Gate Oxide Breakdown

Navid Azizi, Peter Yiannacouras

December 2, 2003

1 Introduction

The gate oxide thickness, (t_{ox}), of Complementary Metal Oxide Semiconductor (CMOS) processes has been steadily been thinning as a result of technology trends. To keep the switching power dissipation of integrated circuits at bay, successive technology generations have relied on reducing the supply voltage. In order to maintain performance, and control short channel effects, however, a reduction in the transistor oxide thickness is required to provide sufficient current drive at the reduced supply voltages. At the $70nm$ technology node, CMOS processes will have oxide thicknesses of $1.2nm$ to $1.6nm$ [1].

Since electric fields in the gate oxide are expected to rise with scaling, the long-term reliability of thin oxides becomes an important concern in modern, deep-submicron $(< 0.10\mu\text{m})$ processes. In this paper the causes, symptoms, and failure models of gate-oxide breakdown will be presented.

Section 2 provides some relevant background. Section 3 explains what gate-oxide breakdown is and what are its causes. Section 4 explains some of the effects of gate-oxide breakdown in both transistors and circuits. In Section 5 different failure models for gate-oxide breakdown are presented. Section 6 goes over the statistics of gate-oxide breakdown in recent technologies. In Section 7 the use of CAD tools to predict gate-oxide breakdown will be explored, and in Section 8 design styles that reduce the probability of gate-oxide breakdown will be presented. Finally Section 9 concludes.

2 Background

This section will go over some information that will be needed to better understand the rest of this paper.

2.1 Transistor Structure

A cross section of a Negative-Channel Metal Oxide Semiconductor (NMOS) transistor is shown in Figure 1. The transistor is a four terminal device. The gate terminal
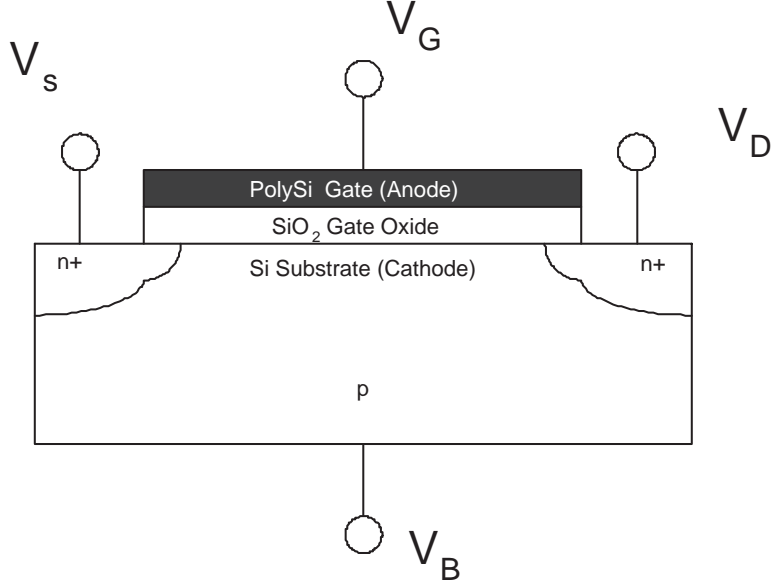
Figure 1: Cross Section of an NMOS

controls whether a channel is created in the p-doped Silicon substrate, connecting the n-doped Silicon drain and the source terminals. The fourth terminal (body) terminally is usually connected to ground in NMOSs or to V_{DD} in Positive-Channel Metal Oxide Semiconductors (PMOSs). The gate oxide is composed of Silicon-Dioxide and serves as an insulator between the gate and substrate. The substrate, is also composed of Silicon, but it is grown rather than crystallized (and therefore called polysilicon). The polysilicon gate is heavily doped to make it a conductor.

For most of the discussion in this paper, the flow of electrons is the tunneling current through the gate from the cathode (substrate) to the anode (gate).

2.2 Oxide Traps

Defects within the gate oxide are usually called traps; they are called traps because the degraded oxide can trap charges [2]. Traps are usually neutral, but quickly become positively charged near the anode, and negatively charged near the cathode [2].

3 Root Cause

Gate-oxide breakdown is defined as the time when a cluster of connected bonds, beginning from a "seed" at one interface of the gate-oxide reaches the opposite interface [3]. In this section, the evolution process of breakdown within gate oxides due to traps is first explained; then different forms of breakdown and their characteristics will be ex-
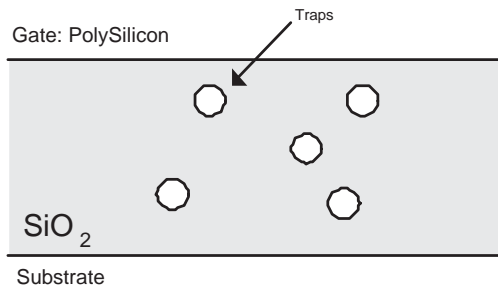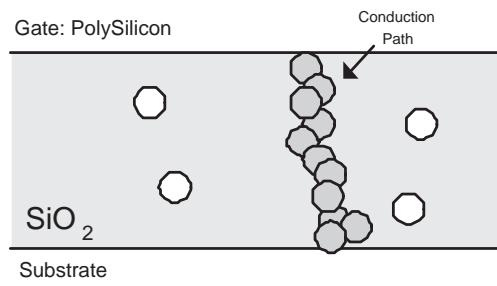
Traps

Gate: PolySilicon

SiO$_2$

Substrate

Figure 2: Formation of Traps in the Gate-Oxide

Conduction Path

Gate: PolySilicon

SiO$_2$

Substrate

Figure 3: Creation of Conduction Path through Traps in the Gate-Oxide

Increased Traps Due to Thermal Damage

Gate: PolySilicon

SiO$_2$

Substrate
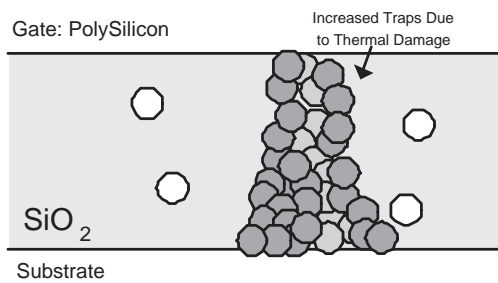
Figure 4: Increased Traps in Gate Oxide after Conduction

Silicon Filament

Gate: PolySilicon

SiO$_2$

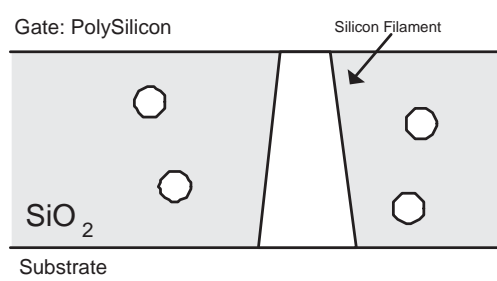Substrate

Figure 5: Cross Section of the Gate Oxide after Hard Breakdown

plained. Finally the different factors that induce the formation of traps within the gate oxide such as the electric field, hot carriers, and radiation will be explained in detail.

3.1 Gate-oxide Evolution Process

Gate-oxide breakdown begins when traps form in the gate-oxide[1]. At first the traps are non-overlapping and thus do not conduct (Figure 2), but as more and more traps are created in the gate-oxide, traps start to overlap creating a conduction path [4]. Once these traps form a conduction path from the gate to the channel, breakdown occurs [4], as shown in Figure 3. This type of breakdown is called Soft Breakdown (SBD).

Once there is conduction, new traps are created by thermal damage, which in turn allows for increased conductance as shown in Figure 4 [5]. The cycle of conduction leading to increased heat to increased conduction leads to thermal runaway [5] and finally to a lateral propagation of the breakdown spot [6]. The Silicon within the breakdown spot starts to melt, and Oxygen is released, and a Silicon filament is formed in the breakdown spot [5]. This type of breakdown is called Hard Breakdown (HBD) and can pictorially been seen in Figure 5.

Some photographs of the gate oxide after breakdown are shown in Figures 6 through 7.

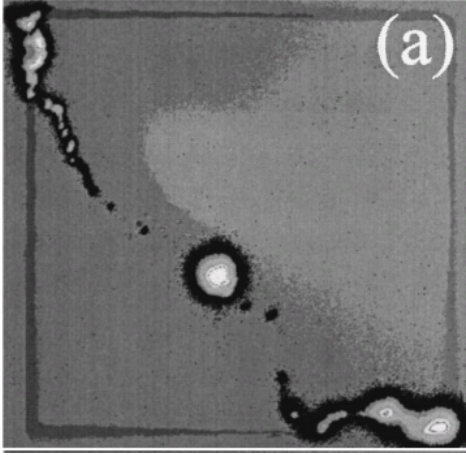[1]The source of these traps may be due to many processes as will be seen in Section 3.2.

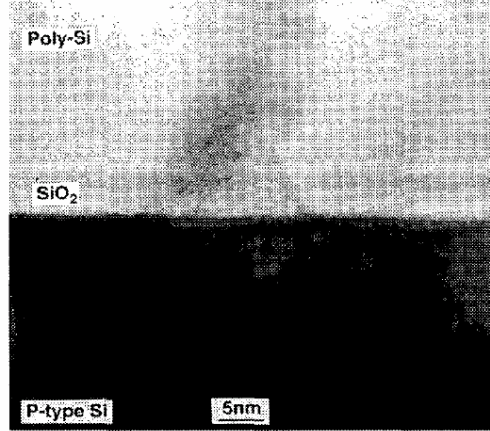Figure 6: EMMI Image of the Gate after Oxide Breakdown[7]

Figure 7: Cross Section of the Gate Oxide after Hard Breakdown [8]

Figure 6 shows the gate-oxide through Emission Microscopy (EMMI); light regions are the areas where photon emission is preferential and shows the areas of gate-oxide breakdown [7]. The dark region in Figure 7 shows the region of breakdown where the Silicon has melted [8]. Figure 8 shows a Transmission Electron Microscopy (TEM) micrograph of the bright spot in Figure 6. Finally Figure 9 shows a TEM micrograph of the substrate below the gate-oxide breakdown; there are many threading dislocations which are characteristic of substrate regions near a gate-oxide breakdown [7].

3.2   Trap Generation

There are many different theories on how defects are generated in the gate-oxide.

The two main models are the Thermochemical model (E model), and the Anode Hole Injection (AHI) model ($1/E$ model), and there has been an ongoing controversy as to which model is correct, as there is data to corroborate both models [9]. Studies have showed that at high stress fields, the $1/E$ model agrees well with experimental data, and at low electric fields the E model shows better agreement with the experimental data than the $1/E$ model.

