Dissipative Tunneling in Two-State Systems at the Si/SiO₂ Interface

D. H. Cobden,¹ M. J. Uren,² and M. Pepper¹

¹Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, United Kingdom
²Defence Research Agency, Malvern, Worcestershire WR14 3PS, United Kingdom

(Received 7 June 1993)

We have observed two-state systems (TSSs) in electrically stressed metal-oxide-silicon field-effect transistors, by studying random telegraph signals at low temperatures. The TSSs are related to defects close to the Si/SiO₂ interface. The asymmetry energy ε is linear in gate voltage, permitting for the first time measurement of the dissipative tunneling rate as a function of ε independently of magnetic field. The electron coupling strength α lies in the range 10⁻³ to 10⁻², allowing comparison with the standard theory in the previously unexplored regime of weak coupling.

PACS numbers: 72.10.Fk, 72.70.+m, 73.40.Qv

A variety of physical processes can be understood in terms of a very simple quantum system which tunnels back and forth between only two quantum states, exchanging energy with its surroundings. In solids, the occurrence of such dissipative two-state systems (TSSs) was first postulated to explain the remarkably universal low-temperature thermal and acoustic properties of glasses [1], and their properties have since been studied in great detail theoretically [2]. The TSS model has recently been used very successfully to understand aspects of noise in electrical transport in metals [3–5]. Liquid helium temperatures are necessary to ensure that all higher energy levels of the system are inaccessible, and that the transition occurs by tunneling rather than activation over a barrier. The behavior of a TSS is then determined by the energy separation ε of its two states, the renormalized tunnel matrix element Δε between them, and a dimensionless coupling constant α which depends on the strength and nature of interactions with the surroundings at temperature T.

The first experimental study of an individual TSS was made by looking at the noise power in a Josephson-junction device [4]. The transition rate γ(ε) between two fluxoid states at ε = 0 showed the predicted power-law dependence on T, γ(ε) ∝ T^2α−1, with α = 1.44. Soon afterwards, random telegraph signals (RTSs) in the conductance of bismuth films were linked to individual scattering defects showing TSS behavior [5]. Here it was found that, for reasons uncertain, ε could be tuned by applying a magnetic field. The results for one TSS were consistent with α = 0.24, giving a remarkable increase in γ(ε) with decreasing T below about 1 K. We report here the discovery and investigation of individual TSSs in silicon metal-oxide-semiconductor field-effect transistors (MOSFETs). The TSSs lie near or at the Si/SiO₂ interface and modulate the channel resistance by changing their scattering cross section for inversion layer electrons, producing RTSs. Their properties are very different from those of electron-trapping defects in similar devices [6], which exchange not only energy but also particles (electrons) with the surroundings. In these new TSSs one has for the first time complete control of ε, which is linear in the gate voltage V_g, independently of magnetic field. The coupling of the TSSs to electrons and phonons is very weak, with α as small as 10⁻³, so that in agreement with theory γ(ε) shows a narrow Lorentzian peak about ε = 0, and γ(0) ∝ 1/T up to 7 K. The narrow range of values of Δε observed implies that only one type of defect is involved. Also, each TSS is highly sensitive to electric or strain fields which affect ε, so that the central peak is often split due to interactions with other defects.

The measurements were made on conventional 1.5 μm process n-channel MOSFETs with electrically active areas of about 1 μm², oxide thickness d_ox = 320 Å, and initial threshold gate voltage 0.7 V, at temperatures between 1.2 and 7 K. The resistance was monitored with a maximum time resolution of about 200 μs, using an ac constant-current technique with a signal level such that the source-drain voltage was not larger than k_BT. Electrical stressing was performed with the source grounded, V_g = 3–5 V, and a positive drain voltage of up to 8 V applied for a second at 4.2 K. This procedure results in a positive shift of the threshold gate voltage due to injection of electrons into the oxide near the drain [7], which increases with stressing drain voltage.

