Electrical Characterization of Defects in Gate Dielectrics

Dieter K. Schroder
Dept. of Electrical Engineering and Center for Solid State Electronics Research
Arizona State University
Tempe, AZ 85287-5706

ABSTRACT
Defects in gate oxides/insulators have been characterized by many techniques, e.g., electrical, electron beam, ion beam, X-rays, neutron activation analysis, electron spin resonance, and others. These methods played a crucial role during the early MOS development to determine the origin of oxide/interface charges that led to unstable MOS devices. Thermally-grown oxides on silicon are now quite well understood and very well controlled during manufacturing. High-K dielectrics, however, have proven to be less well controlled and more difficult to characterize. I give a brief description of the common defects (oxide, border and interface traps) and then describe many of the more common electrical characterization techniques.

1. OXIDE DEFECTS
In this section I will briefly describe the main defects commonly observed in insulators, some in thermally-grown SiO₂, others in non SiO₂ insulators, and some generated during irradiation of MOS devices. I will also briefly touch on some aspects of oxide breakdown as it pertains to oxide defects.

1.1 FIXED OXIDE CHARGE
Fixed oxide charge was identified early during MOS development and was attributed to excess silicon near the SiO₂/Si interface in thermally-grown oxides. The model was that as an oxide grows on a Si wafer, oxygen diffuses through the growing oxide to react at the interface forming SiO₂. This leaves some excess Si in the oxide near the interface. As the growing oxide front moves into the wafer, the excess Si moves with it and defects associated with this excess Si become positively charged during negative bias stress. The fixed oxide charge is not in electrical communication with the underlying silicon. These positively-charged defects were given the symbol Qss and later Qf. They were originally referred to as slow states and later Deal called them fixed surface states.

Qf depends on the final oxidation temperature. The higher the oxidation temperature, the lower is Qf. However, if it is not permissible to oxidize at high temperatures, it is possible to lower Qf by annealing the oxidized wafer in a nitrogen or argon ambient after oxidation. This has resulted in the well-known "Deal triangle", which shows the reversible relationship between Qf and oxidation and annealing. An oxidized sample may be prepared at any temperature and then subjected to dry oxygen at any other temperature, with the resulting value of Qf being associated with the final temperature.

The nature of Qf is still not entirely clear. In one model, some of the trivalent Si interface traps are tied up in some form of Si-Si bonds that do not interact with the adjacent oxide network bonding. During H annealing at moderate temperatures (400-500°C), Si-Si bonds break, one terminated with H and the other free. The free bond interacts with a neighboring oxygen to form an over-coordinated oxygen with a fixed positive charge. When the free bond of the Si atom bonds to a neighboring oxygen atom, the electron can tunnel to the adjacent Si substrate leaving behind a positively charged O³⁺ center.

1.2 MOBILE OXIDE CHARGE
Mobile charge in SiO₂ is due primarily to the ionic impurities Na⁺, Li⁺, K⁺, and perhaps H⁺. Sodium is the dominant contaminant. Lithium has been traced to oil in vacuum pumps and potassium can be introduced during chemical-mechanical polishing. The practical application of MOSFETs was delayed due to mobile oxide charges in the early 1960s. MOSFETs were very unstable for positive gate bias but relatively stable for negative gate voltages. Sodium was the first impurity to be related to this gate bias instability. By intentionally contaminating MOS capacitors (MOS-Cs) and measuring the gate voltage shift after bias-temperature stress, it was shown that alkali cations could easily drift through thermal SiO₂ films. Chemical analysis of etched-back oxides by neutron activation analysis and flame photometry was used to determine the Na profile. The drift has been measured with the isothermal transient ionic current, the
thermally stimulated ionic current, and the triangular voltage sweep methods.6

\[ T = \frac{1000}{T} \]

\[ V_{G} = 10 \text{ V} \]

\[ t_{ox} = 100 \text{ nm} \]

**Fig. 1 Transit times for Na, Li, K, and Cu for an oxide electric field of 10^6 V/cm. Reprinted with permission of John Wiley & Sons, Inc.**

The mobility of oxide contaminants is given by the expression

\[ \mu = \mu_{0} \exp(-E_{A} / kT) \]  

(1)

where for Na: \( \mu_{0}=3.5 \times 10^{-4} \text{ cm}^2/\text{V·s} \) (within a factor of 10) and \( E_{A}=0.44 \pm 0.09 \text{ eV} \); for Li: \( \mu_{0}=4.5 \times 10^{-4} \text{ cm}^2/\text{V·s} \) (within a factor of 10) and \( E_{A}=0.47 \pm 0.08 \text{ eV} \), for K: \( \mu_{0}=2.5 \times 10^{-3} \text{ cm}^2/\text{V·s} \) (within a factor of 8) and \( E_{A}=1.04 \pm 0.1 \text{ eV} \), and for Cu, \( \mu_{0}=4.8 \times 10^{-7} \text{ cm}^2/\text{V·s} \) and \( E_{A}=0.93 \pm 0.2 \text{ eV} \).8

If the oxide electric field is taken as \( V_{G}/t_{ox} \), neglecting the small voltage drops across the semiconductor and gate, the drift velocity of mobile ions through the oxide is \( \nu_{d}=\mu V_{G}/t_{ox} \) and the transit time \( t_{t} \) is

\[ t_{t} = \frac{t_{ox}^{2}}{\nu_{d}} = \frac{t_{ox}^{2}}{\mu V_{G}} = \frac{t_{ox}^{2}}{\mu_{0} V_{G}} \exp(E_{A} / kT) \]  

(2)

Equation (2) is plotted in Fig. 1 for the three alkali ions and for Cu. For this plot the oxide electric field is 10^6 V/cm, a common oxide electric field for such measurements, and the oxide thickness is 100 nm. For thinner or thicker oxides, the transit times change according to Eq. (2). Typical measurement temperatures lie in the 200 to 300°C range and only a few milliseconds suffice for the charge to transit the oxide. Mobile charge densities in the 5 \times 10^9-10^{10} cm\(^{-2}\) range are generally acceptable in integrated circuits.

1.3 OXIDE TRAPPED CHARGE

Oxide trapped charge is due to charge trapped in the oxide. This charge is most commonly electrons and/or holes injected during device operation or during radiation experiments.

1.4 E’ CENTER

The E’ center, usually observed in irradiated MOS devices, consist of two Si atoms joined by a weak, strained Si-Si bond with a missing oxygen atom, sometimes referred to as an oxide vacancy, shown in Fig. 2. It is one of the most dominant radiation-induced defects. E’ centers also pre-exist in oxide films due to the amorphous nature of SiO\(_{2}\) and thermodynamic considerations. Each Si atom is back bonded to three oxygen atoms. It is believed that when a positive charge is captured, the Si-Si bond breaks. Feigl et al. argued that the lattice relaxation is asymmetrical with the positively charged Si relaxing into a planar configuration, away from the vacancy and the neutral Si relaxing toward the vacancy.9 The annealing characteristics of E’ centers have been correlated with positive oxide charge.10
1.5 NEUTRAL ELECTRON TRAPS

Several models have been proposed to explain oxide breakdown. One of these is the electron trap generation model, based on the principles of percolation theory. This model, originally suggested by Massoud and Deaton and later verified by other groups, assumes that neutral electron traps are randomly generated in the oxide during oxide stressing. It is assumed that traps are continuously generated during oxide stress until there is a sufficient number of traps somewhere in the device that a continuous, conducting path is formed across the oxide and breakdown occurs. The percolation model can explain the reduced trap density required for breakdown and the reduced Weibull slope as the oxide becomes thinner. The latter has an important influence on the area dependence of breakdown. If the neutral electron traps capture electrons, they lead to flatband and threshold voltage shifts. That is, in fact, how they are detected, by filling them with electrons.

Oxide breakdown exhibits a surge in current or a sudden drop voltage during stress measurements. It has a “weakest link” character or extreme value statistics. The statistical description is based on the Weibull distribution model in which the cumulative failure probability $F$ is given by

$$F(t_{BD}) = 1 - \exp\left(-\frac{t_{BD}}{\alpha}\right)^{\beta}$$

where $t_{BD}$ is the time-to-breakdown, $\alpha$ the $t_{BD}$ at the 63rd percentile, and $\beta$ the Weibull shape factor or the Weibull slope. Equation (3) is usually written in the form

$$\ln(-\ln(1 - F(t_{BD}))) = \beta \ln\left(\frac{t_{BD}}{\alpha}\right)$$

with a plot of $\ln(-\ln(1-F))$ versus $\ln(t_{BD})$ yielding a straight line. Equations (3) and (4) also apply when the time-to-breakdown is replaced by the charge-to-breakdown $Q_{BD}$. The area dependence is

$$t_{BD}(A_1) = t_{BD}(A_2)(A_2/A_1)^{1/\beta}$$

where $A_1$ and $A_2$ correspond to two different areas. Equation (5) shows that the area dependence is not simply linear because it depends on the shape factor $\beta$, which in turn depends on the oxide thickness, i.e., as $\beta$ decreases, the area dependence becomes stronger. This makes it very important to specify the area during breakdown measurements. The reduction of $\beta$ with decreased oxide thickness is attributed to a reduced number of defects required to trigger a breakdown.

Two techniques have been used to measure the neutral electron trap density. An indirect measure is stress-induced leakage current and a more direct measure is substrate hot electron injection followed by a measurement of the threshold voltage or flatband voltage shift. Both are described in Section 2.

The nature of neutral electron traps is still somewhat ambiguous. A possible defect structure is the following. It is well established that the E’ center is formed by breaking the Si-Si bond in an oxygen vacancy defect, illustrated in Fig. 3(a). The bond breaking is facilitated by capture of a hole (Fig. 3(b)), leaving a positively-charged trap and one Si atom with a dangling orbital containing one unpaired electron. The resonant flipping of the spin of this unpaired electron gives rise to the E’ signal in electron spin resonance. Upon electron capture, the center can return to the E’ center or the electron from one of the Si atoms decays to a ground state by joining the unpaired electron of the other Si atom forming a neutral amphoteric trap (Fig. 3(c)).

Capturing a second electron leaves it negatively charged (3(d)). It is this electron trapping event that gives rise to the threshold voltage shifts associated with filled neutral electron traps following electron injection measurements. Attempts to anneal neutral traps have been
only partially successful. The usual 400-450°C/30 m forming gas anneal only anneals a portion of the traps. High-pressure forming gas was more successful in annealing most of the traps. The hydrogen-anneal model is illustrated in 3(e). Once such a trap is annealed by hydrogen capture, the Si-H bonds may break leading to trap creation, which may be a precursor to oxide breakdown.

Fig. 3 (a) Neutral hole trap (E' center, weak Si-Si bond, oxygen vacancy), (b) positive charge (Eγ' center), (c) neutral electron-hole trap, (d) negatively-charged trap, (e) hydrogen-annealed trap.

