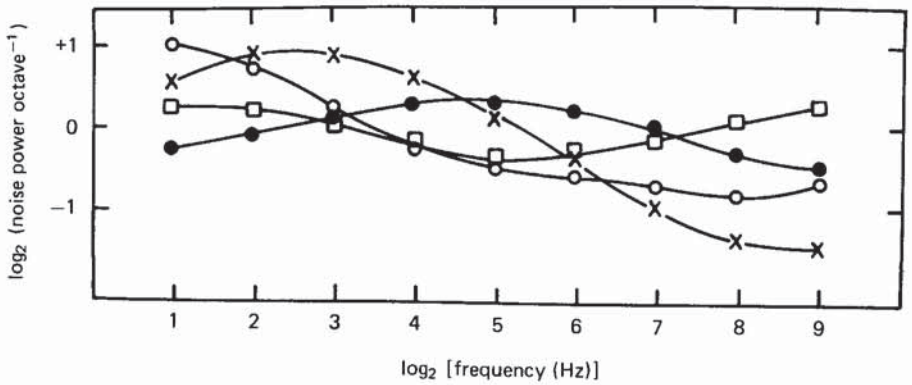

Figure 19. Monte Carlo octave spectrum derived from the fluctuations of 30 two-state systems with randomly picked characteristic frequencies and duty cycles. From Restle *et al.* (1985). © American Physical Society. Reproduced with permission.

versus frequency. Again, true $1/f$ noise would appear as a horizontal trace on this modified plot. Figure 20 shows corresponding power spectra from seven different SOS resistors of area about $1\,\mu m^2$, taken at room temperature. Comparing these two figures, one sees that both sets of spectra show the sample-to-sample variation expected of a limited sampling/summation of two-state systems. In addition, the theoretical spectra more closely conform to a $1/f$ behaviour, indicating that in the experimental data fewer than 30 traps per sample were contributing to the noise. It was found that the general behaviour of the RWRP was not present.

4.3. $1/f$ *noise in small-area MOSFETs*

We investigated the nature of $1/f$ noise in MOSFETs by studying devices whose electrically-active areas ranged from 0·4 to $350\,\mu m^2$ (Uren *et al.* 1985). The devices were conventional silicon-gate n-channel MOSFETs with gate oxide thickness of 40 nm, as discussed in section 3.2.1. On the basis of the model that individual fluctuating trapping centres modulate the conductance in only a small region surrounding the defect site (section 3.2.3), we have seen that equations (4.1) and (4.2) predict that the power spectrum $S_I(f)$ should be proportional to I^2 and inversely proportional to the area A. In figures 21 (*a*)–(*c*) we show the effect on the noise power spectral density of reducing device area. All these measurements were made in the linear regime for a fixed gate voltage in strong inversion. For comparison between differently sized devices, the scaled function $AS_I(f)/I^2$ has been plotted against frequency. For devices from the same wafer this should give $1/f$ spectra that are superimposed on one another. For the largest devices shown in figure 21 (*a*) this behaviour was observed with a variation of less than 20% across the wafer. However, smaller devices with active areas below about $30\,\mu m^2$ departed significantly from the expected behaviour. Figure 21 (*b*) depicts the scaled noise spectra from three devices with active areas of $15\,\mu m^2$. As with the other microstructures, the spectra no longer exhibit smooth $1/f$ behaviour, but significant and consistent irregularities are present so that they no longer superimpose. However, it is clear that by averaging over many such devices the resulting spectrum would be $1/f$-like and smooth, and would exactly match those of figure 21 (*a*). This is shown even more clearly in figure 21 (*c*), where the spectra for three $0{\cdot}4\,\mu m^2$ devices are presented. The spectra are composed from the sum of a few Lorentzians, corresponding to the RTSs observed in the time domain. Again, averaging these scaled spectra over many such devices would lead to the spectra shown in figure 21 (*a*).

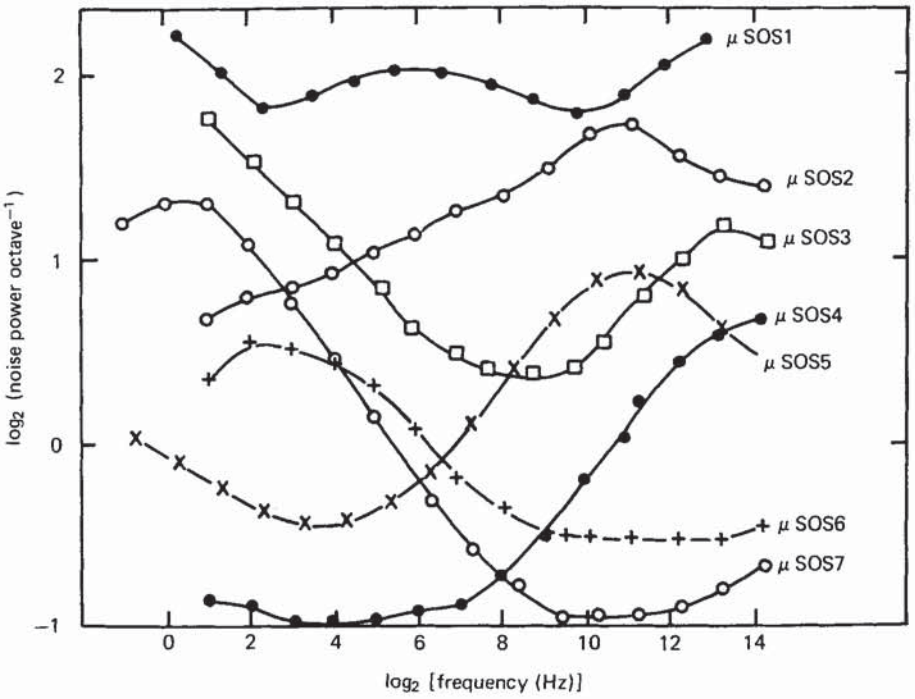

Figure 20. Power spectra at room temperature for seven different SOS samples with approximately same size and geometry. From Restle *et al.* (1985). © American Physical Society. Reproduced with permission.

The behaviour of the amplitude of RTSs as a function of channel conductivity G was discussed in section 3.2.3. There we predicted that $\Delta I_D/I_D$ was a constant in weak inversion and thereafter decreased as the inversion-layer number density was increased. This plateau in $\Delta I_D/I_D$ manifests itself in the power spectrum (Reimbold 1984). Figure 22 shows the measured noise power at $f = 2$ Hz plotted against channel conductivity together with three RTS amplitudes. A plateau in $S_I(f = 2\,\text{Hz})/I^2$ at low values of G is quite evident. Referring to equation (4.1), we see that if the set of values of $(\Delta I_D/I_D)_k$ did not vary as I_D increased and we were always summing over a sufficiently large and constant number of traps then we should expect $S_I(f)/I^2$ to be independent of G. Analysis of the noise power curve using the formula described by Reimbold indicates that the number of active traps is increasing as V_G is increased. However, because the amplitudes of the RTSs decrease rapidly and this effect dominates, the noise power must decrease—as observed. Thus we have another link between the microscopic behaviour of RTSs and the macroscopic power spectrum.

The experiments described in this and the previous two sections provide an elegant demonstration of the full decomposition of the $1/f$ spectrum into its constituent Lorentzian components and the sample-to-sample variation in the power spectra expected from very small systems. Although we have linked carrier trapping events to the generation of $1/f$ noise in MOSFETs, MIM tunnel diodes and SOS resistors, we are still left with a puzzle regarding the mechanism whereby the required wide distribution of time constants arises. In section 5 we shall demonstrate that carrier trapping at individual insulator defects takes place via a thermally activated lattice

reconstruction process. In section 7 we shall show that the measured distributions of activation energies and cross-section pre-factors provide a very plausible explanation for the very wide time-constant dispersion implicated in $1/f$ noise.

5. Capture and emission kinetics of individual defect states

Random telegraph signals (RTSs) generated through the fluctuating occupancy of individual defects provide a unique probe into the trapping dynamics of single defects. In the following subsections we shall review recent work aimed at determining microscopic models of the capture and emission kinetics of individual defects at the Si/SiO_2 interface (section 5.1) and defects in the insulator of MIM tunnel junctions (section 5.2).

5.1. *Charge-carrier trapping into individual defects at the Si/SiO_2 interface*

We investigated the capture and emission kinetics of single defects at the Si/SiO_2 interface by analysing RTSs as a function of temperature and gate voltage (Kirton and Uren 1986, Kirton *et al.* 1989a). In the remainder of this subsection we shall mainly follow our treatment of the problem. First of all, we shall consider the situation in an n-channel MOSFET with a trap level lying close to the Fermi level, as shown in figure 12. Our initial aim is to determine the factors governing the behaviour of the mean capture and emission times of an RTS as a function of temperature.

5.1.1. *Temperature dependence of RTSs in silicon MOSFETs (I): Theory*

Formally, we can write the carrier capture rate for an interface defect as

$$\frac{1}{\bar{\tau}_c} = \int_{E_c}^{\infty} r(E)\,\mathrm{d}E, \tag{5.1}$$

where $r(E)$ is the transition rate (per unit energy) at energy E in the inversion layer. $r(E)$ can be written as the product of a particle flux and cross-section:

$$\frac{1}{\bar{\tau}_c} = \int_{E_c}^{\infty} n(E)v(E)\sigma(E)\,\mathrm{d}E. \tag{5.2}$$

$n(E)$ is the inversion-layer number density (per unit volume per unit energy) at energy E, and $v(E)$ is the carrier velocity at E. We shall use equation (5.2) in the following simplified form:

$$1/\bar{\tau}_c = n\bar{v}\sigma. \tag{5.3}$$

We have made the major assumption that the particle density (per unit volume) in the inversion layer can be represented by the constant value n. $\bar{v}$ is the average thermal velocity of the carriers and σ the average capture cross-section. We shall comment further on this approximation in section 5.1.4.

The relationship between the capture and emission times for a defect which is fluctuating between its n- and $(n + 1)$-electron states is given by equation (2.27 *b*). We repeat it here for convenience:

$$\bar{\tau}_e = \frac{\bar{\tau}_c}{g}\exp\left[-\frac{E(n + 1/n) - E_F}{kT}\right]. \tag{5.4}$$

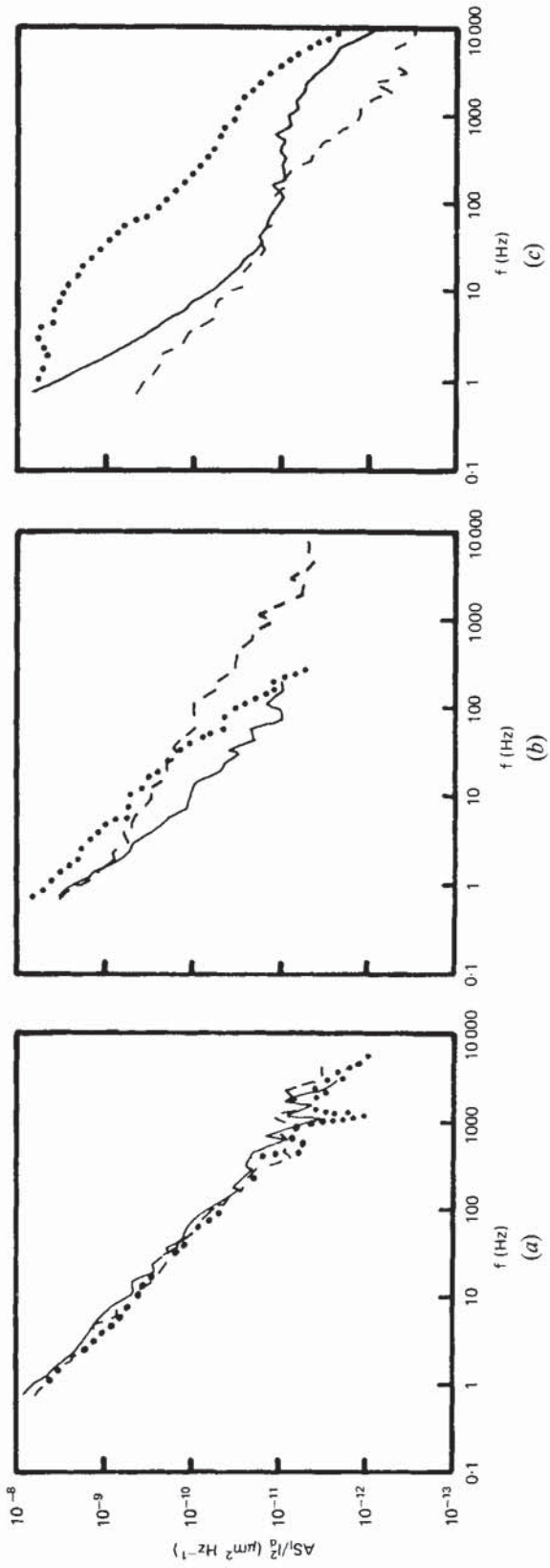

Figure 21. Noise power spectal density of n-channel MOSFETs for $V_G = 2$ V, $T = 293$ K. White noise has been subtracted. (*a*) Three 20 μm × 20 μm devices, active area 350 μm^2, $V_D = 100$ mV, $I_D \approx 3{\cdot}5$ μA; (*b*) three 20 μm × 2 μm devices, active area 15 μm^2, $V_D = 20$ mV, $I_D \approx 15$ μA; (*c*) three 2 μm × 2 μm devices, active area 0·4 μm^2, $V_D = 20$ mV, $I_D \approx 500$ nA. From Uren *et al.* (1985).