Other models include the Hydrogen Release (HR) model, and the role of irradiation and channel hot-carriers. Furthermore, as a start to this section, the role of fabrication defects in the gate-oxide to breakdown will be discussed.

This section will provide the underlying physical interpretations of the model; a mathematical extrapolation of the models to reliability measures will follow in Section 5.
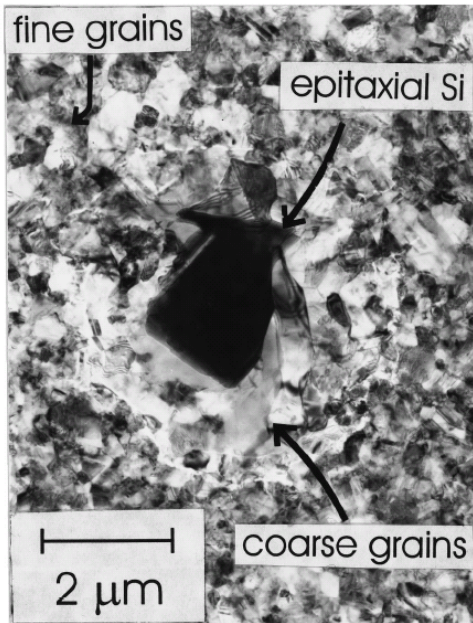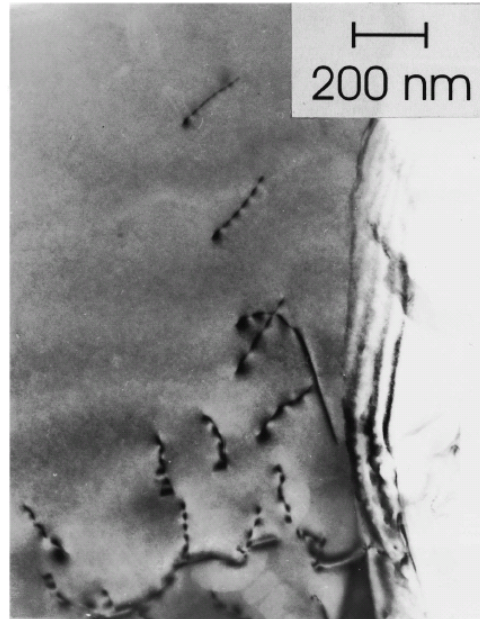
4

Figure 8: TEM Image of Bright Spot in Figure 6[7]

Figure 9: TEM of Substrate near Gate-Oxide Breakdown[7]

3.2.1 Fabrication Defects

Defects can be created in the oxide when the circuit is being fabricated. These defects are heavily dependent on the actual process. Defects can arise due to ion implantation during the creation of the gate, plasma damage by subsequent processing, by mechanical stress due to the isolation structure, or from process contaminants [10].

3.2.2 AHI Model

The AHI model is based on the process of electron injection into the oxide, which generates holes at the anode which then get trapped in the oxide [9]. At high fields, this model shows better agreement to experimental data because, at such fields the electron tunneling is significant, and hole generation dominates over the Thermochemical model discussed in Section 3.2.3 [9].

The process is as follows, at high electric fields, the electrons arriving at the gate have a high kinetic energy ($>$ $8MV/cm$) [11]. When these hot electrons reach the gate electrode they transfer their entire energy to a deep-valence band electron, and then this electron is promoted to the lowest available electron energy state, which is the conduction band edge of the anode [12]. Once the electron reaches the conduction band, it creates a hot hole, which tunnels into the oxide [12].

The holes which enter the oxide, allow for increased current density due to *hole-induced trap generation* [12]. Once the trap has been created, and there is increased
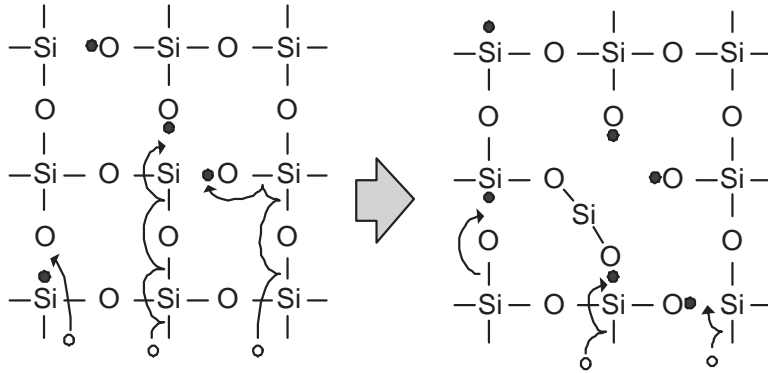
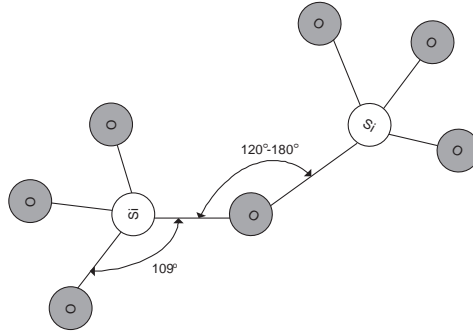Figure 10: SiO$_2$ Bond Breakage due to Hole Current [11]

Figure 11: Chemical Structure of SiO$_2$ [13]

current density, there are more high energy electrons entering the gate which can create more hot holes, and thus there is positive feedback until breakdown occurs [9].

A example of how a hole can create a trap in the oxide is shown in Figure 10, where the holes induce two bond breakages at a single Si atom.

3.2.3   Thermochemical Model

The $1/E$ model presented above ignores important thermal processes which degrade all materials over time [13], and some recent work which covers ranges of temperature and field conditions show better agreement with the E (Thermochemical model) [13].

Amorphous SiO$_2$'s primary structural unit is the SiO$_4$ tetrahedron shown in Figure 11. The angle between O-Si-O is always 109°, but the angle between the bond linking the tetrahedra ranges from 120° to 180° [13].

When the linking bond angle is above 150°, the bond's strength is severely weakened and an oxygen vacancy results where the Si-O-Si bond is replaced with a Si-Si bond as shown in Figure 12 [13]. This oxygen vacancy is thought to be the defect that leads to oxide breakdown [13].

The remaining O-Si bonds within the oxide are highly polar bonds; 70% of the total
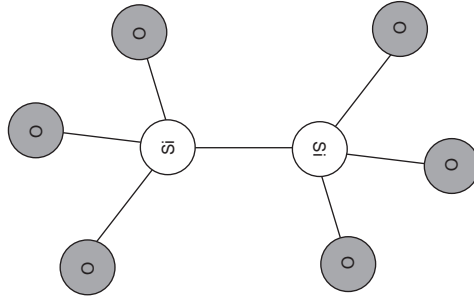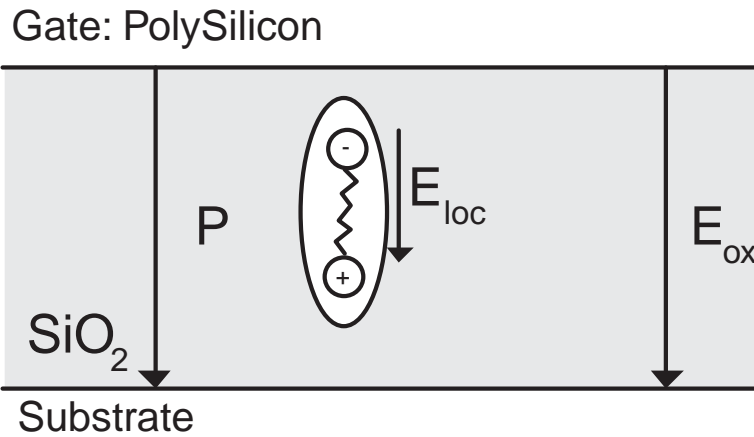
6

Figure 12: Oxygen Vacancy in SiO_2 [13]

Figure 13: Local Electric Field in SiO_2 [13]

energy is due to ionic contribution [13]. Furthermore the Si-Si bond is a very weak bond [13]. Thus when an electric field is applied to the oxide, there is a distortion of the lattice due to the polar nature of the O-Si bond [13]. Furthermore, this distortion induces a polarization $\mathbf{P}$ as shown in Figure 13. Thus each molecule of SiO_2 will not only experience the externally applied electric field, but it will also experience the dipolar field due to to $\mathbf{P}$ [13], and thus the local electric field, E_{loc}, can be significantly larger than the applied field E_{ox}.

Thus the Si-Si bonds are heavily strained due to this large local electric field, and bonds may occasionally gain enough thermal energy to cause the Si-Si bond to break, creating a hole trap, as shown in Figure 14. As noted above the generation of many traps will finally lead to breakdown.

3.2.4 HR Model

The process for the HR model is very similar to the AHI model except that instead of holes being produced at the anode, protons (H^+) are generated and reintroduced into the oxide [14]. There is evidence that the AHI rate is too small to produce the defects
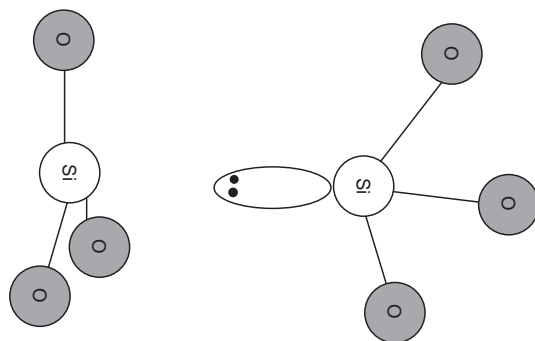
Figure 14: A Hole Trap in SiO_2 [13]

that lead to breakdown, and thus the HR model has been explored [14].

Just as in AHI, high energy electrons tunnel through the oxide, but instead break the Si-H bonds at the interface [15]. Thus a H^+ ion is released into the oxide, and there is a dangling silicon bond at the interface. Once the proton enters the oxide, it reacts with oxygen vacancies (Si-Si) to produce a trap: (Si-Si) + $H^+ \rightarrow$ Si-H^+-Si [14].

3.2.5 Channel Hot-Carriers

While many of the above effects are seen when there is no potential difference between the source to the drain (i.e. source and drain grounded), the presence of hot carriers in the channel may also affect oxide breakdown. Hot carriers are electrons or holes, who under the influence of high lateral fields in the channel gain sufficient energy that they are no longer in equilibrium with the substrate lattice [16].

The hot carriers (electrons in a NMOS) can create an electron-hole pair by impact ionization within the channel [16]. The holes enter the substrate and lead to substrate current, while the electrons enter the gate oxide [16]. As the electron enters the gate, it may create a trap leading to gate-oxide breakdown. It has been showed that the presence of Channel Hot Electrons decrease the time to breakdown for gate oxides [17].