1.6 INTERFACE TRAPPED CHARGE

Interface trapped charge, also known as interface states, interface traps, and fast surface states, exist at the SiO2/Si interface. They are the result of a structural imperfection. Silicon is tetrahedrally bonded with each Si atom bonded to four Si atoms in the wafer bulk. When the Si is oxidized, the bonding configuration at the surface is as shown in Fig. 4(a) and 4(b) with most Si atoms bonded to oxygen at the surface. Some Si atoms bond to hydrogen, but some remain unbonded. An interface trap is an interface trivalent Si atom with an unsaturated (unpaired) valence electron usually denoted by Si3 ≡ Si•, where the “≡” represents three complete bonds to other Si atoms (the Si3) and the “•” represents the fourth, unpaired electron in a dangling orbital (dangling bond). Interface traps, also known as Pb centers,21 are designated as Dn (cm-2eV-1), Qn (C/cm2), and Nn (cm-2). The Pb ESR spectrum was first observed by Nishi22 and later identified by Poindexter et al. as a paramagnetic dangling bond.23,24

On (111)-oriented wafers, the Pb center is situated at the Si/SiO2 interface with its unbonded central-atom orbital perpendicular to the interface and aimed into a vacancy in the oxide immediately above it, as shown in Fig. 4(a). On (100)-oriented Si, the four tetrahedral Si-Si directions intersect the interface plane at the same angle. Two defects, named Pb1 and Pb0 and shown in Fig. 4(b), have been detected by electron spin resonance. A recent calculation suggests the Pb1 center to be an asymmetrically oxidized dimer, with no first neighbor oxygen atoms.25 By 1999, it was unambiguously established that both Pb0 and Pb1 are chemically identical to the Pb center.26 However, there is a charge state difference between these two centers indicating Pb0 is electrically active, while some authors believe the Pb1 to be electrically inactive.27 Interface traps are electrically active defects with an energy distribution throughout the Si band gap. They act as generation/recombination centers and contribute to leakage current, low-frequency noise, and reduced mobility, drain current, and transconductance. Since electrons or holes occupy interface traps, they become charged and contribute to threshold voltage shifts. The surface potential dependence of the occupancy of interface traps is illustrated in Fig. 5.
Interface traps at the SiO₂/Si interface are acceptor-like in the upper half and donor-like in the lower half of the band gap. Hence, as shown in Fig. 5(a), at flatband, with electrons occupying states below the Fermi energy, the states in the lower half of the band gap are neutral (occupied donors designated by “0”), those between midgap and the Fermi energy are negatively charged (occupied acceptors designated by “-”), and those above E_F are neutral (unoccupied acceptors). For an inverted p-MOSFET, shown in Fig. 5(b), the fraction of interface traps between mid gap and the Fermi level is now unoccupied donors, leading to positively charged interface traps (designated by “+”). Hence interface traps in p-channel devices in inversion are positively charged, leading to negative threshold voltage shifts.

![Fig. 5 Band diagrams of the Si substrate of a p-channel MOS device showing the occupancy of interface traps and the various charge polarities for a p-substrate with (a) negative interface trap charge at flatband and (b) positive interface trap charge at inversion. Interface traps are either occupied by electrons (solid circle) or holes, shown by the open circles.](image)

1.7 BORDER TRAPS
In 1980, a committee headed by Bruce Deal established the nomenclature for charges associated with the SiO₂/Si system, i.e., interface trapped, fixed oxide, mobile ionic and oxide trapped charge. In 1992, Dan Fleetwood suggested that this list be augmented by including border traps also designated as slow states, near-interfacial oxide traps, E’ centers, switching oxide traps, and others. He proposed border traps to be those near-interfacial oxide traps located within approximately 3 nm of the oxide/semiconductor interface. There is no distinct depth limit, however, and border traps are considered to be those traps that can communicate with the semiconductor through capture and emission of electrons and/or holes.

Oxide, border, and interface traps are schematically illustrated in Fig. 6(a). Defects at or near the SiO₂/Si interface are distributed in space and energy and communicate with the Si over a wide range of time scales. While for interface traps, the communication of substrate electrons/holes with interface traps is predominantly by capture/emission, for border traps it is mainly by tunneling from the semiconductor to the traps and back. Figure 6(b) shows the flatband diagram with interface and border traps occupied by electrons to the Fermi level E_F. The border traps are shown over a wide energy range for illustrative purposes only. The band diagram in Fig. 6(c) applies immediately after V_G1 is applied, before unoccupied border and interface traps have captured electrons. Interface traps now capture electrons from the conduction band, indicated by (ii) and inversion electrons tunnel to border traps, indicated by (i). Tunnel process (i) is followed by electron capture of lower energy border traps. In Fig. 6(d) interface and border traps up to E_F are occupied by electrons through (i) electron capture and (iii) tunneling. For −V_G2 in Fig. 6(e), electrons tunnel from border traps to the conduction band (iv), interface traps (v) and the valence band (vi). The insulator electric field, shown as constant in these figures, will, of course, distort as the border trap occupancy changes. I have disregarded this change here to bring out the main points.

Inversion electron tunneling is a direct tunnel process with time constant

$$\tau_i \approx \tau_0 \exp(x/\lambda), \quad \lambda = \frac{\hbar}{\sqrt{8m^*\phi_B}}$$

where \(\tau_0\) is a characteristic time (≈10⁻¹⁰ s), \(\lambda\) the attenuation length (≈10⁻⁸ cm), \(m^*\) the tunneling effective mass, and \(\phi_B\) the barrier height at the semiconductor/insulator interface. \(\tau_i\) varies from 0.01 to 1 s (100 to 1 Hz) for \(x\) varying from 1.8 to 2.3 nm. Hence border traps can be determined to a depth of approximately 2.5 nm from the SiO₂/Si interface by measurements for frequencies as low as 1 Hz. Lower frequencies, of course, allow deeper traps to be characterized showing that the trap depth that can be characterized.
depends on the measurement frequency. Such measurements include low-frequency noise, conductance, frequency-dependent charge pumping, and others. The valence band hole tunnel times are longer than for electrons due to the higher effective mass and barrier height. Tewksbury and Lee give a more detailed discussion of tunneling.33

Fig. 6 (a) Schematic of oxide, border, and interface traps, (b) flatband, (c) capture of electrons by interface traps and tunneling of electrons to border traps from conduction band, (d) border and interface trap occupied by electrons (e) electron tunneling from border traps. The solid circles represent occupied and the open circles unoccupied traps.

Tunneling from the conduction band into border traps was questioned, as measurements did not support energy dissipation in the oxide.34 Tunneling from interface traps is a two-step process: the electron must be captured from the conduction band before it can tunnel.35 The capture time is

\[ \tau_c = \frac{1}{\sigma n v_{th} n} \]  

where \( \sigma_n \) is the capture cross section, \( v_{th} \) the thermal velocity and \( n \) the inversion electron density. For strong inversion \( n=10^{18}-10^{19} \) cm\(^{-3} \) and using \( \sigma_n=10^{-16} \) cm\(^2\) and \( v_{th}=10^7 \) cm/s, \( \tau_c=10^{-9}-10^{-10} \) s. Since the capture and tunnel processes proceed in series, to first order the time constant is

\[ \tau_{it} = \tau_c + \tau_t \]  

and the tunnel time constant dominates for all but the shallowest border traps.

1.8 INTERFACE BETWEEN TWO DIFFERENT INSULATORS

For a device consisting of two insulators of thicknesses \( t_1 \) and \( t_2 \) and dielectric constants \( K_1 \) and \( K_2 \) on a semiconductor, charge accumulates at the interface between the two insulators as a result of differing conductivities. When a gate voltage is applied to such a two-layer structure, the two dielectrics will begin to conduct with current densities \( J_1 \) and \( J_2 \). Since the conductivities of the two layers differ from each other, \( J_1 \neq J_2 \) initially, leading to interfacial charge density \( Q \) accumulation at the interface. Eventually, the system reaches steady state with the same current density flowing through the entire gate stack.

This phenomenon is known as the Maxwell-Wagner instability.36 With a Maxwell–Wagner instability, it is not the usual static dielectric constants that determine the electric fields of the two layers in steady state, but rather the requirement that the same current density flow through each. Thus, the field in dielectric 1 may differ from \( VK_2(t_1K_2+t_2K_1) \) by an amount depending on the magnitude of \( Q \). The interfacial charge density \( Q \), and from it the electric fields, could be calculated if the current densities \( J_1 \) and \( J_2 \) were known. Without this knowledge, \( VK_2(t_1K_2+t_2K_1) \) is only an approximation. However, when the dielectric constants of high-K dielectrics are determined experimentally, the interfacial charge is
usually ignored, with the two layers of the gate stack being treated as capacitors in series and the “effective dielectric constants” include the effect of the interfacial charge.

In this brief defect discussion, I have at times referred to radiation-induced defects. Such defects have been fertile ground for studying defects and developing characterization techniques because they can be introduced and annealed at will and their densities can be very high, facilitating measurements. In contrast, oxide defects in conventional ICs are generally of low density and frequently more difficult to characterize. Although radiation-induced defects often differ from those induced during normal IC processing, such processing can introduce its own form of radiation defects. After all, plasma and reactive ion etching and ion implantation are radiation sources that may introduce such defects.

2. MEASUREMENTS

2.1 CAPACITANCE – VOLTAGE

Theory

Capacitance-voltage measurements have played an important role in MOS characterization. They can be found in some of the earliest MOS-related papers. Frankl, in 1961, proposed the use of MOS capacitor C-V curves to analyze such devices. In 1962, Terman used C-V measurements to determine surface state densities. One of the first comprehensive papers was by the Fairchild group in 1964, the same group that played a large role in understanding and developing MOS technology. Figure 7 shows one of those early C-V curves. Barrier heights for various gate metals were characterized with C-V measurements by the same group. A good early overview of the variety of material/device parameters that can be determined from C-V and C-t measurements is given by Zaininger and Heiman. And, of course, the entire MOS capacitor field is very well covered in the well-known book by Nicollian and Brews.

Three insulator-related parameters typically determined with C-V measurements are: oxide charge density, \( \rho_{ox}(x) \), interface trap density, \( D_{it} \), and gate-semiconductor work function difference, \( \phi_{MS} \). The various charges are illustrated in the device cross section in Fig. 8(a). They are determined from the flatband voltage

\[
V_{FB} = \phi_{MS} - \frac{1}{K_{ox} \varepsilon_o} \int_0^{t_o} x \rho(x) dx
\]

(9)

The fixed charge density \( Q_i \) and interface trap charge density \( Q_{it} \) are assumed to be located at the SiO\(_2\)/Si interface and the charge density \( \rho_{ox}(x) \) in the SiO\(_2\), leading to the flatband voltage expression

\[
V_{FB} = \phi_{MS} - \frac{1}{K_{ox} \varepsilon_o} \int_0^{t_o} x Q_i \delta(t_{ox}) dx - \frac{1}{K_{ox} \varepsilon_o} \int_0^{t_o} x \rho_{ox}(x) dx - \frac{1}{K_{ox} \varepsilon_o} \int_0^{t_o} x Q_{it}(\phi_s) \delta(t_{ox}) dx
\]

(10)

where \( t_{ox} \) is the oxide thickness, \( \delta \) the delta function, and \( Q_{it} \) is a function of surface potential \( \phi_s \), as shown in Fig. 5. \( V_{FB} \) can be written as
The oxide charge consist of mobile oxide charge (Na, K, etc.), oxide trapped charge (electrons and holes), and any other charge that may reside within the oxide.