The RTSs measured had fractional amplitudes in resistance ranging from about 0.02% in strong inversion up to 100% in stressed devices in weak inversion. In mildly stressed devices all measurements were reproducible over weeks at 4.2 K, and most RTSs survived thermal cycling to room temperature. The time constants τ₁ and τ₂ for the upper and lower resistance levels, respectively, were obtained from the mean of at least 500 periods of the RTS, giving a standard error of less than 5%. For every data point the standard deviation of the periods was equal to the mean to within a few percent, in agreement with Poissonian statistics. As will be mentioned later, some of the signals did show small deviations from Poissonian behavior as a result of defect interactions. In line with previous reports [6] all RTSs obeyed the relation

\[ \tau_1/\tau_2 = \exp[\eta(V_g - V_{g0})/kT] = \exp(-\Delta E/kT), \]

(1)

where e is the electronic charge and η and V_{g0} are constants for each RTS. The parameter ΔE, which is the
difference in free energy between the states of the system corresponding to the two resistance levels, is therefore given by $\Delta E = -e\eta(V_g - V_{00})$. This linear variation with $V_g$ is equivalent to that of the energy of a dipole of magnitude $e\eta d_{\alpha\alpha}$ aligned parallel to the oxide electric field.

In unstressed devices each of $\tau_1$ and $\tau_2$ always varied monotonically with $V_g$, in a manner consistent with the usual model where a single electron tunnels a distance $\eta d_{\alpha\alpha}$ between the inversion layer and a near-interface oxide defect [6]. $\Delta E$ being the difference between the defect energy level and the Fermi level $E_F$. After electrical stress, new RTSs appeared which usually showed similar behavior. However, one in every few RTSs exhibited a very distinctive, nonmonotonic $V_g$ dependence. Figure 1(a) shows a sweep of $V_g$ over a range where such an RTS is visible, in strong inversion at 4.2 K. The switching rate clearly exhibits a sharp peak at $V_g = 3.33$ V. Figures 1(b) and 1(c) show the profile of a similar resonance in another device at two temperatures, 1.2 and 4.2 K. Figure 1(c) demonstrates that at both temperatures $\tau_1/\tau_2$ varies with $V_g$ according to Eq. (1), with $\eta = 3.0 \times 10^{-3}$ and $V_{00} = 1.567$ V. Figure 1(b) shows the corresponding variation of the mean switching rate $\gamma = (\tau_1\tau_2)^{-1/2}$ with $V_g$. The energy $\Delta E$, deduced from Eq. (1), is plotted along the top axis. The resonance in $\gamma$ is fairly symmetric about $\Delta E = 0$, and it becomes taller and narrower at lower $T$.

In any model involving electron capture from the inversion layer to a defect it is impossible to explain either the existence of such a resonance, which is much narrower than $k_B T$, or the fact that $\gamma$ decreases as $T$ increases at $\Delta E = 0$. On the other hand, the results are completely consistent with the predicted behavior of a two-state system interacting with an electron bath with a coupling constant $\alpha \ll 1$. The geometric mean of the forward and reverse transition rates, $\Gamma_+ \Gamma_-$, of such a system is given by [8]

$$\gamma(\varepsilon) = \sqrt{\Gamma_+ \Gamma_-} = \frac{1}{\sqrt{\Delta^2}} \left\{ \frac{2\pi k_B T}{\hbar} \right\} \left( \frac{e^2/2\pi k_B T}{e^2} \right)^{2a-1} \left| \frac{\Gamma(a+i\varepsilon/2\pi k_B T)}{\Gamma(2a)} \right|^2.$$