3.2.6 Irradiation

Irradiation may also cause traps to occur within the gate oxide [18]. In one experiment a transistor was irradiated with 257MeV ions, and the first test for breakdown showed very unstable behaviour, where the current through the gate oxide was much larger [18]. Subsequent tests showed stable results comparable to that of a fresh transistor. The explanation for the above behaviour is probably due to a single ion being trapped in the oxide, permitting the current to flow across it temporarily [18]. After the first test, the device again becomes stable, probably due to electron-hole recombination [18].

While the radiation may not affect the short term strength of the oxide, it was also shown that after irradiation, breakdown occurs more rapidly when the transistor is stressed due to other mechanisms shown in this section [18]. HBD occurred in irradiated

samples on average in 400s, while breakdown did not occur in 10^5s of stressing in unirradiated samples [18]. Thus the irradiation must have produced some damage in the gate-oxide.

4 Symptoms

In this section the consequences and symptoms of both hard and soft gate-oxide breakdown on transistor characteristics will be explored. The effects of gate-oxide breakdown on circuit performance will also be presented. It will be seen that in digital logic, there may be no symptoms when a single transistor's gate-oxide breaks down, but that the effects are much more visible in analog circuitry.

4.1 Change in Transistor characteristics

As seen in the previous section, the results of HBD and SBD are quite different, and thus they will manifest themselves differently when they occur in transistors and thus they will be dealt with separately.

4.1.1 Hard Breakdown

After HBD has occurred, there is a current path from the gate to the channel. The current path is generally characterized by a gate post-breakdown resistance defined as $R_G = V_G/I_G$ [19]. R_G is such that the current through the gate is orders of magnitude larger than before post breakdown.

Thus to characterize the effects of HBD it becomes important to be able to find R_G. In [19] the authors show that R_G is dependent on the position of the breakdown along the length of the transistor. For a 0.20μm long transistor the post-breakdown resistance versus the position of the breakdown is shown in Figure 15. In the regions where the breakdown occurs over the source and drain extensions, R_G increases linearly [19]. Otherwise, if the breakdown occurs over the channel, R_G is relatively constant [19].

The linear increase in R_G over the source/drain extensions can be explained due to the resistance of the length of the n-doped region in the extension where the breakdown occurred [19].

When the breakdown happens over the channel, it is as if current is injected from the gate through the breakdown path into the channel which then continues to the drain and the source. This effect can be modeled by the two transistors to the right and left in Figure 16. Also when $V_G < 0$ two bipolar transistors are added to the model in Figure 16. to account for electrons that are injected from the gate diffusing through the substrate, and finally collecting at the source or the drain [19].

Regarding the magnitude of the current after HBD, [20] shows that after HBD near the source or drain, there is an increase of two orders of magnitude when the transistor is on, and an increase of six orders of magnitude when the gate voltage is near $0V$. [6]
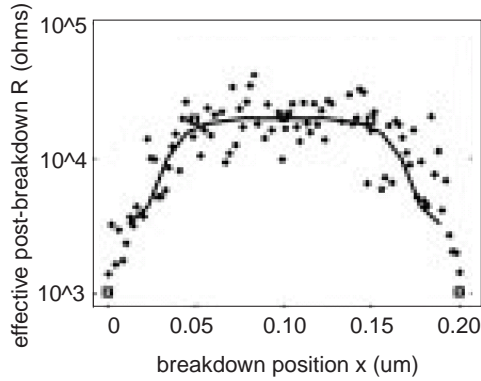
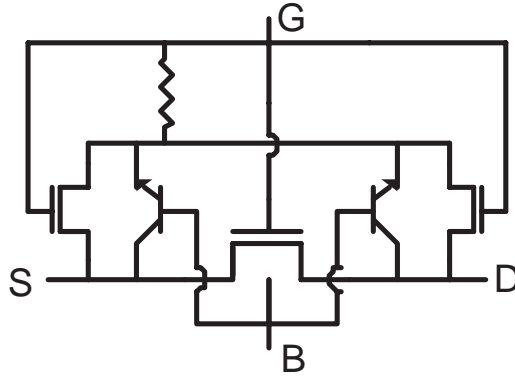Figure 15: Effective Gate Post-breakdown Resistance [19]

Figure 16: Equivalent circuit for an NMOS with HBD to the Channel [19]

also shows increases of gate current of many order of magnitudes depending on the state of the transistor. The specific increase in current depends on many factors including the size of the transistor, and the size and location of the breakdown.

4.1.2 Soft Breakdown

Unlike HBD, SBD generally does not affect the performance of the transistor much, other than increased leakage current in it's off state [21].

With technologies that have thin t_{ox}'s the gate tunneling leakage is large enough that any increase in gate current due to SBD in the transistor's on state may be not important [21]. However, when the transistor is off, and if SBD occurs near the drain extension there is an increase in Gate-Induced Drain Leakage (GIDL) of five orders of magnitude [21]. This increase in GIDL is due to the negative charge trapping in the oxide over the overlap region. SBD at other locations of the gate have very minimal
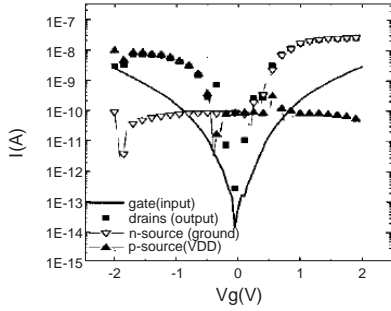
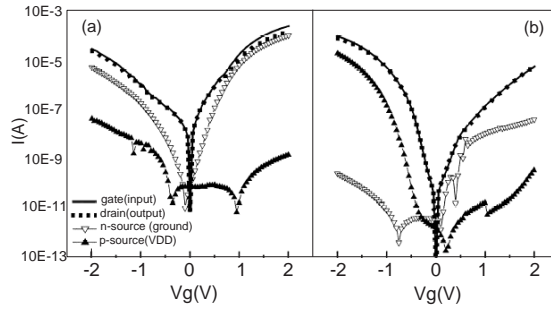Figure 17: IV Characteristics of a Fresh Inverter [23]

Figure 18: IV Characteristics of a Positive (a) and Negative (b) Stressed Inverter [24]

effect on the operation of the transistor.

Also long-channel transistors are less likely to have any SBD effects, as the drain extension forms a much smaller percentage of the total transistor length.

SBD does, however, affect circuit performance for transistors with low W/L. Since the gate area is very small, the SBD region may compose a considerable portion of the gate area. Under such circumstance the transconductance, g_m of the transistor drops by 50% and the saturation current falls to 30% of it's original value [22].

4.2 Change in Circuit characteristics

As seen above, the most visible effect of gate-oxide breakdown is increased gate current In this subsection the effects of gate-oxide breakdown on different circuit types will be explored.

4.2.1 Inverter

In [23] the authors investigate the the effect of breakdown on the operation of an inverter. Both positive and negative constant voltages [23] where applied to stress the inverter, and then the inverter operation was compared to a fresh transistor.

Figure 17 shows the IV characteristics of a fresh inverter, which has not been stressed; the important characteristics of this figure are that at high (low) voltages the NMOS(PMOS) is on and the n-source(p-source) current is high, and that the gate-current is not tracked by any current [23].

Now when this characteristic is compared to a transistor which has been stressed, and broken down, as seen in Figure 18, we see that the drain current very closely matches the gate current, and that the n-source (p-source) current increases by a large amount when the inverter is stressed with a positive (negative) voltage [23].

This change in IV characteristic shows that when the inverter is stressed with positive (negative) voltages, the NMOS (PMOS) is the transistor that is mainly damaged due to breakdown. The authors in [23] do not know exactly why this is the case, but point to the increased current flow through the NMOS (PMOS) when stressed with positive
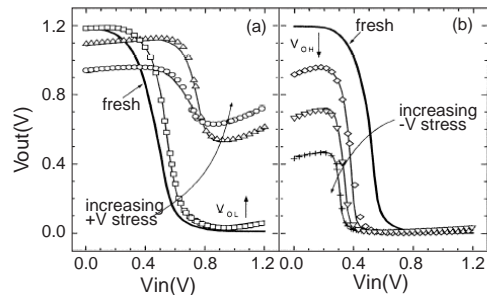
11

Figure 19: Transfer Characteristics of an Inverter. (a) Positive, (b) Negative Stresses [24]

(negative) voltage [23]. Another observation that can be made from Figure 18 is that the breakdown occurs at the drain of the transistor, since the gate current and the drain current of the inverter match closely [23].

Another important inverter characteristic is it's transfer characteristic. Figure 19 shows the transfer characteristic under positive and negative stresses. From Figure 19 (a) it can be seen that when the inverter is stressed by positive voltages, the NMOS breaks down at the drain, and thus at voltages near V_{DD}, there is now a forward biased diode from the input to the output of the inverter and instead of outputting a logic '1', the inverter output steadily increases as the voltage stress is increased [23].

A similar characteristic can be seen in Figure 19 (b), when negative voltages are used to stress the inverter, but this time the PMOS breaks down and a reverse biased diode exists between the input and output of the inverter, thus lowering the output voltage of the inverter at low input voltages [23].

Another change in the transfer characteristic is that it shifts to either the left or the right. This shift is due to a change of V_t which is due to the stresses that are applied to the inverter, but the shift in V_t is not necessarily a result of the gate-oxide breakdown [23].

4.2.2   Digital Logic

Given that the inverter has its characteristic changed, how does breakdown within a digital-logic gate affect the functioning of the complete digital circuit? Figure 20 shows the current drawn and the operating frequency of a ring of inverters forming an oscillator. Every discrete jump in the current in Figure 20 indicates a transistor undergoing a gate-oxide breakdown [25]. Even after gate-oxide breakdown occurring in 12 transistors, the ring oscillator still function at a frequency of around 30% of its original frequency [25]. Figure 21 shows the functioning of the ring oscillator in greater detail after $10700s$ of stress. The output shape is still correct, even though the frequency has dropped by about 15% [25].