For uniform oxide charge density, Eq. (11) becomes

$$V_{FB} = \phi_{MS} - \frac{Q_f}{K_{ox} \varepsilon_o} t_{ox} - \frac{1}{K_{ox} \varepsilon_o} t_{ox} \int_0^{t_{ox}} x \rho_{ox}(x) dx - \frac{Q_{it} (\phi_s)}{K_{ox} \varepsilon_o} t_{ox} \Rightarrow \text{intercept} = \phi_{MS}$$

Equation (12) is plotted in Fig. 9(a) assuming constant $Q_{it}$. This plot is clearly non-linear making it difficult to extract the various charges. However, $\phi_{MS}$ is given by the $V_{FB}$ intercept. Differentiating Eq. (12) with respect to oxide thickness gives

$$\frac{dV_{FB}}{dt_{ox}} = -\frac{Q_f + Q_{it}}{K_{ox} \varepsilon_o} - \frac{\rho_{ox}}{K_{ox} \varepsilon_o} \Rightarrow \text{intercept} = -\frac{Q_f + Q_{it}}{K_{ox} \varepsilon_o}, \text{ slope} = -\frac{\rho_{ox}}{K_{ox} \varepsilon_o}$$

and plotted in Fig. 9(b), yielding $Q_f + Q_{it}$ and $\rho_{ox}$. The fixed oxide charge and interface trap densities cannot be determined independently, only their sum. However, $Q_{it}$ can be measured independently by other techniques.

Frequently the oxide charge density in thermally-grown SiO$_2$ in Eq. (12) is low and can be neglected. In that case, the $V_{FB}$ expression in Eq. (12) becomes linear with $t_{ox}$. An early example is shown in Fig. 10(a), where the thick-oxide $V_{FB} - t_{ox}$ plots show the effect of varying $\phi_{MS}$ and $Q_f + Q_{it}$. Curve A clearly yields a different intercept than the other three because it is p-Si while the other three refer to n-Si.
Curves B to D show the effect of different $Q_f$. A more recent thin-oxide example is shown in Fig. 10(b) where the flatband voltage and gate work functions are extracted from $V_{FB} - t_{ox}$ plots.\(^{44}\)

For two-layer insulators of Fig. 8(b), consisting of a gate, a high-K insulator, SiO$_2$, and a substrate, the flatband voltage becomes

$$V_{FB} = \phi_{MS} - \frac{1}{K_{ox}\varepsilon_o} \int_0^{EOT} x\rho(x)dx$$

where $\phi_{MS}$ is the gate/substrate work function difference, $\rho(x)$ the charge density in the two insulators and at their interfaces and EOT, the equivalent oxide thickness, is

$$EOT = t_{ox} + EOT_{HK} = t_{ox} + \frac{K_{ox}}{K_{HK}} t_{HK}$$

where $t_{ox}$, $t_{HK}$, $K_{ox}$ and $K_{HK}$ are the oxide and high-K insulator thicknesses and dielectric constants. The oxide flatband voltage due to fixed charge density $Q_{f,ox}$ at the SiO$_2$/Si interface and charge density $\rho_{ox}(x)$ in the SiO$_2$ is

$$V_{FB}(SiO_2) = -\frac{1}{K_{ox}\varepsilon_o} \int_0^{EOT} xQ_{f,ox}\delta(EOT)dx - \frac{1}{K_{ox}\varepsilon_o} \int_0^{EOT} x\rho_{ox}(x)dx$$

If the bottom insulator is not SiO$_2$, then $t_{ox}$ is replaced by $EOT_{ins}=(K_{ox}/K_{ins})t_{ins}$, where $K_{ins}$ and $t_{ins}$ are insulator dielectric constant and thickness. Since the charge distribution in insulators is usually not known, it is frequently assumed to be constant, giving

$$V_{FB}(SiO_2) = -\frac{Q_{f,ox}}{K_{ox}\varepsilon_o} EOT - \frac{\rho_{ox}}{2K_{ox}\varepsilon_o} EOT^2 + \frac{\rho_{ox}}{2K_{ox}\varepsilon_o} EOT_{HK}^2$$

Similarly for the high-K insulator with fixed charge density $Q_{f,HK}$ at the HK/SiO$_2$ interface and charge density $\rho_{HK}(x)$ in the high-K insulator, $V_{FB}$ is

$$V_{FB}(HK) = -\frac{1}{K_{ox}\varepsilon_o} \int_0^{EOT_{HK}} xQ_{f,HK}\delta(EOT_{HK})dx - \frac{1}{K_{ox}\varepsilon_o} \int_0^{EOT_{HK}} x\rho_{HK}(x)dx$$

giving

$$V_{FB}(HK) = -\frac{Q_{f,HK}}{K_{ox}\varepsilon_o} EOT_{HK} - \frac{\rho_{HK}}{2K_{ox}\varepsilon_o} EOT_{HK}^2$$

The total flatband voltage is
\[ V_{FB} = \phi_{MS} + V_{FB} (SiO_2) + V_{FB} (HK) \]  

(20)

It is obvious from these equations that it becomes very difficult to determine all four charge components.

To get an idea of the magnitude of each of the components, I consider one particular case. \( t_{ox}=0.5 \) nm, \( t_{HK}=5 \) nm, \( K_{ox}=3.9 \), \( K_{HK}=25 \) ⇒ \( EOT_{HK}=0.78 \) nm, and \( EOT=1.28 \) nm. \( N_{f,ox}=5\times10^{10} \) cm\(^{-2} \) and \( N_{f,HK}=5\times10^{12} \) cm\(^{-2} \) give \( Q_{f,ox}=8\times10^{-9} \) C/cm\(^2 \), \( Q_{f,HK}=8\times10^{-7} \) C/cm\(^2 \), \( N_{o,ox}=10^{16} \) cm\(^{-3} \) and \( N_{o,HK}=5\times10^{19} \) cm\(^{-3} \) ⇒ \( \rho_{ox}=0.0016 \) C/cm\(^3 \) and \( \rho_{HK}=8 \) C/cm\(^3 \). These charge densities correspond approximately to recent experimental data.\(^45,46 \) Substituting into the flatband equations gives

\[ V_{FB} (SiO_2) = \frac{-8\times10^{-3} \times 1.28\times10^{-7}}{3.45\times10^{-13}} - \frac{0.0016 \times 1.64\times10^{-14}}{6.9\times10^{-13}} + \frac{0.0016 \times 6.1\times10^{-15}}{6.9\times10^{-13}} = -3.05 \text{ mV} \]

\[ V_{FB} (HK) = \frac{-8\times10^{-7} \times 0.78\times10^{-7}}{3.45\times10^{-13}} - \frac{8 \times 6.1\times10^{-15}}{6.9\times10^{-13}} = -251 \text{ mV} \]

Clearly the HK charges contribute a much higher flatband voltage shift. Hence, the charges associated with the thin SiO\(_2\) film are frequently ignored during flatband voltage analyses of high-K samples.

Equation (20) is plotted in Fig. 11, where in Fig. 11(a) \( t_{ox} \) is held constant and \( t_{HK} \) is allowed to vary, while in 11(b) \( t_{HK} \) is constant and \( t_{ox} \) varies. The points in Fig. 11(b) are experimental data with the extracted charges shown in the inset. Note the very different \( V_{FB} - EOT \) behavior for the two cases. This is mainly due to the different thicknesses for these two cases. For \( EOT=8 \) nm, in Fig. 11(a) \( t_{ox}=1 \) nm and \( t_{HK}=45 \) nm, while in 11(b) \( t_{ox}=7 \) nm and \( t_{HK}=6 \) nm. The much thicker \( t_{HK} \) in 11(a) has a more severe effect on \( V_{FB} \) than 11(b).

To measure these devices the oxide is sometimes grown to a certain thickness and then etched to varying thicknesses across the wafer. In one method, the oxidized wafer is immersed and slowly withdrawn from a 0.34% HF/H\(_2\)O solution at a constant withdrawal rate yielding a beveled oxide across the wafer.\(^45 \) In another method, the oxide is step etched.\(^47,48,49 \)

Fig. 11 Flatband voltage versus equivalent oxide thickness (a) \( t_{ox} \) fixed, vary \( t_{HK} \) (b) \( t_{HK} \) fixed, vary \( t_{ox} \).

Data in the inset after Kaushik et al. for a TaN/TiN/HfSiO\(_x\)/SiO\(_2\) structure.\(^45 \)

**Interface Traps**

MOS capacitance measurements are made at high and low frequencies. Low-frequency C-V measurements are referred to as quasi-static measurements first demonstrated in 1968-70\(^50 \) to measure interface traps. I have calculated the effects of interface traps on C-V\(_G\) curves to illustrate what one may expect. Fig. 12(a) shows the assumed interface trap density as a function of surface potential for these calculations. This distribution approximates the interface trap density distribution at the SiO\(_2\)/Si interface with \( D_{it,min} \) at midgap. Figure 12(b) shows the surface potential versus gate voltage behavior without and with interface traps. The discontinuity at \( \phi_s=0.4 \) V is the result of the assumption of \( D_{it} \) being donors in the upper half and acceptors in the lower half of the band gap. In real devices there is a more gradual transition. The interface trap density results in a “stretch-out” of the \( \phi_s-V_G\) characteristic because charged interface traps lead to gate voltage shifts.
Figure 12 (a) Interface trap distribution, (b) surface potential versus gate voltage.

\[ N_A = 10^{17} \text{ cm}^{-3}, \; t_{ox} = 5 \text{ nm}, \; D_{it,min} = 2.4 \times 10^{-11} \text{ cm}^{-2}\cdot\text{eV}^{-1}. \]

Figure 13 shows the effect of \( D_{it} \) on (a) low- and (b) high-frequency C-V\(_G\) curves. I assume \( Q_{ox} = 0 \) in these calculations in order to bring out the effects of interface traps. \( Q_{ox} \) would lead to a parallel shift of the curves along the \( V_G \) axis. I assume that at high frequencies, the interface traps do not follow the ac gate voltage and do not contribute a capacitance and the "stretch-out" is solely due to the effect of \( D_{it} \) on \( V_G \). The acceptor states in the upper half of the band gap lead to a positive gate voltage shift in inversion whereas the donor states give the negative \( V_G \) shift in accumulation and depletion. For the low-frequency curve, on the other hand, there is both "stretch-out" and additional capacitance, as the interface traps are assumed to be able to follow the ac gate voltage. Figure 13(c) shows experimental data before and after gate oxide stress, with the stress generating \( D_{it} \). Note how the curve broadens as predicted shifting to the left in region A and to the right in region B indicative of amphoteric interface traps. This curve indicates donor states over slightly more than the lower half of the band gap.

