Chapter 2
MOSFET Operation

The APV25 readout chip is manufactured in a commercial 0.25 μm CMOS technology. To understand how the APV25 will perform in the CMS Tracker environment some knowledge of the operation of the Metal Oxide Semiconductor Field Effect Transistor (MOSFET), the basic building block of the chip, is required. This chapter introduces the properties of the MOSFET and explores how radiation and temperature effects can change its behaviour.

2.1 MOS Device Physics

The MOSFET is essentially a MOS capacitor with two p-n junctions placed adjacent to the channel. CMOS processes use silicon (Si) technology since it is the most mature among all semiconductor technologies. In addition, silicon is easily available, making it the cheapest semiconductor material. It can be doped with impurity atoms, changing its properties. For example, if an arsenic atom with five valence electrons takes the place of one Si atom in the lattice, four of its valence electrons form covalent bonds with surrounding Si atoms and the fifth electron is essentially free for conduction. The silicon is said to be n-type and the arsenic atom is called a donor. When a boron atom with three valence electrons takes the place of a silicon atom, all three valence electrons form covalent bonds and the absence of a fourth electron creates a hole. This is a p-type semiconductor, and the boron atom is an acceptor atom.

The energy at which the probability of finding an electron in a given electronic state is one half is called the Fermi level, E_F. In an n-type (p-type) semiconductor, E_F is closer to the conduction (valence) band. In a very pure slab of silicon the Fermi level is exactly half way between the valence and conduction band. The silicon is in its intrinsic state and its Fermi level is called the intrinsic Fermi level, E_i. The corresponding intrinsic carrier density is n_i ($n_i = 9.65 \times 10^9\ \mathrm{cm}^{-3}$ at room temperature [11]).

2.1.1 Charge Carrier Transport

The motion of charge carriers in MOS devices is the result of two basic transport phenomena, carrier drift and carrier diffusion.

Carrier drift arises in the presence of an electric field. Taking the case of an n-type semiconductor as an example, the charge carriers are electrons. With no applied field,

the thermal motion of an electron can be visualised as a succession of random scattering from collisions with lattice and impurity atoms and other scattering centres. Over a sufficiently long period of time, the net displacement is zero. The average time between collisions is called the mean free time, τ_c. With an applied electric field, ξ, the electron experiences a force of $-q\xi$ and is accelerated in the opposite direction to the field. The additional velocity component introduced is called the drift velocity, v_n. Conservation of momentum leads to the following expression for v_n:

$$v_n = -\left(\frac{q\tau_c}{m_n}\right)\xi \qquad \text{..... eq. (2.1).}$$

where m_n is the effective electron mass which takes into account the influence of interactions with the positive ion cores. The factor $\frac{q\tau_c}{m_n}$ is called the mobility, μ_n (units cm^2/Vs). Various scattering mechanisms contribute to the total mobility of which the most important are lattice scattering from thermal vibrations, μ_L, and impurity scattering, μ_I, when a charge carrier travels past an ionised impurity. The temperature and doping dependence of mobility will be discussed later. The total mobility is given by:

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I} \qquad \text{.....eq. (2.2).}$$

Carrier diffusion arises from the movement of charge carriers from a region of high concentration to a region of low concentration, resulting in a diffusion current. The net current is a combination of drift and diffusion components.

2.1.2 The p-n Junction

The p-n junction is a diode since it allows current to flow only in one direction. It is formed by bringing into contact doped p- and n-type semiconductors, Fig. 2.1.

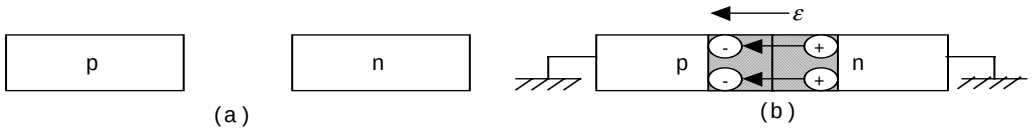


Figure 2.1: The p-n junction.

(a) Uniformly doped p-type and n-type semiconductors before junction is formed.
(b) The electric field in the depletion region.

The large carrier concentration gradients at the junction cause carrier diffusion. Holes from the p-side diffuse into the n-side and electrons from the n-side diffuse into the p-

side. These recombine to remove all free carriers from a small region at the interface called the depletion region. Here, only ionised donors and acceptors remain with a resulting electric field directed from the positive charge to the negative charge.

The junction properties can be modified by applying an external bias. The junction is said to be forward-biased when a positive voltage is applied to the p-side with respect to the n-side. This reduces the depletion layer width and the reduced energy barrier for majority carriers leads to increased diffusion currents. The junction is reverse-biased when a positive voltage is applied to the n-side with respect to the p-side. Reverse bias increases the depletion region width. Fewer carriers have enough energy to diffuse over the higher energy barrier and the diffusion current is strongly suppressed. Only a very small net negative drift current remains. The ideal p-n junction diode equation is:

$$I = I_0\left(e^{\frac{eV_A}{kT}} - 1\right) \qquad \text{.....eq. (2.3).}$$

It is derived using the depletion approximation for an abrupt junction and assuming no external generation sources, no generation or recombination in the depletion region and low-level injection. In practice, eq. (2.3) cannot adequately describe the behaviour of semiconductors with small n_i such as silicon, because of generation and recombination of carriers in the depletion region. Under the reverse-bias condition, the generation current can become significant. For a small forward bias, the recombination current can dominate over the diffusion component.

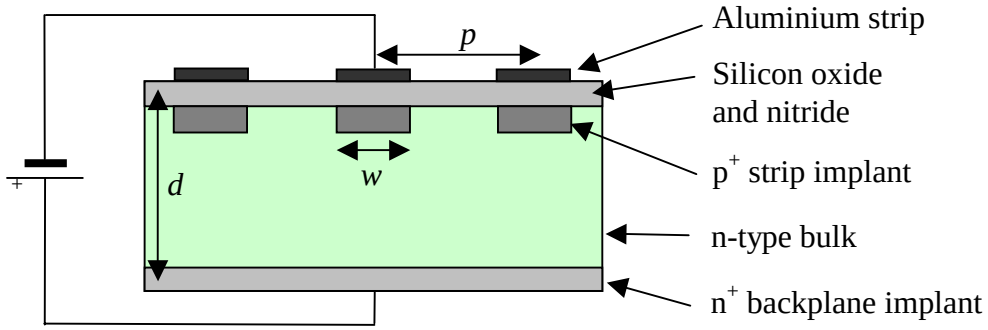


Figure 2.2: Cross-section through an AC-coupled strip detector where w is the strip width, p is the strip pitch and d is the detector thickness [12].

The silicon p-n junction is at the heart of diode detectors that have been used extensively in the physics of radiation detection [13]. The Tracker microstrip sensors are essentially rows of p-n junctions, manufactured by implanting highly doped p^+ strips on an n-type substrate, Fig. 2.2. Particles traversing the sensors generate electron/hole pairs and the resulting current is AC-coupled to the readout electronics.

2.1.3 The MOS Capacitor

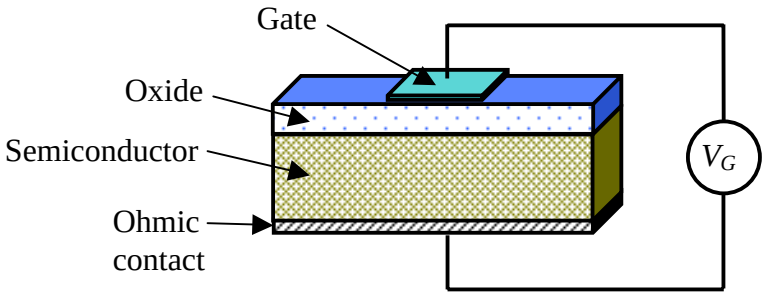


Figure 2.3: Schematic of a MOS capacitor.

In a silicon-based MOS capacitor the gate is usually made of polysilicon, and the oxide is SiO_2, Fig. 2.3. When the semiconductor substrate is p-type silicon, applying a negative voltage to the metal gate ($V_G < 0$) with respect to the ohmic contact leads to an accumulation of majority charge carriers (holes) near the oxide-semiconductor interface. For a small positive gate bias ($V_G > 0$), holes are expelled from the oxide-semiconductor interface leaving negatively charged acceptor ions. As V_G increases, the electron concentration at the oxide-semiconductor interface increases. Eventually, the inversion region is reached where the electron concentration at the interface exceeds the hole concentration.