The reason for this can be seen on Figure 22. Assuming, that the second inverter has
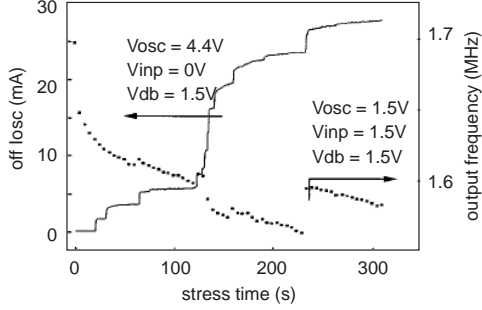
Figure 20: Current Drawn and Operating Frequency of an Oscillator after cumulative stress time [25]
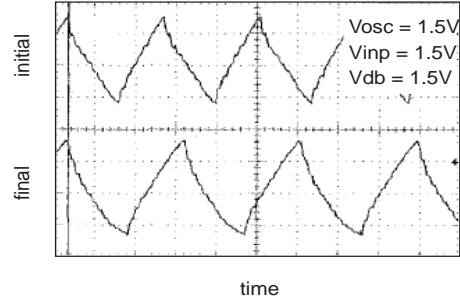

Figure 21: Oscillator Output Waveforms after 10 700 seconds of stress [25]
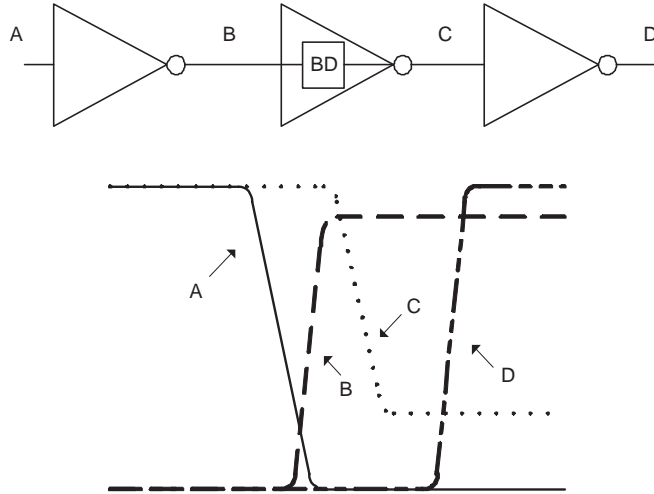
Figure 22: Node Voltages of Inverter Chain with One Gate-Oxide Breakdown

a gate-oxide breakdown at it's NMOS transistor, the voltage at node B does not reach V_{DD} due to the extra loading at the node [23]. Then at node C, due to the gate-oxide breakdown and the diode connection to node B, the voltage does not fall to ground. However, even with these problems, further logic gates down the circuit path will restore the correct digital signals, as seen at node D.

Thus even with gate-oxide breakdowns within a digital circuit, the circuit should continue operating, although it will be slower. Unless the gate affected by the breakdown is on the critical path, a single breakdown will not cause the circuit to malfunction, but the power consumed by the circuit will be larger than normal.

4.2.3 SRAM

Due to the projected large memory content of future System-on-a-Chip devices (71% of die area by 2005 [26]), it is important to understand the effect of gate-oxide breakdown

Figure 23: Possible Locations of Breakdown in SRAM cells
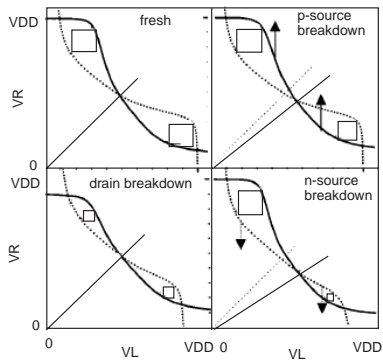
Figure 24: Effect of breakdown on SNM in SRAM cells [27]
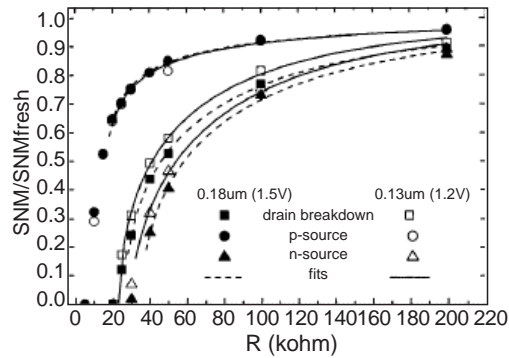
Figure 25: Normalized SNM under Hard-breakdown [27]

on Static Random Access Memory (SRAM) cells.
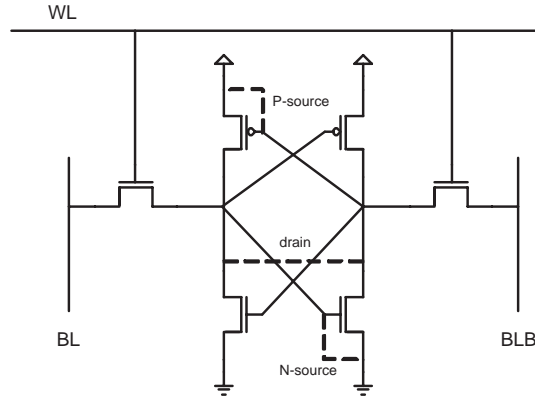
Figure 23 shows the different locations where breakdown can occur. When breakdown occurs, the Static Noise Margin (SNM) of the SRAM cell, which is a measure of the stability of its stability, decreases. Figure 24 shows schematically what happens to the SNM when breakdowns occurs at different locations within the cell. When the breakdown occurs at the n-source, the voltage at the output of the opposite inverter will be lowered due to the increased load [27]. The opposite effect happens when the p-source breakdown occurs, but, a p-source breakdown has less effect than an n-source breakdown since the opposing NMOS is strong enough to combat the effects [27].

Finally, a gate-oxide breakdown at the drain, causes a reduction in the output swing, especially at the high state, since the NMOS is stronger than the PMOS [27]. After empirical analysis, the authors in [27] found that for hard-breakdown, an n-source breakdown affects the SNM of the cell the most, followed closely by breakdown at the drain location; the p-source breakdown has the least effect [27]. For soft breakdown however, a breakdown at the drain has the least effect. Figure 25 shows the SNM of SRAM cells as a function of the ohmic breakdown resistance.

Figure 26: Voltage Gain versus Frequency for LNA [28]
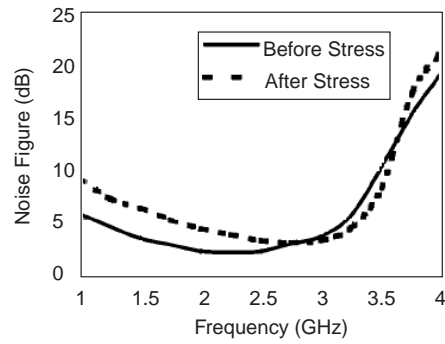
Figure 27: Noise Figure versus Frequency for LNA [28]
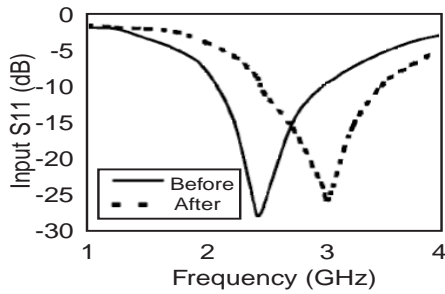
Figure 28: S_{11} Gain versus Frequency for LNA [28]
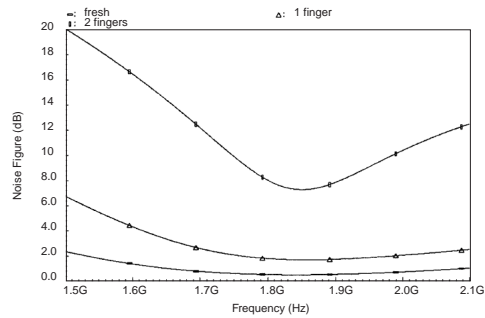
Figure 29: Noise Figure in the Presence of Fingers of Transistor Going Through Gate-Oxide Breakdown [29]
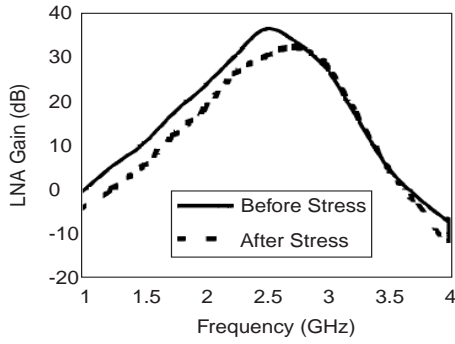
4.2.4 RF Circuitry

Radio Frequency (RF) designs have many different performance measures such as the noise figure, gain, impedance matching and linearity, and a single gate-oxide breakdown can cause the circuit to stop functioning at it's desired operating point. Furthermore, since transistors used in RF designs are usually operating in saturation, where hot-carriers are most active, they are more likely to fail, and thus it is important to understand the effects of gate-oxide breakdown in RF circuits.

Figure 26 shows the gain of a Low-Noise Amplifier (LNA) amplifier operating at an RF frequency of $2.45GHz$ [28], where the gain has been reduced by 5dB (a drop of over 3×) [28]. Furthermore, the noise figure, shown in Figure 27 has increased from 2dB to 3dB [28]. Also the S_{11} plot, which is a way of characterizing the impedance matching, shown in Figure 28, shows that S_{11} minimum has been shiftet by 600MHz, and at the operating frequency has changed from -27dB to -9dB (an increase of 62×) [28].

Finally, since RF devices are usually quite large, and are fingered, different fingers of the device may go through breakdown at different times [29]. Figure 29 shows the noise figure of a LNA after breakdown of 1 or 2 fingers. Thus even a single gate-oxide breakdown can be fatal to the functioning of an RF circuit.

5 Failure Model

Failure modeling for gate oxide breakdown is done through measurement of the (Time Dependent Dielectric Breakdown (TDDB). As seen previously, gate oxide breakdown can occur over time due to trap generation even at normal voltage, and temperature conditions, or it can be caused by external factors such as electrical over stress or Electro Static Discharge (ESD). The TDDB is a measure of the former, which is the focus of this section (as well as the rest of this report). TDDB measures the length of time until breakdown has occurred, with this as our random variable we can derive our failure model.

Before looking at any failure models, it is important to first define what we mean by a failure. Even in the above paragraph we implicitly equated breakdown with failure. The reliability community often does the same, however, recent research has suggested that this is not necessarily the case. As seen previously, the gate oxide can exhibit either a "soft breakdown" or a "hard breakdown" which differ in symptoms only in the magnitude of current leakage. With ultra-thin gate oxides, the normal leakage current is so big that the leakage due to soft breakdown may not be significant [30]. In addition, it has been shown in [31] that some circuits can continue to operate normally even after a hard breakdown. Thus, the failure of a device due to gate oxide breakdown is application specific. In this section we will look at application specific failure only very briefly. Our main focus will be on the failure of the transistor as defined by the presence of a conduction path through the gate oxide. Ultimately we seek a mapping from a set of transistors and usage parameters to an expected failure time.

A failure model is usually built from physical arguments and empirical studies. As usual, there is a divide between physicists, who try to model the process and derive a reliability measure from the model, and engineers, who through experiments try to deduce some predictability. We will discuss both models here.

