Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/*f*?) Noise

K. S. Ralls,[a] W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter,
R. W. Epworth, and D. M. Tennant
Bell Laboratories, Holmdel, New Jersey 07733
(Received 4 August 1983)

Resistance fluctations in submicrometer narrow Si inversion layers are studied over a wide range of temperatures and electron concentrations. Thermally activated switching on and off of discrete resistance increments is observed, caused by the capture and emission of individual electrons at strategically located scatterers (interface traps). The traps have a broad distribution of activation energies, as assumed in accounting for $1/f$ noise in larger devices.

PACS numbers: 73.40.Qv, 05.40.+j, 71.50.+t, 72.70.+m

The various origins of low-frequency ($1/f$) noise in electronic systems are of continuing interest.[1] Quite generally, such noise increases in systems of small physical size. In metal-oxide-semiconductor field-effect transistors (MOSFETs), for example, Mikoshiba[2] recently demonstrated an inverse proportionality between gate area and noise. He identified empirically correlated fluctuations of both mobility and carrier density as the noise source. In this Letter we describe measurements on much smaller (submicrometer) MOSFETs which reveal discrete switching events that account for at least part of the low-frequency noise observed. The switching allows us to study in detail the fluctuating occupancy of a *single* electron trap, interesting both in its own right, and for its connection to explanations of $1/f$ noise in MOSFETs.

Our MOSFET channels are 1 μm long and 0.1 μm wide.[3] The narrow NiCr gate metallization (defined by electron-beam lithography) protects the narrow channel from reactive-ion etching that removes adjacent material, leaving a pedestal consisting of the gate, the 65-nm gate oxide, and an additional 100 nm of the underlying $\langle 100 \rangle$-oriented p-type silicon. The resistance of the electron inversion layer induced at the oxide-silicon interface shows discrete switching events of up to 1% magnitude. No such switching was resolved, although low-frequency noise was observed, for a large device (10 μm$\times$20 μm) fabricated at the same time as other smaller devices that display the discrete behavior.

The measured switching was a genuine resistance change. Typically we measured the ac drain-source voltage V_{ds} induced by an ac current bias, using a lock-in amplifier. The inferred resistance was independent of bias level, so long as V_{ds} was kept small enough ($\sim$1 mV) to prevent device heating at low temperatures.

The discrete switching events, seen in several samples, are illustrated here by data from a single device 0.15 μm wide by 1 μm long. After etching, the device was annealed at 450°C in H_2 for 30 min, resulting in a helium-temperature mobility of 5000 cm^2/V-sec (confirmed by Shubnikov–de Haas magnetoconductance measurements). With repeated use over several weeks the mobility degraded to 2000 cm^2/V-sec, and the low-temperature threshold increased from 1.5 V to greater than 3 V, presumably because of the irreversible capture of electrons in neutral traps.

Figure 1 shows resistance as a function of time for one of the prominent switching sequences observed. It changes randomly between two values, spending an average time $\langle \tau_{on} \rangle$ in the high resistance state, and $\langle \tau_{off} \rangle$ in the low. In our experiment, the window of convenient observation times is limited by lock-in response time at one end, and experimenters' patience at the other. The switching times are a strong function of temperature and gate voltage, so that a given switching feature is observable only over a limited range.

A single such "random telegraph signal" makes a Lorentzian contribution to the noise power spectrum,[4] and a superposition of such Lorentzians with an appropriate distribution of time constants yields a power spectrum proportional to $1/f$. McWhorter[5] noted that such a distribution of trapping times at a semiconductor-oxide interface could arise naturally from a spatially uniform distribution of tunneling depths to traps in the oxide. More recently, Dutta, Dimon, and Horn[6] have demonstrated that thermally activated processes with a broad distribution of activation energies (of order 1 eV) provide a consistent and detailed picture of the temperature dependence of $1/f$ noise in continuous metal films above and below room temperature. Our measurements prove