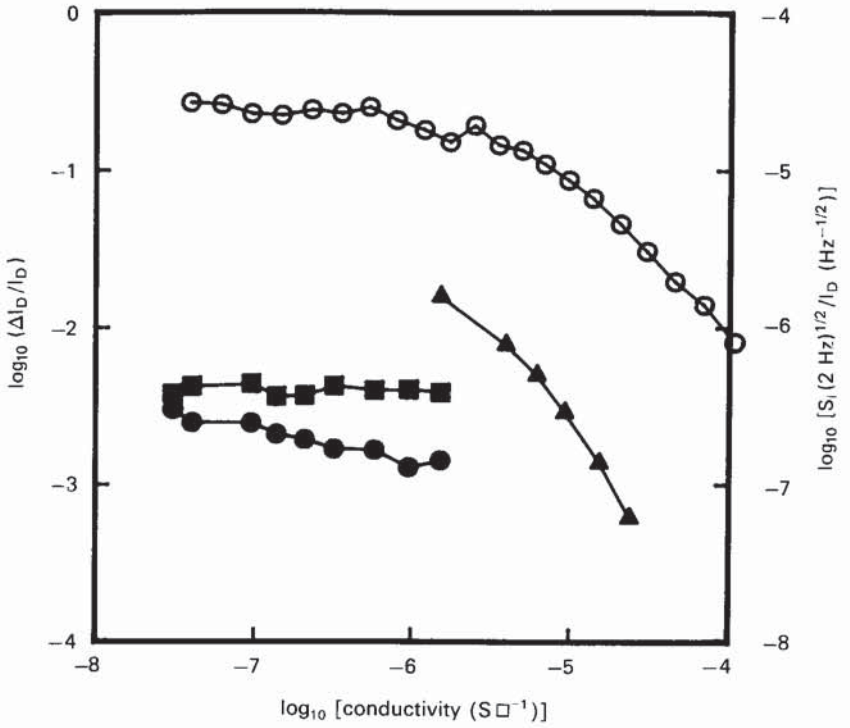

Figure 22. Left-hand scale: fractional change in current *versus* channel conductivity for RTSs from three traps in one 0·4 μm² device (filled symbols). Right-hand scale: noise power spectral density at 2 Hz for a 350 μm² device (open symbols). For both devices V_D = 100 mV and T = 293 K. From Uren *et al.* (1985).

Further, we write the cross-section in the form

$$\sigma = \sigma_0 \exp\left(-\frac{\Delta E_B}{kT}\right). \tag{5.5}$$

In equation (5.5) we are anticipating the experimental findings and taking the cross-section to be thermally activated, with a barrier for capture ΔE_B and a cross-section pre-factor σ_0 (Schulz and Johnson 1978). In fact, the proposed mechanism of charge-carrier capture and emission is through a process directly analogous to the multiphonon mechanism familiar from deep levels in bulk semiconductors (Henry and Lang 1977, Bourgoin and Lannoo 1983). A schematic configuration-coordinate diagram showing the changes in total energy of the system as an electron is transferred from the inversion layer into an interface defect is shown in figure 23. The energy zero in this figure corresponds to the empty trap with an electron at the Fermi level. The dashed curve shows the variation in total energy as the empty defect distorts (plotted against a single normal coordinate). The full curve marked with an open circle shows the same with the electron in the conduction band. The full curve marked with the full circle depicts the variation in total energy of the trap after it has captured the electron. At the cross-over there is strong mixing between the inversion-layer state and the defect state. The non-radiative transition is induced by off-diagonal elements in the Hamiltonian, which induce transitions between vibronic states that differ in electronic energy but have the same total energy. On electron capture, the defect state is well

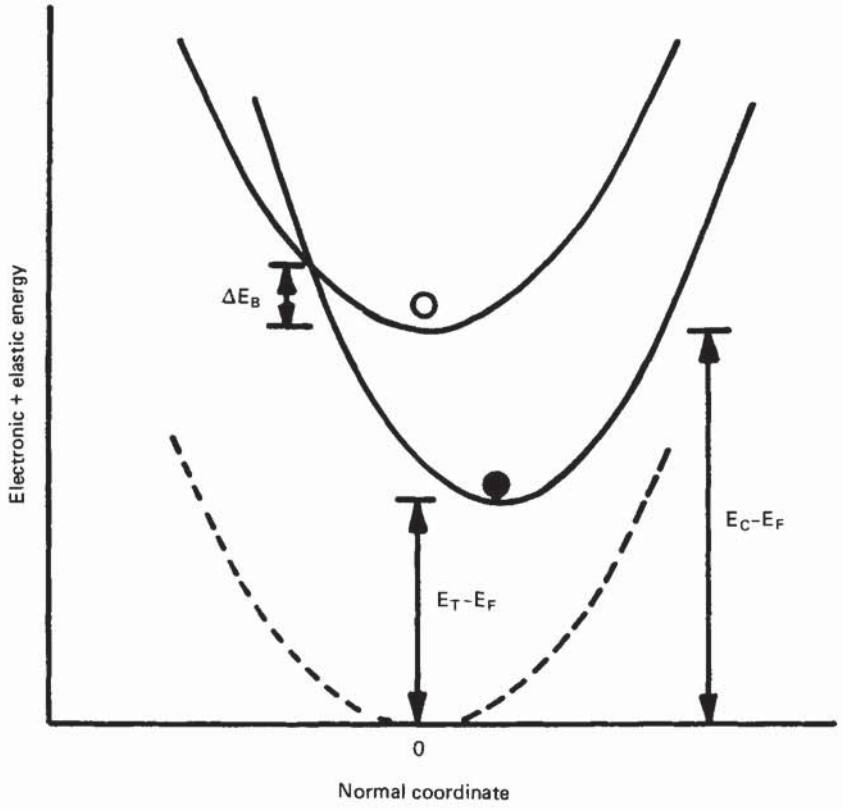

Figure 23. Configuration coordinate diagram: elastic + electronic energies against single normal coordinate. The energy zero of the system corresponds to the empty defect with the electron at the Fermi level. This is shown as the dashed curve. ○ labels the empty trap plus a free electron in the inversion layer. ● marks the filled trap.

away from equilibrium and the excess energy is dissipated by multiphonon emission. Additional confirmation of this model has come from the very recent experiments of Andersson and Engstrom (1989) who observed an energy shift between optical and thermal ionization energies for radiation-induced Si/SiO_2 interface states.

Substituting equation (5.5) into equation (5.3), we obtain

$$\bar{\tau}_c = \frac{\exp(\Delta E_B/kT)}{\sigma_0 \bar{v} n}. \tag{5.6}$$

Equations (5.4) and (5.6) contain all of the physics that we require in order to investigate the temperature dependence of the mean capture and emission times of a single RTS.

Before proceeding further, we need to consider the effects of changing temperature on the band-bending inherent in the MOSFET structure. In table 1 we give representative values of ϕ_s and ϕ_b (see figure 12) for varying temperature with the device operating well into weak inversion. The numbers shown were obtained from a theoretical analysis of our devices based on Brews' (1978) charge-sheet model of the MOSFET. (For computational convenience the flat-band voltage has been set equal

Table 1. Theoretical variation of surface potential ϕ_s, bulk Fermi level ϕ_b and potential across oxide $V_G - \phi_s$ as functions of temperature. The analysis is based on Brews' (1978) charge-sheet model of the MOSFET. The device is operating in weak inversion. $t_{ox} = 40\,\text{nm}$, $N_A = 1{\cdot}1 \times 10^{16}\,\text{cm}^{-3}$.

T (K)	V_G (V)	I_D (pA)	ϕ_s (V)	ϕ_b (V)	$V_G - \phi_s$ (V)
280	1·0	0·322	0·5088	0·3668	0·4912
300	1·0	2·290	0·5094	0·3503	0·4906
320	1·0	12·28	0·5099	0·3337	0·4901
340	1·0	53·76	0·5104	0·3168	0·4896

Table 2. Theoretical variation of surface potential ϕ_s, bulk Fermi level ϕ_b and potential across oxide $V_G - \phi_s$ as functions of temperature. The analysis is based on Brews' (1978) charge-sheet model of the MOSFET. The device is operating in strong inversion. $t_{ox} = 40\,\text{nm}$, $N_A = 1{\cdot}1 \times 10^{16}\,\text{cm}^{-3}$.

T (K)	V_G (V)	I_D (μA)	ϕ_s (V)	ϕ_b (V)	$V_G - \phi_s$ (V)
280	5·0	1·308	0·9077	0·3668	4·0923
300	5·0	1·192	0·8857	0·3503	4·1143
320	5·0	1·091	0·8634	0·3337	4·1366
340	5·0	1·000	0·8405	0·3168	4·1595

to zero so that the theoretical gate voltage is larger than the experimental value by about 0·3 V.) Table 1 shows that as the temperature is increased, the surface potential increases and the field in the oxide decreases. As we shall discuss in section 5.1.4, changes in the oxide field strength and inversion-layer number density affect the dynamics of trapping. It is worth reflecting on the magnitude of these changes. For a 60 K increase in temperature the potential drop across the oxide is reduced by 0·0015 V, corresponding to only a 0·3% change, which is insignificant in practice.

The situation in strong inversion is slightly more complicated. On increasing the temperature, both the valence-band edge in the bulk and the inversion-layer conduction-band edge move away from the Fermi level. The oxide field strength is increased; the current is reduced; and the trap occupancy level moves away from the surface Fermi level. Table 2 shows theoretical estimates of the relevant parameters. Again it is the changing oxide field strength that presents the potential problem; but the change is only at the 1% level for a temperature change of 60 K.