5.1 Percolation Models

It is accepted that the gate oxide break down is related to defect generation, though the exact nature of these defects is debatable. It has also been known that oxide breakdown belongs to the class of weakest-link problem of extreme value statistics and should have a Weibull distribution. These percolation models are physics based models created by physicists. Below we introduce the basics of the percolation model as well as talk about competing models.

5.1.1 Tile Based Percolation Model

Some early research by Suné [32] modeled the the gate oxide as a 2D plane divided up into tiles. There is a potential conduction path corresponding to each tile, but the tile doesn't become a conduction path until some number of defects, n_{bd}, have occurred in that same tile. Defects are spawned randomly over the different tiles, once n_{bd} defects
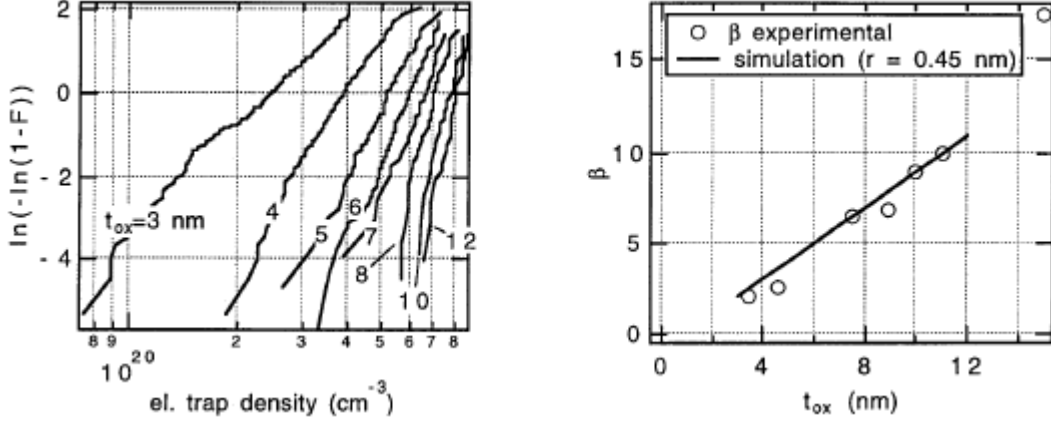
Figure 30: Simulated distributions of electron trap densities for different thicknesses (left) and relationship of Weibull slope β to thickness t_{ox} (right).[4]

have occurred in the same tile, the tile becomes a conducting path and the transistor has experienced gate oxide breakdown. The equation for this model is given in 1

$$\ln(-\ln(1-F)) = \ln(A/a) + \ln(at_{ox}p - \ln(\sum_{n=0}^{n_{bd}-1} \frac{(at_{ox}p)^n}{n!}))$$ (1)

This equation gives us the cumulative distribution function for p, the probability of a single defect occurring, or equivalently, the density of defects in the gate oxide. It has two parameters a the area of each tile, and n_{bd} the number of defects necessary in a single tile for a breakdown. Thus, for a given defect density, there is a corresponding probability of a breakdown occurrence. This model succeeded in relating breakdown to the frequency of defects (or traps), but it was criticized for lacking in predictive power specifically in that it failed to relate n_{bd} to the oxide thickness or composition[3]. Still, this model is indeed a Weibull distribution and was a significant step forward for the percolation model.

5.1.2 Sphere Based Percolation Model

Degraeve et al expanded this model to the 3rd dimension and successfully used it derive a new and improved model[4]. The basic idea was that traps are spawned in random locations in the oxide and a conduction path forms when adjacent traps span the thickness of the gate oxide. This phenomenon was modeled using spheres in a Monte Carlo simulation. When one sphere touches another the spheres conduct. When there is a continuous chain of linked spheres connecting the top and bottom surfaces of the gate oxide, breakdown occurs. The only parameter in this model is the radius of the spheres used. Degraeve used this model to verify the that the distribution of breakdown was Weibull, and also showed the dependence of Weibull slope to thickness (see Figure 30).

17
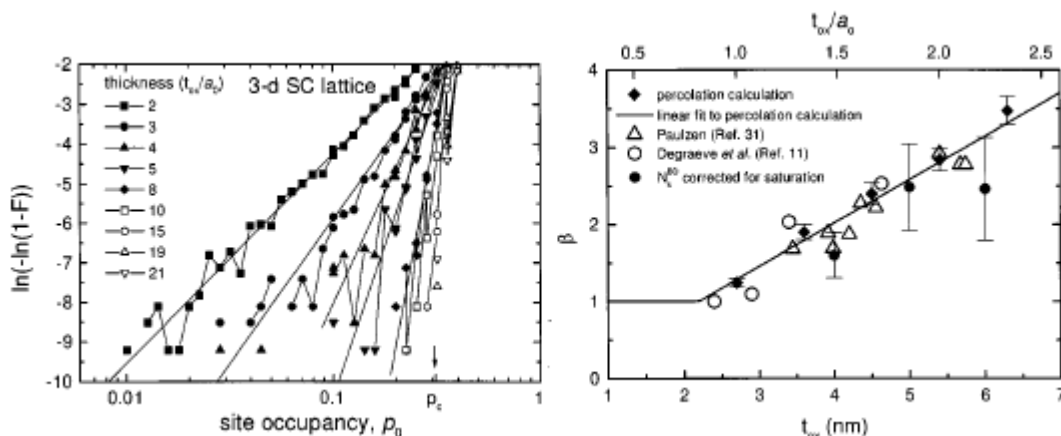
Figure 31: Simulated distributions of election trap densities for different thicknesses (left) and relationship of Weibull slope β to thickness t_{ox} (right) for the cube based Percolation Model by Stathis[3].

In addition, the model successfully accounted for things like non-uniform thickness, and conductance distribution. Because of all these reasons, it is widely accepted that the sphere-based percolation model is the correct break down model.

5.1.3 Cube Based Percolation Model

Other approaches tried doing the same thing with cubes instead of spheres [3]. The success was similar except in one respect. These percolation models imply a lower bound on the thickness of the gate oxide. In the extreme case, if the gate oxide was as thin or thinner than the size of a trap, then breakdown would occur with the first trap generated. Fitting simulated results to expected results, Degraeve deduced the diameter of his spheres were 0.9nm (figure 30) while Stathis measured 2.7nm sides for the cubes. Since transistors with gate oxide thicknesses less than 2nm were seen to be quite reliable, it is difficult to accept that the size of a trap is 2.7nm as seen with the spheres. Because of this, the sphere percolation model reigns.

5.1.4 Analytical Cube Based Percolation Model

With that said we'll now look at a new purely analytical cube based percolation model proposed in [33] by the creators of the tile-based percolation model. First we start with a gate oxide with thickness t_{ox}, and area A_{ox}. We then divide the oxide into cubic cells with length a_0. Next we assume that any cell can become a trap with uniform probability. Hence:

$$F_{cell}(\lambda) = \lambda \tag{2}$$

Now assuming independence, the probability of failure for one column of cells is given by:

18

$$F_{col}(\lambda) = [F_{cell}(\lambda)]^n = \lambda^n \tag{3}$$

where $n = t_{ox}/a_0$, in other words n is the number of cells in a column. If we similarly define N as the number of cells that make up the area, or $N = A_{ox}/a_0^2$, then we obtain the breakdown reliability function as follows:

$$R_{bd} = [1 - F_{col}(\lambda)]^N = [1 - \lambda^n]^N \tag{4}$$

Transforming this into a Weibull plot we have:

$$W_{bd} = \ln[-\ln(1 - F_{bd})] = \ln[-N\ln(1 - \lambda^n)] \tag{5}$$

We can further simplify this expression using the assumption $\ln(1 - \lambda^n) \approx -\lambda^n$ since $\lambda \ll 1$. The result is:

$$W_{bd} = \ln(N) + n\ln(\lambda) \tag{6}$$

Note the equation has only a single parameter, a_0 since this parameter determine both N and n. The random variable in this distribution is λ, the probability of a trap occurrence in a single cell. We can convert this to a function of trap density (as was used by Stathis and Degraeves above) by defining $N_t = \lambda/a_0^3$, where N_t is the trap density per unit volume. As expected, the result is linear in a Weibull plot of W_{bd} and $\ln(N_t)$ as is evident by looking at the equation:

$$W_{bd} = \ln(N) - 3\ln(a_0) + n\ln(N_t) \tag{7}$$

Unfortunately, percolation models stops here. In other words, all percolation models succeed only in relating breakdown to the density of trap occurrences. A relationship between trap occurrences and time is desired in order to find the TDDB. This relationship is difficult to find since it depends on stress conditions, and sample characteristics. In addition, how traps are created and the rate at which they do so are still unknown and controversial. Breakdown projection models are used for this purpose and will be discussed further below. However, for this analysis, the authors assume that λ is some function of Q_{db} given by a power law model:

$$\lambda(t) = \lambda_0 Q_{bd}^\alpha \tag{8}$$

Substituting into 6 we get

$$W_{bd} = \ln(N) + n\ln(\lambda_0^\alpha) + n\alpha\ln(Q_{bd}) \tag{9}$$

With this model we can relate breakdown to the injected charge Q_{bd}. Getting the TDDB only requires knowing the rate at which charge is being injected into the gate. Since the mean breakdown probability is at $F_{bd} = 0.632$ (or $W_{bd} = 0$), we find the mean charge to breakdown to be
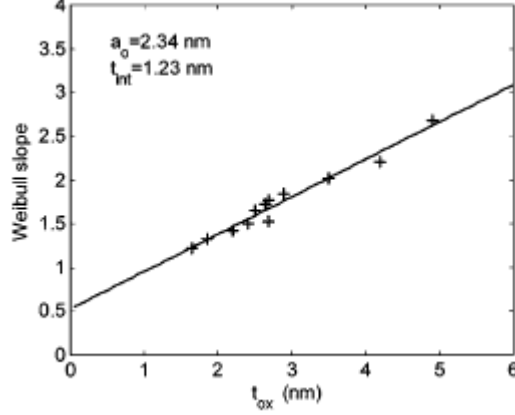
19

Figure 32: Fitting of slope in this analytical model to the results by Stathis[33]

$$< Q_{bd} >= \frac{1}{\lambda_0^{1/\alpha}} exp \left[-\frac{a_0}{\alpha t_{ox}} \ln \left(\frac{A_{ox}}{a_0^2} \right) \right] \tag{10}$$

The breakdown model given by 9 has the Weibull slope $\beta = \alpha \frac{t_{ox}}{a_0}$ by inspection. After adding a constant t_{int} to account for non-zero intercepts seen in actual experiments, we can test this model against the previously seen models. In the cube based model Stathis assumed $\alpha = 1$ and saw that $a_0 = 2.7nm$. Under this model and the same assumptions, we get $a_0 = 2.34nm$ as seen in figure 32. In the sphere based model, Degraeves assumed $\alpha = 0.56$ and got a diameter of 0.9nm compared to the 1.17nm seen for a_0 in this analytic model. Thus this model achieves reasonably good agreement with existing models. This model can also be adapted to account for the fact that not all breakdowns result in device failure. This is done by identifying a fraction α_{fail} of the defected columns that actually cause a failure. We then change equation 3 to $F_{col}(\lambda) = \alpha_{fail} \lambda^n$ and proceed with the rest of the analysis the same way.