The basic theory, developed by Berglund, compares a low-frequency C-V curve with one free of interface traps.\(^{51}\) The latter can be a theoretical curve, but is usually an hf C-V curve determined at a frequency where interface traps are assumed not to respond. "Low frequency" means that interface traps and minority carrier inversion charges must be able to respond to the measurement ac probe frequency. The interface trap density is determined from the If capacitance data according to

\[
D_{it} = \frac{1}{q^2} \left( \frac{C_{ox} C_{If}}{C_{ox} - C_{If}} - C_S \right)
\]

where \( C_{It} \) is related to the interface trap density \( D_{it} \) by \( D_{it} = C_{It}/q^2 \). We should mention here that most publications use \( C_{it} = qD_{it} \). With \( D_{it} \) in the usual units of \( \text{cm}^{-2}\cdot\text{eV}^{-1} \) and \( q \) in Coul, the units for \( C_{it} \) in this expression are \( \text{F/cm}^2\cdot\text{Coul} \), suggesting that the correct definition should be \( C_{it} = q^2 D_{it} \). A simplified approach, proposed by Castagné and Vapaille, uses measured If and hf C-V curves as\(^{52}\)

\[
D_{it} = \frac{C_{ox}}{q^2} \left( \frac{C_{If}}{1 - C_{If}/C_{ox}} - \frac{C_{hf}}{1 - C_{hf}/C_{ox}} \right)
\]

Equation (22) gives \( D_{it} \) over only a limited range of the band gap, typically from the onset of inversion, to a surface potential towards the majority carrier band edge where the ac measurement frequency equals the inverse of the interface trap emission time constant.
Fig. 13 Theoretical (a) low-frequency, (b) high-frequency C-VG (Dit,min=2.4x10^{11} \text{ cm}^{-2}\text{eV}^{-1}), (c) experimental data. \phi_s is the surface potential.

**Border Traps**

C-V_G curves can also be used to gain information on oxide trapped charge. For example, the “hump” sometimes observed in quasi-static data, illustrated in Fig. 14, has been attributed to electron tunneling from the channel into border traps.\(^{53}\) This device has an oxide/nitride/oxide gate insulator and the trap density in the nitride or at the nitride/oxide interface is increased by hot carrier stress. Traps within several kT about the trap energy participate during the ac measurement, leading to an additional capacitance. Its magnitude depends on the measurement frequency. Lower frequencies lead to higher “humps” as carriers can tunnel deeper into the insulator.

High-frequency C-V_G measurements also show a capacitance increase for accumulated n-substrate MOS capacitors and inverted n-MOSFETs with Hf-based insulators, attributed to tunneling from the substrate into border traps.\(^{54}\) It appears that the charge carriers tunneling through the thin interfacial oxide must be electrons due to their smaller effective mass and barrier height compared to holes. The dielectric capacitance increase at low frequencies, typically 1 – 100 kHz, has been proposed as a frequency- and voltage-dependent border trap capacitance C_{bt} in parallel with the interface trap capacitance C_{it}. C_{bt} contributes to the measured capacitance only if the border trap charging/discharging can follow the applied ac signal. The frequency/voltage dependence of C_{bt} yields the tunneling distance from the Si substrate and trap energy depth in the HfO_2.

Related to C-V_G measurements are C-t measurements, commonly used to determine the generation lifetime.\(^{55}\) In this technique, the MOS-C is pulsed into deep depletion and subsequently relaxes to its equilibrium state through electron-hole pair generation. The generation lifetime so determined is a measure of the substrate purity. This technique is well understood but irregularities appear when the flatband voltage changes during the recovery time. This can occur as a result of
carrier injection from the semiconductor into the insulator as, for example, in HfO$_2$. An initial capacitance undershoot has been attributed to electron tunneling into HfO$_2$ traps and interface trap generation.$^{56}$

![Fig. 14 MOSFET normalized capacitance before and after hot carrier injection of 10$^{16}$ electrons/cm$^2$. (Cohen, N.L. et al., Observation and characterization of near-interface oxide traps with C-V techniques, IEEE Trans. Electron Dev. 42, 2004, 1995; © IEEE)](image)

2.2 CONDUCTANCE

**Interface Traps**

The conductance method, proposed by Nicollian and Goetzberger in 1967, yields D$_{it}$ in the depletion and weak inversion portion of the band gap, the capture cross-sections for majority carriers, and information about surface potential fluctuations.$^{57}$ The technique is based on measuring the equivalent parallel conductance of an MOS-C as a function of bias voltage and frequency. The interface traps are detected through the loss resulting from changes in their occupancy due to changes in the gate voltage during the ac measurement.$^{58}$ A small ac gate voltage leads to interface trap occupancy change in an energy interval of a few kT wide centered on the Fermi level by capture or emission of majority carriers. This capture/emission causes energy loss leading to a conductance given by

$$\frac{G_P}{\omega} = \frac{q \omega \tau_{it} D_{it}}{1 + (\omega \tau_{it})^2}$$  \hspace{1cm} (23)

where $\omega = 2\pi f$, $\tau_{it} = R_{it} C_{it}$, the interface trap time constant, given by $\tau_{it} = [v_{th} \sigma p N_A \exp(-q\phi_s/kT)]^{-1}$, and $C_{it} = q^2 D_{it}$. Equation (23) is for interface traps with a single energy level in the band gap. Interface traps at the SiO$_2$-Si interface, however, are continuously distributed in energy throughout the Si band gap leading to a time constant dispersion and giving the normalized conductance as

$$\frac{G_P}{\omega} = \frac{q D_{it}}{2 \omega \tau_{it}} \ln[1 + (\omega \tau_{it})^2]$$  \hspace{1cm} (24)

The conductance is measured as a function of frequency and plotted as $G_P/\omega$ versus $\omega$. $G_P/\omega$ has a maximum at $\omega = 1/\tau_{it}$ and at that maximum $D_{it} \approx 2.9G_P/q\omega$.

For thin oxides, there may be appreciable oxide leakage current. In addition, the device has series resistance leading to

$$\frac{G_P}{\omega} = \frac{\omega (G_c - G_m) C_{ox}^2}{G_c^2 + \omega^2 (C_{ox} - C_c)^2}$$  \hspace{1cm} (25)

where

$$C_c = \frac{C_m}{(1 - r_s G_m)^2 + (\omega r_s C_m)^2}, \quad G_c = \frac{\omega^2 r_s C_m C_c - G_m}{r_s G_m - 1}$$  \hspace{1cm} (26)

where $G_t$ represents the tunnel conductance, $r_s$ the series resistance and $C_m$ and $G_m$ the measured capacitance and conductance.

**Border Traps**
Absent tunneling to border traps, the $G_p/\omega$ peak is symmetrical about the maximum peak frequency. However, a low-frequency conductance ledge is sometimes observed for insulators containing oxide traps. This was first reported by Eaton and Sah who observed a marked asymmetry in their conductance spectrum and attributed this to long time constants arising from carriers tunneling from interface traps into oxide traps. The conductance spectrum in Fig. 15 shows such a low-frequency ledge. Such ledges are usually not observed when measurements are confined to frequencies higher than 0.1-1 kHz.

2.3 BIAS-TEMPERATURE STRESS

In the bias-temperature stress (BTS) method the MOS device is heated to 150 to 250°C, and a gate bias to produce an oxide electric field of around $10^6$ V/cm is applied for 5-10 min. for any mobile oxide charge to drift to one oxide interface. The device is then cooled to room temperature under bias and a C-V$_G$ curve is measured. The procedure is then repeated with the opposite bias polarity. The mobile charge is determined from the flatband voltage shift, according to the equation

$$Q_m = -C_{ox} \Delta V_{FB}$$

To distinguish between oxide trapped charge and mobile charge, a BTS test is done with positive gate voltage. For oxide electric fields around 1 MV/cm mobile charge drifts, but the electric field is insufficient for appreciable charge injection. If the C-V$_G$ curve shifts after BTS, it is due to positive mobile charge. For higher gate voltages, there is a good chance that electrons and/or holes can be injected into the oxide and mobile charge may also drift, making that measurement less definitive. For positive gate voltage, mobile charge drift leads to negative while charge injection leads to positive flatband voltage shifts.

![Fig. 15 Gp/\omega versus f for an MOS-C exhibiting a Dit peak and N$_{ox}$ ledge. The lines indicate the contributions from D$_{it}=1.46\times10^{11}$ cm$^{-2}$eV$^{-1}$ and N$_{ox}=1.3\times10^{17}$ cm$^{-3}$eV$^{-1}$. Reused with permission from M. J. Uren, S. Collins, and M. J. Kirton, Appl. Phys. Lett., 54, 1448 (1989). Copyright 1989, American Institute of Physics.]

2.4 TRIANGULAR VOLTAGE SWEEP

In the triangular voltage sweep (TVS) method the MOS device is held at an elevated, constant temperature of 200 to 300°C and the low-frequency C-V$_G$ curve is obtained by measuring the current in response to a slowly varying ramp gate voltage. TVS is based on measuring the charge flow through the oxide at an elevated temperature in response to an applied time-varying voltage. If the ramp rate is sufficiently low, the measured current is the sum of displacement and conduction current due to the mobile charge. If, at $-V_{G1}$, all mobile charges are located at the gate-oxide interface and at $+V_{G2}$ all mobile charges are located at the semiconductor-oxide interface, the mobile charge is determined by the area under the lf curve according to

$$\int_{-V_{G1}}^{V_{G2}} \left[ \frac{I}{C_f} - \alpha \right] C_{ox} dV_G = \alpha Q_m$$

The hf and lf C-V curves coincide at high temperatures except for the lf “hump”, due to mobile charge...
drifting through the oxide and illustrated in Fig. 16.\textsuperscript{63} Fig. 16(a) illustrates the effect of different mobile charge densities and 16(b) the effect of temperature. Clearly the temperature for this sample must be least 120°C for all of the mobile charge to move.

![Graph of Capacitance vs. Gate Voltage](image1)

(a)

![Graph of Capacitance vs. Gate Voltage](image2)

(b)

Fig. 16 $C_{nf}$ and $C_{hf}$ measured at (a) $T=250°C$ and (b) various temperatures. The mobile charge density is determined from the area between the two curves. Reprinted with permission of John Wiley & Sons, Inc.

![Graph of Ion Current vs. Gate Voltage](image3)

Fig. 17 Ion current normalized by oxide current versus gate voltage. The ion moves to the gate. $T=423°C$, $\alpha=0.513$ V/s. Reused with permission from M. W. Hillen, G. Greeuw, and J. F. Verweij, \textit{J. Appl. Phys.}, \textbf{50}, 4834 (1979). Copyright 1979, American Institute of Physics.

Sometimes two peaks are observed in $I$-$V_G$ curves at different gate voltages. These have been attributed to mobile ions with different mobilities. For an appropriate temperature and sweep rate, high-mobility ions (\textit{e.g.}, Na\textsuperscript{+}) drift at lower oxide electric fields than low-mobility ions (\textit{e.g.}, K\textsuperscript{+}).\textsuperscript{64} Hence, the Na peak occurs at lower gate voltages than the K peak, illustrated in Fig. 17. Such discrimination between different types of mobile impurities is not possible with the bias-temperature method. This also explains why sometimes the total number of impurities determined by the BTS and the TVS methods differ. In the BTS method one usually waits long enough for all the mobile charge to drift through the oxide. If in the TVS method the temperature is too low or the gate ramp rate is too high, it is possible that only one type of charge is detected. For example, it is conceivable that high-mobility Na\textsuperscript{+} drifts but low-mobility K\textsuperscript{+} does not. The TVS method also lends itself to mobile charge determination in \textit{interlevel dielectrics}, since a current or charge is measured instead of a capacitance.