© 1984 The American Physical Society

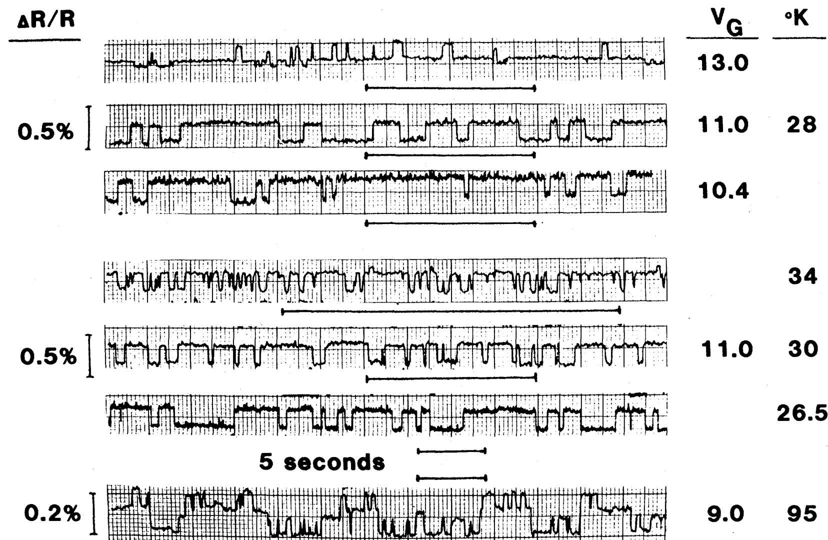

FIG. 1. Resistance switching observed in a small MOSFET in a particular range of temperatures and gate voltages. The duty cycle depends on gate voltage, while the overall rate decreases with temperature (note scale changes). The last trace demonstrates that superposition can lead toward $1/f$ noise.

that the switching we observe is associated with interface traps characterized by a wide range of activation energies, demonstrating a direct connection to such explanations of $1/f$ noise.

Our devices are so small that full superposition is not possible, and individual switching processes can be resolved. When switching is prominent, a fast Fourier transform of computer-logged data yields power spectra with a Lorentzian contribution up to an order of magnitude larger than the general $1/f$ background resulting from superposed contributions of other smaller, less active sources. The last trace in Fig. 1, with two event sequences complexly superimposed, qualitatively displays how a $1/f$ signal can begin to be built up. In larger devices, more sequences of smaller individual effect tend to overlap, eventually becoming impossible to resolve. Although we can characterize prominent switching events in considerable detail, we cannot, of course, *prove* that events of this character account for *all* of the low-frequency noise.

What causes the jumps? Discrete switching dependent on gate voltage implicates individual electrons. Different sets of resistance jumps vary greatly in size, and can be either greater or smaller than the fractional change due to removal of a single channel electron. (For this device, there are ~500 electrons per gate volt above threshold.) Therefore we conclude that the channel mobility is also affected, i.e., individual scatterers are turned on and off, some more strategically located than others. At plausible interface trap densities N_{it} of $10^{11}-10^{13}$ cm^{-2}, our device contains $\sim 10^2-10^4$ such scatterers, consistent with the overall magnitude of effect.

Figure 2 demonstrates that for the feature depicted in Fig. 1 $\langle\tau_{on}\rangle$ and $\langle\tau_{off}\rangle$ depend exponentially on temperature and gate voltage. From the slopes of such plots, we determine apparent activation energies

$$\Delta E_{on,off} = \frac{d\ln\langle\tau_{on,off}\rangle}{d(1/k_BT)}, \tag{1}$$

and define gate-voltage dependences

$$\Delta F_{on,off} = d\ln\langle\tau_{on,off}\rangle/dV_G. \tag{2}$$

Table I describes several such switching fea-

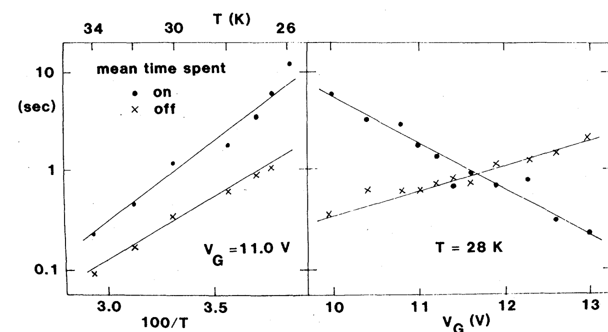

FIG. 2. Exponential dependence of mean lifetimes on inverse temperature and gate voltage for a particular switching sequence. These are identified with capture and emission rates at a single interface trap.

TABLE I. Measured parameters (defined in text) of individual traps. Quantities in parentheses are based on temperatures from estimated electron heating.