5.1.5 Percolation models and Soft and Hard Breakdown

In all the above models, no distinction is made between the types of breakdown whether soft or hard. The assumption is that both forms of breakdown are caused by a linked chain of traps bridging the gate oxide. This assumption agrees with the accepted belief that a hard breakdown is basically a soft breakdown that has experienced thermal runaway (see Symptoms section). This assumption was empirically verified by Sune in [34] where it was seen that the distribution of first soft breakdowns and first hard breakdowns coincide (see figure 33).

20
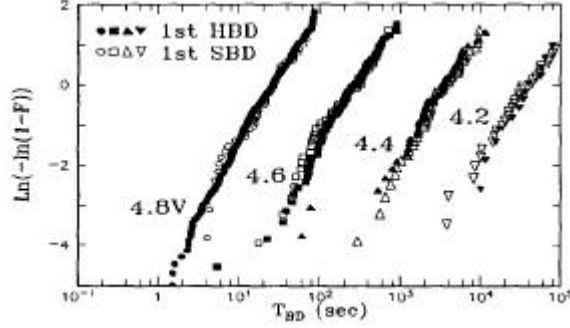
Figure 33: Distribution of 1st Soft and Hard Breakdowns for different voltage stresses seen to coincide[34]
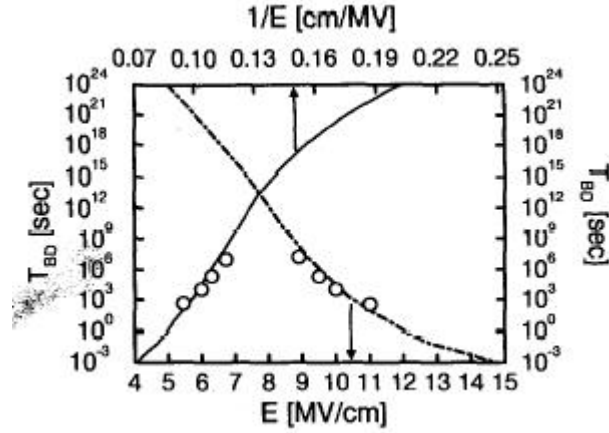
Figure 34: Discrepancy between E and 1/E model for high and low electric fields[35]

5.2 Breakdown Projection Models

The above percolation models relate breakdown to defect generation, but in order to make predictions of a device's lifetime, we need to relate defect generation to test conditions and device characteristics. This is the goal of breakdown projection models. With little known about the actual mechanisms that cause defects, a very heated debate is currently going on about which model is correct. And on top of that, physicists criticize all approaches for being over-simplified engineering approaches [33]. Below we will look at some of them but will focus on the two most competitive models. But before discussing the models themselves, we will introduce the empirical results they are trying to agree with.

5.2.1 E and 1/E Models

Empirically, the TDDB has been seen to behave in one of two ways: Proportional to $exp(E)$ or proportional to $exp(1/E)$ where E is the electric field in the gate oxide. Specifically the two models are characterized according to the following relationships:

$$\text{E model: } \ln(t_{bd}) \propto \frac{\Delta H_0}{k_B T} - \gamma E_{ox} \tag{11}$$

$$\text{1/E model: } \ln(t_{bd}) \propto \frac{\Delta H_0}{k_B T} - G(1/E_{ox}) \tag{12}$$

where ΔH_0 is the activation energy, k_B is Boltzmann's constant, T is the absolute temperature, and γ and G are field acceleration factors. Note that here we use E_{ox} to make explicit that it's the field in the oxide we're talking about. Both models continue to exist since they each work well in their respective regions. Looking at figure 34 found in [35], we see that the E model accurately predicts the breakdown time for low electric fields, while the 1/E model does so for higher electric fields. The debate over which one was correct carried on like the debate over waves and photons, and now a surge of attempts to unify the two indicates that both are indeed correct (we'll look at these below). Several physical models were proposed to explain the results. A physical model would typically choose either the E or 1/E model, and provide a physical justification for it hence lending support to that model. Thus breakdown projection models were traditionally categorized as being either E or 1/E. Of those models, the two top runners were the AHI model and the Thermochemical model which were initially 1/E and E respectively. However that is starting to change since supporters of each side recognize that in order for either the AHI or Thermochemical model to triumph, it need to encompass both E and 1/E models. More details on the AHI and Thermochemical models follow.

5.2.2 AHI Model

The AHI model, or Anode Hole Injection model is based on the dissipation of energy from injected electrons. The model is as follows from [36]:

$$t_{bd} = Q_{bd}/J_n \tag{13}$$

Where t_{bd} and Q_{bd} are the time and charge to breakdown respectively, and J_n is the gate current density which can be calculated from quantum tunneling models. The charge to breakdown is given by

$$Q_{bd} = \frac{Q_p}{a_p \Theta_p} = \frac{Q_p}{a_p} exp\left(\frac{\hat{B}}{E} \Phi_p^{3/2}\right) \tag{14}$$

The breakdown dependence on $exp(1/E)$ is clearly seen in the right most term. All parameters in the equation can be calculated from quantum physics with the exception

of a_p, which is the probability that a hole is generated. Q_p is the critical hole fluence at breakdown known to be 0.1 C/cm$\hat{2}$, and Θ_p is the probability that a generated hole tunnels through the barrier. Since a_p must be determined by curve fitting, the AHI model is not a quantitative model. However once the model is calibrated and this parameter is determined it can be used to accurately predict lifetimes in the 1/E regime.

Recent developments in the AHI model by Alam in [35] shows that the AHI model can account for the change from E to 1/E physics. The critical observation was accounting for minority ionization which can contribute significantly to hole tunneling. After doing so it was seen that in low fields, J_n has a weak field dependence and $a_p\Theta_p$ takes the form $exp(\beta E)$ so that an overall E dependence is seen. In high fields where minority ionization is less significant, J_n has a 1/E dependence and $a_p\Theta_p$ has a weak field dependence for an overall 1/E dependence. In this way, the AHI model can account for both E and 1/E models without changing the underlying mechanisms. Hence the AHI model accounts for all the data seen in figure 34.

5.2.3 Thermochemical model

The only purely quantitative model is the Thermochemical model or TC model. This model is based on dipole moments caused to interact under a field. Electrons or holes don't play a role in this model. The model is as follows from [37]:

$$t_{db} = A_0 \cdot exp\left[\frac{(\Delta H)_0 - (7.2e \cdot \dot{A}) \cdot E_{ox}}{k_B T}\right] \qquad (15)$$

where k_B is Boltzmann's constant, T is temperature, $(\Delta H)_0$ is the enthalpy of activation for trap generation (the breaking of a bond caused by dipole moment), and A_0 is a constant. $(\Delta H)_0$ is estimated from the binding energy of Si-Si to be 450 kJ/mole. The E dependence is clearly seen in the equation. Being purely quantitative this model is easy to verify and as seen in 34, it has good agreement in low fields. The fact that it is the only model which can account for the low E field effects is a credit to the TC model.

A recent extension of the TC model has enabled it to predict both E and 1/E models as well[38]. The basic idea here is that two parallel reactions are going on simultaneously. One is the traditional field-induced bond breakage, the other is another field-induced bond breakage which is catalyzed by a hole capture caused by the current flow. This seems hypocritical since the TC model had always argued that current and electrons didn't play a role. Nonetheless this unifying model was established and is as follows:

Let k_1 be the reaction rate of the traditional TC model:

$$k_1 = v_0 exp\left[\frac{(\Delta H)_0 - a \cdot E_{ox}}{k_B T}\right] \qquad (16)$$

Let the new reaction occurring in parallel with this one be made up of two processes each with separate reaction rates. One is hole capture with reaction rate k_{2a} and the

Table 1: Feature Size and Oxide Thickness

Production year	Minimum feature size (μm)	Equivalent dielectric thickness ($\overset{\circ}{A}$)
1997	0.25	40-50
1999	0.18	30-40
2001	0.15	20-30
2003	0.13	20-30
2006	0.20	15-20
2009	0.07	10-15
2012	0.05	<10

other, with reaction rate k_{2b} is similar to the traditional process except now in the presence of this captured hole which lowers the activation energy by at least half. The reaction rates are

$$k_{2b} \geq v_0 exp\left[\frac{(\Delta H)_0/2 - a \cdot E_{ox}}{k_B T}\right] \tag{17}$$

$$k_{2a} = -\frac{\ln(1/f_{crit})}{t_0 exp[G/E_{ox}]exp[(Q/k_B)(1/300K - 1/T)]} \tag{18}$$

The overall effective reaction rate k_{eff} is then given by

$$k_{eff} = k_1 + \frac{k_{2a}k_{2b}[exp(-k_{2at} - exp(-k_{2b}t)]}{k_{2b}exp(-k_{2a}t) - k_{2a}exp(-k_{2b}t)} \tag{19}$$

So that the final breakdown time is

$$t_{bd} = \frac{\ln(1/f_{crit})}{k_{eff}} \tag{20}$$

When k_{2a} is insignificant compared to k_1 and hence k_{2b}, an E dependence is seen as equation 20 is reduced to equation 15. Conversely, if k_{2a} dominates, in other words the presence of captured holes is significant, then a 1/E dependence is seen as the equation reduces to

$$t_{bd} = t_0 exp\left[\frac{G}{E_{ox}}\right] exp\left[\frac{Q}{k_B}\left(\frac{1}{300K} - \frac{1}{T}\right)\right] \tag{21}$$

Thus the TC model can also account for the change from E to 1/E physics but the agreement is often unsatisfactory [39].

6 Commonality

The problem of gate oxide breakdown is not a freak occurrence like ESD, instead it is a problem inherent in the physics of SiO_2 transistors. The fact is that gate oxides breakdown even from normal use. The smaller the thickness of the gate oxide, the more likely they will breakdown.

Current technological trends geared toward higher performance is driving the gate oxide thickness to shrink. This can be seen in Table 1. However, while better performance drives the gate oxide thickness down, reliability drives the gate oxide thickness up. The bottom line is that chips must be designed so that their gate oxides provide a desirable balance between performance and reliability.