Equation (5.6) can be further expanded by noting that

$$\bar{v} = (8\,kT/\pi m^*)^{1/2}, \tag{5.7}$$

where m^* is the average mass of a carrier in the inversion layer. Previously we have used the RMS velocity (Kirton and Uren 1986); however, the difference between v_{RMS} and $\bar{v}$ is insignificant. The temperature dependence of the number density n can be obtained from the variation of the drain current $I(T)$. Since all of the measurements that we shall report on were carried out in the linear regime where the device is operating in a simple resistive fashion, we can write

$$I(T) = n(T)q\mu(T)V_D t(T)w/l, \tag{5.8}$$

where $\mu(T)$ is the temperature-dependent electron mobility; V_D is the applied drain voltage, and $t(T)$, w and l are the channel thickness, width and length respectively. The thickness was estimated from a classical integration of the surface charge. The temperature dependence of the mobility and inversion-layer thickness can be expressed as (Ando *et al.* 1982)

$$\mu(T) = \mu_0 T^{-3/2} \tag{5.9}$$

$$t(T) = t_0 T. \tag{5.10}$$

Combining equations (5.8)–(5.10), we obtain the temperature variation of the average inversion-layer number density:

$$n(T) = \frac{IT^{1/2}}{q\mu_0 V_D t_0 (w/l)}. \tag{5.11}$$

Incorporating equations (5.7) and (5.11) into equation (5.6), we find

$$\bar{\tau}_c = \frac{q\mu_0 V_D t_0 (w/l) \exp(\Delta E_B/kT)}{\sigma_0 (8kT/\pi m^*)^{1/2} I(T) T^{1/2}}, \tag{5.12}$$

and therefore

$$I(T) T \bar{\tau}_c = \frac{\exp(\Delta E_B/kT)}{\sigma_0 \chi}, \tag{5.13}$$

where χ is a constant and is given by

$$\chi = \frac{(8k/\pi m^*)^{1/2}}{q\mu_0 V_D t_0 (w/l)}. \tag{5.14}$$

Equations (5.4), (5.6) and (5.7) can be combined to give

$$\bar{\tau}_e = \frac{\exp(\Delta E_B/kT) \exp\{-[E(n + 1/n) - E_F]/kT\}}{\sigma_0 g (8kT/\pi m^*)^{1/2} n(T)}. \tag{5.15}$$

In expanding equation (5.15), we shall choose to write the mean inversion-layer number density as

$$n(T) = N_C \exp\left(-\frac{E_C - E_F}{kT}\right), \tag{5.16}$$

where E_C is the energy of the conduction band at the inversion-layer charge centroid (not at the surface) and N_C is the effective density of states. Thus

$$\bar{\tau}_e = \frac{\exp(\Delta E_B/kT) \exp\{[E_C - E(n + 1/n)]/kT\}}{\sigma_0 g (8kT/\pi m^*)^{1/2} N_C}. \tag{5.17}$$

Now N_C can be expressed as $N_{C0} T^{3/2}$ (Sze 1981), and therefore equation (5.17) becomes

$$T^2 \bar{\tau}_e = \frac{\exp[(\Delta E_B + \Delta E_{CT}/kT]}{\sigma_0 g \eta}, \tag{5.18}$$

where

$$\eta = N_{C0} (8k/\pi m^*)^{1/2}, \tag{5.19}$$

$$\Delta E_{CT} = E_C - E(n + 1/n). \tag{5.20}$$

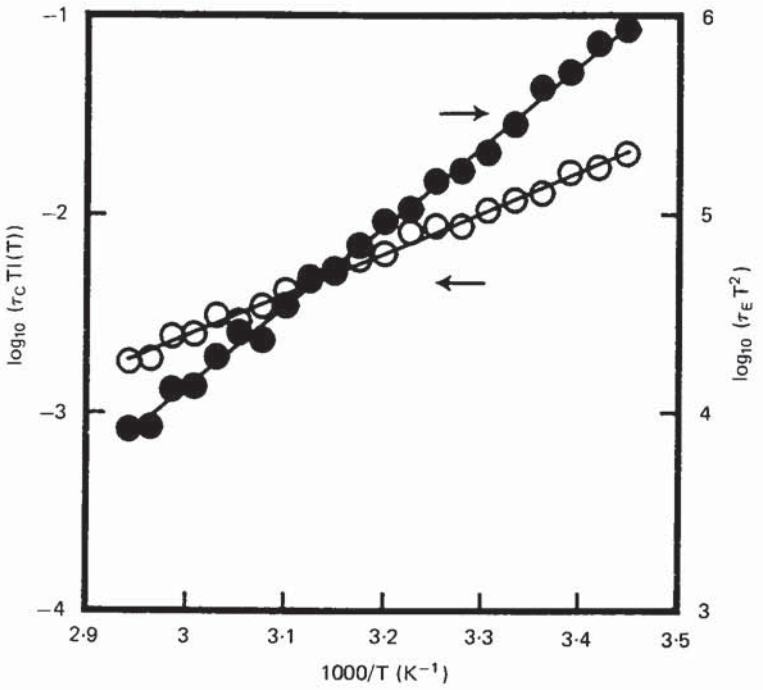

Figure 24. Temperature dependence of capture time (○) and emission time (●) for device H21. $V_G = 3\,V$, $V_D = 50\,mV$. After Kirton *et al.* (1989a).

5.1.2. *Temperature dependence of RTSs in silicon MOSFETs (II): Results*

The two equations to be used in investigating the temperature dependence of the mean capture and emission times are (5.13) and (5.18):

$$I(T)T\bar{\tau}_c = \frac{\exp(\Delta E_B/kT)}{\sigma_0 \chi},$$

$$T^2\bar{\tau}_e = \frac{\exp[(\Delta E_B + \Delta E_{CT})/kT]}{\sigma_0 g\eta}.$$

where equations (2.25), (5.5), (5.14), (5.16), (5.19) and (5.20) define the parameters on the right-hand sides of the above two equations. From plots of $\ln[I(T)T\bar{\tau}_c]$ and of $\ln[T^2\bar{\tau}_e]$ against $1/T$ (see figure 24) the energies ΔE_B and ΔE_{CT} can be obtained and from the intercepts two independent estimates of σ_0 (assuming a nominal value for g of unity) recovered. We found two notable discrepancies using this approach: first the independent estimates of σ_0 differed by typically two to three orders of magnitude; and secondly, the value of ΔE_{CT} was in general around 0·2 eV greater than estimates of $E_C - E_F$, placing the trap occupancy level $E(n + 1/n)$ well below the surface Fermi level. However, if the energy level were at this position then the defect's occupancy would not fluctuate on accessible time scales. Note also that this result cannot be explained by two-electron capture, since it would place the average energy level $\frac{1}{2}[E(2/1) + E(1/0)]$ well below the Fermi level and the defect would remain in its two-electron state.

Engstrom and Alm (1983) noted that in thermal experiments it is Gibbs free-energy changes that are measured, and the energy ΔE_{CT} should be split into its corresponding enthalpy ΔH_{CT} and entropy ΔS components, namely

$$\Delta E_{CT} = \Delta H_{CT} - T\,\Delta S. \tag{5.21}$$

An alternative way of viewing this equation is to say that the trap energy level is temperature-dependent. We have used a first-order expansion and assumed that, within the small temperature range accessed, ΔH_{CT} is constant. Substituting equation (5.21) into equation (5.18), we find that

$$T^2\bar{\tau}_e = \frac{\exp(-\Delta S/k)\exp[(\Delta E_B + \Delta H_{CT})/kT]}{\sigma_0 \eta}, \tag{5.22}$$

where g has been incorporated into the $\exp(-\Delta S/k)$ term. Thus, from the intercept and gradient of the Arrhenius plot of equation (5.22), the change in trap entropy and the enthalpy of ionization can be determined. Table 3 shows the data obtained from ten traps at room temperature.

5.1.3. *Temperature dependence of RTSs in silicon MOSFETs (III): Discussion*

There are two features of the results presented in table 3 that stand out: the wide range, 10^{-20}–$10^{-15}\,\mathrm{cm}^2$, of cross-section pre-factors σ_0; and the wide range, 0·19–0·65 eV, of energy barriers ΔE_B. The range of activation energies is consistent with capture into defects in an amorphous material, with its consequent continuous distribution of trap environments. The pre-factor σ_0 reflects the nature (in terms of symmetry, degeneracy and so on) and overlap of the initial- and final-state wave-functions (Bourgoin and Lannoo 1983). Thus the distribution of values of σ_0 is compatible with the traps being distributed over a range of distances into the oxide. A discussion of the estimation of the trap distance from the interface is presented in the following section.

One correlation present in table 3 is that between σ_0 and ΔE_B. This arises since only defects whose time constants lay within the experimental window 1 ms to 100 s could be observed. Thus as σ_0 increases ΔE_B must increase. This leads to a marked selectivity in the defects investigated. In section 7 we shall discuss further this observation and its implications for the widths of the distributions of σ_0 and ΔE_B. In addition, we show that the multiphonon model is able to explain the wide distribution of time constants implicated in $1/f$ noise as well as the magnitude of the power spectrum.

A particularly interesting aspect of the results is the determination of the entropies of ionization ΔS of individual Si/SiO_2 interface defects (Engstrom and Grimmeiss 1989); a representative value for ΔS is about $5k$. In order to place these results in context, it is useful to consider, as an illustrative example, the temperature dependence of the band gap of silicon. As discussed by Van Vechten and Thurmond (1976), the forbidden gap ΔE_{CV} of a semiconductor corresponds to the increase in Gibbs free energy upon increasing the number of electron–hole pairs n_p by one at constant temperature and pressure:

$$\Delta E_{CV} \equiv \left.\frac{\partial G}{\partial n_p}\right|_{T,P}.$$

Table 3. Data obtained from the investigation of ten room-temperature traps observed in 0·4 μm² n-channel MOSFETs. The device numbers and the gate voltage of the devices are indicated in columns 1 and 2 respectively. Column 3 gives the temperature range over which the defects were measured and column 7 the temperature at which the cross-section σ (column 8), trap binding energy ΔE_{CT} (column 9) and position of Fermi level ΔE_{CF} (column 10) were evaluated. ΔE_{CT} and ΔE_{CF} are measured with respect to the charge centroid in the inversion layer. σ_0 (column 4) is the cross-section pre-factor, ΔE_B (column 5) is the activation energy for capture and ΔS (column 6, in units of Boltzmann's constant k) the entropy of ionization. $G\ (=I_D l/V_D w)$ is the approximate channel conductivity at which the measurements were performed. The final column gives the estimated distance of the trap into the oxide, see section 5.1.4.

Device	V_G (V)	Temperature range (K)	σ_0 (cm^2)	ΔE_B (eV)	$\Delta S/k$	Temperature (K)	σ (cm^2)	ΔE_{CT} (eV)	ΔE_{CF} (eV)	G ($S\,\square^{-1}$)	d (nm)
G19	1·00	250–290	$1{\cdot}34 \times 10^{-17}$	0·263	7·20	260	$1{\cdot}08 \times 10^{-22}$	0·180	0·150	$1{\cdot}6 \times 10^{-6}$	16·5
H21	3·00	290–340	$2{\cdot}10 \times 10^{-19}$	0·411	11·8	310	$4{\cdot}40 \times 10^{-26}$	0·076	0·080	$3{\cdot}6 \times 10^{-5}$	1·1
H28	0·82	295–350	$7{\cdot}40 \times 10^{-16}$	0·496	6·47	320	$1{\cdot}16 \times 10^{-23}$	0·211	0·217	$6{\cdot}6 \times 10^{-7}$	19·1
H25	1·00	310–350	$3{\cdot}60 \times 10^{-16}$	0·569	8·24	325	$5{\cdot}50 \times 10^{-25}$	0·136	0·163	$4{\cdot}6 \times 10^{-6}$	6·8
G34	1·15	270–350	$6{\cdot}40 \times 10^{-15}$	0·538	5·31	320	$2{\cdot}19 \times 10^{-23}$	0·151	0·181	$2{\cdot}3 \times 10^{-6}$	14·5
G8	0·86	275–350	$8{\cdot}40 \times 10^{-20}$	0·186	10·7	320	$9{\cdot}90 \times 10^{-23}$	0·218	0·253	$1{\cdot}6 \times 10^{-7}$	1·3
H12	1·45	305–360	$5{\cdot}80 \times 10^{-15}$	0·645	7·65	330	$8{\cdot}35 \times 10^{-25}$	0·102	0·143	$8{\cdot}3 \times 10^{-6}$	1·4
G19	3·67	250–310	$3{\cdot}80 \times 10^{-15}$	0·480	5·85	270	$4{\cdot}25 \times 10^{-24}$	0·078	0·048	$4{\cdot}6 \times 10^{-5}$	0·9
H15	1·00	290–340	$3{\cdot}59 \times 10^{-15}$	0·511	4·40	315	$2{\cdot}44 \times 10^{-23}$	0·216	0·198	$1{\cdot}1 \times 10^{-6}$	20·5
H4	1·90	275–325	$2{\cdot}04 \times 10^{-15}$	0·593	5·20	300	$2{\cdot}24 \times 10^{-25}$	0·076	0·108	$1{\cdot}5 \times 10^{-5}$	1·0

ΔE_{CV} may be decomposed into standard enthalpy and entropy components. It is an experimental fact that the band gap of silicon decreases with increasing temperature; the entropy change has a surprisingly large value of $2{\cdot}9k$ at room temperature. The reason for the entropy change lies in the fact that on creation of an electron–hole pair an electron is removed from a bonding valence-band state and placed in an anti-bonding conduction-band state. This process softens the lattice, leading to a decrease in phonon-mode frequencies. In particular, as shown by Brooks (1955),

$$\Delta S_{CV}(T) \approx k \sum_i \ln\left(\frac{\omega_i}{\omega_i'}\right) \quad \text{for } kT > h\omega_i,$$

where ω_i and ω_i' represent the lattice vibration frequencies before and after creation of an electron–hole pair. Heine and Van Vechten (1976) concluded that the large value of ΔS for silicon was the result of each n–p pair softening the lattice vibrations by an amount equivalent to the removal of several bonds.