The gate voltage at which the minority carrier concentration in the inversion layer is equal to the majority carrier concentration in the bulk semiconductor is called the threshold voltage, V_{th}, given by:

$$V_{th} = 2\phi_F + \frac{\varepsilon_r^s x_{ox}}{\varepsilon_r^{ox}} \sqrt{\frac{4eN_A}{\varepsilon_r^s \varepsilon_0} \phi_F} \qquad \text{…….eq. (2.4) for p-type Si.}$$

$$V_{th} = 2\phi_F - \frac{\varepsilon_r^s x_{ox}}{\varepsilon_r^{ox}} \sqrt{\frac{4eN_D}{\varepsilon_r^s \varepsilon_0} (-\phi_F)} \qquad \text{……eq. (2.5) for n-type Si.}$$

where ε_r^s, ε_r^{ox} are the relative permittivities of Si and SiO_2 respectively, ε_0 is the permittivity of free space, N_A and N_D are acceptor and donor concentrations, e is the elementary charge and ϕ_F is the electrostatic potential at the Fermi level w.r.t the intrinsic level. ϕ_F is given by:

$$\phi_F = \frac{kT}{e} \ln\left(\frac{N_A}{n_i}\right) \qquad \text{…...eq. (2.6) for p-type Si.}$$

$$\phi_F = -\frac{kT}{e} \ln\left(\frac{N_D}{n_i}\right) \qquad \text{……eq. (2.7) for n-type Si.}$$

where k is the Boltzmann constant and T is the temperature in Kelvin.

2.1.4 The MOSFET

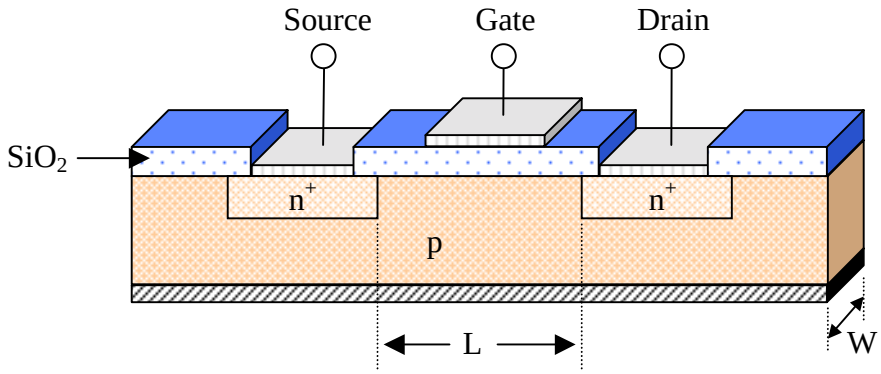


Figure 2.4: Schematic diagram of an n-channel MOSFET.

In the n-channel MOSFET known as NMOS, Fig. 2.4, the drain and source diffusions (n^+ implants) adjacent to the conducting channel are heavily doped with ions of the opposite polarity to that of the ions in the channel or bulk. Contacts are made to the source, drain and gate. Parameters of interest are the gate-to-source voltage, V_G, the drain-to-source voltage, V_D, and its corresponding current, I_D. An additional contact can also be made to the bulk and a voltage applied here will combine with the gate voltage to change the properties of the channel.

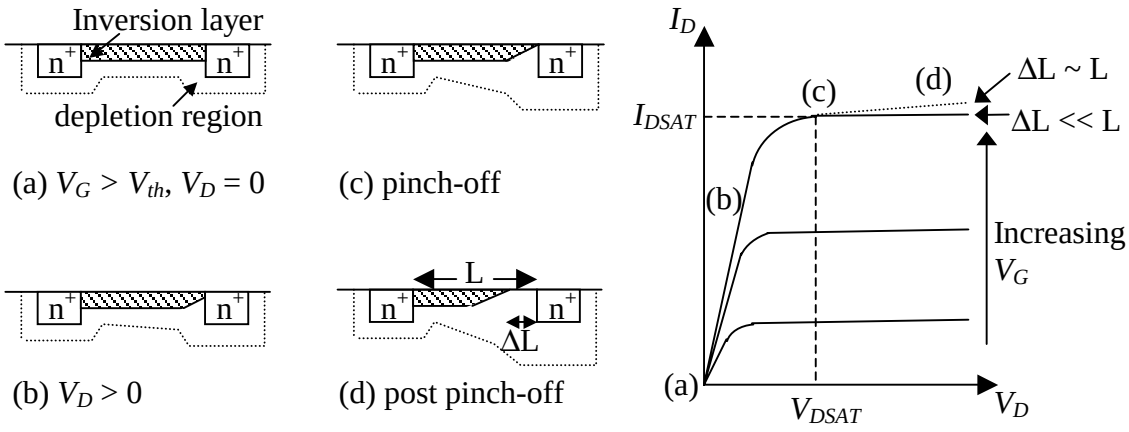


Figure 2.5: $V_G > V_{th}$ in (a), (b), (c) and (d).

The schematic diagrams depict the change in shape of both the inversion layer and the depletion layer in an n-channel MOSFET as V_D increases. The I_D-V_D curve shows how I_D varies with V_D.

When no voltage is applied to the gate, the region between the source and the drain immediately below the oxide contains excess holes and very few electrons, effectively an open circuit since no current can flow between the drain and the source. With the application of a sufficiently positive V_G, an inversion layer forms, enabling a current to

flow between the drain and the source. The greater V_G is, the stronger the inversion in the conducting channel and the higher the conductance between source and drain.

For a fixed value of V_G above threshold, the magnitude of the drain-to-source current flowing through the conducting channel is controlled by V_D. For a small positive V_D, electrons flow from the source to the drain and the channel acts as a resistance so I_D is proportional to V_D and the transistor is said to be operating in the linear region, Fig. 2.5 (b). As the drain voltage is increased, it eventually reaches a point at which the inversion layer width is reduced to zero near the drain implant. This is the pinch-off point beyond which the current remains essentially the same, Fig. 2.5 c). As the drain voltage is further increased, the inversion layer retreats away from the drain implant and the channel length decreases from a length L to a length $L - \Delta L$, Fig. 2.5 d). In practice, for smaller channel lengths, I_D continues to increase slowly beyond the pinch-off point.

To describe the drain current, I_D, an expression is required for each region of operation. In the linear region of operation below pinch-off, I_D is given by:

$$I_D = \frac{W\mu_n C_{ox}}{L}\left[(V_G - V_{th})V_D - \frac{V_D^2}{2}\right] \qquad \text{.....eq. (2.8).}$$

At pinch-off, $I_D = I_{DSAT}$, $V_D = V_{DSAT} = V_G - V_{th}$ and substituting this information into eq. (2.8) leads to:

$$I_{DSAT} = \frac{Z\mu_n C_{ox}}{2L}(V_G - V_{th})^2 \qquad \text{.....eq. (2.9).}$$

Beyond pinch-off, in the region of operation also known as saturation, the drain current does not increase and eq. (2.8) can be used, substituting I_D for I_{DSAT} and V_D for V_{DSAT}.

2.1.4.1 MOSFET Transconductance

The transconductance, g_m, is given by the following equations:

$$g_m = \frac{\partial I_D}{\partial V_G} = \frac{W\mu_n C_{ox}}{L}V_D \qquad \text{.....eq. (2.10) in the linear region.}$$

$$g_m = \frac{W\mu_n C_{ox}}{L}(V_G - V_{th}) \qquad \text{.....eq. (2.11) in the saturation region.}$$

The channel conductance is essentially zero in saturation for an idealised MOSFET and in the linear region it is given by:

$$g_d = \frac{\partial I_D}{\partial V_D} = \frac{W\mu_n C_{ox}}{L}(V_G - V_{th}) \qquad \text{.....eq. (2.12).}$$

2.1.4.2 MOSFET Subthreshold Current

The subthreshold current is the drain current for $V_G < V_{th}$ where the conducting channel is only weakly inverted. It is dominated by diffusion instead of drift and as a consequence, it decreases exponentially with decreasing V_G below V_{th}:

$$I_D \approx e^{q(V_G - V_{th})/kT} \qquad \text{……eq. (2.13).}$$

The subthreshold region of operation is important in digital circuits where transistors act as on/off switches because it describes where the switch changes state. It is not so important in analogue circuits where transistors are operated in strong inversion.

2.1.5 Technological Trends

Over the past forty years the benefits of higher density, lower power consumption, improved performance and reduced costs as transistors reduce in size have encouraged continued growth in the CMOS technology sector. In 1965 Gordon Moore predicted that the number of transistors on integrated circuits would double each year. Ten years later, his predictions were observed to be correct and his initial statement acquired the status of a law, called Moore's law (with a modification in the time scale from one year to nearly two years). Table 2.1 illustrates the difference between two generations of Intel Pentium processes.

Table 2.1: Comparison of two Intel Pentium process generations.

