Gate Oxide Reliability: Physical and Computational Models

Andrea Ghetti

1 Introduction

Since its birth, the microelectronics industry has been characterized by the continuous struggle to find new technological processes that allow the reduction of the physical dimensions of the devices integrated in a single chip of silicon. As matter of fact, since the invention of the first integrated circuit (IC) the number of single devices per chip has kept doubling every 18 months, that corresponds to a steady exponential growth over the last 30 years. Such shrinking process is driven by the fact that smaller device operate at higher speeds and allow the integration of more and more complex circuits of the same area of silicon making each single function less and less expensive. However, the operating voltage does not scale with the same pace, hence the electric fields inside the devices keep increasing. This leads to a degradation of the device performance over time even during normal operation. Therefore, it is necessary to guarantee that microelectronics product performance remains within the customer's specifications for a determined period of time. This is the concept of reliability.

The large majority of the microelectronics products are bases on the Metal-Oxide-Semiconductor (MOS) transistor that is schematically shown in Fig. 1. Two heavily doped regions are formed in a semiconductor substrate to make the source and drain extensions. The gate electrode is built between source and drain over an insulator layer of silicon dioxide (or simply "oxide"), and controls the conduction between source and drain through the electric field across the oxide. When no bias is applied to the gate with respect to the substrate, source and drain are isolated. On the contrary, if the applied gate voltage is high enough, a thin conductive layer of electrons is induced in the substrate, and connects source and drain (channel). In this condition, if also a drain voltage is applied, current can flow from source to drain.

There are many electrical parameters that such a device is required to keep during its working life, and there are many physical phenomena that can degrade them. This Chapter deals only with the reliability of the oxide layer. From the above description, it appears clearly that the operation of the MOS transistor is based on the insulating properties of the oxide layer. As any other dielectric material, there is a maximum field that makes silicon

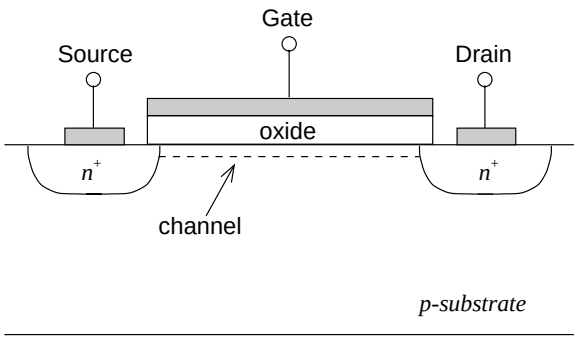

Fig. 1. Schematic structure of a n-MOS transistor

dioxide lose its insulating properties (breakdown) as it is applied (dielectric strength). Obviously, MOS devices operate at lower field. However, if a lower electric field is applied for a long enough time, the oxide slowly degrades (wear out), and eventually breaks down anyway. This phenomenon called time dependent dielectric breakdown (Tddb) is an important parameter for MOS device reliability and it is the object of this Chapter.

Usually, time dependent dielectric breakdown is divided in two categories: extrinsic and intrinsic breakdown. Breakdown is defined intrinsic when it is related only to the oxide structure, while it is extrinsic when it is due to defects that can be present because of the many technological steps needed to make an IC. Extrinsic and intrinsic breakdown have different characteristics, and, usually, are characterized in different ways. However, with the continuous reduction of the oxide thickness and increase of the electric field, intrinsic breakdown has become the most likely problem as far as oxide reliability is concerned.

In summary, this Chapter is focused on the intrinsic reliability of the gate oxide of MOS devices. Intrinsic oxide reliability is a very complex matter to tackle. There are still many phenomena to be understood. A number of models have been proposed to explain some of the feature of oxide breakdown, but a comprehensive model is still lacking. In this Chapter, we will review the main physical models that have been proposed about intrinsic oxide breakdown. We will try to show pros and cons of each model, possible computational implementation or practical methodology to predict oxide breakdown they enable, and what it is still missing.

This Chapter is arranged as follows. Section 2 details the concept of reliability applied to gate oxide in the microelectronics industry, also introducing the essential elements that are common to all reliability models. Sections 3,4,5 describe in details these elements. In particular, Sect. 5 addresses also the issue of the reliability projection provided by the different models. The different types of breakdown are discussed in Sect. 6. Finally, Sect. 7 draws some conclusions.

2 Gate oxide reliability

2.1 Basic statistics concepts

Time dependent dielectric breakdown is a statistical phenomenon: two identical devices subjected to the same stress break down at different times. Therefore it can only be described in statistical terms. For a device it is possible to define only a failure probability as a function of time, or, more usefully, a set of devices (such as all the transistors or functions of an IC) can be described by a distribution function, and the time to failure of the whole set (lifetime) can be defined as some average value of this distribution.

