Electronic Structure of Inversion Layers in Many-Valley Semiconductors

M. J. Kelly* and L. M. Falicov
Department of Physics,† University of California, Berkeley, California 94720
(Received 1 June 1976)

We outline a new theory of the electronic structure of inversion layers in many-valley semiconductors which includes broken-symmetry many-body effects. This theory describes charge-density waves which account for the experimentally observed (i) degeneracy and (ii) occupancy of the electron levels, and (iii) the high cyclotron mass in the (111) surface n-type inversion layers of silicon. It reduces to earlier theories at the Si (100) surface, where theory and experiment are in good agreement.

Recent experiments[1-3] on n-type inversion layers in silicon at the (111) surface in the metal-oxide-semiconductor configuration have produced a series of well-characterized results which cannot be explained in terms of existing (essentially one-electron) theories.[4-6] In these theories, the inversion-layer electron wave functions are written as[7]

$$\psi(\vec{r}) = \sum_v \alpha_v F_v(\vec{r}) \varphi_v(\vec{r}), \tag{1}$$

where $\varphi_v(\vec{r})$ is the Bloch function at one of the (100) conduction band minima, $F_v(r)$ is the envelope function which satisfies the effective-mass Schrödinger equation, and the summation extends over all valleys v.[7] The coefficients α_v determine the normalized linear combination of valleys. For the Si (111) surface, where the six valleys are equivalent, six independent and degenerate wave functions can be constructed. Experimental data[1,3] reveal for Si (111) a valley degeneracy of 2 ± 0.2 instead of 6. Existing theories, which include longitudinal-transverse mass mixing,[4] predict a cyclotron mass for Si (111) of $0.358 m_e$, while the experimental values[1] are somewhat higher at $(0.40 \pm 0.03) m_e$.

The present theory incorporates correlations and exchange effects between different valleys. It is similar in spirit to the approach of Penn,[8] although our many-valley Hamiltonian requires a very different treatment. As a result, the ground state of the system, over a wide range of values for the electronic interaction, is that of either a charge-wave or a spin-density wave, with the attendant correlations pushing the experimentally unoccupied states well above the Fermi level.

We solve the problem in two stages. We first follow earlier theories and study, self-consistently or variationally, the ordinary paramagnetic state. For ν identical valleys,[9] described by a normal effective mass m_3, density-of-states transverse mass m_d, smooth envelope function[10]

$$F(\vec{r};\lambda,\vec{k}_{\parallel}) = A^{-1/2}(2\lambda^{3/2}) z e^{-\lambda z} \exp(i\vec{k}_{\parallel}\cdot\vec{r}), \tag{2}$$

and a total occupancy of n electrons per unit area, the total energy per unit area of the system in the Hartree approximation is

$$E_T = \frac{\hbar^2 \pi n^2}{2\nu m_d} + \frac{\hbar^2 n \lambda^2}{2 m_3} - \frac{enV}{2} + \frac{15\pi e^2 n^2}{16 \epsilon_s \lambda}, \tag{3}$$

where V is the gate voltage and ϵ_s is the static dielectric constant for silicon.[11] The energy is minimized with respect to λ and n, subject to Maxwell's equations,

$$n = \frac{V}{4\pi e(\delta/\epsilon_0 + 3/2\lambda\epsilon_s)}, \tag{4}$$

for oxide thickness δ and oxide dielectric constant ϵ_0. For typical values, λ varies approximately as $n^{1/3}$ and is such that for $n \approx 3.6 \times 10^{12}$ cm^{-2}, $\lambda \sim 0.06$ Å^{-1}.

In the second stage, we treat the electron-electron interaction as a two-dimensional many-body problem. It is worth pointing out here that the inversion-layer geometry shares a number of common features with layer materials where charge-density-wave ground states are known to exist.[12] If we label the six valleys by creation operators $a^\dagger$, $b^\dagger$, $c^\dagger$, $d^\dagger$, $e^\dagger$, and $f^\dagger$ and by their respective destruction operators a, b, etc., the model Hamiltonian is

$$\mathcal{H} = \sum_{k\sigma} [\epsilon_a(k) a_{k\sigma}^\dagger a_{k\sigma} + \text{corresponding terms for } b,\ c,\ d,\ e, \text{ and } f]$$
$$+ \sum_{qkk'\sigma\sigma'} U_1 [a_{k\sigma}^\dagger a_{k-q,\sigma} b_{k'\sigma'}^\dagger b_{k'+q,\sigma'} + \text{other distinct combinations from } a,\ b,\ c,\ d,\ e,$$
$$\text{and } f \text{ taken two at a time}]$$