\subsection{2.5 DEEP LEVEL TRANSIENT SPECTROSCOPY}

\textit{Bulk Traps}
Deep-level transient spectroscopy (DLTS), introduced by Lang in 1972, is commonly used to determine densities, energies, and capture cross sections of bulk and interface traps in semiconductors and sometimes for border traps. The device capacitance, current, or charge is measured as a function of time after pulsing the device between zero/forward and reverse bias. The traps capture electrons/holes which they subsequently emit, leading to the time-dependent capacitance

\[
C(t) = C_0 \left[ 1 - \frac{N_T}{2N_D} \exp \left( -\frac{t}{\tau_e} \right) \right]
\]

with the electron emission time constant \( \tau_e \) depending on temperature as

\[
\tau_e = \frac{\exp((E_c - E_T)/kT)}{\gamma_n \sigma_n T^2}
\]

where \( \gamma_n \) is a constant for a given semiconductor. Determining the time constant at various temperatures allows the energy level \( E_T \), density \( N_T \), and capture cross section \( \sigma_n \) to be determined.

Example DLTS spectra of iron-contaminated Si are shown in Fig. 18. Iron forms Fe-B pairs in boron-doped p-type Si with a DLTS peak at \( T \approx 50 \) K. When the sample is heated at 180-200°C for a few minutes, the Fe-B pairs dissociate into interstitial iron \( \text{Fe}_i \) and substitutional boron and the DLTS peak for \( \text{Fe}_i \) occurs around \( T \approx 250 \) K. After a few days the interstitial iron again forms Fe-B pairs and the “Fe-B” peak returns while the “Fe-B” peak shrinks as shown in Fig. 18.

**Interface Traps**

The instrumentation for interface trapped charge DLTS is identical to that for bulk deep-level DLTS, but the data interpretation differs because interface traps are distributed in energy through the band gap. We illustrate the interface trapped charge majority carrier DLTS concept for the MOS-C in Fig. 19(a). For a positive gate voltage most interface traps are occupied by majority electrons for n-substrates (Fig. 19(b)). A negative gate voltage drives the device into deep depletion, causing electrons to be emitted from interface traps (Fig. 19(c)). Although electrons are emitted over a broad energy spectrum, emission from interface traps in the upper half of the band gap dominates. DLTS is very sensitive, allowing interface trap density determination in the mid 10^9 cm^-2eV^-1 range.

Interface trap characterization by DLTS was first implemented with MOSFETs. Being three-terminal devices, they have an advantage over MOS capacitors. By reverse biasing the source/drain and pulsing the gate, majority electrons are captured and emitted without interference from minority holes that are collected by the source-drain. This allows interface trap majority carrier characterization in the upper half of the band gap. With the source-drain forward biased, an inversion layer forms, allowing interface traps to be filled with minority holes. Minority carrier characterization is then possible and the lower half of the band gap can be explored. This is not possible with MOS-Cs because there is no minority carrier.
source other than thermal generation.

\[ E_c \]
\[ E_v \]
\[ E_F \]

(a)

Fig. 19 (a) Flatband, (b) majority carrier capture and (c) majority carrier emission from interface traps.

MOS capacitors are, nevertheless, used for interface trap characterization. Unlike the conductance technique, DLTS measurements are independent of surface potential fluctuations. The derivation of the capacitance expression is more complex for MOS-Cs than it is for diodes. We quote the main results whose derivations can be found in Johnson and Yamasaki et al. For \( q^2 D_{it} = C_{it} << C_{ox} \) and \( \delta C = C_{it}(t_1) - C_{it}(t_2) << C_{hf} \)

\[
\delta C = \frac{C_{hf}^2}{K_\epsilon N_D C_{ox}} \int_{-\infty}^{\infty} D_{it} \left( e^{-t_2/\tau_e} - e^{-t_1/\tau_e} \right) dE_{it}
\]

(31)

where \( \tau_e \) is the electron emission time, \( t_1 \) and \( t_2 \) the sampling times, and \( E_{it} \) the interface trap energy. If \( D_{it} \) varies slowly in the energy range of several \( kT \) around \( E_{it,max} \), it can be considered reasonably constant and can be taken outside the integral of Eq. (31) leading to

\[
D_{it} = -\frac{K_\epsilon N_D C_{ox}}{kT C_{hf}^2 \ln(t_2/t_1)} \delta C
\]

(32)

If the sample contains bulk as well as interface traps, it is possible to differentiate bulk traps from interface traps by the shape and the peak temperature of the DLTS plot. Interface trap densities determined by DLTS and quasi-static C-V are shown in Fig. 20.

Border Traps

DLTS has also been used to characterize border traps by using large amplitude filling pulses, allowing electrons not only to be captured by \( D_{it} \) but also to tunnel to border traps. This effect was observed for electron tunneling, but not for holes attributed to the higher barrier for holes at the SiO\(_2\)/Si interface. The spatial and energetic border trap distribution is obtained by varying the pulse width and amplitude. In this tunnel DLTS it is possible to distinguish between interface and border traps.
2.6 CHARGE PUMPING

Interface Traps

In the charge pumping (CP) method, originally proposed in 1969,\textsuperscript{73} the MOSFET source and drain are tied together and either grounded or slightly reverse biased. The time-varying gate voltage (square, triangular, trapezoidal, sinusoidal, or trilevel) is of sufficient amplitude for the surface under the gate to be driven between inversion and accumulation. The CP current is measured at the substrate, at the source/drain tied together, or at the source and drain separately. For the oxide-semiconductor band diagram in Fig. 21, the interface traps are shown by the filled circles representing electron-occupied interface traps and the open circles representing unoccupied traps. In Fig. 21(a) with the device in inversion, most of the interface traps are occupied by electrons. I show the Fermi level as a single level $E_F$ for simplicity. Of course, the device being in non-equilibrium during the CP measurement should be represented by two quasi-Fermi levels.

Fig. 21 Band diagrams for (a) inversion (b) flatband and (c) accumulation.
The transitions are discussed in the text.

When the gate pulse falls from its \textit{high} to its \textit{low} value during its finite transition time, most channel electrons drift to source and drain and those electrons on interface traps near the conduction band are thermally emitted into the conduction band (Fig. 21(b)) and also drift to source and drain. Those electrons on interface traps deeper within the band gap do not have sufficient time to be emitted and remain trapped. Once the hole barrier is reduced (Fig. 21(c)), holes flow to the surface where some are captured by those interface traps still occupied by electrons and all interface traps are occupied by holes. As the device reverts back to inversion some of the holes near the valence band are emitted to drift to the substrate (not shown) while those remaining on interface traps are annihilated by electron capture (Fig. 21(a)). Hence, only a fraction of interface traps, those toward the center of the band gap, are not emitted.
and participate in capture or recombination processes and lead to CP current. This energy interval depends on the rise and fall times of the CP waveform. The waveforms can be constant base voltage in accumulation and pulsing with varying voltage amplitude $\Delta V$ into inversion or varying the base voltage from inversion to accumulation keeping $\Delta V$ constant. The current saturates for the former, while for the latter it reaches a maximum and then decreases.

From Shockley-Read-Hall recombination statistics the occupancy of interface traps is determined by carrier capture and emission. Considering capture processes, the variation of the occupancy during one charge pumping cycle is

$$
\Delta F = \frac{[1 - \exp(-c_n / 2f)][1 - \exp(-c_p / 2f)]}{1 - \exp[-(c_n / 2f) - (c_p / 2f)]}
$$

(33)

where $c_n$ and $c_p$ are the capture coefficients for electrons and holes ($c_{n,p}=\sigma_{n,p}\nu_{th}$), $\sigma_{n,p}$ the capture cross sections and $\nu_{th}$ the thermal velocity, and $f$ the CP frequency. The charge pumping current then becomes

$$
I_{cp} = qAf \int_{E_{low}}^{E_{high}} E dE
$$

(34)

where $E_{high}$ and $E_{low}$ are the Fermi energies for high and low gate bias.

The basic charge pumping technique gives an average value of $D_n$ over the energy interval $\Delta E$.

Various refinements have been proposed to obtain energy-dependent interface trap distributions. For a sawtooth waveform, the recombined charge per cycle, $Q_{cp}=I_{cp}/f$, is given by

$$
Q_{cp} = 2qkTD_{it} f A_G \ln \left( \frac{\sigma_{n,\nu_{th}} \sigma_{p} \sqrt{1 - \zeta}}{\nu_{th} \nu_{th}} \right) \left( \frac{V_{FB} - V_{T}}{\Delta V_{GS} \ln 2} \right)
$$

(35)

where $T_{it}$ is the average interface trap density, $\Delta V_{GS}$ the gate pulse peak-peak amplitude, and $\zeta$ the gate pulse duty cycle. The slope of a $Q_{cp}$ versus log($f$) plot yields $D_n$ and the intercept on the log($f$) axis yields $(\sigma_n\sigma_p)^{1/2}$. By varying the gate waveform rise and fall times, one obtains the interface trap energy distribution. Han et al. using this technique for SiON/HfO$_2$ insulators, found higher $D_n$ in the upper half than in the lower half of the band gap. This also agreed with $V_T$, mobility, and subthreshold slope data, e.g., $V_T$ shift of n-MOSFETs was higher than the flatband voltage shift for these HfO$_2$ samples.

Fig. 22 Trilevel charge pumping waveform and corresponding band diagrams.
Reprinted with permission of John Wiley & Sons, Inc.

The interface trap distribution through the band gap and capture cross-sections can be determined with the trilevel waveform with an intermediate voltage level $V_{step}$, illustrated in Fig. 22. At point (a), the device is in strong inversion with interface traps filled with electrons. As the waveform changes to (b)
electrons are emitted from interface traps. The gate voltage remains constant to point (c). For $t_{\text{step}} > \tau_e$, where $\tau_e$ is the emission time constant of interface traps being probed, all traps above $E_T$ have emitted their electrons and only those below $E_T$ are available for recombination when holes come in to recombine with the electrons at point (d) on the waveform. This gives a charge pumping current that saturates as $t_{\text{step}}$ increases. For $t_{\text{step}} < \tau_e$, fewer electrons have time to be emitted and more are available for hole recombination giving a correspondingly higher charge pumping current.