Process	Year	Gate length [μm]	No. of transistors [$\times 10^6$]	Die size [mm^2]
P648	1989	1	1.2	79
P858	2000	0.18/0.13	42	217

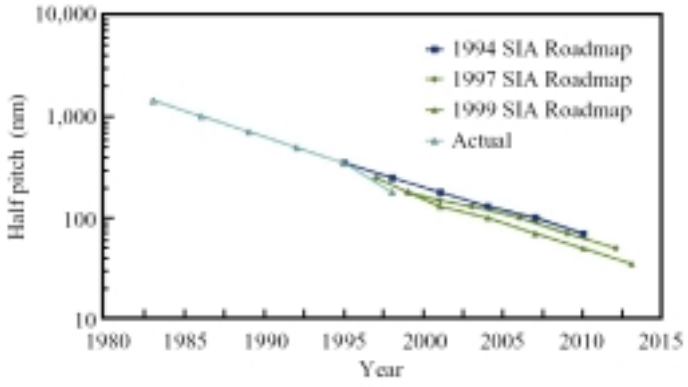
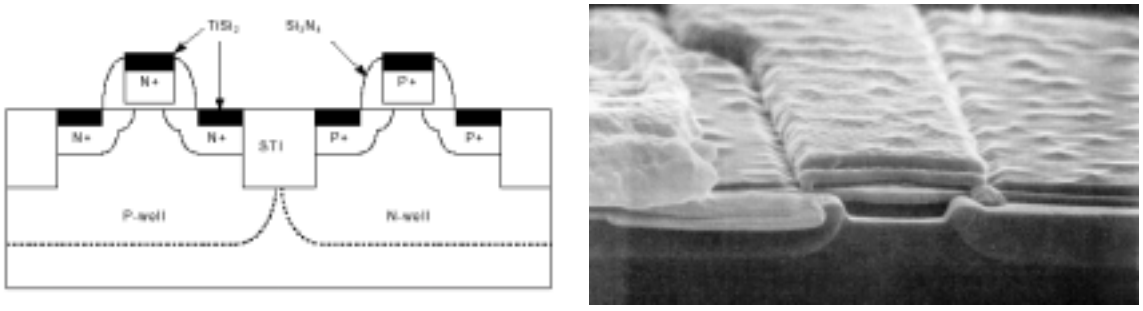


Figure 2.6: Trends in lithographic resolution.

Half pitch represents the minimum size of lithographic features on a chip. (SIA-Semiconductor Industry Association) [14].

The reduction in transistor dimensions can be attributed to improvements in the lithographic resolution and in the design of transistors. Fig. 2.6 illustrates the trends in

lithographic resolution. One major consequence of these is that the conducting channel becomes shorter. This leads to the so-called short-channel effects, affecting the operation of MOSFET transistors and leading to a departure from the idealised behaviour outlined previously. For example instead of remaining constant, the drain current increases with drain voltage beyond V_{DSAT}. If the channel length approaches the depletion layer widths of the source and drain junctions, a condition known as punch-through occurs, where the gate loses control of the current. Minimising the short-channel effects can be achieved by reducing transistor dimensions and voltages by a scaling factor α, which ensures that all internal electric fields remain the same as those of long-channel MOSFETs. As the gate oxide is made thinner, tunnelling effects become more pronounced and will eventually limit the size of transistors.



(a) Schematic cross-section of a 0.25 μm process [15].

(b) Scanning electron microscope photograph of an NMOS transistor.

Figure 2.7: Modern CMOS technologies.

Fig. 2.7 shows a schematic cross-section of a 0.25 μm CMOS process along with a photograph of a MOS transistor.

2.1.6 Noise

Noise, in the broadest sense, can be defined as any unwanted disturbance that obscures or interferes with a desired signal [16]. In cases of interest here, the term noise refers to spontaneous fluctuations in the current passing through, or the voltage developed across, semiconductor bulk materials or devices. These spontaneous fluctuations which are related to the discrete nature of charge carriers set a lower limit to the quantities to be measured or the signal to be amplified.

The CMS microstrip detectors will produce current signals, which have to be read and amplified by the APV25. Amplification is the first stage in signal processing to ensure that later stages do not significantly contribute to the noise [17]. The overall signal-to-

noise ratio is therefore determined by the noise due to the amplifier, ideally arising in the input transistor of the preamplifier [18].

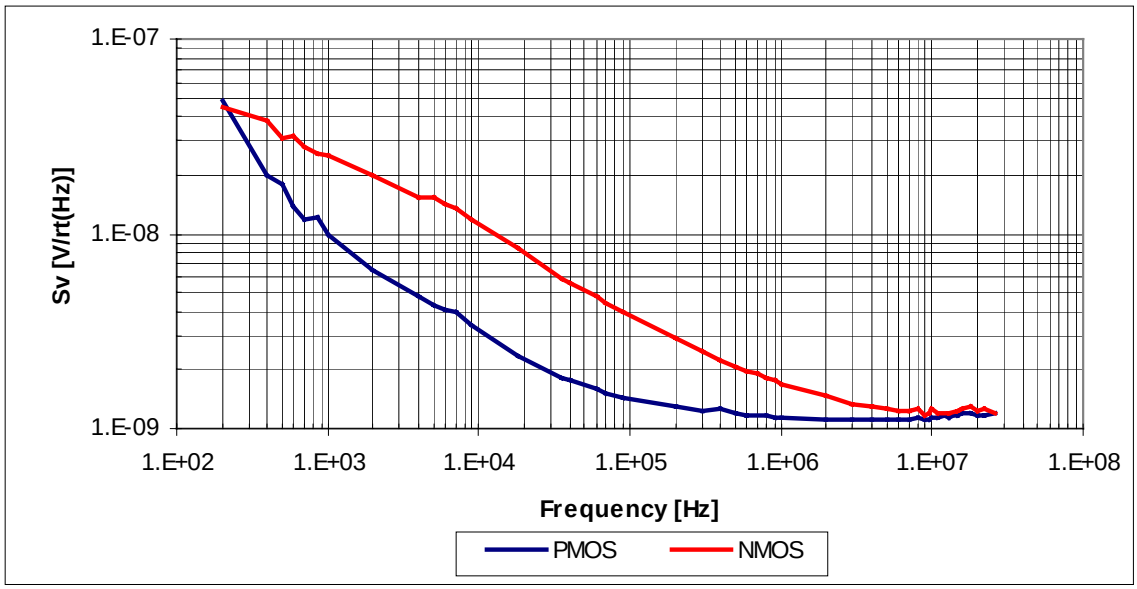


Figure 2.8: Noise spectrum of PMOS and NMOS transistors from a 0.5 μm process.

The typical voltage noise spectral density of MOSFETs is illustrated in Fig. 2.8. Two distinct frequency regions can be identified, with different noise behaviour in each. At low frequencies, the dominant component of the noise is frequency dependent. At high frequencies, the noise is essentially frequency independent. The two regions can be thought of as separated by a 'corner frequency', f_c. Values from several Hz to several hundred kHz are common for this parameter. As is apparent in Fig. 2.8, the corner frequency is higher for NMOS than it is for PMOS.

2.1.6.1 1/f Noise

The noise dominating at low frequencies is called 1/f noise because the power spectral density for the current caused by this type of noise is practically proportional to 1/f. It is also known as flicker noise and has historically been explained as free-carrier exchange with traps at interfaces, which cause a fluctuation of carrier density. The model explains the near-linear increase of 1/f noise with decreasing frequency. This model was first proposed by McWorther and is better adapted for noise at interfaces and therefore MOSFET devices [19].The increase in noise is a result of a distribution of characteristic exchange times related to the traps' spatial distance into the oxide and/or energy difference from the semiconductor band edges [20]. The further a trap is from the SiO_2/Si interface, the lower its frequency response. Hooge proposed another model based on mobility fluctuation, which is more consistent with noise in bulk semiconductors [21].

The gate referred voltage noise is given by:

$$S^2_{V,1/f} = \frac{A_f}{f^{\alpha}} \qquad \text{……eq. (2.14).}$$

where f is the frequency, α is an experimental coefficient known as the flicker noise exponent (typically ~ 1) and A_f represents the voltage noise in units of V^2 at its intercept with the y-axis. A_f is given by:

$$A_f = \frac{K_f}{(C_{ox})^2 WL} \qquad \text{…….eq. (2.15).}$$

where C_{ox} is the gate capacitance per unit area, W and L are the nominal channel width and length respectively, and K_f is a technology-dependent parameter known as the flicker noise coefficient. Eq. (2.15) shows that the 1/f noise component is dependent on process technology and transistor dimensions. For the transistors shown in Fig. 2.8, A_f is around 10^{-14} V^2. For the 0.25 μm technology used to manufacture the APV25, K_f is of the order of 10^{-25} V.C [22].