To answer the question of how common gate oxide breakdown is, it is simply as common as designers allow it to be. Therefore, in terms of consumer products, a device is not very likely to breakdown because of gate oxide breakdown, assuming that designers have designed the oxide thickness to die well after other failures have already occurred. Any gate oxide breakdowns caused by ESD occur as common as ESD occurs (see the project on ESD). Gate oxide breakdowns caused by incorrect fabrication of the circuit are normally filtered out at burn-in. These breakdowns are associated with the infant mortality region of the bathtub curve.

7 Prediction

Clearly predicting gate oxide breakdown is complicated, and worse yet, controversial. As seen in the previous section, percolation models were used to verify that gate oxide breakdown occurs because of trap generation. The AHI and TC models extend this idea and account for the trap generation in order to predict the time to breakdown. However, the only quantitative model, the TC model, is the least favourite of the two since in general the AHI model has better agreement with experimental data. As long as modeling the breakdown involves experiment and curve fitting, it is impossible for these to be automatically incorporated into a CAD tool.

There are several parameters that need to be known in order to predict breakdown:

- Device Usage Parameters
    1. Temperature
    2. Voltage stress
    3. Current injection
    4. Field applied

- Physical Factors
    1. Process-dependent constants
    2. Physics constants (activation energies, tunneling probabilities, etc.)
    3. Area of oxide

4. Thickness of oxide

- Failure Determination

    1. Desired breakdown time
    2. Desired probability of breakdown time
    3. Dependence of failure on breakdown
    4. Choice of breakdown model and determination of associated constants
    5. An accurate method to extrapolate data obtained (can't test for 10 years)

Many of these parameters may not be fully understood by the designer. Certainly usage characteristics could be determined by the end user of the device. Failure determination and the selection of an appropriate model is likely to be the most difficult parameters. However, what makes the process so complicated is that each of the listed parameters are not independent of the others. The field depends on the voltage and thickness, the choice of model depends on the field, the physics constants can depend on thickness, etc.

With the models still at an adolescent stage and currently still in debate, using the failure models is an unattractive method for reliability prediction. Current prediction is done using the following procedure outlined by Wu et al. in [40].

1. Collect failure data for accelerated voltages and elevated temperatures using Constant Voltage Stressing (CVS) or Constant Current Stressing (CCS).

2. Using a "typical lifetime projections" for a given oxide thickness, extrapolate the time to breakdown from the stressed voltages to the use condition. Similarly for temperature. Also, if the area of the tested oxides is different than the area of the gate oxide being analyzed, extrapolate the data to account for this.

3. The lifetime at higher failure probabilities must be projected to the lower cumulative failure probabilities using the Weibull breakdown distribution.

This process must be repeated for every different gate oxide used, hence it does not lend itself to automation. Manual experiments and extrapolation is the core of the process. Still there has been at least one effort to introduce a tool that can automate some of this process.

Figure 35 shows Tu's attempt to predict process-induced gate oxide damage in [41]. This process uses the 1/E AHI model and is a module that attaches to the BERT (Berkeley Reliability Tool) framework. The proposed simulator can simulate gate oxide failures through processing, burn-in, and circuit operation. Doing so requires that the design is already completed, however, the authors claim that with increased knowledge of gate oxide breakdown, a tool for "designing-in-reliability" can be developed. The simulator requires the results of oxide breakdown tests as an input in order to develop a relationship between X_{eff} (effective oxide thickness) and $D(X_{eff})$ (the area density of defects of size X_{eff}).
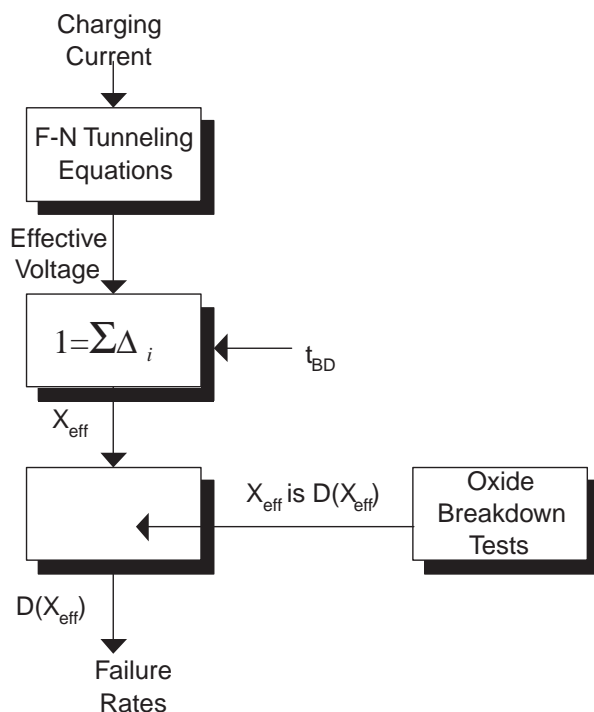
Figure 35: Prediction process proposed in [41]

To conclude this section, we'll state that design phase predictions of gate oxide breakdown are currently impractical. The only solution is an iterative trial-and-error method as seen in both of the above processes. A tool which, given certain parameters, generates a reliable oxide thickness is far from production. Progress in this area is pending further developments in gate oxide reliability.

8   Protection

As has been seen throughout this paper, the electric field through the gate oxide is an important factor leading to gate oxide breakdown. Gate oxide breakdown is either directly related to the field through the E model, or indirectly related to the electric field through the $1/E$ model due to the high energy electrons that are present in a high electric field. It has been seen that hot-carriers can also affect the time to gate oxide breakdown.

In the subsections that follow, two designs are shown where first the voltage differential on the oxide is guaranteed not to exceed the process guidelines even with the gate underdrive and second the amount of time a transistor spends in saturation is minimized to reduce hot-carriers.
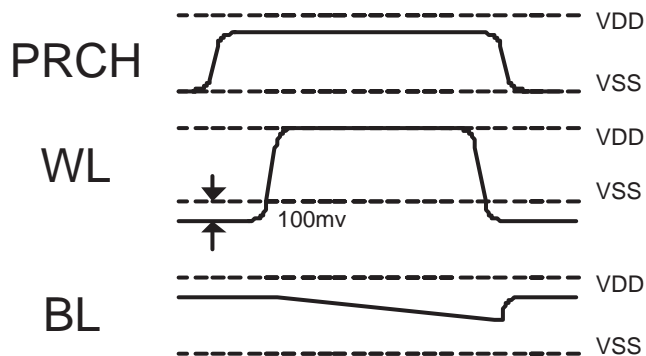
Figure 36: Waveforms in Bitline Leakage Reduction Scheme

8.1 Bitline Leakage Reduction Technique

Since leakage current is becoming a large fraction of total power consumption, leakage reduction techniques have become important. Furthermore, since leakage current is proportional to the number of on-chip transistors and SRAM cells compose the vast majority of on-chip transistors, many leakage reduction techniques have been devised to lower leakage in SRAM cells.

One particular technique to reduce leakage in SRAM cells is to underdrive the word-line when the cell is not being read, thus reducing V_{GS}, and lowering the subthreshold current to the bitlines. This technique, if used by itself, would however, reduce the reliability of the oxide since V_{GD} and V_{GS} of the non-leaking transistor would be larger than V_{DD}.

To alleviate the problem, two power supplies are used on chip, where the voltage of the cell and the bitlines is reduced by the underdrive factor, thus allowing the maximum voltage across the gate of the pass-transistors to remain at V_{DD} [42]. Waveforms for this bitline reduction technique are shown in Figure 36.

The reduced power supply to the cell, however, leads to lower stability, which has been recovered by an increased cell area [42]. Thus, this design shows that circuit area has been sacrificed for oxide reliability.

8.2 RF Circuit

As discussed in Section 4.2.4, RF circuitry is very sensitive to gate oxide breakdown [43], and thus it is important to reduce the possibility of the circuit being vulnerable to such conditions as that may lead to oxide breakdown.

Figure 37 shows a Gilbert Mixer which is used with RF circuitry. In it's active mode, the mixer has a g_m stage (see Figure 37) where the transistors are normally operated with a V_{dsat} of 200mV and and a V_{ds} of 550mV [43]. Since hot-carrier damage to the oxide is exponentially related to $V_{ds} - V_{dsat}$, those transistors are susceptible to much oxide damage [43].
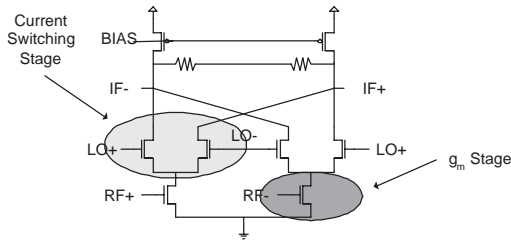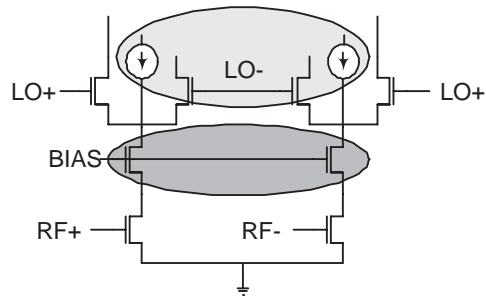
Figure 37: CMOS Gilbert Mixer

Figure 38: New CMOS Gilbert Mixer to Improve Reliability

Also the transistors within the current-switching stage carry much current and are also biased in saturation, and are thus also vulnerable to gate oxide breakdown [43].

To reduce the possibility of hot-carriers, and thus gate-oxide breakdown, a new design for the Gilbert Mixer was created and shown in Figure 38 [43]. First two extra transistors, that are carefully biased, are added to the g_m stage, thus reducing V_{ds} across the four transistors in the g_m stage [43]. Also two extra PMOS current sources are added which allow for the common-mode current needed to bias the g_m stage to bypass the current-steering stage [43].

9 Conclusion

Gate oxide breakdown is a big concern in transistor reliability. With current technology trends continuing to shrink the gate oxide thickness, gate oxide breakdown is becoming more and more prevalent. Though some circuits can continue normal operation even after a few gate oxide breakdowns, this is clearly not a general case and breakdown in a critical transistor will cause device failure.

Gate oxide breakdown comes in two forms, soft and hard breakdown. The physical characteristics of each are reasonably well understood. Hard breakdown is simply a more severe case of soft breakdown where thermal runaway has resulted in a larger conduction path. It was proven that both are caused by the same conditions though the actual cause is still debatable.