A typical $I_{\text{cp}}$ versus $t_{\text{step}}$ plot in Fig. 23 shows the $I_{\text{cp}}$ saturation and the $t_{\text{step}} = \tau_e$ breakpoint. By varying $V_{\text{step}}$ one can probe interface traps through the band gap. $D_{it}$ is determined from the slope of the $I_{\text{cp}}$ versus $t_{\text{step}}$ curve according to the expression:

$$D_{it} = -\frac{1}{qkTA_G f} \frac{dI_{\text{cp}}}{d\ln t_{\text{step}}}$$  \hspace{1cm} (36)$$

Fig. 23 (a) $I_{\text{cp}}$ as a function of $t_{\text{step}}$ showing $\tau_e$ at the point where $I_{\text{cp}}$ begins to saturate. 

The charge pumping current is assumed to be due electron-hole pair recombination at interface traps. For thin oxides, gate current adds to the charge pumping current. The gate oxide leakage current can exceed $I_{\text{cp}}$. Fig. 24 shows the effect of gate oxide leakage current on $I_{\text{cp}}$. At sufficiently low frequencies, the gate leakage current dominates and can be subtracted from the total current.

Fig. 24 Charge pumping current versus base voltage for two voltage pulse heights before and after gate leakage current correction. $t_{\text{ox}}=1.8$ nm, $f=1$ kHz. (Bauza, D., Extraction of Si-SiO$_2$ interface trap densities in MOS structures with ultrathin oxides, IEEE Electron Dev. Lett. 23, 658, 2002; © IEEE)
**Border Traps**

In the presence of interface and border traps, carriers can tunnel into border traps as the charge pumping waveform frequency is reduced and \( Q_{cp} = I_{cp} / f \), which is constant during conventional interface trap CP, increases. Paulsen et al. characterized such traps by charge pumping experiments in which the recombined charge increases at low frequencies.\(^{81}\) This is attributed to charging/discharging of oxide traps within a tunneling distance of the \( \text{SiO}_2/\text{Si} \) interface. In their experiments they determined the trap distances to be 1.2-1.5 nm, based on tunneling time constants of \( 10^{-7} \) to \( 10^{-3} \) s.

This effect can be modeled by letting the oxide trap capture cross section be depth dependent\(^{82}\)

\[
\sigma_{n,p}(x) = \sigma_{n,p}(0) \exp(-x / \lambda)
\]  

(37)

where \( \sigma_{n,p}(0) \) is the interface electron/hole capture cross section and \( \lambda \) the attenuation length in Eq. (6).

The variation of the border trap occupancy during one charge pumping cycle, \( \Delta F_{ox} \), is that in Eq. (33) with a spatially varying \( \sigma_{n,p} \). The total charge pumping current then becomes

\[
I_{cp} = qAf \left( \int_{E_{high}}^{E_{low}} \int_{E_{high}}^{E_{low}} \Delta F_{ox} D_{bt}(E) dx dE \right)
\]

(38)

where \( D_{bt} \) is the border trap density (cm\(^{-3}\) eV\(^{-1}\)). \( D_{bt} \) is determined from a \( Q_{cp} \) vs. log\( f \) plot according to

\[
D_{bt} = -\frac{1}{q^2 A \lambda \Delta \phi_s} \frac{dQ_{cp}}{d \log f}
\]

(39)

where \( \Delta \phi_s \) is the surface potential change during one period and \( Q_{cp} \) the charge pumped per cycle. The distance over which oxide traps can be probed for CP where only the inversion part of the pulse leads to tunneling into traps is\(^ {83}\)

\[
x_m \approx -\lambda \ln \left( \frac{\sigma_n V_{th}}{2f} \right) \approx 2.2 \text{ nm}
\]

(40)

for \( \sigma_n = 10^{-16} \text{ cm}^2 \) and \( f = 1 \text{ kHz} \). This shows that traps to distances around 1-2 nm from the insulator/semiconductor interface can be probed. Example trap distributions for \( \text{SiO}_2 \) and \( \text{SiO}_2/\text{Al}_2\text{O}_3 \) insulators are shown in Fig. 25 illustrating the higher trap density in \( \text{Al}_2\text{O}_3 \) compared to \( \text{SiO}_2 \). A complication in these measurements is the possible gate insulator leakage current at low CP frequencies.

---

Fig. 25 Insulator trap density versus insulator depth from the insulator/Si interface for \( \text{Al}_2\text{O}_3 \) and \( \text{SiO}_2 \). (Jakschik S. et al., Influence of \( \text{Al}_2\text{O}_3 \) dielectrics on the trap-depth profiles in MOS devices investigated by the charge-pumping Method, *IEEE Trans. Electron Dev.* **51**, 2252, 2004; © IEEE)

In the Amplitude sweep Charge Pumping (ACP) technique, the base voltage is held constant and the pulse amplitude is gradually increased.\(^ {84}\) The charge-per-cycle increases with gate voltage amplitude if there is insulator trapping. By using charging and discharging pulses it is possible to obtain spectroscopic trap information. The trap energy levels in the insulator are determined by the \( \text{SiO}_2/\text{Si} \) interface acting as a retarding potential analyzer and sweeping the
discharge voltage to more negative voltages. This method showed the defect energy levels to be close to the Si conduction band.

There is sometimes a geometric component during CP measurements, due to those channel carriers unable to drift to source/drain during the accumulation portion of the pulse. Being unable to return to S/D from where they originated, they are injected into the substrate to recombine. This component is maximized by using long-channel MOSFETs and fast rise/fall times. Under these conditions the long channel is rapidly pinched-off, forcing a large fraction of the inversion charge to recombine in the Si where it is measured as a substrate current. By grounding both source and drain, the inversion charge density is \( N_{\text{inv},S/D} \), by grounding the source and floating the drain it is \( N_{\text{inv},S} \), given by

\[
N_{\text{inv},S/D} = N_{\text{inv}} - 2N_{S/D}, \quad N_{\text{inv},S} = N_{\text{inv}} - N_{S/D}
\]

where \( N_{\text{inv}} \) is the true inversion charge density and \( N_{S/D} \) is the geometric component. From Eq. (41)

\[
N_{\text{inv}} = 2N_{\text{inv},S} - N_{\text{inv},S/D}
\]

yielding the true \( N_{\text{inv}} \). This technique was used to determine the mobility for SiO2/HfO2 gate insulators, eliminating trapping effects.

2.7 MOSFET SUBTHRESHOLD SLOPE

Interface Traps

The effects of \( D_{\text{it}} \) on the MOSFET subthreshold \( I_D-V_G \) characteristics are illustrated in Fig. 26. Similar to the broadening of the C-V \( G \) curves in Fig. 13, the subthreshold characteristic also broadens and only the surface potential region between about midgap (\( \phi_s = \phi_F \)) and inversion can be probed. The drain current of a MOSFET in the subthreshold regime for drain voltages higher than about 4kT/q is

\[
I_D = \frac{W\mu_{\text{eff}}(kT/q)^2}{L} \sqrt{qK_e \varepsilon_p N_A} \exp\left(\frac{q(V_{GS} - V_T)}{nkT}\right)
\]

where \( n = 1 + (C_b + C_{it})/C_{ox} \), where \( C_b \) is the bulk (substrate) capacitance.

![Fig. 26 Theoretical \( I_D-V_G \) curves for \( D_{\text{it}}=0 \) and \( D_{\text{it},\text{min}}=2.7 \times 10^{10} \text{cm}^{-2}\text{eV}^{-1} \).](image)

The usual subthreshold plot is log(\( I_D \)) versus \( V_G \), with subthreshold swing \( S \) being that gate voltage necessary to change the drain current by one decade, given by

\[
S = \frac{\ln(10)nkT}{q} \approx \frac{60nT}{300} \text{ mV / decade}
\]

with \( T \) in Kelvin. The interface trap density is

\[
D_{\text{it}} = \frac{C_{ox}}{q^2} \left( \frac{qS}{\ln(10)kT} - 1 \right) - \frac{C_b}{q^2}
\]

requiring an accurate knowledge of \( C_{ox} \) and \( C_b \). The slope also depends on surface potential fluctuations. This is the reason that this method is usually used as a comparative technique in which the subthreshold swing is measured, then the device is degraded and remeasured. The \( D_{\text{it}} \) change is given by...
\[ \Delta D_{it} = \frac{C_{ox}}{\ln(10)q^2kT} \left( S_{after} - S_{before} \right) \]  

(46)

The subthreshold MOSFET curves are shown in Fig. 27 before and after stress, causing a threshold voltage shift and a slope change.

![MOSFET subthreshold characteristics before and after stress](image)

Fig. 27 MOSFET subthreshold characteristics before and after MOSFET stress. The change in slope results in a stress-generated \( \Delta D_{it}=5\times10^{11} \text{ cm}^2\text{eV}^{-1} \). The dashed curve is the “After stress” curve shifted to the left to coincide with the “Before stress” curve at \( I_D=10^{-13} \text{ A} \) to bring out the slope change. Reprinted with permission of John Wiley & Sons, Inc.

**Oxide Traps**

Subthreshold measurements are also made to determine oxide charge densities. When the surface potential coincides with the Fermi level, as shown in Fig. 28(a) by \( \phi_s=\phi_F \), interface traps in the upper and lower half in the band gap are neutral, and neither contributes to a gate voltage shift. The corresponding gate voltage is \( V_{mg} \), which is typically the gate voltage at \( I_D=0.1-1 \text{ pA} \). Increasing the gate voltage from \( V_{mg} \) to \( V_T \) fills interface traps in the upper half of the band gap with electrons (Fig. 28(b)). \( \Delta E \) usually covers the range from mid-gap to strong inversion. Since at mid-gap the interface traps do not contribute any voltage shift, a shift of \( V_{mg} \) must be due to oxide trapped charge according to \(^{87}\)

\[ \Delta V_{ot} = V_{mg2} - V_{mg1} \text{ and } \Delta N_{ot} = \frac{\Delta V_{ot} C_{ox}}{q} \]  

(47)

![Band diagrams for midgap and threshold voltages](image)

Fig. 28 Band diagrams for midgap and threshold voltages. Reprinted with permission of John Wiley & Sons, Inc.

**Border Traps**

The threshold voltage of high-K dielectric MOSFETs frequently exhibits an instability which has been explained by charging pre-existing defects in the HfO2 in HfO2/SiO2 dual dielectrics with a strong dependence on gate bias and charging (discharging) time.\(^{88}\) The measured drive current decay in n-
MOSFETs is caused by the continuous increase in $V_T$ due to a build up of negative charge, caused by trapping in the HfO$_2$. Comparing the pulsed technique with the dc measurement methods, shows the conventional dc techniques to underestimate the charging effects in SiO$_2$/HfO$_2$ dual layer gate dielectrics. An example of such dielectric charging is shown in Fig. 29, clearly illustrating the hysteresis for the transient measurements due to charge trapping/detrapping. Information about border traps can be gained from such measurements. The pulsed techniques was recently used to determine the capture cross sections in HfO$_2$ as $10^{-16}$-$10^{-14}$ cm$^2$.