For reliability characterization we are interested in the following statistics concepts:

- *Cumulative Distribution Function* (CDF), or Probability distribution function, or simply Distribution function $F(t)$: is the fraction of population that has failed before time t. CDF is normalized to 1, and usually it is expressed as a percentage.
- *Reliability Function* $R(t) = 1 - F(t)$: is the fraction of population that survives until time t.
- *Probability Density Function* (PDF): is the percentage of failing devices at any interval of time dt:

$$f(t) = dF(t)/dt \tag{1}$$

- *Failure (or hazard) rate* $h(t)$: is the rate at which a unit is expected to fail, given that is has survived until time t:

$$h(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)} \quad . \tag{2}$$

A particularly useful CDF to describe the statistical properties of oxide breakdown (see Sec. 4) is the *Weibull* distribution family [1]. A Weibull distribution is described by the following expressions:

$$h(t) = \frac{\beta}{\alpha}\, t^{\beta-1} \tag{3}$$

$$f(t) = \frac{\beta}{\alpha}\, t^{\beta-1}\, e^{-\frac{t^\beta}{\alpha}} \tag{4}$$

$$R(t) = e^{-\frac{t^\beta}{\alpha}} \tag{5}$$

$$F(t) = 1 \; - \; e^{-\frac{t^\beta}{\alpha}} \tag{6}$$

where α and β are two parameters. By appropriate choice of the two parameters, a wide range of hazard curves can be described. Figure 2 shows the various functions obtained for typical values of α and β. A very popular and useful way to plot a Weibull distribution is to plot the so-called Weibull

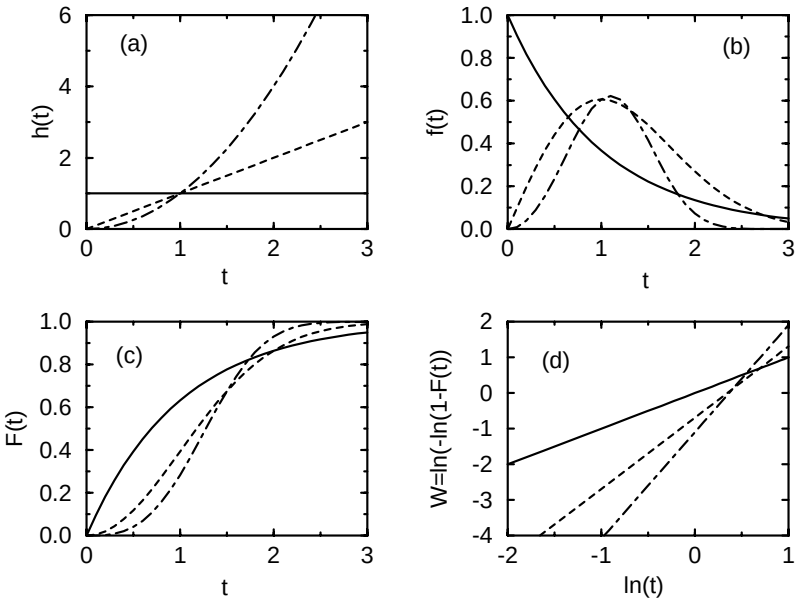

Fig. 2. Reliability functions for the Weibull distribution for different values of α and β. Solid line: $\alpha = \beta = 1$, dashed line: $\alpha = \beta = 2$, dot-dashed line: $\alpha = \beta = 3$. (a) hazard function; (b) density function; (c) failure distribution; (d) Weibull plot.

number $W \equiv ln(-ln(1-F(t))$ as a function of $ln(t)$. Because of (6), the plot will result in a straight line with slope β (Fig. 2.d):

$$W \equiv ln(-ln(1-F(t)) \;=\; \beta\, ln(t) \;-\; ln(\alpha) \quad . \tag{7}$$

The slope β is a very important parameter for lifetime extrapolation as shown in Sect. 4. The parameter α is also called modal value.

As mentioned before, time to breakdown (T_{BD}) cannot be described by a single number, since it is statistically distributed. Therefore we must use some average value of this distribution. Usually, T_{BD} is taken as the time corresponding to same specific value of the failure rate (p):

$$F(T_{BD}^{(p)}) = p \ . \tag{8}$$

In the case of the Weibull distribution, p is usually taken as 63%, that corresponds to $W = 0$ in the Weibull plot. In this Chapter, we will always indicate with T_{BD} the time to breakdown corresponding to $F = 63\%$, unless otherwise noted.