2.1.6.2 Thermal Noise

Thermal noise (also called white noise, Johnson noise or Nyquist noise) is the type of noise best characterised for the MOSFET. It is dominant at high frequency and independent of frequency. The origin of this noise can be traced to the random thermal motion of carriers in the channel. This is similar to the mechanism in resistors where the thermal noise spectral density is given in voltage by *4kTR*, *R* being the resistance. The derivation for the channel thermal noise of a MOSFET is more complex than that of a resistor since the operating region of the device has to be taken into account. In saturation and strong inversion, the following expression is used for the thermal noise:

$$S^2_{V,White} = \frac{8}{3} kT \frac{\Gamma}{g_m} \qquad \text{……eq. (2.16).}$$

where k is the Boltzmann constant, T is the temperature in Kelvin, Γ is the excess noise factor (typically ~ 1) and g_m is the transconductance given earlier, eq. (2.11).

2.2 Consequences of Radiation on MOSFET Operation

Severe degradation of the operational characteristics of MOS devices and circuits can result from exposure to ionising radiation. Ionising radiation is simply defined as that with an energy above the threshold to break atomic bonds and create electron/hole pairs in the materials of interest, which are Si and SiO_2 in silicon-based MOS devices.

Photons or charged particles such as electrons, protons, or atomic ions can all cause ionisation.

Another class of radiation damage exists, displacement damage, where atoms in a material are displaced from their original locations by the incident radiation. Displacement damage is primarily caused by hadrons and can result in a reduction in the gain of bipolar transistors through a deterioration of the minority carrier lifetime in the silicon substrate. MOS devices are not affected by changes in the minority carrier lifetime and are therefore relatively insensitive to displacement damage.

Ionising radiation effects in MOS devices can be divided into two categories:

- total dose effects which are due to an accumulation of ionising radiation over a period of time in the oxide layers in MOS devices and circuits.
- transient radiation effects (or Single Event Effects, SEE) due to the generation of photocurrents in the silicon substrate by high-dose-rate radiation such as the passage of highly energetic particles.

2.2.1 Total Dose Radiation Effects

When a MOSFET is exposed to high-energy ionising radiation, electron/hole pairs are created uniformly throughout the oxide. These carriers can be trapped in the oxide and at the Si/SiO_2 interface [23]. The number of electron/hole pairs created in a material is directly proportional to the amount of energy absorbed by that material. The total dose is commonly expressed in units of rads, equal to 10^{-2} joules absorbed per kilogram of the specified material (1 rad = 10^{-2} Gy = 100 erg.gm^{-1}). All the effects characterising the parameter evolution of the MOSFET after irradiation are grouped under the term annealing.

Following electron/hole pair generation in the oxide by ionising radiation, electrons rapidly drift out of the oxide since the electron mobility is high (low field mobility at T = 300 K is $\mu \sim 20$ cm^2.V^{-1}.s^{-1}), leaving holes, which have much lower mobility ($\mu \sim 10^{-4}$–10^{-11} cm^2.V^{-1}.s^{-1}), trapped in the oxide. The electron/hole pair creation energy in SiO_2 is around 18 eV [24]. Radiation-generated electrons play a minor role in initial recombination processes but are otherwise generally ignored in discussions of charge trapped in the oxide. Not only are they highly mobile but long term trapping rates are up to six orders of magnitude less than those for holes.

Much work has been done to characterise charge generation and recombination, hole transport and charge trapping in SiO_2 in thick oxides (d_{ox} > 20 nm), a good summary of which can be found in [25]. Many models have been used to describe experimental data

such as the columnar recombination model (when charge pairs are deposited in dense columns and recombination is a strong process), the Onsager geminate recombination model (when charge pairs are created far apart on average and recombination is a weak process) and the continuous-time random walk (CTRW) model for hole transport. This work successfully identified the characteristics of oxide trapped charge such as the field, temperature and oxide thickness dependence and the location and type of traps. The hole traps are local oxygen vacancy defects such as strained Si-Si bonds that can capture holes and become various forms of a defect know as the *E'* centre. The accumulated knowledge has lead to the adoption of measures to harden thick oxides for use in radiation environments.

Improvements in oxide quality (fewer contaminants, low-growth-temperature) and the scaling of CMOS technologies, in particular the reduction of d_{ox}, have led to a reduction in oxide trapped charge and an increase in the inherent radiation hardness of MOS devices [26]. For $d_{ox} > 10$ nm the effects of radiation generated charge generally drop as the square of the oxide thickness. For $d_{ox} < 10$ nm, most of the holes are removed from the oxide within seconds to minutes and the models developed to describe hole generation, transport and trapping in thicker oxides are no longer valid [27]. The main processes of hole removal are described in the following paragraph.

Immediately after oxide trapped charge is created it begins to be neutralised. This process is dependent on time, temperature and electric field and can be accounted for by invoking two mechanisms [25], Fig. 2.9:

- the tunnelling of electrons from the silicon into oxide traps,
- the thermal emission of electrons from the oxide valence band into oxide traps.

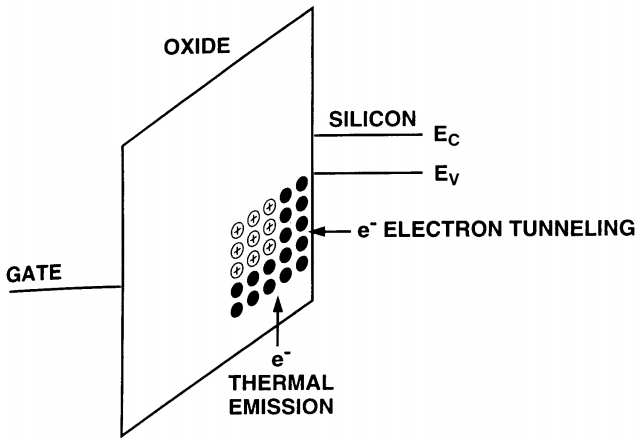


Figure 2.9: Schematic diagram illustrating the neutralisation of oxide-trap charge by electron tunnelling from the silicon and by thermal emission of electrons from the oxide valence band.

The first mechanism can be depicted as a tunnelling front moving into the oxide. If this front is defined as the position corresponding to the maximum rate of tunnelling, $x_m(t)$, then the distance of the front into the oxide at a given time t is given by:

$$x_m(t) = \frac{1}{2\beta}\ln\left(\frac{t}{t_0}\right) \qquad \text{……eq. (2.17).}$$

where β is the tunnelling barrier height parameter, and t_0 is a time scale parameter [25]. This simple model is not complete since after a given distance in the oxide, there are fewer trapped holes in certain oxides and the rate of annealing decreases with time.

At temperatures above 150 ^{o}C, hole removal occurs via the thermal process. Trapped hole annealing depends on both the energy level and spatial distribution of traps which are heavily process-dependent.

The 0.25 μm process used in the manufacturing of the APV25 has an oxide thickness of 5.5 nm, allowing neutralisation of oxide traps during the process of ionisation by tunnelling from both sides of the oxide. Work done by Benedetto *et al.* has shown that, for 5.3 nm oxides, the tunnelling front from the silicon meets that from the gate close to the middle of the oxide about 1 s after hole generation and trapping [28].

The lack of oxygen at the SiO_2/Si interface introduces strained and uncompleted bonds. These can trap charge from the Si conduction and valence bands and are the so-called interface traps. The pre-irradiation interface-trap density is usually too small to cause any significant effect (< 10^{10} traps/cm^2). However, ionising irradiation generates additional interface traps which can affect the operation of MOS devices. Interface traps are located within one or two atomic bond distances (0.5 nm) from the silicon lattice. Unlike oxide trapped charge which is always positive, interface trapped charge can be positive, neutral or negative. Because of this, interface traps are identified as either donors or acceptors. For NMOS transistors, interface traps contribute a net negative charge and are therefore acceptors. For PMOS transistors, interface traps contribute a net positive charge and are therefore donors.

A trivalent Si atom bonded to three other Si atoms, the P_b centre, has been identified as the defect responsible for interface traps in work by Lenahan and Dressendorfer [29].

The most common technique for measuring interface traps in transistors is the subthreshold technique. It consists of analysing changes in the subthreshold slope or swing, S, of the I_D-V_G curve. S is defined as the change in V_G required to reduce the transistor current by one decade.

The change in interface trap density, ΔD_{it}, following an irradiation is given by:

$$\Delta D_{it} = \left[C_{ox} / kT \ln(10)\right]\left(S_{Drad} - S_{Do}\right) \qquad \text{.....eq. (2.18).}$$

where S_{Do} and S_{Drad} are subthreshold swings measured before and after irradiation, respectively [25].