![Graph showing ID-VG characteristics](image)

**Fig. 29** $I_D$-$V_G$ dc (open circles) and transient (100 µs pulse width, rise and fall times) characteristics showing higher transient current and hysteresis due to carrier trapping/detrapping. (Kerber, A. et al., Characterization of the $V_T$-instability in SiO$_2$/HfO$_2$ gate dielectrics, IEEE Int. Rel. Phys. Symp. 41, 41, 2003; © IEEE)

### 2.8 DC-IV

The DC-IV method is a dc current technique to determine $D_{it}$ illustrated in Fig. 30(a). With the source S forward biased, electrons injected into the p-well diffuse to the drain to be collected and measured as drain current $I_D$. Some electrons recombine with holes in the p-well bulk (not shown) and some recombine with holes at the surface below the gate with only the surface-recombining electrons controlled by the gate voltage. The recombined holes are replaced by holes from the body contact leading to body current $I_B$.

![MOSFET configuration for DC-IV measurements](image)

**Fig. 30** (a) MOSFET configuration for DC-IV measurements and (b) cross-sections showing the space-charge regions and the encircled surface generation regions.

Reprinted with permission of John Wiley & Sons, Inc.

The electron-hole pair surface recombination rate depends on the surface condition. With the surface in strong inversion or accumulation, the recombination rate is low. The rate is highest with the surface in depletion. The body current is given by

$$\Delta I_B = qA_G n_s s_r \exp(qV_{BS}/2kT)$$  \hspace{1cm} (48)

where $s_r$ is the surface recombination velocity given by

$$s_r = (\pi/2)\sigma_n v_n \Delta N_B$$  \hspace{1cm} (49)
with \( \sigma_0 \) the capture cross-section (assuming \( \sigma_n = \sigma_p = \sigma_0 \)).

Fig. 31 DC-IV measured body currents. Control wafer and stressed with –12 mA/cm\(^2\) gate current density. \( V_{BS}=0.3 \) V, \( W/L=20/0.4 \) \( \mu \)m, \( t_{ox}=5 \) nm. (Guan, H. et al., Nondestructive DCIV method to evaluate plasma charging in ultrathin gate oxides, *IEEE Electron Dev. Lett.* 20, 238, 1999; © IEEE)

When the gate voltage exceeds the flatband voltage, a channel forms between S and D and the drain current will increase significantly. For \( V_{GB}=V_T \), the \( I_D - V_{GB} \) curve saturates. If charge is injected into the oxide, leading to a \( V_T \) shift, the drain current will also shift. It is this shift that can be used to determine oxide charge. The interface trap density determined with the subthreshold slope method samples the band gap between midgap and strong inversion while the DC-IV body current samples the band gap between subthreshold and weak accumulation, i.e., surface depletion. By varying the gate voltage, different regions of the device are depleted (Fig. 30(b)) and those regions can be characterized, allowing spatial \( D_n \) profiling. Experimental DC-IV data are shown in Fig. 31 for a MOSFET before and after gate current stress.\(^9\) A clear peak is observed at maximum surface recombination around \( V_{GB}=0 \).

### 2.9 STRESS-INDUCED LEAKAGE CURRENT (SILC)

An effect frequently observed in thin electric field-stressed oxides is an enhanced gate oxide current - stress-induced leakage current, defined as the increase of oxide leakage current after high-field stress (\( \approx 10^{-12} \) MV/cm) compared to before stress and first reported in 1982.\(^9\) It is typically observed at low to moderate oxide electric fields (\( \approx 4-8 \) MV/cm) and increases markedly for thinner oxides. However, the SILC decrease for oxides thinner than about 5 nm, is believed to be due to reduced trap generation rates in thin oxides. Different models have been proposed to explain SILC: interface-state generation, bulk-oxide electron-trap generation, non-uniformities or weak spot formation in the oxide films, trapped holes injected from the anode. SILC can be best explained by the generation of neutral electron traps in the oxide, allowing more current to flow through the oxide layer by these traps acting as “stepping stones” by trap-assisted tunneling.\(^9\) The charge-to-breakdown is inversely related to the defect creation initial rate which can be measured at low injected charge where the defect density increases linearly with the injected charge.\(^9\) Hence, SILC can be used to obtain the oxide defect generation rate. Example SILC curves are shown in Fig. 32.\(^9\) SILC is usually not detected by time-zero or time-dependent oxide breakdown measurements. It has been proposed as a reliability monitor.\(^9\)

SILC has been used to determine the nitride trap density in Si/SiO\(_2\)/Si\(_3\)N\(_4\)/SiO\(_2\)/Si (SONOS) devices, where the oxide adjacent to the substrate is typically thicker (\( \approx 3-5 \) nm) than the oxide (\( \approx 1 \) nm) in SiO\(_2\)/high-K dual insulators to improve data retention in non-volatile memories. Unlike a MOS-C, the SILC in a SONOS exhibits a two-stage time dependence with the first stage exhibiting dc-like SILC characteristic and the second stage following a \( 1/t \) time dependence.\(^9\) The first stage leakage current is limited by stress-created oxide traps and the second stage by Frenkel-Poole emission of nitride trapped electrons. Through appropriate data analysis, the nitride trap density of \( 7 \times 10^{12} \) cm\(^{-2}\)eV\(^{-1}\) was found to be nearly constant with energy.
2.10 SUBSTRATE HOT ELECTRON INJECTION (SHE)

Substrate hot electron injection is the injection of electrons from Si into an oxide at low electric fields. One application of SHE is to make neutral electron traps “visible”, by charging them with trapped electrons. It has been shown that stressing gate oxide at moderate to high oxide electric fields (5-12 MV/cm) produces traps in the oxide. Since many of these traps are not detected by conventional measurements, e.g., threshold voltage shifts, it is believed that they are electron generated, but neutral when unoccupied by electrons. Once occupied by electrons, they lead to threshold and flatband voltage shifts and can be easily characterized. Electron injection into SiO₂ was studied initially by Ning because such injected electrons lead to device degradation and instability and is a powerful tool for studying electron and hole traps in thermal SiO₂ films.

![Diagram](image)

Fig. 33 Hot electron injection from leakage current in (a) and forward-biased diode in (b); (c) band diagram showing hot electron injection.

The concept of the measurement is illustrated in Fig. 33. With source and drain grounded, a positive gate voltage normally leads to electron channel formation for grounded substrate. When the substrate is reverse biased with respect to source/drain, the channel disappears. It is possible, however, for thermally-generated electrons accelerated towards the Si/SiO₂ interface, to have sufficient energy to surmount the barrier φₜ at the interface, shown in Fig. 33(a) and (c). These electrons have moderate energy and there is a good chance that they become trapped in the oxide. A problem with the arrangement in Fig. 33(a) is the low density of electrons for injection. The electrons are thermally generated, i.e., the
leakage current of the gate-induced space-charge region/substrate junction. Ning proposed to use a MOSFET with the p-substrate replaced by a p-epitaxial layer grown on an n-substrate. The resulting np junction can then be forward biased to inject electrons into the p-layer. Such a structure is a natural portion of MOSFETs fabricated in a p-well on an n-substrate. It is also possible to form an np junction beside the source or drain and forward bias it, shown in Fig. 33(b), or surrounding the MOSFET with an n-type implant. Either configuration allows independent control of the oxide electric field and the electron injection current.

2.11 CONSTANT GATE VOLTAGE OXIDE STRESS
Gate oxide integrity (GOI) characterization determines the quality of the gate oxide and does provide trap densities. I mention it briefly here, because it is important to determine the gate oxide quality. GOI is measured by one of several methods. In the time-zero method, the gate voltage is swept and the gate current is measured until the oxide breaks down. In the time-dependent methods, the gate voltage is held constant (CVS) and the gate current is measured, or the gate current is held constant (CCS) and the gate voltage is measured, or the gate current is stepped and the gate voltage is measured. For thin oxides it has been shown that the time-to-breakdown, $t_{BD}$, increases with decreasing oxide thickness for constant current stress while it decreases for constant voltage stress. For CCS, the oxide voltage changes during the measurement and critical defect density required for breakdown change. This makes the CVS technique the more appropriate method for GOI determination.

2.12 1/f NOISE
Noise measurements can be used to characterize semiconductors. The recent review papers by Wong and Claeys/Mercha/Simoen give a good overview of the present state of noise theory and measurements. At high frequencies, thermal noise and shot noise dominate. Both of these noises are fundamental in nature, forming an intrinsic lower noise limit. At low frequencies, flicker or 1/f noise dominates. Generation-recombination (G-R) noise can also occur in this frequency range. It is characterized by a Lorentzian spectrum with a constant plateau at $f< f_c$ and a $1/f^2$ roll off beyond the characteristic frequency $f_c$. In contrast to the fundamental thermal and shot noise, 1/f and G-R noise depend on material and semiconductor processing. Only 1/f noise is discussed here since it is sensitive to interface and border traps.

Low-frequency or flicker noise, first observed in vacuum tubes over eighty years ago, dominates the noise spectrum at low frequencies. It gets its name from the anomalous plate current “flicker”. Flicker noise is also called 1/f noise, because the noise spectrum varies as $1/f^n$, where the exponent $n$ is very close to unity. Fluctuations with a 1/f power law have been observed in many electronic materials and devices, including homogenous semiconductors, junction devices, metal films, liquid metals, electrolytic solutions, Josephson junctions, and even in mechanical, biological, geological, and even musical systems. Two competing models have been proposed to explain flicker noise: the McWhorter number fluctuation theory and the Hooge mobility fluctuation theory with experimental evidence to support both models. Christensson et al. were the first to apply the McWhorter theory to MOSFETs, using the assumption that the necessary time constants are caused by the tunneling of carriers from the channel into traps located within the oxide. Popcorn noise, sometimes called burst noise or random-telegraph-signal (RTS) noise, is a discrete modulation of the channel current caused by the capture and emission of a channel carrier.

RTS presents a serious problem to analog and digital devices as devices are scaled down. For example, in floating gate non-volatile memories the capture/emission of even a single electron affects the threshold voltage and drain current. RTS-induced drain current produces significant variation in “read” current. The threshold voltage change $\Delta V_T = q/A C_{ox}$ increases as the area, $A$, decreases. The trap properties leading to RTS can be studied by statistical analysis. One analysis procedure is based on the difference in the statistical properties of discrete Markovian telegraph fluctuations and Gaussian background noise. The average statistical lifetimes and amplitudes of the telegraph signal are then determined in an iterative way, allowing for analyzing “noisy” random telegraph signals with low ratio between the signal amplitude and the intensity of the background noise that cannot be analyzed by the classical approach. Separation of the time record enables an in-depth analysis of the spectral properties of the background noise observed together with the telegraph fluctuations.
The MOSFET current is proportional to the product of mobility \( \mu_{\text{eff}} \) times the charge carrier density or number \( N_s \). Low-frequency fluctuations in charge transport are caused by stochastic changes in either of these parameters, which can be independent (uncorrelated) or dependent (correlated). In most cases, fluctuations in the current, or more specifically in the \( \mu_{\text{eff}} \times N_s \) product are monitored, which does not allow the separation of mobility from number effects and therefore obscures the identification of the dominant 1/f noise source.