$$+\sum_{qkk'\sigma\sigma'}[U_{2ab}a_{k\sigma}{}^{\dagger}b_{k'+q,\sigma}b_{k'\sigma'}{}^{\dagger}a_{k-q,\sigma'}+\text{other distinct combinations from } a,\ b,\ c,\ d,\ e,\ \text{and } f \text{ taken two at a time}]$$

$$+\sum_{qkk'\sigma\sigma'}\tfrac{1}{2}V_1[a_{k\sigma}{}^{\dagger}a_{k-q,\sigma}a_{k'\sigma'}{}^{\dagger}a_{k'+q,\sigma'}+\text{corresponding terms for } b,\ c,\ d,\ e,\ \text{and } f], \tag{5}$$

where, in addition to the six one-electron terms [$\epsilon_a(k)$, etc.], we have included six intravalley (V_1) and fifteen intervalley (U_1) direct interaction terms as well as the fifteen intervalley (U_2) exchange terms. While the direct terms are dominated by the Coulomb repulsion, to a very good approximation V_1 and U_1 are equal (i.e., they are the electron-electron Coulomb integrals for two electrons either in the same or in different valleys). The exchange integrals (U_{2ab}, etc.) take different values if the two valleys ab are along the same axis (e.g., [100] for $U_{2\alpha}$) or at right angles to each other (e.g., [100] and [010] for $U_{2\beta}$). This is illustrated in Fig. 1.

By defining the correlation and occupation quantities

$$A_\sigma\equiv\textstyle\sum_k\langle a_{k\sigma}{}^{\dagger}a_{k\sigma}\rangle\ ,\quad A_{\uparrow\downarrow}\equiv\sum_k\langle a_{k\uparrow}{}^{\dagger}a_{k\downarrow}\rangle\ ,\quad \Delta_{ab\sigma}\equiv\sum_k\langle a_{k\sigma}{}^{\dagger}b_{k\sigma}\rangle\ ,\quad \Delta_{ab\uparrow\downarrow}\equiv\sum_k\langle a_{k\uparrow}{}^{\dagger}b_{k\downarrow}\rangle\ ,\ \text{etc.} \tag{6}$$

we can write, following standard procedure,[8] the total energy of the system in the unrestricted Hartree-Fock approximation

$$\begin{aligned}\langle\mathcal{H}\rangle_{\mathrm{HF}}&=\textstyle\sum_k F_k E_k\\&\quad-U_1[(A_\uparrow+A_\downarrow)(B_\uparrow+B_\downarrow)-|\Delta_{ab\downarrow}|^2-|\Delta_{ab\uparrow}|^2-2|\Delta_{ab\uparrow\downarrow}|^2+\text{other distinct pairwise combinations}]\\&\quad+\{U_{2ab}[A_\uparrow B_\uparrow+A_\downarrow B_\downarrow+2A_{\uparrow\downarrow}B_{\downarrow\uparrow}-|\Delta_{ab\uparrow}|^2-|\Delta_{ab\downarrow}|^2-2\,\mathrm{Re}(\Delta_{ab\uparrow}\Delta_{ba\downarrow})]\\&\qquad+\text{other distinct pairwise combinations}\}\\&\quad-V_1[A_\uparrow A_\downarrow-|A_{\uparrow\downarrow}|^2+\text{corresponding terms for } b,\ c,\ d,\ e,\ \text{and } f],\end{aligned} \tag{7}$$

where F_k is the occupation of the state of energy E_k; E_k is obtained from the solution of a 12×12 secular equation derived from the equations of motion of the real quasiparticle creation and destruction operators.[8]

The many possible self-consistent states are characterized by the nonvanishing parameters defined by Eq. (6). A nonvanishing Δ describes a density wave of vector $\vec{q}$ equal to the surface projection of the relevant intervalley k-space separation. The possible solutions are numerous, while many more can be ruled out. For example, triple spin- or charge-density waves ($\Delta_{ac\uparrow}=\Delta_{bd\uparrow}=\Delta_{ce\uparrow}=\Delta_{df\uparrow}=\Delta_{ea\uparrow}=\Delta_{fb\uparrow}\neq0$; $\Delta_{xy\uparrow}=\pm\Delta_{xy\downarrow}$) are always unstable for any reasonable values of the parameters V_1, U_1, and U_2. In Fig. 2, we present the phase diagram for the system as a function of $U_{2\beta}$ and U_1, for $V_1=U_1$ and three values[13] of the ratio $(U_{2\alpha}/U_{2\beta})$. The various U and V interactions are in fact functions of λ, and hence functions of n. But if they are considered as independent parameters, the total energy (7) scales with n^2.

The results of the calculation, as given in Fig. 2, show the following:

(1) For small values of the U and V, the ordinary paramagnetic state is stable.