It is generally accepted that gate oxide breakdown is caused by trap generation forming a conduction path across the gate oxide . The percolation model is used to model this process and is seen to be quite successful especially in the sphere based model. The percolation model does not account for the cause of the traps, and this is where the controversy lies in gate oxide breakdown. The two main theories are the AHI model and Thermochemical model. The AHI model claims traps are caused by energy dissipation from electrons tunneling through the gate oxide. The AHI model is quite successful when the oxide field is strong but is not a purely quantitative model, it requires curve fitting to determine its parameters. The Thermochemical model claims

traps are caused by dipole moments interacting with the gate oxide field. This model is purely quantitative but good agreement with experimental data is only seen in the low field regime. Unifying models have emerged which claim that the cause is some combination of the AHI and Thermochemical models and at low and high fields one dominates over the other.

Due to this ongoing controversy and the fact that there is no purely quantitative model which applies across a usable space of stress conditions, predicting gate oxide breakdown is impossible at the design level. Designs are done on a trial and error basis and are tested using accelerated testing. Experience is the only thing that can aid a designer. There are, however, techniques for reducing the probability of a breakdown occurrence and for minimizing the effects of gate oxide breakdown.

References

[1] *2002 International Technology Roadmap for Semiconductors.*

[2] D. J. Dumin. Oxide wearout, breakdown, and reliability. In *Oxide Reliability. A summary of Silicon Oxide Wearout, Breakdown and Reliability.* World Scientific, 2002.

[3] J.H Stathis. Percolation models for gate oxide breakdown. *Journal of Applied Physics*, 86(10):5757–5766, Nov 1999.

[4] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H.E. Maes. A consistent model for the thickness dependence of instrinsic breakdown in ultra-thin oxides. In *Electron Devices Meeting*, pages 863–866, 1995.

[5] H.C. Lin, D.Y. Lee, C.Y. Lee, T.S. Chao, T.Y. Huang, and T. Wang. New insights into breakdown modes and their evolution in ultra-thin gate oxide. In *International Symposium on VLSI Technology*, pages 37–40, 2001.

[6] Michel Depas, Tanya Nigam, and Marc M. Heyns. Soft breakdown of ultra-thin gate oxide layers. *IEEE Transactions on Electron Devices*, 43(9):1499–1504, Sept. 1996.

[7] S. Lombardo, F. Crupi, A. La Magna, and C. Spinella. Electrical and thermal transiet during dielectric breakdown of thin oxides in metal-SiO$_2$-silicon capacitors. *Journal of Applied Physics*, 84(1):472–479, July 1998.

[8] H Uchida, S. Ikeda, and N. Hirashita. An accurate discrimination method of gate oxide breakdown positions by a new test structure of MOS capacitors. In *International Conference on Microelectronic Test Structures*, pages 229–232, 2001.

[9] Chenming Hu and Qiang Lu. A unified gate oxide reliability model. In *IEEE International Reliability Physics Symposium*, pages 47–51, 1999.

[10] K. Shiga, J. Komori, M. Katsumata, A. Teramoto, and M. Sekine. A new test structure for evaluation of extrinsic oxide breakdown. In *International Conference on Microelectronic Test Structures*, pages 197–200, 1998.

[11] Takayuki Tomita, Hiroto Utsunomiya, Yoshinari Kamakura, and Kenji Taniguchi. Hot hole induced breakdown of thin silicon films. *Applied Physics Letters*, 71(25):3664–3666, December 1997.

[12] Klaus F. Schuegraf and Chenming Hu. Hole injection SiO_2 breakdown model for very low voltage lifetime extrapolation. *IEEE Transaction on Electron Devices*, 41(5):761–767, May 1994.

[13] J.W. McPherson and H.C. Mogul. Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in sio_2 thin films. *Journal of Applied Physics*, 84(3):1513–1523, August 1998.

[14] Jordi Suñé and Ernest Wu. A new quantitative hydrogen-based model for ultra-thin oxide breakdown. In *Symposium on VLSI Technology*, pages 97–98, 2001.

[15] T. Pompl, H. Wurzer, M. Kerber, and I. Eisele. Investigation of ultra-thin gate oxide reliability behavior by seperate characterization of soft breakdown and hard breakdown. In *IEEE International Reliability Physics Symposium*, pages 40–47, 2000.

[16] Cheng T. Want, editor. *Hot Carrier Design Considerations for MOS Devices and Circuits*. Van Nostrand Reinhold, 1992.

[17] Felice Crupi, Ben Kaczer, and Guido Groeseneken. New insights into the relation between channel hot carrier degradatin and oxide breakdown in short channel nMOSFETs. *IEEE Electron Device Letters*, 24(4):278–280, April 2003.

[18] A. Cester. Wear-out and breakdown of ultra-thin gate oxides after irradiation. *Electronics Letters*, 38(19):1137–1139, Sept. 2002.

[19] Ben Kaczer, Robin Degraeve, An De Keersgieter, Koen Van de Mieroop, Veerle Simons, and Guido Groensenekn. Consistent model for short-channel nMOSFET after hard gate oxide breakdown. *IEEE Transaction on Electron Devices*, 49(3):507–513, March 2002.

[20] W. Kirklen Henson, Nian Yang, and Jimmie J. Wortman. Observartion of oxide breakdown and its effects on the characteristics of ultra-thin-oxide nMOSFET's. *IEEE Electron Device Letters*, 20(12):605–607, Dec. 1999.

[21] T. Pompl, H. Wurzer, M. Kerber, R.C.W. wilkins, and I.Eisele. Influence of soft breakdown on NMOSFET device characteristics. In *IEEE International Reliability Physics Symposium*, pages 82–87, 1999.

[22] A.Cester, S. Cimino, A. Paccagnella, G. Ghidini, and G. Guegan. Collapse of MOSFET drain current after soft breakdown and its dependence on the transistor aspect ration W/L. In *IEEE International Reliability Physics Symposium*, pages 189–195, 2003.

[23] R. Rodriguez, J.H. Stathis, and B.P. Linder. Modeling and experimental verification of the effect of gate oxide breakdown on CMOS inverters. In *IEEE International Reliability Physics Symposium*, pages 11–16, 2003.

[24] R. Rodriguez, J.H. Stathis, and B.P. Linder. A model for gate-oxide breakdown in CMOS inverters. *IEEE Electron Device Letters*, 24(2):114–116, Feb. 2003.

[25] Ben Kaczer, Robin Degraeve, Mahmoud Rasras, Koen Van de Mieroop, Philippe J. Roussel, and Guido Groensenekn. Impact of MOSFET gate oxide breakdown on digital circuit operation and reliability. *IEEE Transaction on Electron Devices*, 49(3):500–507, March 2002.

[26] Jay Abraham. Overcoming timing, power bottlenecks. *EE Times*, page 58, April 28 2003.

[27] R. Rodriguez, J.H. Stathis, B.P. Linder, S. Kowalczyk, C.T. Chuang, R.V. Joshi, G. Northrop, K. Bernstein, A.J.Bhavnagarwala, and S. Lombardo. The impact of gate-oxide breakdown on SRAM stability. *IEEE Electron Device Letters*, 23(9):559–561, Sept. 2002.

[28] Qiang Li, Jinlong Zhang, Wei Li, Jiann S. Yuan, Yuan Chen, and Anthony S. Oates. RF circuit performance degradation due to soft breakdown and hot-carrier effect in deep-submicrometer CMOS technology. *IEEE Transaction on Microwave Theory and Techniques*, 49(9):1546–1551, Sept. 2001.

[29] Hong Yang, Jiann S. Yuan, Yi Liu, and Enjun Xiao. Effect of gate-oxide breakdown on RF performance. *IEEE Transaction on Device and Meterials Reliability*, 3(3):93–97, Sep. 2003.

[30] W. Loh, B. Cho, D, Chan, and J. Zheng. Localized oxide degradation in ultra-thin gate dielectric and its statistical analysis. In *IEEE Transactions on Electron Devices*, pages 967–972, 2003.

[31] R. Degraeve, B. Kaczer, A. De Keersgieter, and G. Groeseneken. Impact of MOSFET oxide breakdown on digital circuit operation and reliability. In *IEDM Technical Digest*, pages 553–556, 2000.

[32] J. Suné, I. Placencia, N. Barniol, E. Farrs, F. Martn, and X. Aymerich. On the breakdown statistics of very thin SiO_2 films. *Thin Solid Films*, 185:347–362, 1990.

[33] J Sune. New physics-based analytic approach to the thin-oxide breakdown statistics. *IEEE Intl. Electron Device Letters*, 22(6):296–298, 2001.

[34] J. Sune, E.Y. Wu, D. Jimenez, R.P. Vollertsen, and E. Miranda. Understanding soft and hard breakdown statistics, prevalence ratios and energy dissipation during breakdown runaway. *IEEE Intl. Electron Device Meeting*, pages 6.1.1–6.1.4, 2001.

[35] M.A. Alam, J. Bude, and A. Ghetti. Field acceleration for oxide breakdown-can an accurate anode hole injection model resolve the E vs. 1/E controversy? *IEEE Intl. Reliability Physics Symposium*, pages 21–26, 2000.

[36] Yee-Chia Yeo, Qiang Lu, and Chenming Hu. MOSFET gate oxide reliability: anode hole injection model and its applications. *International Journal of High Speed Electronics and Systems*, 11(3):849–886, 2001.

[37] J. W. McPherson. Intrinsic time-dependent dielectric breakdown in SiO_2 dielectrics. *Oxide Reliability Summary of Silicon Oxide Wearout, Breakdown, and Reliability*, pages 135–171, 2001.

[38] K. P. Cheung. A physics-based, unified gate-oxide breakdown model. *IEEE Intl Electron Device Meeting*, pages 719–723, 1999.

[39] K. P. Cheung. Thin gate-oxide reliability - the current status. *Intl Nano Symposium*, 2001.

[40] E. Wu, W. Abadeer, L.-K. Han, S.-H. Lo, and G. R. Hueckel. Challenges for accurate reliability projections in the ultra-thin oxide regime. *IEEE International Reliability Physics Symposium*, 1999.

[41] R. Tu, J. King, and H. Shin. Simulating process-induced gate oxide damage in circuits. *IEEE Transactions on Electron Devices*, 44(9), 1997.

[42] Y. Ye, M. Khellah, D. Somasekhar, A. Farhang, and V. De. A 6 GHz 16 Kbytes L1 cache in a 100nm dual-v_t technology using a bitline reduction (BLR) technique. *Journal of Solid-State Circuits*, 38(5):839–842, May 2003.

[43] Qiang Li, Wei Li, Jinlong Zhang, and Jiann S. Yuann. Soft breakdown and hot carrier reliability of CMOS RF mixer and redesign. In *IEEE Radio Frequency Integrated Circuits Symposium*, pages 399–402, 2002.