Returning now to the entropy terms of table 3, we see that similar considerations to those above apply. The first point to note is that in the filled state shown in figure 23 the distortion has minimized the total energy of the defect and stabilized the lattice as much as possible. Thus the relaxed electron configuration is as 'bonding' as possible (Van Vechten and Thurmond 1976). On the release of the electron back into the inversion layer, the main contributions to the increase in entropy are the following: the softening of the lattice in the immediate vicinity of the defect; the placing of the electron in the antibonding conduction-band state; and the change in trap degeneracy which gives rise to a contribution of magnitude $k \ln g$. In section 5.1.4 we shall see that these changes in entropy are very sensitive to modifications in the trap environment brought about by changes in applied gate voltage.

Finally, we should like to consider one aspect of the trapping phenomena that we have so far overlooked: the relative roles played by electron and lattice temperature in determining the dynamics of electron capture and emission. In this respect we are now in a position to comment on the work of Jackel *et al.* (1985). In our analysis so far, we have assumed that the lifetimes of the empty and filled states depend on the motion of atomic masses. In silicon at helium temperatures the electrons and phonons are only weakly coupled. This means that by applying a source–drain voltage the electron temperature can be raised above the lattice temperature, thus allowing one to separate out the effects of electron and lattice temperature on the capture and emission times.

Figures 25 and 26 show the measured dependence of the capture and emission times of a single trap as a function of gate voltage, lattice temperature and electron temperature. At a fixed gate voltage the capture and emission times fall as the lattice temperature is increased: this is exactly the same behaviour as observed at room temperature and shown in figure 24. The interesting feature to note in figure 26 is that, whereas the capture time shows a weak dependence on electron temperature, the emission time is independent of electron temperature.

A speculative configuration-coordinate diagram that accounts for these findings is shown in figure 27 (see also Skocpol (1987)). In this figure we have plotted the total energy of the system—defect plus inversion-layer electron—as a function of the normal coordinate of the defect and the energy of the electron. Thus our energy zero corresponds to the free electron residing at the Fermi level and the defect unoccupied. The series of broken-line curves marked with open circles represents the variation of total energy of the empty defect and electron as the electronic energy is increased. The

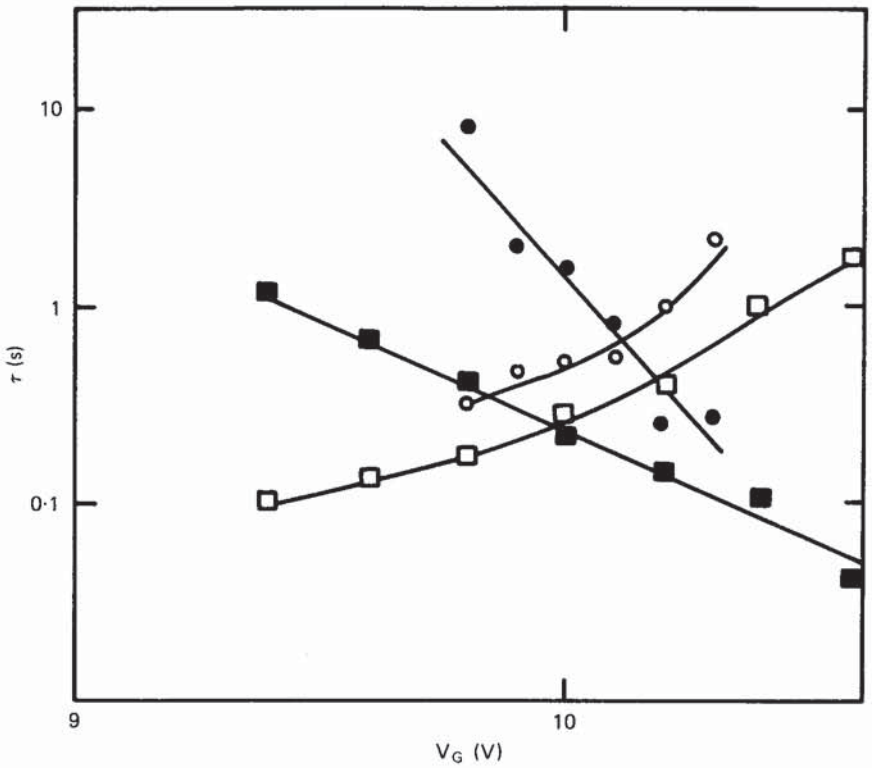

Figure 25. Mean lifetime for capture (solid symbols) and emission (open symbols) *versus* gate voltage for the same trap at two lattice temperatures: ●, ○, 5 K; ■, □, 15 K. After Jackel *et al.* (1985). © Springer, New York. Reproduced with permission.

full curve marked with the filled circles represents the total energy of the defect with its trapped electron.

As discussed by Baraff *et al.* (1980) and Skocpol (1987), the transition from the n-electron to the $(n + 1)$-electron state of the defect will proceed via all states which can supply the extra electron needed: see equation (5.2). Clearly, the actual rate will depend on the probability that a given state is occupied, its overlap with the trapped state, and the activation energy out of the state. The activation energy for capture is given by the energy difference between the minimum of a given broken curve and its intersection with the full curve. It is evident from the figure that this energy shows a weak dependence on electron temperature: as drawn, it decreases slightly as the electron temperature increases. Similarly, the occupation of the higher-lying states increases as the electron temperature increases. These two factors taken together provide a basis for explaining the weak dependence of the capture rate on electron temperature.

Now the activation energy for emission is given by the energy difference between the minimum of the full curve and its intersection with a broken curve. Clearly this energy increases as the filled trap tries to interact with one of the higher-lying electronic states; therefore, the minimum activation energy will dominate and corresponds to re-emission of the electron into an empty inversion-layer state close

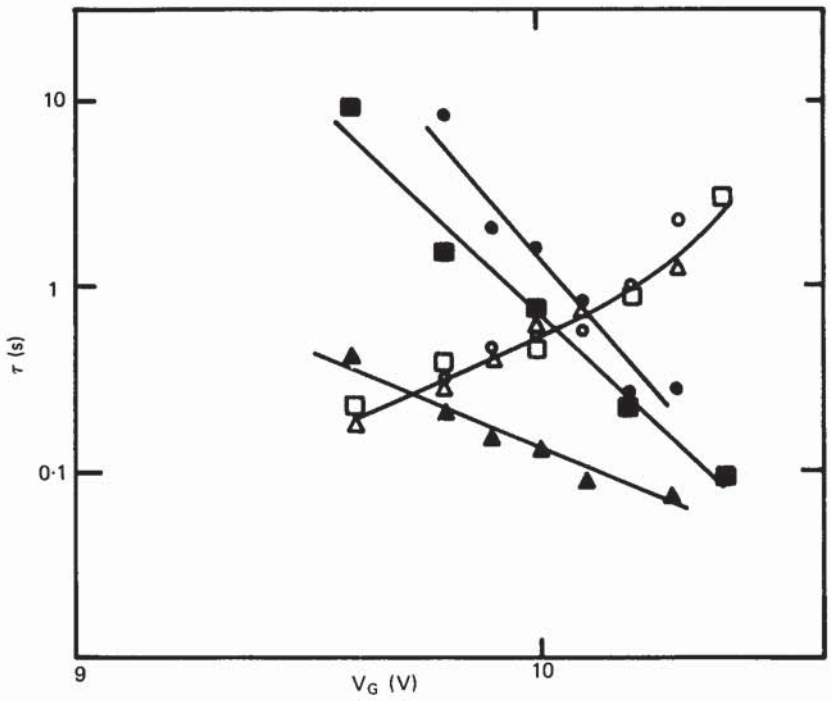

Figure 26. Mean lifetime for capture (solid symbols) and emission (open symbols) *versus* gate voltage for the same trap at different drain–source voltages corresponding to different electron temperatures: ●, ○, 0 mV (about 5 K); ■, □, 4 mV (about 15 K); ▲, △, 12 mV (about 25 K). The emission rate is independent of electron temperature. After Jackel *et al.* (1985). © Springer, New York. Reproduced with permission.

to the Fermi level. This activation energy does not therefore depend on electron temperature, apart from minor effects electron heating may have on the availability of empty states close to the Fermi level. Thus the multiphonon process provides a simple description of the capture and emission processes at both helium and room temperatures. It is evident that further work is called for at cryogenic temperatures. In particular, one obvious experiment will be to investigate the dependence of the measured activation energy for capture on electron temperature in order to test the model of figure 27.

5.1.4. *The gate-voltage dependence of trapping into individual oxide defects in MOSFETs (revisited)*

We should now like to turn our attention again to the behaviour of RTSs as gate bias is varied. In section 3.3.2 we discussed the behaviour of the mark-space ratio of the RTS with gate voltage and considered the arguments for and against the standard two-level RTS being due to multi-electron trapping. Here, we wish to take the single-electron capture model and address the following three areas: the gate-voltage dependence of the capture time; estimates of the trap depth into the oxide for the device operating around threshold; and the behaviour of the emission time with gate voltage.

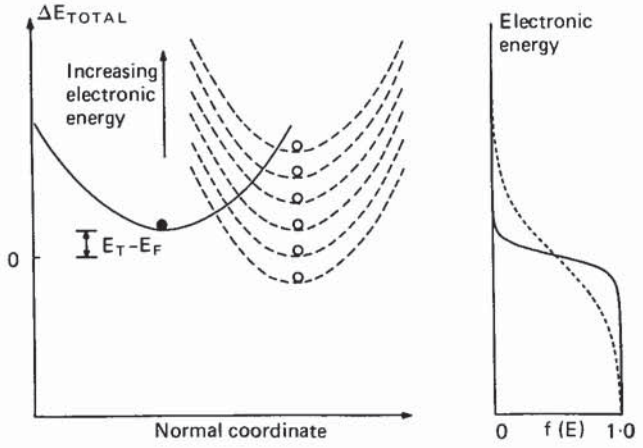

Figure 27. (*a*) Speculative total-energy diagram to explain the results of figures 25 and 26 (see also Skocpol (1987)). The energy zero corresponds to the empty defect with an inversion-layer electron residing at the Fermi level. The series of broken curves represents a range of accessible electron states around the Fermi level. The capture probability depends on the occupation of these initial states, the activation energy out of them, and the overlap with the final state. A representation of the change in initial-state occupancy with electron temperature (source–drain field) is shown in (*b*). As drawn, the average activation energy for capture would appear to decrease. The emission probability is dominated by the activation energy for emission, and this will be determined by the intersection of the full curve with unoccupied states close to the Fermi level. This will show very little dependence on source–drain field.

On the basis of equations (5.2) and (5.3), it would be expected that the change in capture time on increasing gate voltage would come about predominantly from changes in n, the inversion-layer number density (per unit volume). This assumes an invariant cross-section. Equation (5.8) predicts that as the gate bias increases, n is roughly proportional to the current I. Figure 28 shows for one particular RTS that as the gate voltage is changed from 2.25 to 3.5 V the current increases by a factor of 1·8, while there is a corresponding decrease in $\bar{\tau}_c$ by a factor of 15·7. Thus the capture time is decreasing much more rapidly than predicted by equations (5.3) and (5.8). This is quite a general phenomenon and is by no means restricted to the RTS of figure 28. This particular RTS was singled out for detailed investigation because it showed a more marked effect than most. To account for the data of figure 28 on the basis of equations (5.3) and (5.8), it is required that the product μt decrease by a factor of nine over the gate-voltage range 2·25–3·5 V. However, the inversion-layer thickness changes by no more than a factor of two (Ando *et al.* 1982) and the mobility is found to change by 8%. Thus variations in μ and t alone are unable to account for the data presented in figure 28.