(2) For very large values of the parameters, the stable solution is the ferromagnetic occupation of one valley. This solution maximizes the kinetic energy term, but the electron-electron contribution to the total energy vanishes.

(3) At intermediate values of U_1, ordinary charge-density waves (CDW) and spin-density waves (SDW) are stable. The former exist for values of $U_2<0$, while the latter exist for $U_2>0$.

(4) These density waves (DW) are of two types, and driven respectively by the $U_{2\alpha}$ (opposite valleys) and $U_{2\beta}$ ("right angle" valleys) interactions. We label them DWα and DWβ and they can be either CDW or SDW. The DWα states couple two

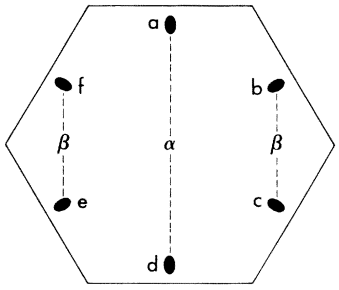

FIG. 1. The two-dimensional (111) silicon surface Brillouin zone. The electron valleys are indicated and labeled. The two different exchange interactions $U_{2\alpha}$ and $U_{2\beta}$ are indicated; they give rise to the corresponding density-wave couplings DWα and DWβ.

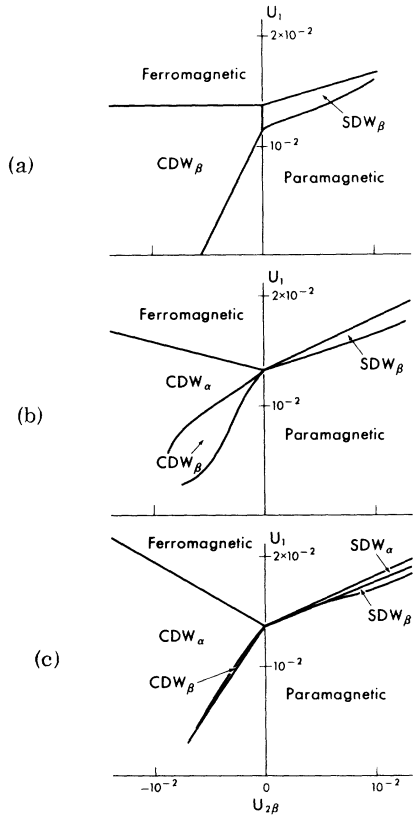

FIG. 2. Phase diagrams for the (111) Si/SiO_2 inversion layer ground state. In all cases $U_1 = V_1$. Axes are in eV for a normalizing area of 10^{-12} cm^2. The three diagrams correspond to different ratios $r \equiv U_{2\alpha}/U_{2\beta}$. (a) $r = 0$; (b) $r = 0.5$; (c) $r = 1$.

opposite valleys (see Fig. 1 for the corresponding q vector) and push one of them below the Fermi level; the other four valleys remain unoccupied and uncoupled. The DWβ states couple four valleys in two pairs, pushing two of them below the Fermi level; the other two valleys remain uncoupled. Energy splittings between occupied and empty (nonbonding and antibonding) bands are typically 10–15 meV. Details related to this and other optical properties will be given elsewhere. All states above have been tested for stability against the onset of an additional density wave.

(5) The DWα states have an occupation multiplicity of 1 and are defined by

$$\Delta_{ad\uparrow} = \pm\,\Delta_{ad\downarrow} = \tfrac{1}{4}nA, \quad A_\uparrow = A_\downarrow = D_\uparrow = D_\downarrow = \tfrac{1}{4}nA,$$

with all other occupations and correlations vanishing.

(6) The DWβ states have an occupation multiplicity of 2 and are given by

$$\Delta_{bc\uparrow} = \Delta_{ef\uparrow} = \pm\,\Delta_{bc\downarrow} = \pm\,\Delta_{ef\downarrow}$$

$$\cong \frac{nA}{8} - \text{correction } O\left(\frac{\epsilon(k)}{U_1}\right)^2,$$

$$B_\sigma = C_\sigma = D_\sigma = F_\sigma = \tfrac{1}{8}nA,$$

with all other occupations and correlations equal to zero.

