Observation of Resonant Tunneling in Silicon Inversion Layers

A. B. Fowler, G. L. Timp,[a] J. J. Wainer, and R. A. Webb
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 29 April 1986)

Measurements of the temperature and carrier-density dependence of the strongly localized conductance of short silicon inversion layers are reported. At the lowest temperatures we observe well-isolated, large conductance peaks whose width and temperature dependence are only consistent with resonant tunneling and are inconsistent with Mott hopping. Several new features are observed which we believe may be the result of Coulomb interactions.

PACS numbers: 71.55.Jv, 73.40.Qv

Silicon-inversion-layer metal-oxide-semiconductor field-effect transistors (MOSFET's) have proven to be one of the most versatile experimental systems in which the properties of a two-dimensional electron gas can be probed in detail.[1] In these devices the electron concentration can be varied by adjustment of the voltage on the metal gate electrode. In one sample a wide variety of conduction processes can be studied from the strong localization regime up to the nearly metallic regime.

Recently[2-6] experiments on very small silicon structures have revealed new behavior in the transport properties. Near threshold in these devices the number of energy states available for conduction near the Fermi energy can approach unity at low temperatures. Under these conditions the conductance has been shown to exhibit strong fluctuations (or large isolated peaks) as a function of carrier concentration. Since the number of energy states available for conduction is extremely sparse, it was thought by Azbel, Hartstein, and Divincenzo[7] that these fluctuations were associated with resonant tunneling through localized sites, a subject first discussed by Lifschitz and Kirpichenkov.[8] However, an experimental study[3] of the temperature dependence of many of the conductance peaks showed the peaks to be exponentially activated at low T; at higher T, where the peaks overlapped in gate voltage, the conductance was found to be exponential in $-(T_0/T)^{1/2}$, where T_0 is a characteristic temperature of 10–30 K. This behavior was inconsistent with the predictions of the models of resonant tunneling, but was well described by percolation models for hopping conductance in a quasi 1D system, including the critical-path models studied by Lee[9] and by McInnes and Butcher.[10]

All models for transport agree that at $T=0$ the conduction must be dominated by resonant tunneling while at finite temperatures both resonant tunneling and hopping can occur in parallel. The characteristic temperature at which resonant tunneling dominates over hopping for a 1D sample is of the order of[7,11] $T_0(L_0/2L)^2$, where L_0 is the localization length, and L is the sample length. For the samples used to date,[2,3] $T_0 \sim 20$ K, $L_0 \sim 0.1$ μm, and $L = 10$ μm, corresponding to temperatures less than 10 mK, below the range of previous work. In this Letter we present new results on the conduction processes in short MOSFET devices ($L \sim 10L_0$) that clearly demonstrate the existence of resonant tunneling. Several new features are observed in the shape of the conductance peaks which may result from Coulomb interactions.

Azbel, Hartstein, and Divincenzo[7] and Stone and Lee[11] have presented alternative models of resonant tunneling in marginally conducting samples. The former predict that, at very low temperatures, the conductance peaks increase in amplitude as $G \propto 1/T$ (until a saturation is reached), while at intermediate temperatures the peak conductance increases with increasing temperature with a form similar to Mott hopping in 1D, $\ln(G) \propto -(T_0/T)^{1/2}$, but with a different T_0 than in Mott's picture. The widths of the peaks increase with temperature in both regimes. Alternatively, Stone and Lee[11] predict a constant amplitude and peak width with increasing temperature until $k_B T$ exceeds Γ, the natural linewidth of the resonance. Above this temperature the amplitude decreases and the width broadens with increasing T in such a way as to keep the integrated intensity constant. Both theories predict that at high temperatures variable-range hopping will dominate the conductance. Both ignore electron-electron interactions.

We expect that in the case of resonant tunneling the conductance should be of the form[12,13]

$$G(\mu) \propto \frac{2e^2}{\pi h} \frac{\int dE\, t(E)\, df(\mu,E)/dE}{\int dE\, [1-t(E)]\, df(\mu,E)/dE}$$

for a 1D channel, where f is the Fermi function, $t(E)$ is the transmission coefficient as a function of energy, and μ is the chemical potential. Clearly, this would have to be modified for a multiple channel or a 2D sample. Here we assume that the transmission has some natural width which may arise from various sources.[12] For $t(E) \ll 1$ the lower integral approaches unity and for low temperatures, $G \sim t(E_0-\mu)$, where E_0 is the center of the resonance