We carried out further investigations to measure σ, σ_0 and ΔE_B as functions of V_G (Kirton *et al.* 1989a): the results are shown in figures 29 (*a*)–(*c*). In producing these figures, the formalism of equations (5.13) and (5.22) was used with a classically derived, V_G-dependent inversion-layer thickness. The mobility was taken to be a function of temperature only. Figure 29 (*a*) shows that the very strong dependence of

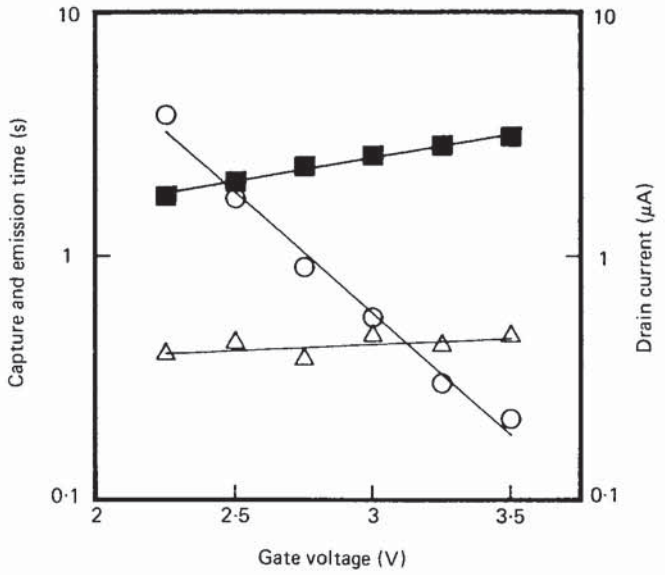

Figure 28. Gate-voltage dependence of the capture (○) and emission (△) times for device H21 at $T = 320$ K. (■) indicates the drain current. From Kirton *et al.* (1989a).

$\bar{\tau}_c$ on V_G is accounted for by a monotonically increasing capture cross-section. On inspecting figures 29 (*b*) and (*c*), we see that this increase in σ is brought about by an increasing σ_0, with ΔE_B effectively constant throughout the range of V_G.

In a bulk solid, where the free carriers uniformly bathe a defect, the definition $\bar{\tau}_c = (\sigma \bar{v} n)^{-1}$ is perfectly sensible. In the case of a defect in the oxide of an MOS structure there are the following two problems: first, the inversion-layer charge is displaced from the defect site; and second, as the gate voltage changes so does the electric field in the inversion layer. As the electric field strength increases, the inversion-layer charge-density peak moves closer to the interface, thus increasing wavefunction overlap. Our original approximation of a uniform inversion-layer charge density becomes less valid. In addition, the changing oxide field strength also lowers the tunnelling barrier. The net result is an increase in the value of σ_0. To quantify this explanation at room temperature will be a very difficult task, but it may be possible from the results of low-temperature studies, where only one sub-band is filled and for which model calculations of the electronic states are more tractable.

As we discussed in section 3.2.2, Ralls *et al.* (1984) showed that by differentiating equation (3.1) with respect to gate voltage and evaluating the change in $E_T - E_F$ and the surface potential ϕ_s for a given change in V_G, the distance d of the trap into the oxide could be easily estimated. We show below that exactly the same result can be obtained from the variation of ΔE_{CT} with gate voltage. The general point that we now wish to make is that while the prescription of Ralls *et al.* produces perfectly sensible values for d (e.g. up to 2 nm) when the device is operating in strong inversion (and in very weak inversion), anomalously large values (e.g. up to 20 nm) are obtained when the same analysis is applied to traps active around threshold, see table 3.

To understand how this comes about, consider figures 30 (*a*) and (*c*), which show the variation in the measured value of ΔE_{CT} and the calculated value of ϕ_s for two

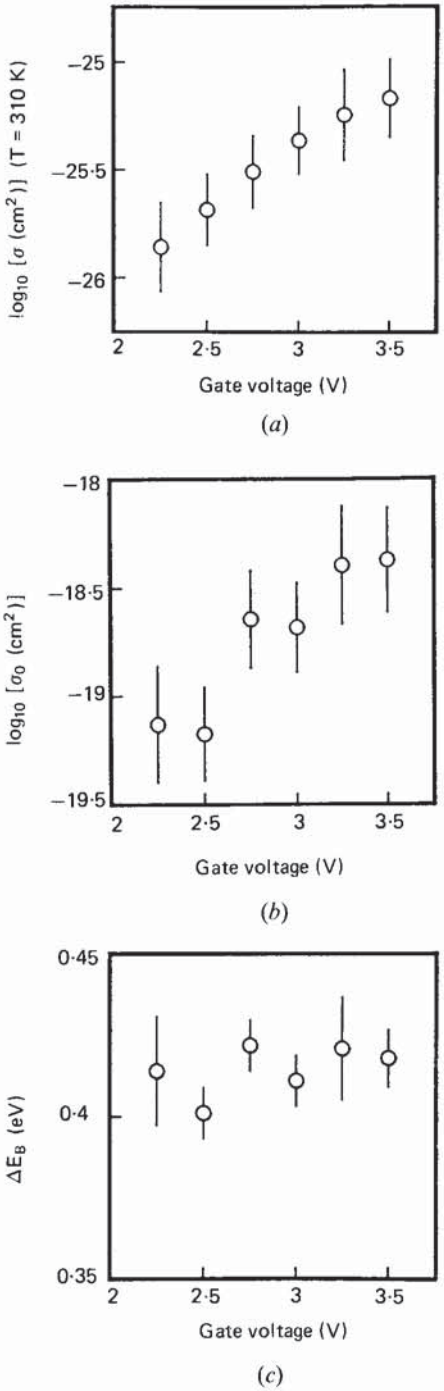

Figure 29. Gate-voltage dependence of cross-section σ, cross-section pre-factor σ_0 and capture activation energy ΔE_B for device H21. After Kirton *et al.* (1989a).

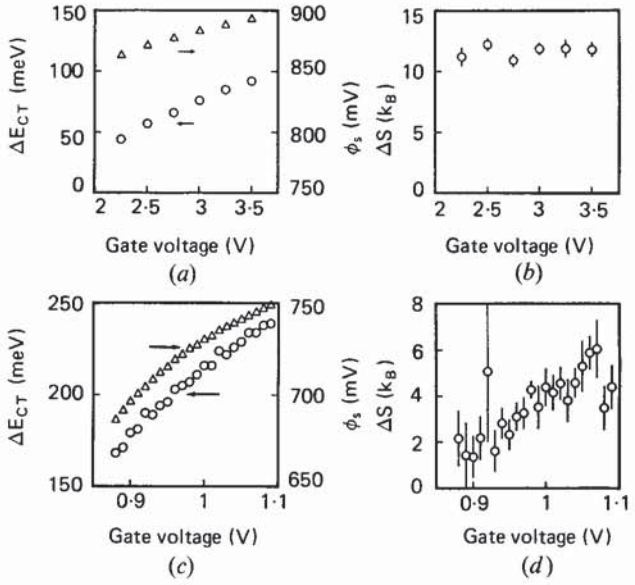

Figure 30. (*a*) and (*c*) show the gate-voltage dependence of trap energy level ΔE_{CT} and surface potential ϕ_s; (*b*) and (*d*) show the gate-voltage dependence of trap entropy ΔS. (*a*) and (*b*) are for device H21 in strong inversion and (*c*) and (*d*) are for device H15 around threshold. V_T = 0·85 V. After Kirton *et al.* (1989a).

defects active in strong inversion and around threshold respectively. ΔE_{CT} represents the trap binding energy with respect to the energy of the charge centroid of the inversion layer. From figure 30 (*a*), we can see that for the strong inversion trap for a δV_G of 1250 mV, ϕ_s changes by 30 mV and ΔE_{CT} by 48 meV. In strong inversion the field is fully screened by the inversion layer, so the potential change at the charge centroid is $\frac{1}{2}\,\delta\phi_s$ (Brews 1978). Thus if the trap were located directly at the interface, and assuming the trap energy level moved rigidly with changing electrostatic potential, then ΔE_{CT} would change by 15 meV ($=q\,\delta\phi_s - \frac{1}{2}q\,\delta\phi_s$). The increase in ΔE_{CT} above this value is due to the trap being located in the oxide such that it experiences the changing oxide field strength. With an oxide thickness t_{ox} of 40 nm, $d = t_{ox}\,(48 - 15)/(1250 - 30) = 1{\cdot}1$ nm—a perfectly reasonable value. Around threshold and below, where there is little screening by the inversion layer, the potential at the charge centroid changes at nearly the same rate as at the surface. So ΔE_{CT} for a trap at the interface would be very small and, applying the same analysis as above to the data for the trap of figure 30 (*c*), we find $d = 71t_{ox}/(210 - 63) = 19{\cdot}3$ nm—a totally unreasonable value if there is to be a finite chance of tunnelling to the inversion layer. The conclusion to be drawn is that in the case of traps observed near the gate threshold voltage, the trap binding energy ΔE_{CT} is changing as a function of gate voltage over and above what one would expect from simple electrostatics alone.

The trap binding energy ΔE_{CT} is calculated from the relation $\Delta E_{CT} = \Delta H_{CT} - T\,\Delta S$: ΔH_{CT} is the trap binding enthalpy and is obtained from the difference in the gradients of the emission and capture data; ΔS is the change in trap entropy on removing an electron from the defect and placing it in the inversion layer. Figures 30 (*b*) and (*d*) show that for the defect active in strong inversion ΔS is constant over a gate voltage range of 1·25 V, while for the threshold trap ΔS changes by a factor of about

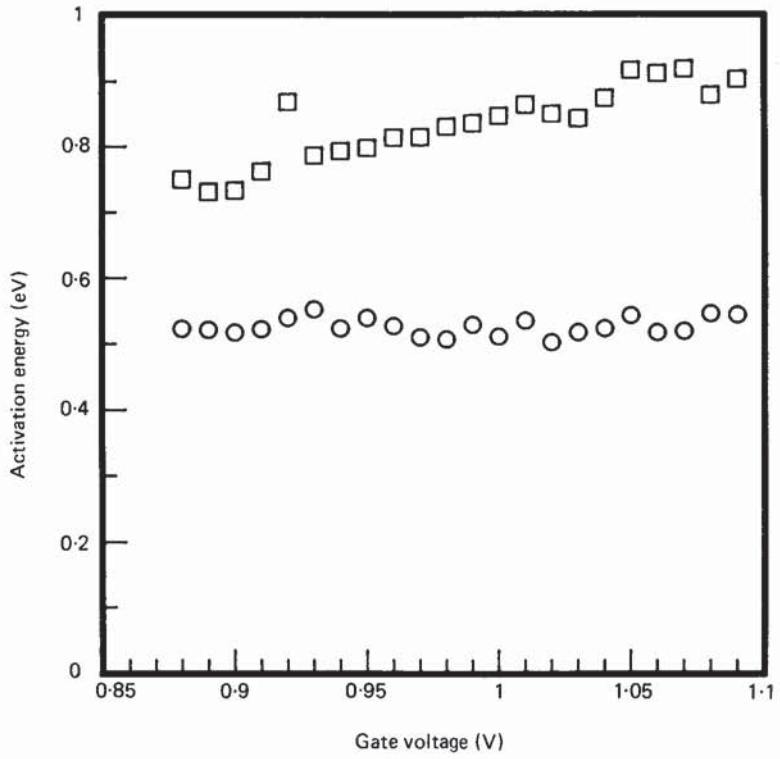

Figure 31. Measured activation energies against gate voltage for trap in device H15 operating around threshold: ○, activation energy for capture, ΔE_B; △, activation energy for emission, $\Delta H_{CT} + \Delta E_B$.

2·5 for a δV_G of 0·21 V. Figure 31 shows the strong dependence on gate voltage of the emission activation energy. Since ΔH and ΔS are recovered from temperature-dependent data obtained at fixed gate voltage, each individual value is little affected by surface-potential fluctuations. However, these fluctuations are being screened out by increasing gate voltage; in strong inversion they are fully screened by the inversion-layer charge (Brews 1975a, b). Around threshold, one might therefore expect to see a change in the activation energy for capture ΔE_B, as the fluctuations are minimized. The measured emission activation energy is the sum of ΔE_B and ΔH_{CT}. It is then possible that a steadily increasing (or decreasing) ΔE_B could account for the larger than expected variation in emission activation energy with gate voltage. For device H15, figure 31 shows that there is no obvious trend in ΔE_B as gate voltage is varied. Thus it seems very unlikely that surface-potential fluctuations can be invoked to explain the results of figures 30 (*c*) and (*d*).