(7) We have calculated, in all cases, the cyclotron masses by determining areas of the Fermi distribution and their derivatives with respect to the Fermi energy. We find for reasonable values of U and V [discussed below in paragraphs (8) and (9)], CDWα gives $m^* \equiv 0.358m_e$ while CDWβ yields $m^* = 0.405m_e$ at the paramagnetic boundary decreasing slowly to $m^* = 0.375m_e$ in Fig. 2(a) at the ferromagnetic boundary. All these masses include those many-body effects arising from the unrestricted Hartree-Fock approximation. Dynamical corrections, which we have not included, may modify these values.[14,15]

(8) The value of $U_1 = V_1$ is heavily dominated by the repulsive Coulomb interaction. With a screening length s^{-1}, the envelope function (2) yields

$$U_1(q=0) = \frac{e^2\pi}{\epsilon_s A}\left[\frac{128\lambda^6}{s(4\lambda^2 - s^2)^3} + \frac{\lambda(1.5s^4 - 20s^2\lambda^2 + 120\lambda^4)}{(s^2 - 4\lambda^2)^3}\right].$$

With s^{-1} of the order of typical interelectron spacings,[16] and a normalizing area $A = 10^{-12}$ cm^2, U_1 takes values $U_1 \lesssim 10^{-2}$ eV.

(9) The electron-electron exchange contribution to U_2 is positive and small, but the electron-phonon contribution is negative and large,[17,18] and can be even larger in absolute value than U_1. The ratio $U_{2\alpha}/U_{2\beta}$ is $\lesssim 0.5$ as crudely estimated from the ratio of the relevant phonon frequencies, and with $|U_{2\beta}| \sim U_1$ we are clearly in the CDWβ regime.

(10) Higher order correlation effects may change (or renormalize) the values of the interaction parameters.[8] We cannot at present estimate these changes, but assume them not to be drastic.

(11) We have, of course, not examined all possible self-consistent states. It is conceivable that ferrimagnetic states, spiral arrangements, and complicated mixed DW's may exist and be stable. However, the states we have found seem to explain satisfactorily the observed experimen-

tal data.

(12) The presence of either a DWα or a DWβ lowers the C_{3v} symmetry of the Si (111) surface to a mere mirror symmetry. This leads to the formation of domains with three different orientations of the density wave. Experiments performed under controlled uniaxial conditions (e.g., strains or fields) should be able to test the validity of our theory.

(13) The theory outlined above proceeds in a similar way for other Si surfaces. For the (100) surface, the asymmetry of the effective masses and the cubic character of Si make the paramagnetic state the stable one over a very wide range of interaction parameters; and existing theories[4-6] should apply to them. The many-body aspects in this case should be the ordinary ones of the two-dimensional normal electron gas.[14,15]

---

*IBM Postdoctoral Fellow. Permanent address: Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, England.

†Work supported in part by the National Science Foundation through Grant No. DMR72-03106-A02.

[1]T. Neugebauer, K. Von Klitzing, G. Landwehr, and G. Dorda, Solid State Commun. 17, 295 (1975).

[2]A. A. Lakhani and P. J. Stiles, Phys. Lett. 51A, 117 (1975).

[3]D. C. Tsui, private communication, and Bull. Am. Phys. Soc. 21, 334 (1976).

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

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

[6]For recent reviews, see F. Stern, Crit. Rev. Solid State Sci. 4, 499 (1974); G. Landwehr, *Festkörperprobleme XV* (Pergamon, New York, 1975), p. 49.

[7]The structure of these theories is derived from J. M. Luttinger and W. Kohn, Phys. Rev. 97, 869 (1955).

[8]D. R. Penn, Phys. Rev. 142, 350 (1966).

[9]We follow the notation of Ref. 5 throughout.

[10]Strictly, $F(\boldsymbol{r})$ has a z-dependent phase factor, but this can be incorporated into the effective mass parameters as in Eqs. (8)–(13) of Ref. 4. In all these equations, A is the area of the sample.

[11]An image potential term should also be included in Eq. (3), as pointed out by F. Stern, Phys. Rev. Lett 30, 278 (1973). This is a small correction of no relevance for our purposes.

[12]J. A. Wilson, F. J. Di Salvo, and S. Mahajan, Adv. Phys. 24, 117 (1975).

[13]The apparent discrepancy between the graphs in Fig. 2 for $U_{2\alpha}=U_{2\beta}=0$ arises from the lack of analyticity of the total energy with respect to the ratio $U_{2\alpha}/U_{2\beta}$. The only stable phases in this case are either the ferromagnetic or the paramagnetic one. For some values of U_1, however, the CDW and SDW are infinitesimally close.

[14]C. S. Ting, T. K. Lee, and J. J. Quinn, Phys. Rev. Lett. 34, 870 (1975).

[15]B. Vinter, Phys. Rev. Lett. 35, 1044 (1975), and references therein.

[16]Detailed discussion of the values of the parameters will be published elsewhere.

[17]M. L. Cohen, Phys. Rev. 134, A511 (1964).

[18]The electron-electron interaction via phonons is known to be the dominant intervalley exchange effect; it leads to superconductivity in some many-valley degenerate semiconductors, but is large and attractive even in materials which are not superconducting.