It is more common to deal with the arithmetic mean of the transition rates, which contains an extra factor of $\cosh(e/2k_B T)$, because it is related to the associated noise power. However, if $\Gamma_+$ and $\Gamma_-$ are measured directly, as in RTS experiments, this added complication is unnecessary. The renormalized tunnel matrix element $\Delta_\alpha$ is related to the bare element $\Delta$ by $\Delta_\alpha = \Delta(\omega_c)^{a/2}$, where $\omega_c$ is the bath cutoff frequency. If $\alpha \ll 1$ then $\Gamma_\alpha = \Delta$ and $\Gamma(a) \approx a^{-1}$, and for $|\varepsilon| \leq k_B T$, $|\Gamma(a+i\varepsilon/2\pi k_B T)|^2 \approx |\Gamma(a)|^2 [1 + (e/2\pi k_B T)^2]^{-1}$ [9], so Eq. (2) reduces to

$$\gamma(\varepsilon) \approx \frac{h\Delta^2}{2\pi k_B T} \frac{1}{1 + (e/2\pi k_B T)^2}.$$

The form of this equation reflects the fact that the intrinsic level width $h\Delta$ is smeared out to an effective width $2\pi a k_B T$ by interactions with the electron bath [10]. A plot of $\gamma(\varepsilon) T$ against $e/k_B T$ should give a Lorentzian of half-width $2\pi a$ and height $h\Delta^2/2\pi k_B$, independent of $T$.

Figure 2 shows $\gamma T$ plotted against $\Delta E/k_B T$ for the resonant RTS of Fig. 1(b) at four temperatures. The solid line is the best fit of Eq. (3) to all the data points, identifying $\gamma$ with $\gamma(\varepsilon)$ and $\Delta E$ with $\varepsilon$. This yields $\alpha = 7 \times 10^{-3}$ and $\hbar\Delta/k_B = 3.7 \times 10^{-5}$ K, and the agreement of the fit over 3 orders of magnitude, including the correct scaling with temperature, is convincing evidence that the RTS is due to a defect acting as a TSS with $\alpha \ll 1$.

From the fits in Fig. 2 it is apparent that $\gamma(0) \sim T^{-1}$ up to 7 K. This implies that the characteristic phonon-coupling temperature [11] is above 7 K in this system, compared with about 1 K in bismuth films. Also, $\alpha$ is 2 orders of magnitude smaller in the MOS system. This may be explained by its quadratic dependence on the density of electron states at $E_F$ and on the strength of the scattering potential [5,10]. The bismuth films contain both electrons and holes, each at around $10^{19}$ cm$^{-3}$ [5], while the silicon inversion layer contains only electrons, at
a density typically 10 times lower. Assuming the defect is located within a few Å of the Si/SiO₂ interface, where the inversion layer electron wave function almost vanishes, the scattering potential is likely to be weaker in this case as well.

Not all resonant RTSs exhibit a single isolated peak in γ as a function of ∆E. Figure 3(a) shows results for another RTS at 4.2 K, where the central resonance has a smaller satellite peak. The central peak is fitted well by Eq. (3) with α =2.1×10⁻³ and h/Δk_B=3.4×10⁻⁶ K, and the smooth curve, which fits both peaks reasonably, was obtained by adding another Lorentzian with the same value of α but only one-fifth the amplitude and displaced in ∆E by −35 μeV. The RTS of Fig. 3(b) has a more complex behavior still, exhibiting structure on different scales. Some of the satellite peak structure (open circles) was only present for a few hours before it suddenly vanished, but the remaining structure (filled circles) was completely reproducible over one week at 4.2 K. The sharp peaks close to ∆E=0 are best fitted by α ≈1.5×10⁻³ and h/Δk_B=4×10⁻⁶ K. It is tempting to relate the peaks at nonzero ∆E to the alignment of higher energy levels of the two states of the TSS. However, the separation of the sharpest peaks can be as small as 30 μeV, which is orders of magnitude below typical excitation energies for individual atomic defects, and the arrangement of peaks is very variable—no two resonant RTSs show the same structure in γ. In fact, the peak splitting in Fig. 3(a) can be understood if the true instantaneous value of α at fixed V_g depends on the configuration of another defect in the device, which interacts with the TSS defect through either electric or strain fields. If this other defect switches rapidly between two states, then ε jumps between two values separated by some coupling energy Δε, and the measured rate γ is a mixture of γ(ε) and γ(ε+Δε), giving two peaks in γ as a function of ∆E separated by Δε. It is likely that the complex structure in Fig. 3(b) results from simultaneous interactions with several other switching defects. In support of this explanation, for some multiple-peak RTSs the mean transition rate near a peak in γ was observed to change occasionally and suddenly, as if modulated by another slower switching defect. This causes a small deviation of the RTS from Poisson statistics, as noted earlier.