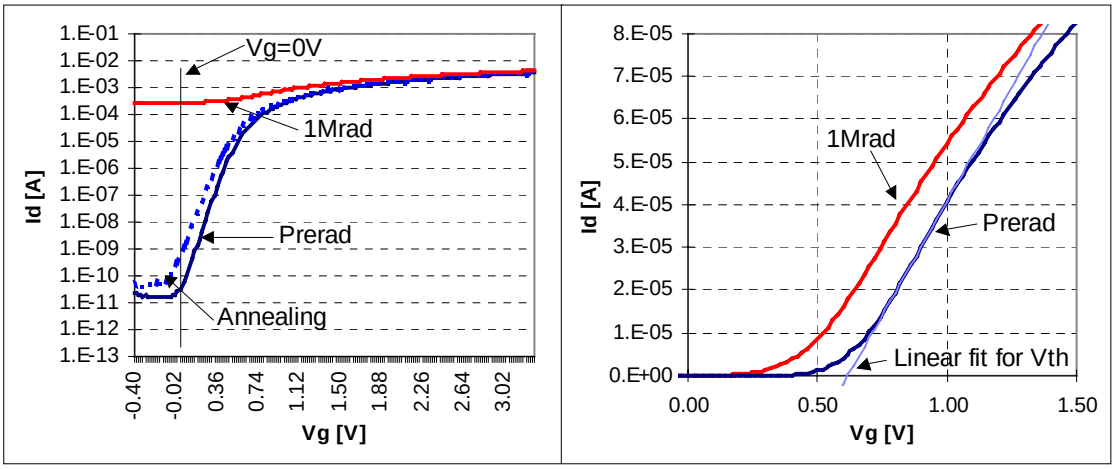
Interface trap buildup occurs on time scales much slower than oxide trapped charge buildup and continues to evolve after irradiation. It is dependent on many variables such as dose, dose rate, electric field, temperature and processing. Many workers have reported a D^X dependence on dose with X ranging from 0.65 to 0.95 for d_{ox} > 45 nm. For very thin oxides, interface trap buildup is observed to be strongly suppressed. Saks *et al.* have reported a ΔD_{it} which is very much smaller in very thin oxides (d_{ox} < 12 nm) compared with that predicted from the D^X dependence observed in thicker oxides [30]. This is attributed to the tunnelling of radiation-induced holes out of the oxide (equivalent to electrons tunnelling into oxide) before conversion of the holes to interface states. The buildup of interface traps is strongly temperature dependent. Below 100 K, electrons and holes are essentially 'frozen' and interface trap buildup is completely inhibited. Unlike oxide trapped charge, interface trapped charge does not anneal at room temperature [25]. Significant annealing is observed only for T > 100 ^{o}C.

The consequences of oxide trapped charge and interface trapped charge on the electrical characteristics of the MOSFET are the following:

- threshold voltage shift,
- decrease in mobility,
- increase in noise,
- increase in leakage current.

Effects on the thin gate oxide are responsible for the first three consequences. Charge trapped in the lateral and field oxides used to isolate drain and source implants and to isolate adjacent MOSFETs causes the increase in leakage current.

Fig. 2.10 shows the effects of leakage current and threshold voltage shift on the electrical characteristics of a MOSFET.



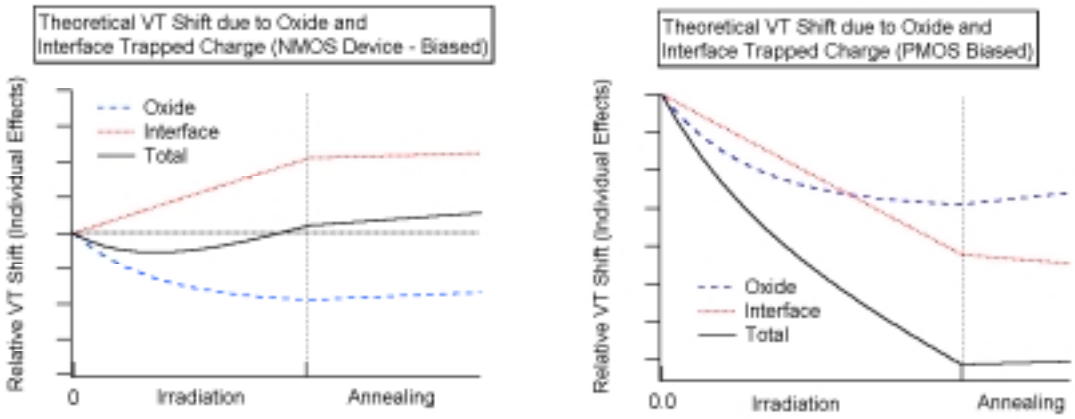
(a) *(b)*

Figure 2.10: Drain current as a function of gate voltage.

(a) in logarithmic scale as a function of gate voltage to show the leakage current, and (b) in linear scale to show the threshold voltage shift.

2.2.1.1 Threshold Voltage Shift

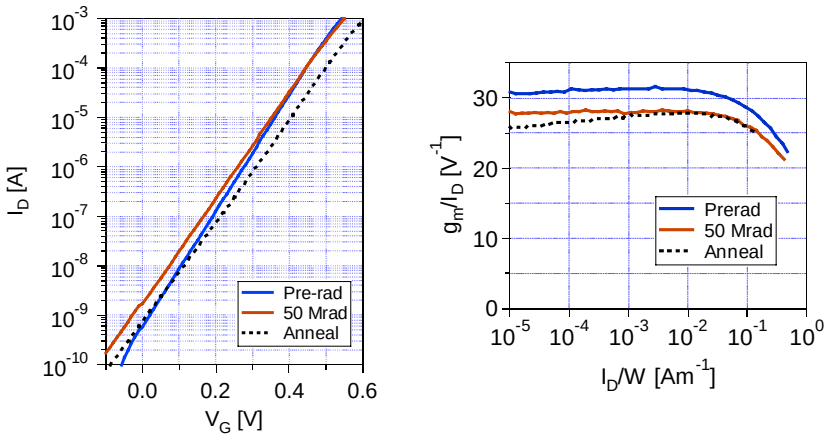
The trapped charge in the oxide and at the Si/SiO_2 interface causes a shift in the threshold voltage. Oxide trapped charge is always positive and causes a negative shift in the threshold voltage for NMOS and a positive shift (in absolute value) in the threshold voltage for PMOS. Interface trapped charge causes a positive threshold voltage shift in both NMOS and PMOS. Since oxide trapped charge is positive for both N- and PMOS transistors, oxide trapped charge and interface trapped charge compensate each other for n-channel transistors and have a cumulative effect for p-channel transistors, Fig. 2.11.



(a) *(b)*

Figure 2.11: V_{Th} shift in biased (a) NMOS, and (b) PMOS transistors [31].

2.2.1.2 Decrease in Mobility



(a) *(b)*

Figure 2.12: NMOS transistor (a) Subthreshold swing and (b) Transconductance.

Fig. 2.12 illustrates the change in subthreshold swing and transconductance due to irradiation. Interface trap formation is responsible for the decrease in the mobility of carriers by introducing additional Coulomb scattering centres in the conduction channel. Mobility degradation is more pronounced for NMOS than it is for PMOS. Transconductance, which is directly related to mobility, also decreases with irradiation.

2.2.1.3 Increase in Noise

Transistor noise is particularly important for analogue applications. Although 1/f noise is related to traps at the interface, in NMOS devices a good correlation has been shown between the increase in 1/f noise after irradiation and charge trapping in the oxide in defects of the *E'* type located near the interface called border traps [32]. Under normal operating bias conditions (+ve bias for NMOS and –ve bias for PMOS), the 1/f noise of both n- and p-channel transistors decreases through post-irradiation annealing [33].

At high frequencies, a radiation-induced decrease in mobility and transconductance can lead to an increase in noise. It has also been reported that the excess noise factor, Γ (in short-channel devices it accounts for hot carrier effects due to high electric fields, when carriers in the inversion layer are not in thermal equilibrium with the Si substrate), can account for most of the increase in thermal noise [22]. The mechanisms behind the radiation-induced increase in Γ have not been identified or reported in the literature.

2.2.1.4 Increase in Leakage Current

A build up of charge in the oxides used to isolate drain and source implants can lead to a leakage current between those implants and can cause the device to be switched on when there is no external applied bias on the gate.

