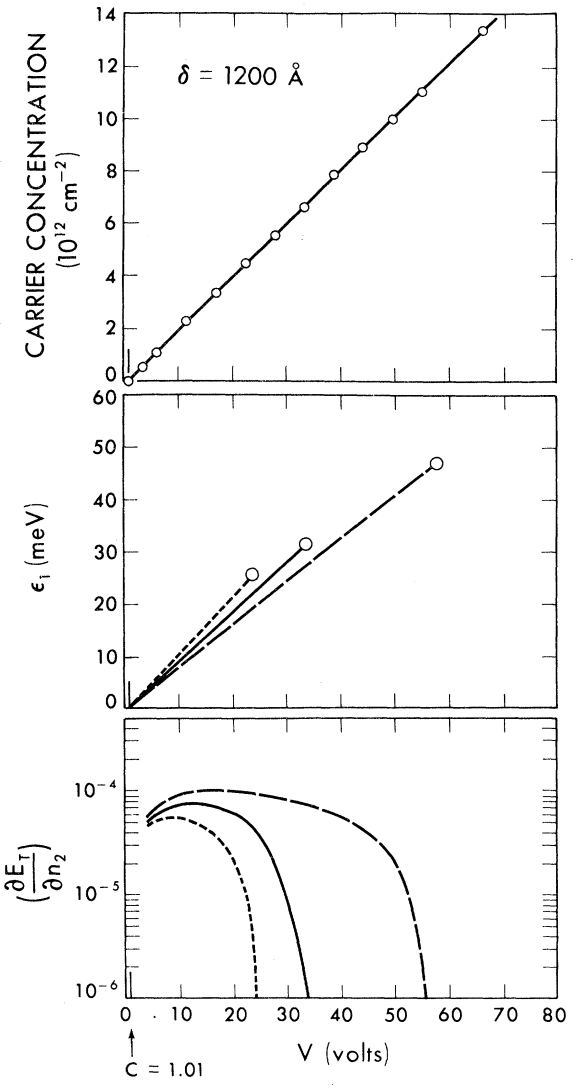

FIG. 1. Results for an MOS sample with an oxide layer of $\delta = 1200$-Å thickness. The heavy carrier concentration (upper diagram), the position of the top of the hole subband (middle diagram) and the generalized susceptibility $\delta E_T/\delta n_2$ (lower diagram) as a function of gate bias. The upper diagram is for all transverse masses. The middle and lower diagrams are for $m_{TH} = 0.45 m_0$ (short dashes), $m_{TH} = 0.52 m_0$ (full line), and $m_{TH} = 0.60 m_0$ (long dashes). A second subband appears when $\delta E_T/\delta n_2 = 0$.