energy. The magnitude and shape of the conductance peak will be temperature independent. For $k_BT > \Gamma$ there will be a transition to a lower peak conductance value. As shown by Stone and Lee,[11] the peak amplitude should decrease as $1/T$ and the line should broaden with $d\ln G(\mu)/d\mu = \pm(k_BT)^{-1}$. The change in chemical potential can be related to the change in gate voltage through the gate capacitance and the density of states. In a MOSFET, $dn_s/dV_G \sim \epsilon/e\delta$, where ϵ is the oxide permittivity and δ is the oxide thickness. The density of states in these devices is generally reduced from the full 2D density of states because the localized states are in a band tail.[1] For a (100) Si surface $D(E) = dn_s/dE = a \times 1.6 \times 10^{14}$ eV^{-1} cm^{-2}, where $a < 1$. Consequently $d\mu/dV_G = (dE/dn_s)(dn_s/dV_G) = a^{-1}1.4 \times 10^{-2}$ eV/V for our samples with a 10-nm-thick SiO_2 oxide. It should be noted that this procedure assumes a constant density of states which may not be valid. In a band tail the number of states decays exponentially to zero with μ and for small μ the states should be discrete.

The channel geometry in the devices used in these experiments are about 20 times wider ($W = 1$–1.75 μm) and 20 times shorter ($L = 0.5$ μm) than those previously examined.[2,3] Therefore, we expect about the same number of states within k_BT of the Fermi energy and the density of peaks should be sparse enough to resolve individual peaks with little or no overlap at low temperatures as in Ref. 3. In contrast to previous work, we expect that resonant tunneling should dominate the conductance because of the short channel length. Unlike conventional resonant-tunneling experiments[14] on oxide barriers where the electrode voltages are varied to find structure,[15] in our experiments the chemical potential is swept through the localized states by using the voltage on the gate electrode. Thus the transmission peaks occur as a function of μ. The voltage between the source and drain electrodes was always kept small (0.5–5 μV) relative to the peak width.

Figure 1 displays the drain-source conductance as a function of gate voltage for a 1.0×0.5-μm^2 device over a range of 0.554 to 0.566 V corresponding to a range of μ of $168a^{-1}$ μeV. There are several well-defined and isolated peaks, but each peak has structure associated with it. Several of these peaks were examined as a function of temperature. In all of our low-temperature data, the largest value of the conductance of any peak was $0.1e^2/h$. In Fig. 2 we show amplitudes of two different peaks from a 1.0-μm-wide, 0.5-μm-long sample and a 1.75×0.5-μm^2 sample as a function of $1/T$. At low temperatures both samples have a constant peak amplitude, but as the temperature increases the amplitude of the wider device decreases with increasing T (until the Mott-hopping background dominates) while the amplitude of the narrower device only increases with T. The later behavior is consistent with the observation that the Mott-hopping background is becoming important at 150 mK. In Fig. 3 we display the temperature dependence of the slope $d\ln(G)/dV_G$ of both peaks. At low T, the wider device has a larger slope (narrower width) than the narrower device, and the wider sample exhibits broadening at temperatures above 30 mK while the narrower one remains temperature independent below 100 mK. At high temperatures the slopes of both peaks are consistent with $1/T$ behavior. The inset to Fig. 1 displays the gate-voltage dependence of one of our peaks at low temperatures on a logarithmic scale. The structure as

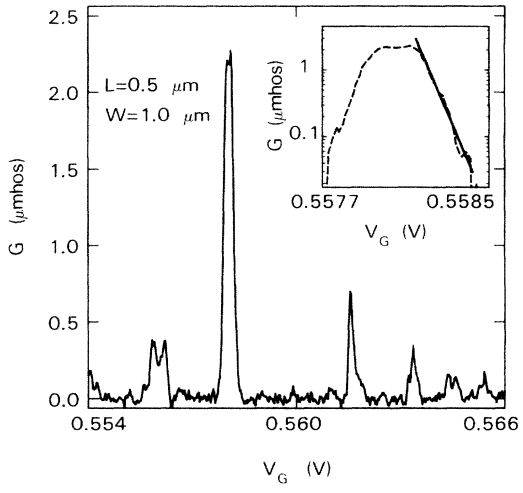

FIG. 1. Conductance as a function of gate voltage for one of our devices at 22 mK. Inset: Conductance of the largest peak on a logarithmic scale with an expanded gate-voltage scale. The straight line is representative of the quality of the fits used to determine the slope.

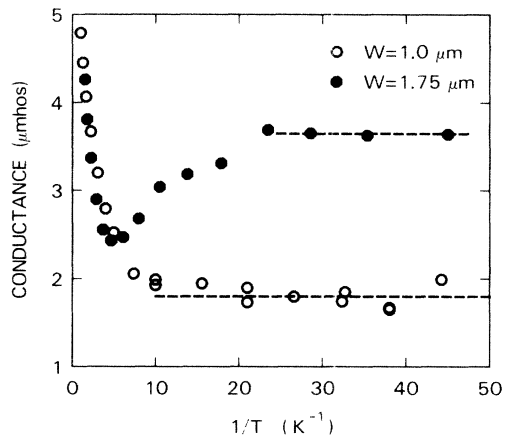

FIG. 2. Temperature dependence of the maximum value of the peak conductance for two 0.5-μm-long devices. The dashed lines are average values of low-temperature conductance.