Observation range				Gate voltage dependence				Activation energy	
T (K)	V_G (V)	Polarities	R/R (%)	$\Delta F_{\rm off}$ (1/V)	$\Delta F_{\rm on}$ (1/V)	dE_T/dV_G (meV/V)	d_0 (nm)	$\Delta E_{\rm off}$ (meV)	$\Delta E_{\rm on}$ (meV)
101–111	3.8–5.3	−	0.3	−1.4	2.2	28	1.9	210	280
48–56	2.2–5.5	+	0.1	0.47	−1.0	6	0.3	70	70
26–34	10–13	+	0.2	0.57	−1.0	3	0.1	26	39
(12–20)	3.5–4.5	−	0.7	−2.9	4.3	10	0.5	(16)	(9)
4.2	10.3–10.9	+	0.2	4.7	−7.5	5	0.2	···	···

tures sufficiently prominent and well isolated to be studied in detail in our device. As noted above, the percentage size does not correlate with gate voltage (electron density).

Data of this sort represent an unusual[7] degree of specificity about electron traps. We can identify $\langle\tau_{\rm off}\rangle$ and $\langle\tau_{\rm on}\rangle$ directly with the emission and capture times τ_e and τ_c, respectively, whenever (+ in Table I) a positive scattering center seems neutralized by electron capture at high gate voltages. The reverse identification holds whenever (− in Table I) a negative scattering center turns on by electron capture. Both polarities are observed. Ordinarily, emission and capture rates of ensembles are measured separately in different types of experiments or interrelated by assumptions about detailed balance.[8]

These data reveal the spatial location of the trap, for example, provided that we assume that it has a single local energy level E_T in equilibrium with the channel electrons, which have Fermi level $E_{\rm F}$. In a MOSFET at small $V_{\rm ds}$, $E_{\rm F}$ throughout the channel is set by the potential of the drain and source contacts. Varying of the gate potential results in spatial variation of the energy of both bands and localized traps, and hence their equilibrium occupancy. In our device, above threshold, the conduction-band edge at the interface lies near $E_{\rm F}$ and shifts by ~ 2.2 meV per volt on the gate. The depleted bands further in the semiconductor are screened by the inversion layer and shift by much less. The bulk of the potential drop occurs in the 65-nm oxide, resulting in energy shifts of ~ 15 meV per gate volt for each nanometer from the interface. Thus sensitivity to V_G provides direct information about the spatial location of traps. By detailed balance (equal numbers of up and down jumps) we obtain

$$\frac{\tau_c}{\tau_e} = \frac{(1-f_T)}{f_T} = g\exp\left(\frac{E_T - E_{\rm F}}{k_{\rm B}T}\right), \tag{3}$$

where f_T is the trap occupancy function and g is the trap degeneracy.[9] From (3), we have

$$\frac{dE_T}{dV_G} = k_{\rm B}T\frac{d}{dV_G}\ln\left(\frac{\tau_c}{\tau_e}\right) = -k_{\rm B}T\,|\Delta F_{\rm on} - \Delta F_{\rm off}|. \tag{4}$$

If the gate-voltage sensitivity is greater than 2.2 meV per volt, then the trap must be in the oxide at a distance

$$d_0 = (-dE_T/dV_G - 2.2)/15 \text{ nm}. \tag{5}$$

Data in Table I show that the traps lie 0.2–2 nm into the oxide, consistent with available information about interface traps.[10] The same argument shows that under flat band conditions, the trap energy E_T lies *above* the conduction-band edge, an important regime not previously investigated.

Current knowledge about interface traps[10] is limited to states with E_T in the forbidden gap, which show a U-shaped distribution, rising in density near the conduction-band edge. An activation energy $E_c - E_T$ is required to reach the conduction-band edge E_c, and the capture cross section is defined as

$$\begin{aligned}\sigma_n &= (\tau_c v_{\rm th} n)^{-1}\\ &= (\tau_e v_{\rm th} N_c)^{-1}\exp[(E_c - E_T)/k_{\rm B}T],\end{aligned} \tag{6}$$