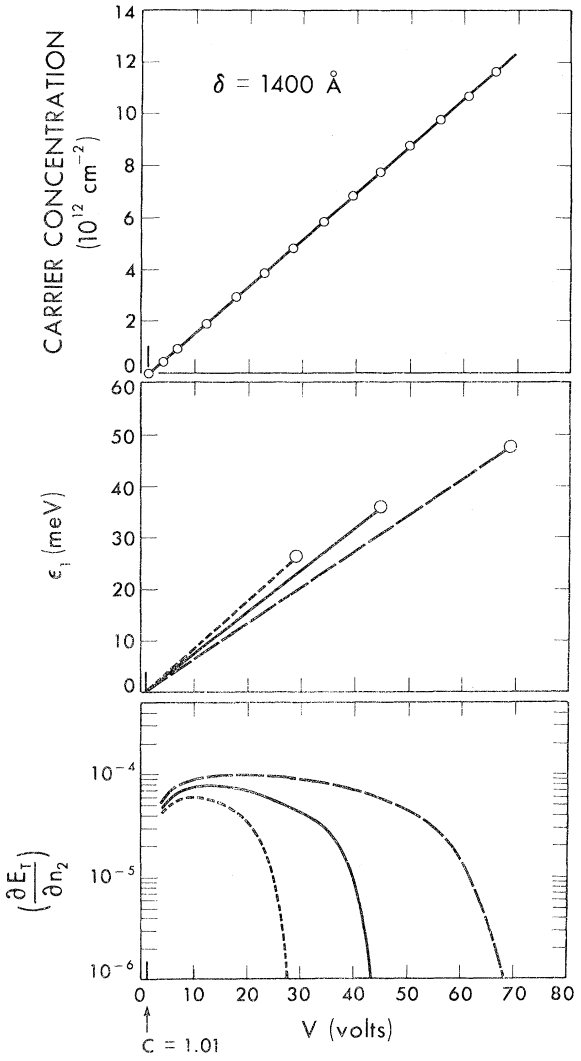

FIG. 2. Results for an MOS sample with an oxide layer of δ=1400-Å thickness. Notation same as for Fig. 1.

equations. The carriers move only in the half-space $z \geq 0$ and in a self-consistent potential, which satisfies Poisson's equation. The electrostatic part of this potential satisfies Maxwell's equations, continues into $-\delta \leq z \leq 0$ with the proper boundary conditions, and reaches the value $-V$ at $z=-\delta$. Since the problem is separable and the potential depends only on z, we have

$$\Psi_{n\nu}(\vec{r}) = \frac{1}{(L_x L_y)^{1/2}} e^{i(k_x x + k_y y)} f_{n\nu}(z) \ . \tag{5}$$

It must be pointed out that the total energy E_T which goes into G contains in the second term the interaction of the carriers with the "outside" field—i.e., the field that would appear if the carriers were not present—and, in the third term, the interaction of the carriers with the field generated by the other carriers—i.e., the field due to the charge on the surface of the semiconductor which clamps the potential at $z=-\delta$ and $z=L$. We have not included the electromagnetic field energy, proportional to the integral of the square of the electric field, which is negligibly small.

(j) The function $f_{n\nu}(z)$ describes a bound state in the O-S interface region. We have approximated, for the highest eigenvalue of each carrier, $f_\nu(z)$ by

$$f_\nu(z, \lambda_\nu) = Az\, e^{-\lambda_\nu z} \tag{6}$$

and we have treated λ_ν as a variational parameter. Suitable modifications for other states are straightforward but we have not attempted to calculate such cases. The analytic expression (6) allows one to calculate a potential and an energy in closed form. Various authors have previously used such a form,[7,8,14,15] thus taking advantage of the resulting analyticity.

The problem now can be solved in the standard fashion: Given the input parameters (δ, V, C, m_{LH}, m_{TH}, m_{TL}) we place holes only in the highest energy eigenstate of each carrier in the z direction. These states are determined by the variational parameters λ_H and λ_L. Along the other two directions x and y we fill the states according to Fermi statistics up to a maximum level. This involves two carrier-density parameters n_H and n_L (carriers per unit surface area). We minimize (4) with respect to the λ's and the n's, subject to the proper boundary conditions for the electrostatic potential. The complete calculation is done analytically up to a self-consistency set of nonlinear equations, which then must be solved numerically. Some results follow.

(i) For $V<C=1.01$ V, no self-consistent bound state is occupied by carriers.

(ii) At $V \gtrsim C$ always the carriers with the heavier longitudinal mass m_{LH} are the first to produce an occupied subband. This result has been found before repeatedly.[7-13]

(iii) There is a voltage range $C \leq V \leq V_2$ for which only one subband of only one mass (the heaviest) is occupied. We have studied that region in detail for $m_{LH}=0.52$, values of $\delta=1200$ Å (Fig. 1) and $\delta=1400$ Å (Fig. 2), and varying V. In the top diagram of the figures we plot the carrier concentration n_H as a function of applied voltage (n_H is independent of the transverse mass m_{TH}). In the middle diagram we plot the energy ϵ_1 of the top of the subband: It crosses the Fermi level at $V=C$ and increases linearly with V. Its rate of increase depends on the transverse mass m_{TH}. In Figs. 1 and 2 we plot three curves corresponding to three values of m_{TH}: $0.45m_0$, $0.52m_0$, and $0.60m_0$. For the same m_{LH} and m_{TH}, both n_H and ϵ_1 decrease with increasing δ.