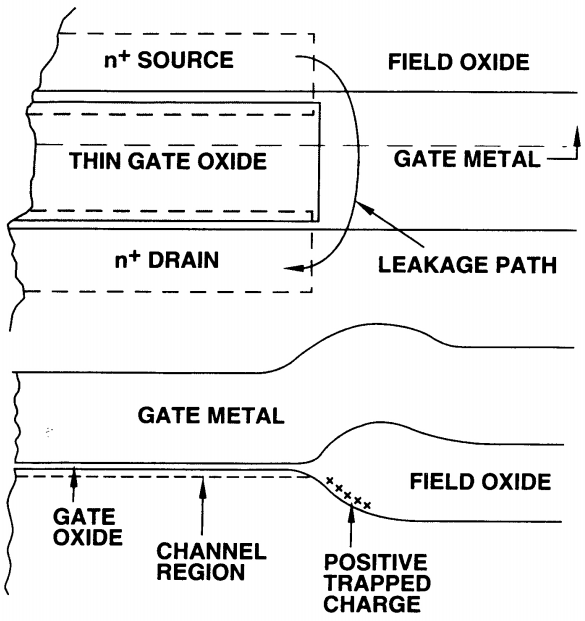


Figure 2.13: Cross-section of a parasitic field oxide transistor showing the primary leakage current paths [34].

Fig. 2.13 illustrates the primary leakage path due to a buildup of positive charge in the lateral (field) oxide for a device using the LOCOS (local oxidation of silicon) isolation scheme. The process used in the manufacturing of the APV25 employs the STI (Shallow Trench Isolation) scheme.

Field oxides are much thicker than gate oxides. Unlike gate oxides, which are routinely grown by thermal oxidation, field oxides are produced using a wide variety of deposition techniques (e.g. Chemical Vapour Deposition, CVD). Thus, the trapping properties of a field oxide may be poorly controlled and can be considerably different than for a gate oxide.

The threshold voltages of parasitic field oxide transistors are initially very large. As radiation-induced oxide charge builds up in a field oxide, it causes a decrease of the threshold voltage of the field oxide parasitic transistor on a p-type substrate (equivalent to an n-channel field oxide transistor). If the buildup of charge is large enough, excessive leakage current can flow from the source to the drain of the gate-oxide transistors and between transistors. This will greatly add to the static supply leakage current of a circuit and, sometimes, functionality is lost.

Because of the nature of trapping in the field oxide (positive charge trapping), increase in leakage current due to irradiation only applies to NMOS transistors.

Special layout techniques can be used to cut the source-to-drain leakage paths which are present at the edges of standard n-channel transistors. One such technique consists of drawing a gate around the drain when designing n-channel transistors, thus eliminating

the edges of the transistor [35, 36, 9]. This is illustrated in Fig. 2.14. Transistors designed with these layout features are known as edgeless or enclosed transistors.

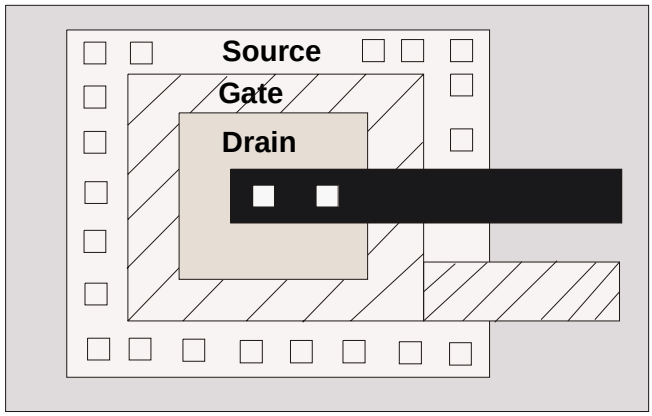


Figure 2.14: Edgeless transistor.

Leakage paths that, under the field oxide, carry current between neighbouring n-channel transistors and between differently biased n^+ diffusions can be cut by the use of channel stops. This is achieved by surrounding the n-channel transistors with p^+ guard rings.

The APV25 design incorporates both enclosed transistors and p^+ guard rings.

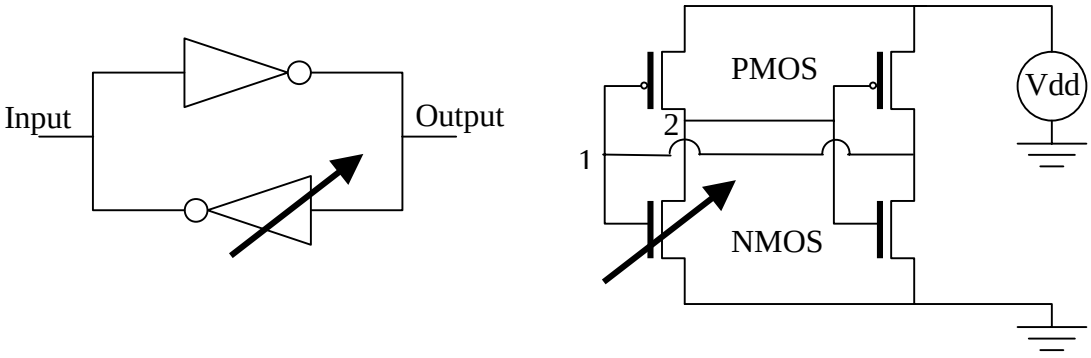
Significant differences in radiation hardness have been reported in transistors of different geometry (channel length and width) [37]. Studies on a process with a gate thickness of 25 nm show that transistors with shorter gate lengths tend to show more negative threshold voltage shifts during irradiation but less positive shifts during post-irradiation annealing than transistors with longer channel lengths. Simple scaling theory cannot account for these effects which occur because of differences in trapped-charge densities among devices of different sizes. It is difficult to predict whether the radiation hardness dependence on transistor geometry is significant for processes with thin gate oxides where annealing via tunnelling dominates. A considerable difference has also been observed in the radiation hardness of commercial technologies from different manufacturers since these have little interest in identifying and controlling technology parameters that affect it, such as gate oxide quality [38].

2.2.2 Single Event Effects (SEE)

Single event effects (or transient radiation effects) depend on the rate at which the ionising dose is delivered to a circuit rather than on the total dose delivered. A relatively short pulse in a MOS circuit can be induced by a high-density ionisation track created by the passage of a charged particle, leading to various failure modes. A distinction is usually made between failure modes that are permanent (hard errors) and those that are non-destructive (soft errors). There are three main categories of SEE: Single Event Upset (SEU), Single Event Gate Rupture (SEGR) and Single Event Latchup (SEL).

2.2.2.1 Single Event Upset (SEU)

SEU is a soft error affecting both dynamic and static registers that store logic states. It is characterised by the deposition of sufficient charge on a circuit node to cause the inversion of a logic state. SEUs pose a threat to the normal operation of the digital parts of the APV25 chip because they can cause logic errors in control circuits and bit errors in data transmission by corrupting the digital header address. A detailed knowledge of SEU behaviour under LHC conditions is necessary in order to plan how to operate the chips (e.g. hard resets) in CMS.



(a) Schematic representation. *(b) Circuit description.*

Figure 2.15: A memory cell composed of two cross coupled inverters [39]. *The arrow represents a heavy ion strike.*

Fig. 2.15 depicts a memory cell composed of two cross-coupled inverters and its circuit description. It is designed to have two stable states, one representing a stored '0' and the other a stored '1'. In each state, two transistors are switched on and two are turned off.

A heavy ion strike on one of the inverters can produce a transient change in the voltage appearing at the input of the other. If the input voltage falls below the threshold for switching, the memory cell changes state [40]. Depending on the state of the cell, injection of charge above a critical value, Q_{crit}, at points 1 or 2 will cause the state of the cell to invert. The sensitive volume of the memory cell, defined as that volume in which charge generated by ionisation can be collected fast enough to cause an upset, can be approximated by the drain implant volume of the 'off' NMOS transistor and that of the source of the 'off' PMOS transistor. An incident particle must have a linear energy transfer (LET) high enough to generate charge above Q_{crit} and must strike the sensitive volume for an upset to occur. The LET for a given particle incident on a material is a measure of the rate of energy transfer in that material and is given by:

$$LET = \frac{dE}{dx} \times \frac{1}{\rho} \qquad \text{…….eq.(2.19).}$$

where $\frac{dE}{dx}$ is the energy loss per unit length and ρ is the material density.

The cross-section, σ, for SEU is defined at normal incidence as:

$$\sigma = \frac{N_{events}}{\Phi} [\text{cm}^2] \qquad \text{.....eq. (2.20).}$$

where Φ is the total incident particle fluence, and N_{events} is the number of SEU events counted during the test [41].

SEU results are usually given in plots of cross-section vs. particle LET, which is illustrated in Fig. 2.16. LET_{th} is the minimum particle LET required for SEU and σ_{sat} is the saturating cross-section. The experimental points are often fitted with a Weibull curve using the following expression [42]:

$$\sigma = \sigma_{sat}\left\{1 - \exp\left[-\left(\frac{LET - LET_{th}}{W}\right)^{S}\right]\right\} \qquad \text{.....eq. (2.21).}$$

This allows for the extraction of σ_{sat} and of LET_{th}. W and S are fitting parameters without physical meaning.