where $v_{\rm th}$ is the thermal velocity, n is the electron density, and N_c is the effective density of states at the conduction-band edge. The latter relation depends on detailed balance. Near the band edge σ_n decreases dramatically,[11] and may itself appear thermally activated (although there is dispute even about the sign of the activation energy).[12] Our data show directly that both τ_e and τ_c have an activated temperature dependence, even though $E_T > E_c$. [The activation energies measured for emission and capture differ slightly because they are measured at fixed gate voltage, not fixed ratio ("duty cycle").] The observed "activation" energies cover a wide range, with

the "shallowest" traps being observably active at the lowest temperatures. The "deepest" trap seems furthest from the interface, but more extensive data would be required to establish a valid correlation. It is tempting to model the process simply as activation over an electron potential barrier between the trap and the silicon conduction band, so that the gate-voltage dependence of the emission rate could be accounted for by field-induced barrier changes. This seems incorrect, however, because electron energy barriers of millivolt height together with nanometer thickness should be dominated by *nonactivated* tunneling. This problem arises with particular clarity because we are investigating the previously unexplored interface states which lie *above* the conduction-band edge, allowing direct escape by tunneling. One solution to this dilemma is to imagine a potential barrier in some system configuration space, in which a lattice distortion, for example, is required for capture.[13] Another solution might be a direct calculation of inelastic tunneling processes.[14] Clearly additional theoretical work is called for, as well as more extensive experimental observations. In any case, the opportunity to observe the behavior of a single trap represents a fresh approach to a problem of longstanding scientific and technological interest.

The authors gratefully acknowledge helpful conversations with P. M. Downey, J. P. Gordon, D. M. Fleetwood, J. R. Brews, and D. V. Lang. One of us (K.S.R.) is the recipient of a Bell Labs Graduate Research Program for Women fellowship.

[(a)]Present address: Department of Physics, Cornell University, Ithaca, N.Y. 14853.

[1]For a diverse overview see A. van der Ziel, Adv. Electron. Electron Phys. 49, 225 (1979); D. A. Bell, J. Phys. C 13, 4425 (1980); F. N. Hooge, T. G. M. Kleinpenning, and L. K. J. Vandamme, Rep. Prog. Phys. 44, 479 (1981); P. Dutta and P. M. Horn, Rev. Mod. Phys. 53, 497 (1981).

[2]H. Mikoshiba, IEEE Trans. Electron Devices 29, 965 (1982).

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

[4]S. Machlup, J. Appl. Phys. 25, 341 (1954).

[5]A. L. McWhorter, in *Semiconductor Surface Physics*, edited by R. H. Kingston (Univ. of Pennsylvania Press, Philadelphia, 1957), p. 207.

[6]P. Dutta, P. Dimon, and P. M. Horn, Phys. Rev. Lett. 43, 646 (1979).

[7]Switching events have also been seen in *large* silicon junction field-effect transistors (JFET's) and attributed to switching of bulk traps in depletion regions: K. Kandiah, M. O. Deighton, and F. B. Whiting, in *Sixth International Conference on Noise in Physical Systems*, edited by P. H. E. Meijer, R. D. Mountain, and R. J. Soulen, National Bureau of Standards Special Publication No. 614 (U.S. GPO, Washington, D.C., 1981), p. 75.

[8]For example, D. V. Lang, H. G. Grimmeiss, E. Meijer, and M. Jaros, Phys. Rev. B 22, 3917 (1980).

[9]E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology* (Wiley, New York, 1982), pp. 37 and 363.

[10]Ref. 9, Chaps. 7, 8, and 16.

[11]H. Dueling, E. Klausmann, and A. Goetzberger, Solid State Electron. 15, 559 (1971).

[12]E. Kamieniecki, N. Gomma, A. Kloc, and R. Nitecki, J. Vac. Sci. Technol. 18, 883 (1981); T. Katsube, K. Kakimoto, and T. Ikoma, J. Appl. Phys. 52, 3504 (1981).

[13]C. H. Henry and D. V. Lang, Phys. Rev. B 15, 989 (1977); E. Kamieniecki, Appl. Phys. Lett. 35, 807 (1979); K. L. Ngai and C. T. White, J. Appl. Phys. 52, 320 (1981).

[14]F. P. Heiman and G. Warfield, IEEE Trans. Electron. Devices 12, 167 (1965); M. Schulz, in *Insulating Films on Semiconductors—1979*, edited by G. G. Roberts and M. J. Morant, IOP Conference Proceedings No. 50 (Institute of Physics, London, 1980), p. 87.