The key to understanding this data probably lies in a consideration of the changes that take place in the local environment of the trap as the gate voltage is increased. In strong inversion, the electron number density (per unit area) is high—typically $10^{12}\,\mathrm{cm}^{-2}$—and the trapped charge is fully screened by the inversion-layer charge. However, around threshold, the image charge of the trapped electron is shared between the gate, the channel and the depletion region, and the number density in the channel is very sensitive to changes in gate bias. Thus the location of the image charge

is changing rapidly and so constitutes a rapidly changing local environment for the defect. This gives rise to modifications in the bonding and dynamical properties of the trap, which are reflected in V_G-dependent enthalpy and entropy terms respectively. A further test of this model would be to fully investigate the behaviour of d for traps operating in very weak inversion. Under such operating conditions, the trapped electron is screened by the gate and depletion-layer edge. When V_G changes, the inversion-layer charge is still too low to modify the screening field, and it would be expected that ΔH and ΔS would show no anomalous behaviour. Table 3 shows that for the weak inversion trap in device G8, d has a value 1·3 nm, which is encouraging. However, we have as yet been unable to investigate in detail many RTSs in this regime.

A consequence of the effects we have just described is that $\bar{\tau}_e$ becomes gate-voltage dependent. As shown in figure 28, $\bar{\tau}_e$ rises only very slowly as the gate bias is increased. In general, $\bar{\tau}_e$ is found to show all variations from none at all to a sensitivity equal and opposite to that exhibited by $\bar{\tau}_c$. Using equations (3.1) and (5.3) and the fact that $n = N_C \exp[-(E_C - E_F)/kT]$, where N_C is the effective density of states in the conduction band, one finds

$$\frac{\bar{\tau}_e(V_{G2})}{\bar{\tau}_e(V_{G1})} = \frac{\sigma(V_{G1})}{\sigma(V_{G2})} \exp\left[\frac{\Delta E_{CT}(V_{G2}) - \Delta E_{CT}(V_{G1})}{kT}\right]. \tag{5.23}$$

As we have already seen, $\sigma(V_{G2}) > \sigma(V_{G1})$ and $\Delta E_{CT}(V_{G2}) > \Delta E_{CT}(V_{G1})$ for $V_{G2} > V_{G1}$. The behaviour of $\bar{\tau}_e$ depends on how these opposing effects cancel out. One would anticipate that the exponential term would generally dominate, corresponding to all of our observations up to this point. However, if the σ term were to dominate then $\bar{\tau}_e$ would decrease as V_G increased.

5.2. *Charge-carrier trapping into individual defects in the insulator of metal–insulator–metal tunnel junctions*

In section 4.1 we described the work of Rogers and Buhrman (1984) on the decomposition of $1/f$ noise in Nb_2O_5 tunnel junctions into its individual fluctuating components. Here we wish to describe their investigations into the nature of the dynamics of these fluctuations (Rogers and Buhrman 1985).

In sufficiently small-area devices the capture of a single electron in the insulator gives rise to a discrete and measurable change in the tunnel resistance. As for the case of carrier trapping at Si/SiO_2 interface states, the process can be described by the two time constants $\bar{\tau}_c$ and $\bar{\tau}_e$, representing the time for electron capture (trap filling) and electron emission (trap emptying) respectively. In the frequency domain the power spectrum of a resistance change ΔR is given by (see equation (2.19 *a*))

$$S_R(f) = \frac{4(\Delta R)^2}{(\bar{\tau}_e + \bar{\tau}_c)[(1/\bar{\tau}_e + 1/\bar{\tau}_c)^2 + (2\pi f)^2]}. \tag{5.24}$$

The total integrated power S_I (equation (2.20)) is then

$$S_I = \frac{(\Delta R)^2}{(\bar{\tau}_e + \bar{\tau}_c)(1/\bar{\tau}_e + 1/\bar{\tau}_c)}. \tag{5.25}$$

Writing $(\Delta R)^2 = S_0$ and

$$\frac{1}{\tau_{eff}} = \frac{1}{\bar{\tau}_e} + \frac{1}{\bar{\tau}_c} \tag{5.26}$$

so as to follow the notation of Rogers and Buhrman, we find

$$S_1 = \frac{S_0 \tau_{\text{eff}}}{\bar{\tau}_e + \bar{\tau}_c} = \frac{S_0 \bar{\tau}_e \bar{\tau}_c}{(\bar{\tau}_c + \bar{\tau}_c)^2}. \tag{5.27}$$

From the roll-off frequency (section 2.3) of the Lorentzian power spectrum of the fluctuations, Rogers and Buhrman were able to determine τ_{eff} for a given set of operating conditions. Above temperatures of about 15 K, they found $1/\tau_{\text{eff}}$ to increase in a thermally activated manner; at the same time S_1 varied much more slowly with temperature: see figure 32. The behaviour is consistent with both $\bar{\tau}_e$ and $\bar{\tau}_c$ being thermally activated, see equations (5.26) and (5.27). They chose to write the temperature dependence of $\bar{\tau}_c$ and $\bar{\tau}_e$ in the form

$$\frac{1}{\bar{\tau}_i} = \frac{1}{\tau_{0i}} \exp\left(-\frac{E_{Bi}}{kT}\right). \tag{5.28}$$

Using equations (5.26)–(5.28), they obtained a fit to the data shown as the solid lines in figure 32. From the fits, they recovered values for the activation energies E_{Bi}, attempt frequencies τ_{0i} and the resistance change S_0.

A particularly interesting and important result obtained by Rogers and Buhrman is the temperature dependence of τ_{eff} at very low temperatures, as shown in figure 33. The key feature to note is that below about 15 K the kinetics of charge trapping and de-trapping become very weakly dependent on temperature. Particular care was taken to ensure that the effect was real and not due to some local heating effect.

The model developed by Rogers and Buhrman to explain the data shown in figures 32 and 33 is shown in the inset of figure 32. Essentially, the point they make is that the kinetics of capture and emission are governed by ionic reconfigurations at a single point defect. Thus the inset shows the total energy of the defect for two different geometrical configurations of the atoms surrounding the defect site. They noted that the figure bore a strong resemblence to the two-well tunnelling configurations invoked by Anderson *et al.* (1972) and Phillips (1972) to explain the anomalous low-temperature specific heat capacity of glassy materials. (These configurations will be discussed in more detail in section 6 when we consider individual defect metastability.) They then argued that while the defect is in the first total-energy minimum it is unable to capture an electron; but that on transfer to the second minimum, capture immediately takes place. Thus the two wells shown in figure 32 correspond to different total charge.

The temperature dependences of the data can then be understood as follows. At high temperatures, transfer between minima occurs via thermal activation over the barrier; but at low temperatures tunnelling will dominate. Since ionic transfer is considered to be the rate-limiting process, τ_{eff} would be expected to show both thermally activated and weakly temperature-dependent regions. Furthermore, since in the barrier region there are very strong electrochemical gradients, Rogers and Buhrman consider that displacement of the defect by a distance d would dramatically affect its occupancy; thus causing instantaneous electron capture or emission.

The model represented in the inset to figure 32 is highly schematic. A more complete description would show the total energy of the defect, which incorporates electronic energy plus elastic energy, plotted against some general reaction coordinate describing the displacements of the atoms in the neighbourhood of the defect

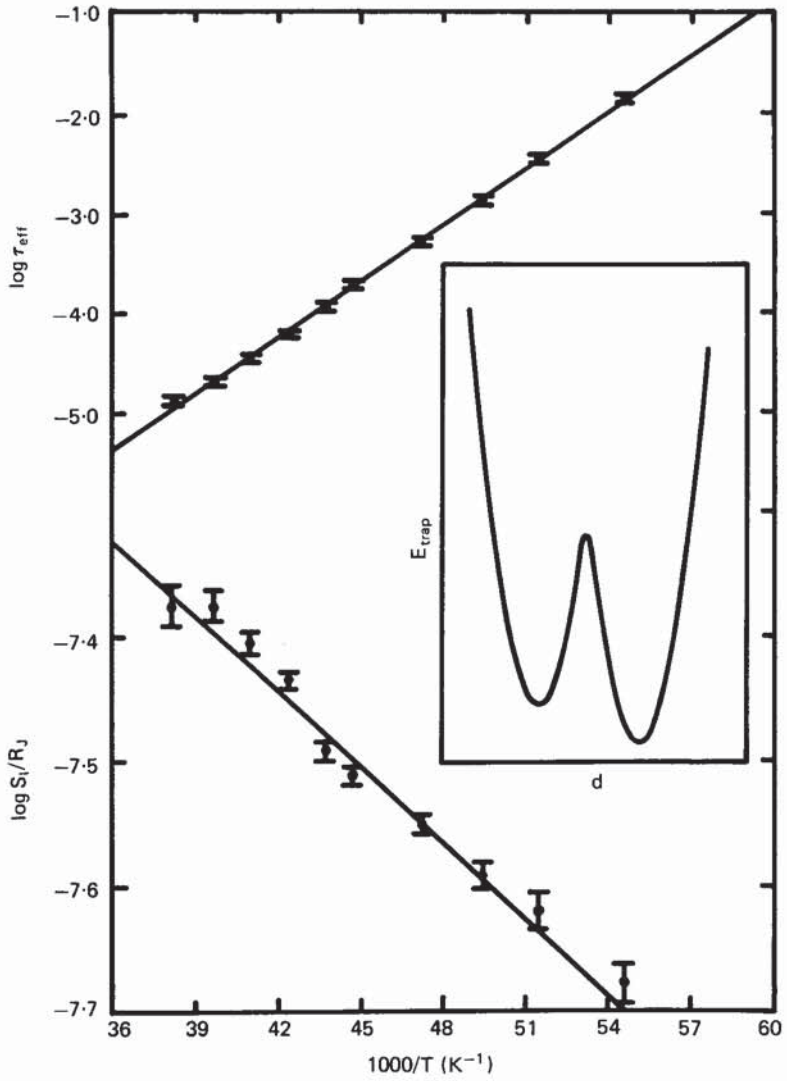

Figure 32. Plot of log τ_{eff} and log S_1 against $1/T$, showing thermally activated behaviour. Solid lines are fits assuming equations (5.27) and (5.28). The inset shows a two-well model describing the rate-limiting kinetics. From Rogers and Buhrman (1985). © American Physical Society. Reproduced with permission.

(Stoneham 1975). Rogers and Buhrman interpreted the displacement d quite literally: they tried to correlate it with the movement of an oxygen vacancy by a distance d through the tunnel junction and found a reasonable value of approximately 0·3 nm. In fact, d describes the change in some generalized reaction coordinate as the defect flips from one minimum-energy configuration to another. In the construction of total-energy diagrams of the various charge states of defects, careful account has to be taken of the energy of particles either added to or removed from the Fermi reservoir: see for example Baraff *et al.* (1979) and Uren *et al.* (1988).

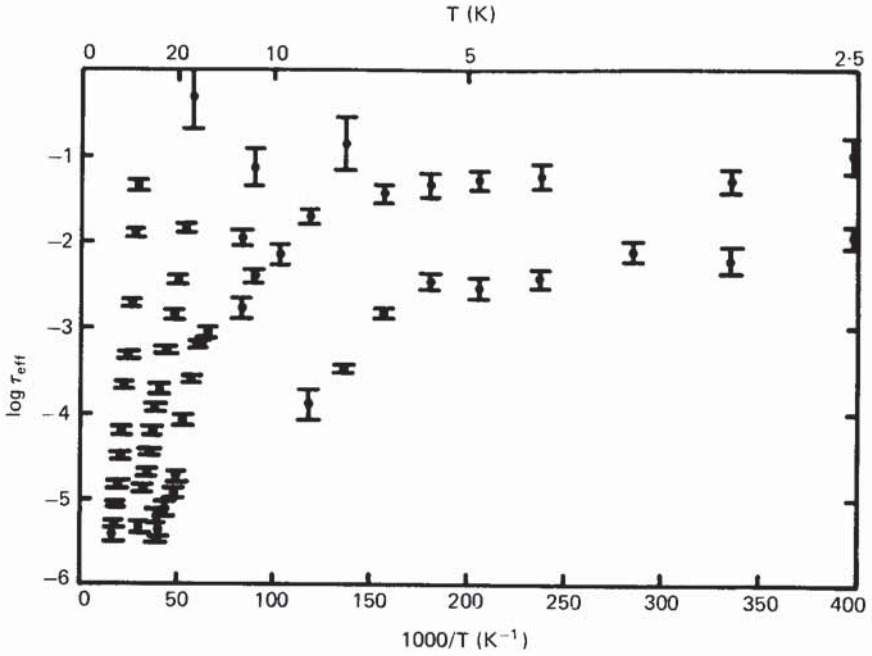

Figure 33. Typical data set for τ_{eff}, showing the abrupt change from thermally activated behaviour above to non-activated behaviour below $T \approx 15\,K$. From Rogers and Buhrman (1985). © American Physical Society. Reproduced with permission.