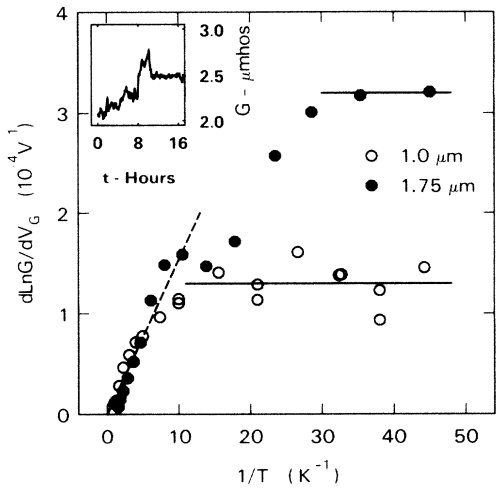

FIG. 3. Temperature dependence of the logarithmic slopes of the data displayed in Fig. 2. Dashed line was used in the determination of the correction to the density of states. Solid lines are the values of the low-temperature slope used in the calculation of the intrinsic width. Inset: Time dependence of a peak conductance at 38 mK for the narrower device.

a function of gate voltage mentioned above is evident in this plot. The slopes are not the same on both sides of the curve and the data plotted in Fig. 3 were obtained by use of the side of the peak which showed less structure.

The high-temperature variations of the slopes, $d\ln(G)/dV_G$, are consistent with a T^{-1} temperature dependence over the range studied. Comparing the rate of change with $1/T$, we find $d\mu/dV_G = 0.13(\pm 0.06)$ eV/V for both samples. This implies that $a \sim 0.11$ or that the density of states at this point in the band tail is reduced by about a factor of 9 from that of the 2D conduction band.

Azbel, Hartstein, and Divincenzo[7] emphasized that for a resonant center to have a large transmission coefficient it must be within a localization length L_0 of the center of the sample. From fitting the temperature dependence of the conductance away from a resonance (nearest conductance minimum) at high temperatures by 2D variable-range hopping, we determined that the localization lengths were approximately 42 nm ($T_0 = 364$ K) and 28 nm ($T_0 = 792$ K) for the narrow and wide samples, respectively. These are similar to estimates based on previous work.[16] Therefore, the states that should contribute to resonant tunneling should be within a 40-nm band near the center of the sample. If the resonances are temperature broadened, the widths at half maximum should be of the order of $2k_BT$. The number of states in such a band is $4k_BTL_0WD(E) = 0.8a \sim 0.1$ state at 50 mK. It is unlikely that every peak would overlap another; it is improbable that the structure observed on every isolated peak is from overlap of different peaks. The limiting low-temperature slopes displayed in Fig. 3 can also be used to estimate the intrinsic widths of the peaks. Converting the gate-voltage scale to an energy scale we find that the intrinsic widths of the peaks are 38 mK (3 μeV) and 100 mK (8.6 μeV) for the wide and narrow samples, respectively. The onset of temperature dependence of the linewidth data is in excellent agreement with the data displayed in Fig. 2 for the onset of temperature dependence of the peak heights.

Although both the localization lengths and the sample lengths are similar, the fact that the low-temperature conductance is different for the two samples can be understood in terms of the spatial locations of the resonant tunneling sites. In the case of resonant tunneling, the maximum conductance G_0 is obtained when the resonant site is exactly at the center of the sample. If the resonant state is not at the center we expect[7,11] the conductance to be $G = G_0\exp(-2\beta/L_0)$, where β is the distance from the center of the sample. With the assumption that $G_0 \sim e^2/h$, this implies that the resonant state for the wide sample is 43 nm from the center while the narrow sample is 48 nm from the center. This estimate assumes a 1D channel. In our measurements of the electric field dependence of the resonant conduction, we observe both nonlinear and rectifying behavior in the I-V curves over the voltage range $-40\ \mu V < V_{SD} < 40\ \mu V$ which is the same region over which similar effects have previously[3,5] been observed in much longer devices.

Thus we find qualitatively the behavior expected of resonant tunneling. The observed behavior is quantitatively different from that seen in longer devices and attributed to Mott hopping. However, there are still some serious questions.

The first is that none of the peaks was a pure Lorentzian; all exhibited a fine structure as a function of V_G. As argued above, we do not expect more than about a 10% chance of overlap of peaks. Second, in order to get reasonably reproducible data it was necessary to make very slow sweeps of many hours over a very narrow range of chemical potential at each temperature. Even then the data did not reproduce perfectly, although with increasing T the reproducibility improved. The lack of reproducibility is evident in a plot of the time dependence of the conductance of a peak as shown in the inset to Fig. 3. Here one may see both small high-frequency and large low-frequency fluctuations that can be as large as 30% of the total conductance. The magnitude of the fluctuations with time decreases with increasing temperature.