Multiple-scattering quantum interference effects are small in these devices at 4.2 K. The elastic and inelastic scattering lengths are both at least an order of magnitude less than the device size, and the RTS amplitudes are not noticeably affected by a magnetic field of up to 4 T, which is much greater than the correlation field for universal conductance fluctuations [12]. Still, it may be that modulation of the local charge density as a function of V_g by quantum interference [5,13] produces small fluctuations in α which cause some deviations of γ from the ideal Lorentzian form of Eq. (3), such as those apparent in Fig. 2, even in the absence of interactions with other defects. On the other hand, the change in scattering cross section, estimated from the RTS amplitudes in strong inversion, is consistent with the local interference model [14] assuming the total cross section is at the unitary limit.

This large change in cross section without change in charge requires the TSS defects to be very close to the inversion layer electrons. As stress-related defects are unlikely to lie in the silicon, the TSSs are therefore probably associated with the Si/SiO₂ interface. The peak switching rate at 4.2 K is never less than 200 s⁻¹ and rarely exceeds the experimental bandwidth of 5000 s⁻¹, giving values of h/Δk_B lying within about a decade centered on 10⁻⁵ K. This strongly suggests that the microscopic system involved is similar, but not identical, in every case.
The TSS defect should have two similar atomic/electronic configurations with a low probability for tunneling between them. The values of $\eta d_{\alpha}$ for six TSSs all lay in the range $\pm 1$ Å, which is too small for tunneling of an electron between two separate, randomly located oxide defects, but could be determined by random fixed charges which follow changes in the lattice around the TSS as it reconfigures. Some of the oxygen-vacancy related $E'$ centers in SiO$_2$ contain a hydrogen atom or a trapped hole which can sit on either of two silicon atoms [15]. If these acted as TSSs, the random values of $V_{\phi 0}$ and the range of $\Delta$ observed would reflect the sensitivity of $\epsilon$ and $\Delta$ to the particular bonding configuration around the two silicon atoms, as well as to the electric and strain fields present. However, the large lattice relaxation predicted for $E'$ centers should make tunneling of the hydrogen atom or hole impossible at low temperatures. It is possible in fact that the TSS defect has not been detected previously.

As in the metal systems, the requirement $h \Delta \ll (\epsilon$ or $ak_BT)$ for incoherent tunneling [2] is fulfilled here. However, since $a$ can be as small as $10^{-3}$, and $\epsilon$ can be accurately tuned to zero, by going to lower temperatures it may be possible to violate this condition and explore the coherent regime. Moreover, in the MOS system the electron density of states at $E_F$ can be changed drastically by applying a magnetic field strong enough to form Landau levels, and with a substrate bias it is also possible to vary $E_F$ at constant oxide electric field, so it should also be possible to vary $a$ at fixed $\epsilon$.

In summary, one species of defect generated by electrical stress at the Si/SiO$_2$ interface behaves as a two-state system at low temperatures, giving rise to random telegraph signals with transition rates resonant in gate voltage. The dynamics of these flexible new experimental systems have been investigated as a detailed function of asymmetry energy, providing strong support for the existing theory in the previously unexplored regime where the coupling constant $\alpha \ll 1$.

We thank N. O. Birge, D. E. Khmelnitskii, M. J. Kirtton, A. J. Leggett, G. McMullan, Y. Oowaki, A. K. Savchenko, and I. Zailer for helpful discussions, and we acknowledge the support of the United Kingdom Science and Engineering Research Council.

[9] This can be seen from the equality $|\Gamma(a+i\nu)|^2 = |\Gamma(a)|^2 [1 + y^2(a+n)^2]^{-1}$, where for $a \ll 1$ and $y \ll 1$ the $n=0$ term dominates.