The configuration-coordinate diagram of figure 23 may provide an alternative way of interpreting Rogers and Buhrman's data without the need to invoke defect metastability. If we let the total-energy curve enclosing the open circle represent the total energy of the empty defect with a free electron residing at the Fermi reservoir of the emitting contact then the curve marked with the filled circle represents the variation of the total energy of the defect plus a trapped electron. At high temperatures, for electron capture to take place, both electron tunnelling and lattice vibrations are required; thus the capture time is thermally activated. At low temperatures the multiphonon capture process becomes only weakly dependent on temperature, see for example Henry and Lang (1977), Ridley (1978) and Bourgoin and Lannoo (1983), in accord with the findings of Rogers and Buhrman. The barriers for capture and emission will also depend strongly on the applied bias across the junction, which is also consistent with the experimental findings.

6. Complex telegraph noise

So far, we have discussed what might conveniently be called normal RTS behaviour. By this, we mean that the RTS resulting from single-electron capture and emission events is well behaved; the up and down times obey Poisson statistics; and the RTS itself is stable for times extending to $\pm\infty$. We have shown that $1/f$ noise in, for example, MOSFETs and MIM structures is composed of a summation of these RTSs (two-level systems).

In this section we wish to consider recent investigations that have found significant departures from this conventional behaviour. We begin (in section 6.1) with a discussion

of the work of Restle *et al.* (1985, 1986). Through a careful statistical analysis of the noise in small silicon-on-sapphire (SOS) and GaAs resistors, these workers found significant departures from Gaussian statistics and were able to make remarkably accurate predictions concerning the origin of the effects. Following on from this (sections 6.2, 6.3 and 6.6), we shall discuss the small fraction (about 4%) of RTSs in MOSFETS that exhibit anomalous and complex switching behaviour. In section 6.4 we shall describe the work of Farmer *et al.* (1987, 1988) on complex RTS behaviour in tunnelling currents through small-area MOS structures.

$1/f$ noise in metals is generally outside the scope of this review, and the interested reader is referred to the work of Weissman (1988). However, recent work by Ralls and Buhrman (1988) on the noise in metallic microstructures is of such importance that we have included a short section (section 6.5) devoted to a cursory survey of the most important features of noise in metals and the implications of their findings.

It is worth noting that complex burst noise was observed in commercial electronic devices some time ago: see for example Knott (1978, 1980). Figure 34 shows the switching recorded in the base current of a PNP bipolar transistor. Three major current levels are present, and a small amplitude RTS waveform is present at all these levels. The high-frequency noise is greatest at the intermediate level. K. Knott (1988, personal communication) has observed even more complex signals in which the rise time of the RTS contains discrete structure. While dislocations have been implicated in its production, burst noise remains poorly understood. It has also been little investigated, despite the fact that it may provide unique insight into the complex behaviour of defects in the solid state.

6.1. *Non-Gaussian noise phenomena in small silicon-on-sapphire and gallium arsenide resistors*

We have already briefly touched upon the work of Restle *et al.* (1985) in section 4.2 in connection with the generation of $1/f$ noise from a summation of two-level systems. Here we wish to discuss a particularly interesting aspect of their work: namely, their statistical analyses of the noise in testing for non-Gaussian processes.

By a Gaussian process, we are referring to one in which the probability distribution of amplitudes is normal. Such processes are completely specified by the mean and covariance function $\langle x(t_1)x(t_2)\rangle$. The importance of Gaussian behaviour lies in its implications for the stationarity of the noise source: for a statistically stationary process (in the wide sense), the mean is independent of time and the covariance depends only on the difference between observation times. In principle, one cannot verify that a particular noise pattern is Gaussian without checking on an infinite number of higher-order correlation functions. In practice, a few statistical properties are picked on the assumption that non-Gaussian behaviour would show up in those properties if it were present. The simplest test is the distribution of noise voltage itself; it has been found to be Gaussian in all but a few cases. The model of $1/f$ noise that we have advocated thus far, namely, a superposition of a large number of parallel, independent noise generating events with a spread of characteristic times, satisfies the assumptions of the central-limit theorem and thus predicts Gaussian noise (Nelkin and Tremblay 1981).

Restle and co-workers developed sensitive statistical tests with a view to testing for non-Gaussian phenomena. The techniques are described well in their papers; here we shall cover only the bare essentials of the method. The basic idea was to divide the

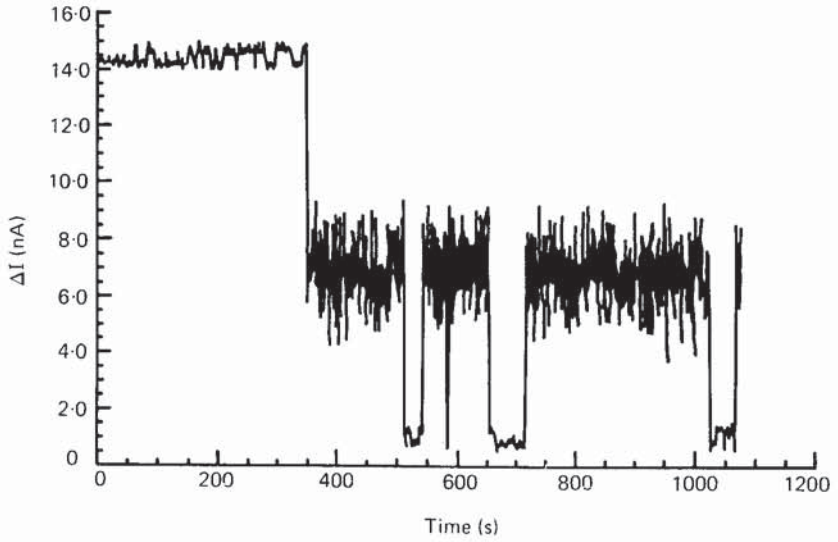

Figure 34. Current fluctuations measured in the base current of a PNP bipolar transistor (2N3799); $T = 115$ K, base current 650 nA. Figure courtesy of K. Kandiah.

noise power into octaves and to evaluate the correlation coefficient $r_{i,j}$ between noise power in octaves i and j. In this way, a covariance matrix was evaluated. When suitably normalized, the diagonal elements of the matrix should be equal to unity and the off-diagonal elements should tend to zero for a Gaussian noise process.

Table 4 shows the covariance matrix obtained from a Monte Carlo simulation of 30 random two-level systems. (The power spectrum corresponding to this simulation is shown in figure 19.) Extra variance shows up in those regions in which the spectrum arises from a small number of systems with characteristic frequencies below the observed band. Since with 30 RTSs the requirements of the central-limit theorem are only barely justified, the average diagonal term is about 8% higher than for Gaussian noise and the average inter-octave correlation coefficient is 0·07, or about 0·11 for adjacent octaves.

Table 4. Covariance matrix for the Monte Carlo simulation, which appears as dots in figure 19. From Restle *et al.* (1985). © American Physical Society. Reproduced with permission.

Octave	1	2	3	4	5	6	7	8	9
1	0·88								
2	0·04	1·07							
3	0·02	0·11	1·01						
4	0·04	0·06	0·05	0·99					
5	−0·01	0·00	−0·02	0·03	1·03				
6	−0·03	0·00	−0·01	0·04	0·06	1·12			
7	0·02	−0·02	0·03	−0·01	0·07	0·06	1·13		
8	−0·03	−0·01	−0·05	0·04	0·01	0·03	0·11	1·27	
9	0·04	−0·03	0·04	−0·02	0·04	0·01	0·09	0·11	1·36

Table 5. Normalized covariance matrix for μSOS 7 measured from 1000 transforms. The frequency range covered is 53 Hz–4·8 kHz. From Restle *et al.* (1985). © American Physical Society. Reproduced with permission.

Octave	4	5	6	7	8	9	10
4	1·14						
5	0·16	1·10					
6	0·09	0·10	1·14				
7	0·09	0.09	0·10	1·17			
8	0·00	0·07	−0·01	0·07	1·10		
9	0·10	0·05	−0·02	0·07	0·11	1·09	
10	−0·03	−0·07	−0·03	0·02	0·06	0·10	1·00

Most small silicon-on-sapphire (SOS) samples that were investigated showed small reproducible deviations from Gaussian behaviour: see the results for μSOS 7 in table 5. For the small non-Gaussian effects the terms of the covariance matrix were independent of sampling rate, consistent with the computer simulations for two-level systems. However, two samples, namely μSOS 1 and μSOS 3, exhibited extremely non-Gaussian noise (table 6). In addition, the normalized noise variance was found to have a sampling-rate dependence not consistent with the simple two-level system model: the normalized variance of the noise power in an octave increased monotonically with the number of frequency bins of the FFT contained in the octave (figure 35), that is, with the length of time taken to make the measurement. The expectation of the normalized variance would be 1·0 for Gaussian noise, and for non-Gaussian noise due to a small number of two-state systems the normalized variance would not increase with the number of bins per octave.

This dependence of normalized power on number of bins implies a slow variation of the power with time. Figure 36 shows the results obtained by Restle *et al.* for the integrated noise power in a given octave as a function of time obtained from a Johnson noise source μSOS 7, which is approximately Gaussian, and μSOS3, which shows non-Gaussian behaviour. It is clear from this figure that μSOS 3 shows significant low-frequency amplitude modulations. Restle *et al.* concluded that, while their results broadly confirmed the model involving superposition of two states, the amplitudes of some prominent two-state (RTS) components were not constant in time but showed modulation, with these modulations having a messy spectrum of a $1/f$ form. The likely origin of the modulations was thought to be slow structural reconfigurations in the glassy SiO_2.