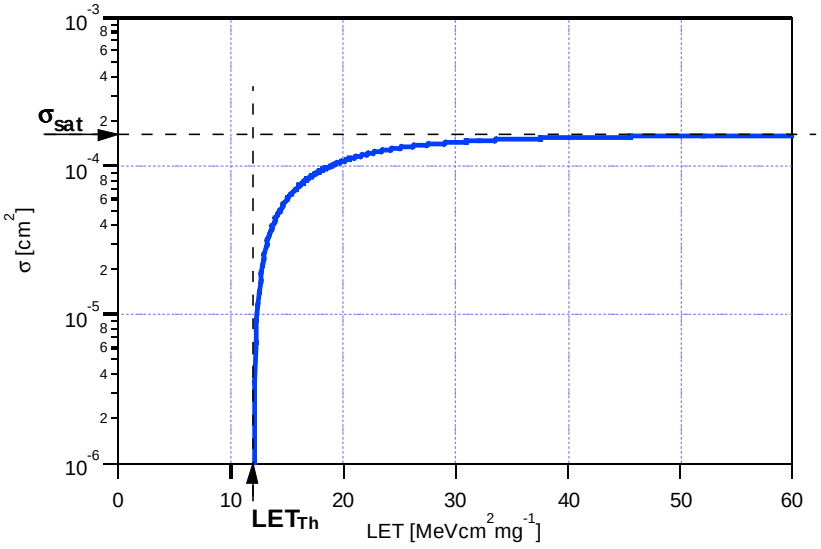


Figure 2.16: A typical cross-section curve.

2.2.2.2 Single Event Gate Rupture (SEGR)

A Single Event Gate Rupture (SEGR) is caused when a highly ionising particle traverses the gate oxide of a MOS device and induces the creation of a dense ionised track. This can lead to an electrical short through the oxide, causing permanent damage.

SEGRs are normally associated with power MOSFETs operating under high electric fields. However, their importance has to be assessed for the CMS environment.

Various models have been developed to try and explain the SEGR mechanism. One of the models which fits experimental data well is based on the concept of a conducting cylinder surrounding the track of the ionising particle.

The conducting pipe model is a simple semi-empirical model which hypothesises a linear dependence of inverse electric field on LET:

$$E_{CR} = \frac{E_0}{1 + L/B} \qquad \text{.....eq. (2.22).}$$

where E_{CR} is the critical field to rupture, E_o is the breakdown field, L is the particle LET and B is a fitting parameter which is technology dependent.

The increase in oxide electric field as technologies scale raises the concern that single event gate rupture (SEGR) could become a severe problem for advanced integrated circuits, such as those used in space applications [43].

Sexton *et al.* [44] have explored the dependence of SEGR on oxide thickness for oxide thicknesses between 6 nm and 18 nm. They report that the critical field, defined as the field at which gate rupture is observed at a given LET, increases with decreasing oxide thickness at any given LET. The thinnest oxides had critical fields above 7 MV/cm (LET = 80 MeV.cm^2/mg), indicating the potential for improved SEGR tolerance for advanced technologies. Massengill *et al.* report similar trends in their sub-5 nm oxides [45]. This increase in critical oxide field is attributed to reduced defect creation by hot carriers in the oxide.

Following their work on thin oxides, Sexton *et al.* have examined the impact of ion damage in 5 nm and 7 nm oxides to determine whether SEGR is a single ion effect [46]. They conclude that SEGR (identified by a large leakage current at low voltage levels) and precursor ion damage (noticeable by a gradual increase in leakage current at high voltage levels) are largely unrelated effects. This supports a single ion model for SEGR. They have observed that the capacitors with 5 nm oxides are unexpectedly difficult to rupture. In a few cases, an SEGR characteristic similar to the thicker 7 nm oxides is observed. However, in most cases no SEGR characteristic is observed.

A paper by Johnston *et al.* [47] addresses the issue of the dominance of soft, rather than hard, breakdown in oxides below 6 nm. Capacitors with oxide thicknesses of 4.5 nm and 7.5 nm have been studied. The results obtained show increasing critical field for decreasing oxide thicknesses, in agreement with the trend observed in [44]. The authors report that no true oxide shorts are observed in either oxides, but breakdown in the thinner oxides produces much lower current conditions. Localised regions within the oxide, which are more sensitive to the breakdown process, are taken to be the reason for the requirement of multiple hits before the onset of breakdown. The authors state that the ions have to strike a small critical region of the device when the applied field is high enough to cause breakdown in this region. The results are further evidence for a single

ion interaction mechanism, with no dependence on residual damage from previous ion strikes. Although the leakage currents observed are very small, they are large enough to cause circuit failure if they occur internally in small-area devices within VLSI circuits.

The work reported above has been carried out at high fields with relatively low ion fluences (maximum of 10^8 cm^{-2}) [44, 46, 47]. In a study of the cumulative damage produced by high radiation doses at low oxide fields during irradiation, Ceschia *et al.* [48] have observed soft breakdown phenomena at zero fields under irradiation with high LET ions. Radiation-induced leakage current (RILC), a manifestation of a trap assisted electron tunnelling conduction mechanism [49, 50], is observed for irradiations with low LET ions. Local oxide damage due to the overlap of different ion tracks, because of the high ion fluences (greater than 10^8 cm^{-2}) used, is quoted as the likely cause of the occurrence of radiation-induced soft breakdown (RSB). RILC and RSB have different field dependences, with RILC being a maximum at zero fields whilst RSB is a minimum at zero fields. The authors have observed a large increase in leakage current when RSB is activated, with an LET threshold for RSB activation in their 4 nm oxides of around 20 $MeV.cm^2.mg^{-1}$.

In some very rare cases, nuclear interactions with tungsten (used in circuits for the connection between silicon and metal layers) can lead to an LET of 80 $MeV.cm^2.mg^{-1}$ [51]. However, most of the knock-on atoms generated by the passage of charged particles in the CMS environment will have LETs rarely exceeding that of the Si ion. Fig. 2.17 shows that the maximum LET for a Si ion in silicon is $\sim 3.5 \times 10^4/(2.33 \times 10^3)$ = 15 $MeV.cm^2.mg^{-1}$.

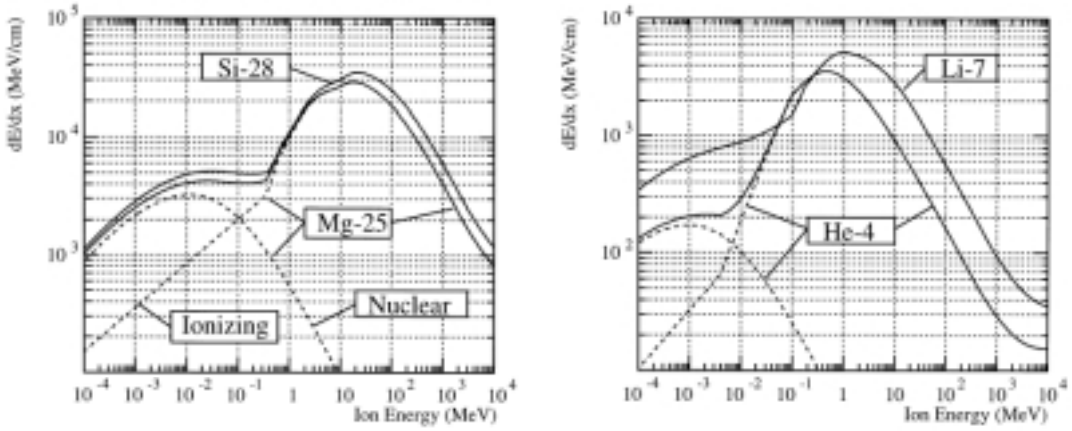


Figure 2.17: Energy loss of some ion species in silicon as a function of the kinetic energy of the ion, from [52].

Fig. 2.18 shows some results by previous workers along with lines representing the maximum electric field across the APV25 oxide and the maximum LET expected in the CMS radiation environment, indicating that the APV25 should be SEGR-free.