The explanation of these phenomena may be related to the experiments of Rogers and Buhrman[17] and Skocpol *et al.*[4] In the entire sample there are other states which may or may not be occupied. A Fermi

function describes the average occupation but not the fluctuations from equilibrium. At 50 mK there are about four states within $\pm 5k_BT$, which includes all states which range from 6×10^{-3} probability of being occupied to 6×10^{-3} probability of being unoccupied. Distributions of the electrons in this ensemble will take place continuously at a maximum rate of $10^{13}\exp[-(T_0/T)^{1/3}]$ or about 100 Hz at 20 mK with a large range of lower frequencies.

If the occupation of the states changes, the Coulomb interactions with the resonant center will change. Each configuration will give a different Coulomb interaction. If the bare Coulomb interaction is considered and if an electron is placed on a site 200 nm from the center of the resonance the shift in energy could be as much as 3 meV, whereas the structure is of order 5 μeV. The samples are made with 10-nm oxides with a metal gate above the oxide. If one calculates the screened interaction assuming the electrons are 4 nm inside the silicon, the energy shift of a resonant state 200 nm away is then only 5 μeV. At 50 mK the chance of a level at $\pm 5kT$ within a distance of 200 nm is about 1%. The dipolar interaction falls off as r^{-3} so that most of the states will give a much smaller shift. These screened Coulomb effects are consistent with the observed spacing of the structure. Another possible explanation for the structure in the peaks is associated with the (100) valley splittings in silicon. The best theoretical estimates[1] for the energy splitting are on the order of 50 μeV which is much larger than both the observed spacing of the structure and k_BT.

In conclusion, studies of the temperature dependence of the width and amplitude of the conductance peaks in small MOSFET's have clearly shown the behavior expected of resonant tunneling and are inconsistent with the Mott-hopping behavior observed in longer devices but with the same number of available energy states for conduction. The observation that every well-isolated resonant peak has additional fine structure was unexpected and the explanation is likely to be associated with Coulomb interactions.

The samples described here were supplied by M. Wordemann. The impetus for exploring the effects of configurational changes was supplied by a very perceptive comment by W. A. Harrison. We have greatly benefitted from discussions with M. Büttiker, R. Landauer, P. A. Lee, D. DiVincenzo, and S. Washburn.

[a]Now at AT&T Bell Laboratories, Holmdel, NJ 07733.

[1]T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 54 (1982).

[2]A. B. Fowler, A. Hartstein, and R. A. Webb, Phys. Rev. Lett. **48**, 196 (1982).

[3]R. A. Webb, A. Harstein, J. J. Wainer, and A. B. Fowler, Phys. Rev. Lett. **54**, 1577 (1985).

[4]W. J. Skocpol, L. D. Jackel, R. E. Howard, and C. A. Fetter, Phys. Rev. Lett. **49**, 951 (1982).

[5]R. F. Kwasniak, M. A. Kastner, J. Melngalis, and P. A. Lee, Phys. Rev. Lett. **52**, 224 (1984).

[6]C. C. Dean and M. Pepper, J. Phys. C **5**, L617 (1982).

[7]M. Ya. Azbel, Solid State Commun. **45**, 527 (1983), M. Ya. Azbel, A. Harstein, and D. P. Divincenzo, Phys. Rev. Lett. **52**, 1641 (1984), and Physica (Amsterdam) **127B**, 252 (1984).

[8]I. M. Lifschitz and V. Ya. Kirpichenkov, Zh. Eksp. Teor. Fiz. **77**, 989 (1979) [Sov. Phys. JETP **50**, 499 (1979)].

[9]P. A. Lee, Phys. Rev. Lett. **53**, 2042 (1984).

[10]J. A. McInnes and P. N. Butcher, J. Phys. C **18**, L921 (1985).

[11]A. D. Stone and P. A. Lee, Phys. Rev. Lett. **54**, 1196 (1985).

[12]H. L. Engquist and P. W. Anderson, Phys. Rev. B **24**, 1151 (1981).

[13]R. Landauer, in *Localization, Interaction, and Transport Phenomena in Impure Metals,* edited by B. Kramer, Y. Brunseraede, and G. Bergmann, Springer Tracts in Solid State Sciences Vol. 61 (Springer, Heidelberg, 1985), p. 38.

[14]S. J. Bending and M. R. Beasley, Phys. Rev. Lett. **55**, 324 (1985).

[15]J. Halbritter, J. Appl. Phys. **58**, 1320 (1985).

[16]S. Pollitt, Commun. Phys. **1**, 207 (1976).

[17]C. T. Rogers and R. A. Buhrman, Phys. Rev. Lett. **55**, 859 (1985).