The voltage noise spectrum density is

\[
S_V(f) = \frac{q^2 k T \lambda}{\alpha W L C_{ox}^2 f} (1 + \sigma \mu_{\text{eff}} N_s)^2 N_{bt} \tag{50}
\]

where \( \lambda \) is the tunneling parameter, \( \mu_{\text{eff}} \) the effective carrier mobility, \( \sigma \) the Coulombic scattering parameter, \( N_s \) the density of channel carriers, \( C_{ox} \) the gate oxide capacitance/unit area, \( W L \) the gate area, and \( N_{bt} \) the border trap density (cm\(^{-2}\)eV\(^{-1}\)) near the interface. In weak inversion, the channel carrier density \( N_s \) is very low (10\(^7\)-10\(^{11}\) cm\(^{-2}\)), so that the mobility fluctuation contribution becomes negligible and the second term in the bracket can be neglected.

It is assumed that the free carriers tunnel to traps in the oxide with a tunneling time constant, which varies with distance \( x \) from the interface. The tunneling parameter and time constant are given in Eq. (6). For a trap at 1 nm from the semiconductor interface, \( \tau \approx 0.5 \) s or \( f = \frac{1}{2 \pi \tau} \approx 0.3 \) Hz. Hence, a distribution of traps in the oxide gives rise to a wide range of frequencies and can explain the 1/f dependence. 1/f noise shows sensitivity to the wafer orientation, which correlates with interface trap density. Figure 34 shows an example of low frequency noise before and after annealing with the anneal reducing the interface trap density and the 1/f noise.\(^{113}\) The main advantage of using a noise-based technique is that it can be applied even to very small area devices, which is not possible with capacitance-based DLTS, and it is well suited for characterization of electrically active defects in devices with no substrate contact, such as Si-wire transistors, carbon nanotubes, and others. Low-frequency noise has become a FA characterization technique, but the noise measurement equipment is quite complex.\(^{114}\) A recent low-frequency noise study of the role of fluorine in HfSiON gate insulators showed that the noise for both n- and p-MOSFETs was reduced by fluorine due to passivation of traps close to valence and conduction bands.\(^{115}\) By combining low-frequency noise with gate leakage currents measurements, information on the trap location within the insulator was gained.

![Fig. 34 Low-frequency noise spectra before and after annealing. W/L=10 \( \mu \)m/0.8 \( \mu \)m, \( t_{ox}=3.3 \) nm, \( V_G-V_T=-1.05 \) V, \( V_D=-0.005 \) V. (Ahsan, AKM and Schroder, D.K., Impact of post-oxidation annealing on low-frequency noise, threshold voltage, and subthreshold swing of p-channel MOSFETs, IEEE Electron Dev. Lett. 25, 211, 2004; © IEEE)](image)

2.13 ELECTRON SPIN RESONANCE
Electron spin resonance (ESR), also known as electron paramagnetic resonance (EPR), detects species with unpaired electron spins. However, it is almost always possible to render a defect ESR active by
charge injection. Thus, almost all atoms or molecules capable of charge capture can be observed in ESR after injection of the appropriate charge carriers. I give a more detailed discussion of ESR than other methods, because it is not commonly used except by a few experts and is not well understood by many “electrical characterization” engineers. Since ESR is limited to atoms or molecules with unpaired spins, most electronic materials are invisible to ESR. However, when ESR detects a species it is a powerful tool to characterize such defects. It proved its worth when it showed the nature of SiO2/Si interface traps to be P b centers. Its sensitivity is generally poorer than electrical techniques, but it is usually better than Auger electron spectroscopy, secondary ion mass spectrometry, X-ray crystallography and others.21 Although I-V, C-V, G-V, etc. techniques are generally more sensitive than ESR, they lack specificity in determining the atomic nature of the point defects. Hence ESR has clear advantages because it provides atomic identification of point defects and enables distinction between the various types of positive, negative or amphoteric charge traps.

The technique depends on the transitions induced between spin states by a magnetic field perpendicular to electromagnetic energy in the form of microwave signals at frequencies of typically 9-10 GHz. The spin states are induced by the large magnetic field while the transitions between spin states are induced by the radio frequency electromagnetic energy. The magnetic field-induced spin state splitting, proportional to the magnetic field, is illustrated in Fig. 35. A radio frequency signal incident on the sample containing the defect under study, leads to a transition from one spin state to another. Quantization of spin angular momentum leads to the energies of an isolated unpaired electron of

\[ E = \pm \frac{1}{2} g \mu_B B_0 \]  

where \( g \) is the free electron g-value (2.00232), \( \mu_B \) the Bohr magneton \( (9.27 \times 10^{-24} \text{ J/T}) \), and \( B_0 \) the magnetic field. The energy \( E \) is the Zeeman energy. The energy level separation \( \Delta E \) is

\[ \Delta E = hf = g \mu_B B_0 \]  

where \( f \) is the microwave frequency.

The measured deviation from the free electron g-value is very important in ESR measurements, because it provides information about the paramagnetic center’s electronic structure. An unpaired electron “feels” the external magnetic field \( B_0 \) and the effects of local magnetic fields. Hence the effective magnetic field is

\[ B_{\text{eff}} = B_0 (1 - \sigma) \]  

where \( \sigma \), which can be positive or negative, accounts for the local field effects. Hence, Eq. (52) becomes

\[ \Delta E = hf = g \mu_B B_{\text{eff}} = g \mu_B B_0 (1 - \sigma) = g \mu_B B_0 \]  

\( g \) is determined during ESR measurements by knowing \( B_0 \) and \( f \) at resonance, i.e.,

\[ g = \frac{hf}{\mu_B B_{\text{res}}} \]  

where \( B_{\text{res}} \) is the magnetic field at resonance, indicated in Fig. 35.

For a g-factor different from \( g \), the electron magnetic moment/angular momentum ratio, the electron has gained or lost angular momentum, since the magnetic moment (Bohr magneton) is constant. The energy gain/loss is due to spin-orbit coupling. With spin-orbit coupling well understood, the g-factor change provides information about the nature of the atomic or molecular orbital containing the electron. The measured g-value deviates from \( g \) because the local magnetic field differs from the applied magnetic field due to spin-orbit coupling and/or electron-nuclear hyperfine interactions.

Two measurement approaches are possible: (i) the magnetic field is constant and the microwave frequency is varied and (ii) the magnetic field is varied and the frequency is constant until resonance is achieved in either case. The latter is preferred. The ESR signal looks qualitatively like that in Fig. 36(a), showing a peak at a certain magnetic field. To determine the relevant magnetic field at resonance it is easier to plot the differentiated ESR signal versus magnetic field, as also illustrated in Fig. 36(a), because the signal is detected with a lock-in amplifier. Instead of the magnetic field at the peak, one simply determines the zero-point crossing, indicated by the point.
The ESR spectrometer consists of a frequency-stable radiation source provided by a Gunn diode and a magnet which is slowly swept to record a spectrum. By far the most widely used frequency is X-band, about 9.5 GHz, due to its relatively moderated magnetic fields and inexpensive Gunn diode microwave sources. Higher frequencies require higher magnetic fields, leading to higher sensitivity. The most common higher frequency is 35 GHz with a wavelength of 8 mm. Microwave components and samples are thus correspondingly much smaller than for X-band measurements.\textsuperscript{116} The sample is located in a resonant cavity with dimensions matching the wavelength of the radiation so that a standing wave pattern is set up in it. Varying the magnetic applied field to bring the sample to resonance, leads to microwave power absorption by the sample. Magnetic field modulation is essential in ESR to allow for lock-in detection with sensitivities in the mid $10^{10}$ cm$^{-2}$ range. The magnetic field is modulated at typically 100 kHz.

ESR has been extensively used to characterize defects in insulators and other materials. As mentioned earlier, it led to the identification of the nature of interface traps at the SiO$_2$/Si interface. Early irradiation studies used ESR to identify various radiation-induced defects.\textsuperscript{117} Defect spectra in SiO$_2$ are easily separated from those due to defects in Si since the resonance line shapes of SiO$_2$ defects are independent of magnetic field direction, whereas those in Si are not. The magnetic field orientation is due to the physical defect location at the interface (interface of a crystalline material, hence repeated oriented lattice structure) or in the oxide (randomly oriented amorphous material). \textit{Negative bias temperature instability} (NBTI) is a significant reliability issue in p-MOSFETs. It manifests itself as a threshold voltage and drain current degradation and interface trap generation is one of the culprits. ESR was used early on during NBTI recognition to follow P$_x$ center generation by negative gate voltage stress.\textsuperscript{118}

Corona charging is used routinely in semiconductor characterization.\textsuperscript{119} The method consists of depositing corona charge on the surface of the sample and measuring a variety of material/device parameters. The corona charge replaces the metal/poly-Si gate. On oxidized samples, it can generate oxide electric fields as high as 10-15 MV/cm. Since the deposited ions are thermalized by their passage...
through a long air path, corona charging would seem to constitute an almost ideal non-damaging biasing tool. For positive biasing, the charging process involves hydrogenic species that can interact with the Si/SiO₂ system. The impact of room temperature corona charging biasing on thermal (100) and (111) Si/SiO₂ has been probed by ESR. Both structures fully passivated in H₂ as well as exhaustively H-depleted ones were investigated. At least five types of ESR-active defects were generated most likely due to the interaction of atomic hydrogen released during the corona process with the Si/SiO₂ system. However, for lower corona-induced oxidized electric fields, this biasing technique does not damage the sample.

A variation of ESR is spin-dependent recombination (SDR). In conventional Shockley-Read-Hall recombination, electrons and holes are captured by deep-level impurities during a recombination event. In SDR, the semiconductor device is placed in a strong magnetic field which polarizes the spins of the dangling bond electrons as well as the conduction electrons and the holes. Consider a silicon dangling bond defect with an unpaired electron when it is electrically neutral. Let the SRH recombination process begin with electron capture by the defect. If both the dangling bond and electron have the same spin orientation, the conduction electron cannot be captured, because two electrons may not occupy the same orbital with the same spin quantum number. Thus, placing the semiconductor sample in the magnetic field modulates the trap capture cross-section by a combination of magnetic field and microwaves.

SUMMARY
First I have discussed some of the more common defects observed in MOS devices, divided into oxide, border, and interface traps. “MOS”, of course, is a misnomer, since the gate in today’s devices is not a metal and the insulator is rarely pure SiO₂. While the defects in pure SiO₂ are quite well understood, that is not the case for high-K/oxide insulators. Not only are high-K dielectrics more complex, there is an additional interface in such insulators and the interface and bulk defects are usually higher and less well understood. Then I have provided a fairly complete set of electrical characterization methods. In most cases, these techniques, first developed for SiO₂, were extended to more complex dielectrics. For some of the newer techniques I provide a brief theory, but for the well-known methods I only give appropriate references. Wherever possible, I have given application to interface, oxide and border trap measurements. Other techniques may well be developed in the future, but for the most part existing techniques, some of which have been around since the 1960s, have served the IC community well and will continue to do so. Some of the techniques must be used with caution for advanced ICs because gate oxide leakage current and other non-idealities make data interpretation more difficult.

REFERENCES
43 Werner, M.W., The work function difference of the MOS-system with aluminum field plates and polycrystalline silicon field plates, *Solid-State Electronics* 17, 769, 1974.