Restle and co-workers extended their investigations to small-area gallium arsenide resistors (Restle *et al.* 1986). As in the SOS samples, they found spectra that were essentially composed of a superposition of two-state systems. In two of the smaller samples, after certain surface treatments, they discovered that the noise became very non-Gaussian with the variance in the noise power per octave as much as eight times the variance of the Gaussian noise. The noise power in a given octave band in these samples was measured as a function of time to generate a second spectrum: see figure 37. The second spectrum shows that the low-frequency power modulations responsible for the increased variance of the noise power have an approximately $1/f$ power spectrum. Thus one can identify two $1/f$ noise processes: the first is due to a simple summation of single two-state processes as exemplified by the RTS; and the second is due to a modulation of those individual two-state sites. Restle *et al.* (1986) suggested that the second mechanism was caused by individual sites being prevented

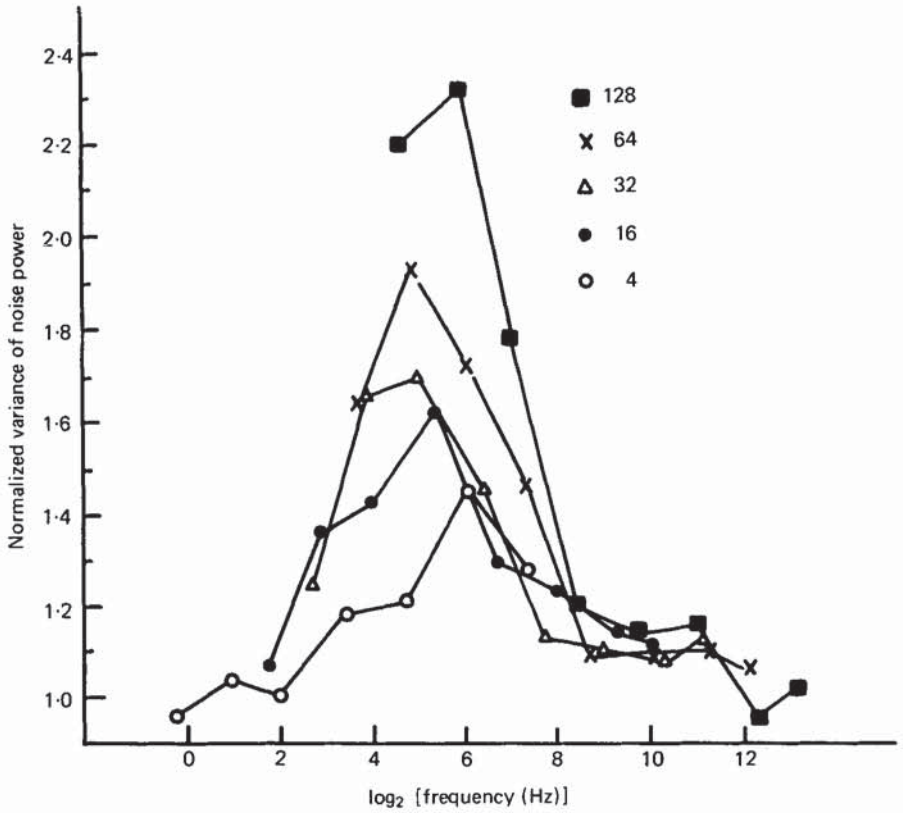

Figure 35. Each line in the figure shows the normalized variance of the noise power of μSOS 3 as a function of frequency, with the number of frequency bands per octave held constant. The different points on a single line were obtained by changing the sampling rate. The line with the highest peak corresponds to data in which 128 frequency bins were summed to measure the power in an octave. ■, 128 bins octave^{-1}; ×, 64; △, 32; ●, 16; ○, 4. From Restle *et al.* (1985). © American Physical Society. Reproduced with permission.

Table 6. Normalized covariance matrix for μSOS 1 measured from 1000 transforms. The frequency range covered is 0·5–45 Hz. From Restle *et al.* (1985). © American Physical Society. Reproduced with permission.

Octave	4	5	6	7	8	9	10
4	0·96						
5	0·03	1·02					
6	0·02	0·05	1·14				
7	0·06	0·05	0·05	1·14			
8	0·03	0·02	0·05	0·17	1·22		
9	0·02	0·04	0·07	0·12	0·26	2·01	
10	0·01	−0·02	0·12	0·14	0·25	0·37	1·66

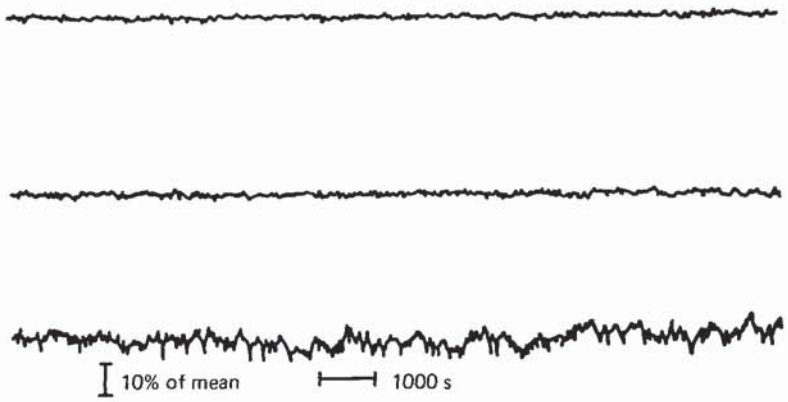

Figure 36. The three traces show the RMS noise in about one octave as a function of time. The top trace was taken from a Johnson noise source, the middle from the approximately Gaussian μSOS 7 and the bottom from the non-Gaussian μSOS 3. Each source was adjusted to about the same mean RMS noise. Low-frequency amplitude modulations are evident in μSOS 3. From Restle *et al.* (1985). ©American Physical Society. Reproduced with permission.

from fluctuating for approximately half of the time. In section 6.2 we shall verify the remarkable accuracy of this prediction when we show the existence of this effect on a single RTS from a silicon MOSFET.

6.2. *Complex telegraph noise in MOSFETs (I): Two-level signals*

In figure 38 we depict the unique switching behaviour found in the drain current of a small-area MOSFET by Uren *et al.* (1988). This complex behaviour was found in a small proportion of RTSs: in a search in which 320 RTSs were observed, 12 were found to be anomalous. Within the time window from t_1 to t_2 the rapidly-switching RTS shows the conventional behaviour corresponding to fluctuations in occupancy of an individual interface defect. As we have previously seen, most RTSs have precisely this form for times extending to $\pm\infty$; the times in the up and down states correspond to single-electron capture and emission respectively. During periods such as that from t_2 and t_3 shown in figure 38 the rapid switching completely disappears and the RTS maintains its low level. It thus appears that the fast-switching RTS is modulated in time, with the envelope of modulation itself being an RTS.

The initial observations of this phenomenon were in two devices operating in weak inversion, indicating that the likely cause was either two traps on the same percolation channel or Coulombic interactions between pairs of traps in close proximity (spatially and energetically). The percolation model is restricted to the weak-inversion regime, since it requires inhomogeneous current flow due to partially screened charge centres at the interface. This model was quickly discarded since the majority of these complex signals were subsequently observed in strong inversion where percolation is not a major current-carrying mechanism.

The Coulombic model requires two traps with relatively fast and slow time constants to be located close together and with both their occupancy levels residing close to the Fermi level. The rapid switching shown in figure 38 is then due to the

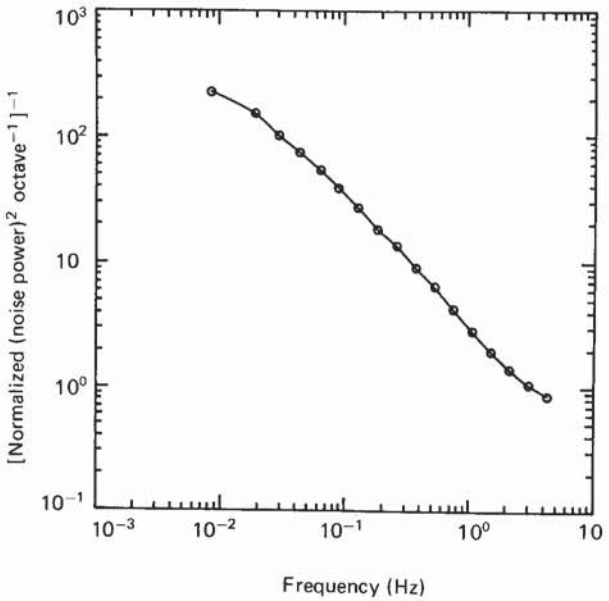

Figure 37. Second spectrum taken on the noise in an octave band centred at 10 kHz in GaAs sample no. 2. The Gaussian prediction of 1·0 has been subtracted. The $1/f$ form of the spectrum was stable in time and did not depend on which octave of the ordinary spectrum was chosen for analysis. After Restle *et al.* (1986). © American Physical Society. Reproduced with permission.

fluctuations in occupancy of the fast trap. On electron capture into the slow trap, the occupancy level of the fast trap will be lifted higher in energy by an amount that can be estimated from simple electrostatics. If the fast trap's occupancy level is moved by several kT, the rapid fluctuations will cease and the current will remain at its low level, as in the period from t_2 to t_3, until such time as the slow trap releases its trapped electron and the fast trap is free to fluctuate in occupancy again.

A variant of the Coulombic model was proposed by B. Dierickx (1987, personal communication). He pointed out that such signals could arise from two closely spaced defects lying at different distances into the oxide. Carriers would be unable to tunnel directly to the deeper defect on accessible time scales, and charge transfer would take place from the shallower trap. On occupation of the deep defect, all fluctuations of the faster trap would cease. It should be noted that this charge transfer must involve lattice relaxation as well as tunnelling, since time periods such as that from t_2 to t_3 are found to be strongly thermally activated.

The key feature of the anomalous RTS in figure 38 that these Coulombic models must explain is the presence of two and not three or four current levels. During those periods in which the slow trap has captured an electron no observable fluctuations of the fast trap take place, otherwise three current levels would be present. In general, no third level was seen (but see sections 6.3 and 6.6). This requires the following set of conditions: either the channel in the immediate vicinity of the fast trap is completely blocked (inversion charge excluded), so that the fluctuations are still occurring but are not observable; or else the occupancy level of the fast trap is moved several kT away from the Fermi level so the trap's occupancy does not fluctuate.

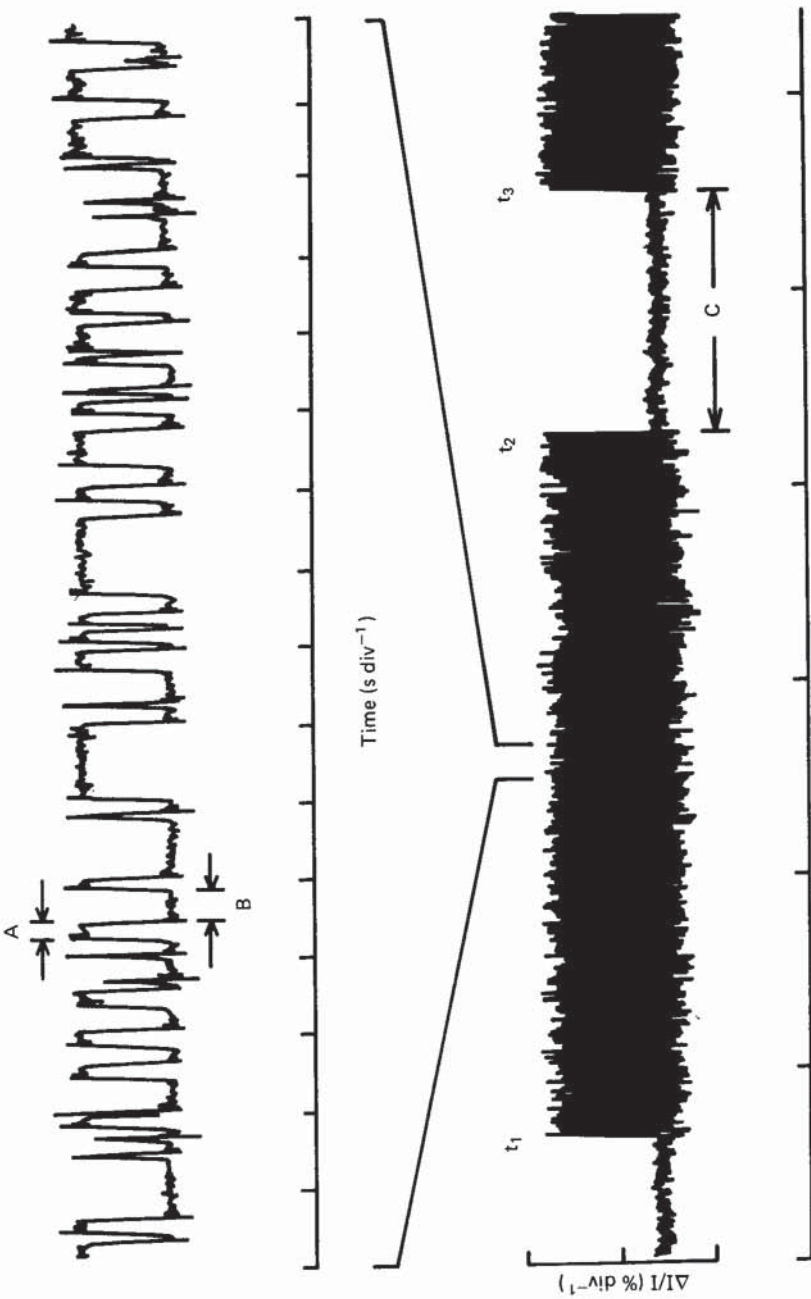

Figure 38. Fluctuations in current against time, showing a rapidly switching RTS modulated by an envelope of the same amplitude. The upper trace is an expansion of part of the lower trace. $V_G = 1{\cdot}0$ V, $V_D = 10$ mV, $I_D = 10{\cdot}3$ nA and $T = 293$ K. From Uren *et al.* (1988).