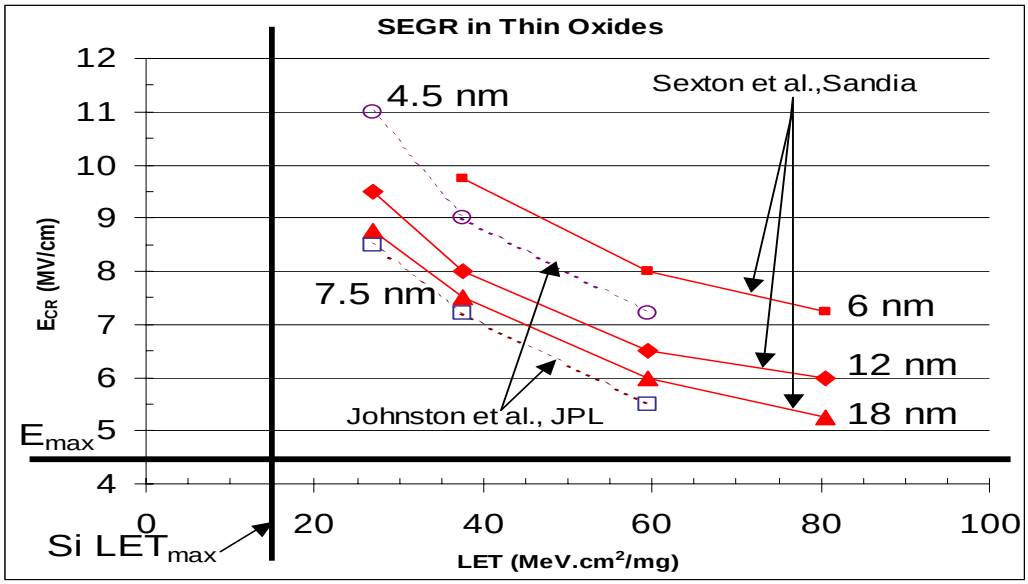


Figure 2.18: Critical electric field as a function of particle LET.

The horizontal bold solid line indicates the electric field across the oxides in an APV25. The vertical line indicates the maximum LET expected in the CMS Tracker.

2.2.2.3 Single Event Latchup (SEL)

Inherently, CMOS technologies consist of adjacent n- and p-type regions. A parasitic SCR (semiconductor controlled rectifier) can be formed by p-n-p-n structures which can be present in CMOS circuits. Under normal conditions, this parasitic device is switched off. It can be turned on by a triggering current such as that produced by ionising radiation. This is known as latchup and causes the circuit to turn fully on, creating a short across the device until it burns up or the power is cycled. Modern commercial CMOS processes are designed to be less susceptible to electrically-induced latchup from transients at input/ouput pins and power supply sequencing but are not necessarily immune to radiation-induced latchup.

The use of Shallow Trench Isolation (STI), thin epitaxial substrates and retrograde wells reduces latchup sensitivity in commercial technologies [53]. Theory predicts that the minimum triggering conditions should depend on the reciprocal of the epitaxial thickness and the thin (< 3 μm) epitaxial substrates of most commercial devices do not latch up for LET values of ~ 100 $MeV.cm^2.mg^{-1}$ [43].

The design rules that ensure the APV25 is total-dose radiation tolerant (enclosed NMOS transistor layout and p^+ guard rings which cut leakage current paths) provide additional

protection against SEL. Static and dynamic registers implemented with these design rules in a 0.25 μm technology did not latch up during irradiation tests up to the maximum LET of 89 MeV.cm^2/mg [54]. There is no evidence to suggest that SEL could be a problem for 0.25 μm technologies in the CMS radiation environment.

2.3 Temperature Effects

The APV25 chip has been extensively tested by research groups involved with the production and testing of components in the readout chain, mostly at room temperature. To limit radiation damage to the silicon microstrip detectors, the CMS Tracker will be operated at –10 ^{o}C. Although temperature effects are more pronounced for the silicon microstrip detectors and components such as the laser drivers, it is important to understand their consequences on the operation of the readout chip.

2.3.1 Temperature Dependence of Threshold Voltage

An expression for V_{th}, eq. (2.4), was given in section 2.1.3. It can be seen that the sensitive term is ϕ_F, which varies with temperature as eq. (2.6). The following expression is used for n_i:

$$n_i = \sqrt{N_C N_V}\, e^{\left(-E_g / 2kT\right)} \qquad \text{……..eq. (2.23).}$$

where N_C, N_V and E_g all change with temperature [11, 20, 55].

Fig. 2.19 shows the variation of threshold voltage with temperature for temperature-dependent values of n_i taken from [2.9].

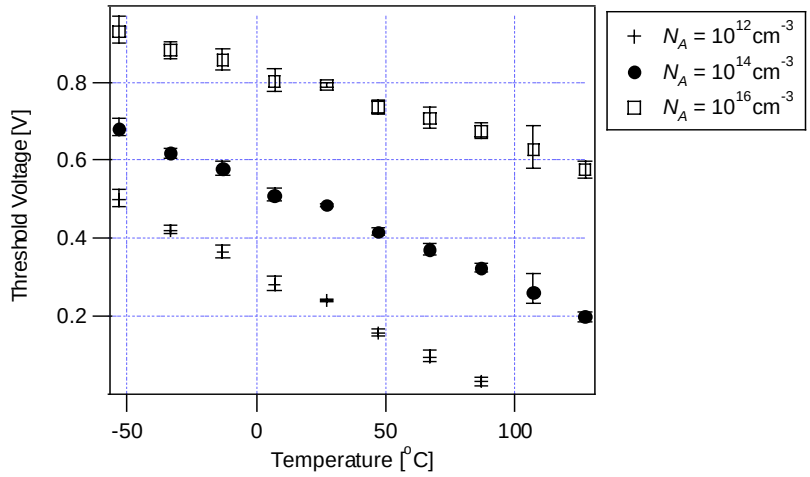


Figure 2.19: Variation of V_{th} with temperature for an NMOS transistor for different values of N_A.

As can be seen in Fig. 2.19, V_{th} decreases with increasing temperature for an NMOS transistor. V_{th} decreases in absolute value with increasing temperature for PMOS.

Usually, information on the processing details such as N_A is not available and the following expression is used to quantify the change of V_{th} with temperature [56]:

$$V_{th}(T) = V_{th}(T_0) - \alpha(T - T_0) \qquad \text{……..eq. (2.24).}$$

In eq. (2.24), $\alpha \approx 1\text{mV/}^{\text{o}}\text{C}$ but is process-dependent and has to be determined, for example by measuring individual transistors on test structures manufactured alongside chips on a wafer.

2.3.2 Temperature Dependence of Mobility

In Section 2.1.1, mobility was seen to be related to lattice scattering and impurity scattering. Above absolute zero, lattice atoms are subjected to thermal vibrations which are responsible for lattice scattering. Thermal vibrations increase with increasing temperature and lattice scattering dominates at high temperatures leading to a decrease in mobility with increasing temperature. The mobility term due to lattice scattering, μ_L, decreases in proportion to $T^{-3/2}$[57].

Ionised impurities are responsible for impurity scattering, where a charge carrier is deflected due to the Coulomb interaction. This deflection is less pronounced the higher the temperature since carriers move faster and spend less time near impurities. Impurity scattering is therefore less significant at higher temperatures. The mobility term due to impurity scattering, μ_I, varies as $T^{3/2}/N_T$, where N_T is the total impurity concentration [58]. The temperature dependence of mobility is represented by the following expression:

$$\mu_n(T) = \mu_n(T_0)\left(\frac{T}{T_0}\right)^X \qquad \text{…....eq. (2.25).}$$

where the factor X takes into account the compensating contributions of lattice and impurity scattering. Typically $X \approx -1.5$ but it is process-dependent and must be derived from data.

2.3.3 Temperature Dependence of Transconductance and Noise

Transconductance and noise are related to mobility by the following expressions:

$$g_m(T) \propto \mu_n(T) \qquad \text{…....eq. (2.26).}$$

$$S_V \propto \sqrt{\frac{T}{g_m}} \qquad \text{…….eq. (2.27).}$$

Eq. (2.27) suggests that the noise should decrease with decreasing temperature.

2.4 Summary

Two major types of radiation damage affect CMOS processes: total dose ionising radiation effects and single event effects (SEU, SEL, SEGR). The current trend in device scaling introduces inherent total dose radiation hardness to commercial CMOS technologies. In particular, tunnelling into the thin gate oxide of the latest processes neutralises the oxide trapped charge induced by ionising radiation. Leakage current paths between drain and source and between adjacent transistors can be reduced through the adoption of special design rules such as enclosed NMOS transistor layouts and the extensive use of p^+ guard rings. These design rules along with common features of modern CMOS processes such as shallow trench isolation, thin epitaxial substrates and retrograde wells ensure that SEL sensitivity is not a concern for CMS electronics. SEGR is a problem which usually affects the operation of power MOSFETs where the electric fields across the oxide are high. Experimental data suggest that SEGR susceptibility actually decreases with decreasing oxide thickness and that this damage mechanism should not pose a threat to the operation of the 0.25 μm CMOS process. SEUs will occur in many of the CMS electronics components and must be accounted for in the final systems. One important parameter for SEUs is the rate at which they occur since this determines the reloading or resetting of components that are affected.

The need to monitor the change in the electrical characteristics of MOS devices with temperature was identified. A crucial parameter in the design of the CMS silicon microstrip tracker is the signal-to-noise ratio. This is expected to improve in the CMS environment (compared with room temperature measurements) since the APV25 chip will be operated at –10 ^{o}C.