(iv) In order to calculate V_2, the gate-bias voltage, at which a second carrier gets trapped, we have evaluated the derivative of the total energy E_T [Eq. (4)] with respect to n_2, the concentration of the second carrier ($n_2 = n_L$ in our case) at $n_2 = 0$ and λ_H and n_H at the equilibrium value. The parameter λ_L was chosen so as to minimize the derivative. This is a susceptibility-type calculation which gives a stability criterion: If $\delta E_T/\delta n_2$ is positive, the system is stable against introduction of a second set of carriers; $\delta E_T/\delta n_2$ is negative, the system is unstable. Parameters for which

$$\frac{\delta E_T}{\delta n_2} = 0 \tag{7}$$

give the critical values at which a second set of carriers gets localized and a second subband appears. In Figs. 1 and 2—third diagram—the values of $\delta E_T/\delta n_2$ are plotted for the same parameters as before. We can see that V_2 is a very sensitive function of both m_{TH} and δ and can easily show variations of up to a factor of 3. For $\delta = 1200$ Å and $m_{TH} = 0.45 m_0$ the second subband (light mass) appears at a bias $V_2 = 23$ V, while for $\delta = 1400$ Å and $m_{TH} = 0.60 m_0$ the second subband appears at about $V_2 = 69$ V.

(v) As an incidental interesting remark, we find that if we calculated $\delta E_T/\delta n_2$ for $n_1 = n_L$ and $n_2 = n_H$, i.e., for the case in which arbitrarily we allow the light carriers to get localized before the heavy ones, we find the system always unstable, $\delta E_T/\delta n_2 < 0$.

(vi) We have not attempted a numerical solution for both n_H and n_L nonvanishing ($V > V_2$). This can be achieved in principle—albeit laboriously—if experimental interest in such a case does arise.

---

*Juan March Foundation (Spain) Fellow.

†Work supported in part by the National Science Foundation under Grant No. DMR72-03106-A01.

[1]K. von Klitzing, G. Landwehr, and G. Dorda, Solid State Commun. 14, 387 (1974).

[2]A. A. Lakhani, P. J. Stiles, and Y. C. Cheng, Phys. Rev. Lett. 32, 1003 (1974).

[3]A. Kamgar, P. Kneschaurek, G. Dorda, and J. F. Koch, Phys. Rev. Lett. 32, 1251 (1974).

[4]G. Abstreiter, P. Kneschaurek, J. P. Kotthaus, and J. F. Koch, Phys. Rev. Lett. 32, 104 (1974).

[5]S. J. Allen, Jr., D. C. Tsui, and J. V. Dalton, Phys. Rev. Lett. 32, 107 (1974).

[6]K. F. Komatsubara, K. Narita, Y. Katayama, N. Kotera, and M. Kobayashi, J. Phys. Chem. Solids 35, 732 (1974).

[7]F. Stern, Phys. Rev. Lett. 21, 1687 (1968).

[8]F. Stern, Phys. Rev. B 5, 4891 (1972).

[9]C. B. Duke, Phys. Rev. 159, 632 (1967).

[10]J. A. Appelbaum and G. A. Baraff, Phys. Rev. B 4, 1235 (1971); 4, 1246 (1971).

[11]G. A. Baraff and J. A. Appelbaum, Phys. Rev. B 5, 475 (1972).

[12]F. F. Fang and W. E. Howard, Phys. Rev. Lett. 16, 797 (1966).

[13]F. Stern and W. E. Howard, Phys. Rev. 163, 816 (1967).

[14]N. Garcia and L. M. Falicov (unpublished).

[15]See, for instance, C. Kittel, *Introduction to Solid State Physics*, 4th ed. (Wiley, New York, 1971), p. 384.