As an example, Fig. 3 shows an experimental determination of the oxide failure distribution in the typical Weibull plot. Data for the devices with larger area lay on two straight lines with very different slope. The part of the distribution with the smaller slope is due to extrinsic failures. It is present

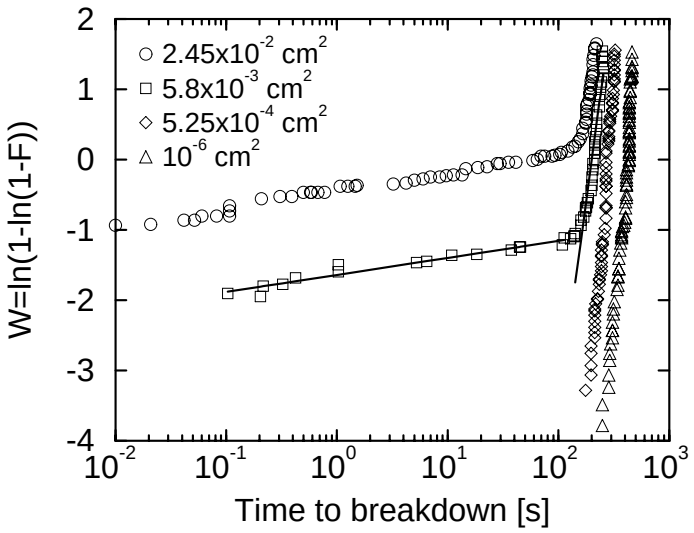

Fig. 3. Weibull plot of the breakdown distribution of capacitors with $t_{ox} = 11nm$ but different area. The two lines are a linear fit of the extrinsic and intrinsic part, respectively. Reproduced from [2].

only in the samples with the larger area because in large area it is more likely to find extrinsic defects. Reducing the oxide thickness (t_{ox}), the time to intrinsic breakdown decreases more than the extrinsic one, and the intrinsic Weibull slope becomes smaller too, making intrinsic breakdown the dominant failure mechanism in ultra thin oxides.

2.2 The industry problem

As mentioned before, reliability is to guarantee that product performance remains within customer's specifications for a determined period of time. From the point of view of oxide breakdown, the usual definition adopted by the microelectronics industry is the following [3]: "After 10 years of operation at the nominal conditions (voltage and temperature) at most 100 devices per million can be broken".

This definition, if taken literally, implies several difficulties. The most severe is the time required to perform such a check. Obviously, it is not possible to carry out qualification test for ten years. Therefore, it is necessary to perform some kind of accelerated test, and then extrapolate to the real operating conditions. And it is precisely to perform this extrapolation procedure that physical and computational models of oxide breakdown have been developed.

There many way to accelerate degradation of an oxide. The most straight forward is to applied a larger voltage (or current) than the nominal one (see Sec. 3). However, this voltage scaling is the most critical since differ-

ent degradation mechanisms providing different voltage dependence of oxide breakdown can be active at the different voltages. And it is in this field that most of the investigation carried out so far is focused because of its importance.

Another way to accelerate degradation is to increase the temperature. Empirically it has been seen that increasing the temperature determines a shorter time to breakdown. However, the temperature dependence of breakdown is the feature less investigated of the degradation phenomena. It involves the microscopic process of defect creation in a complex structure such as the amorphous silicon dioxide whose physical modeling is a difficult task to carry out. Usually the temperature dependence is described by an Arrhenius type law where the activation energy has to be found empirically, although recently experimental finding about a non-Arrhenius dependence have also been reported [4–6].

Moreover, also the adoption of samples with area larger than real device reduces time to breakdown simply because the number of possible failure spot increases. But then, an area scaling is necessary.

It must also be noticed that it is difficult and time consuming to test millions of devices to verify such low failure rate as that required by the industry standard definition above. Generally much smaller population is used to estimate oxide breakdown distribution. Thus, only time to breakdown corresponding at high failure rates (such as 50% or 63%) is usually measured in lab tests that must be then scaled to the required failure rate (percentile scaling).

In summary, to extrapolate oxide lifetime from accelerated lab tests to the real operating conditions a number of scaling operation are necessary. Area and percentile scaling are well defined if the statistical distribution is known. Therefore this operation is influenced by the uncertainty whit which the distribution is know due to the limited number of tested devices (see Sec. 4). Temperature scaling it is still not well known, and it is usually described by an Arrhenius type law. There are instead a lot of models about voltage scaling that are described in more details in the following of this Chapter.

2.3 General model

All the models about oxide intrinsic breakdown have a common denominator that is suggested by the following experimental evidence. Let's consider a MOS device with ultra thin oxide (i.e. $t_{ox} = 4nm$) subjected to a constant voltage stress (CVS). At the beginning, the virgin device features the initial $I - V$ characteristic with the typical exponential dependence on the applied bias as shown in Fig. 4. Then, the stress bias is applied. During the first phase of stress: (a) the gate current slowly changes (see Fig. 5); (b) some charge is trapped inside the oxide layer as shown by flat-band or threshold voltage shift (this is also one of the reasons the gate current changes in time); (c) interface traps are created as pointed out by charge pumping and other

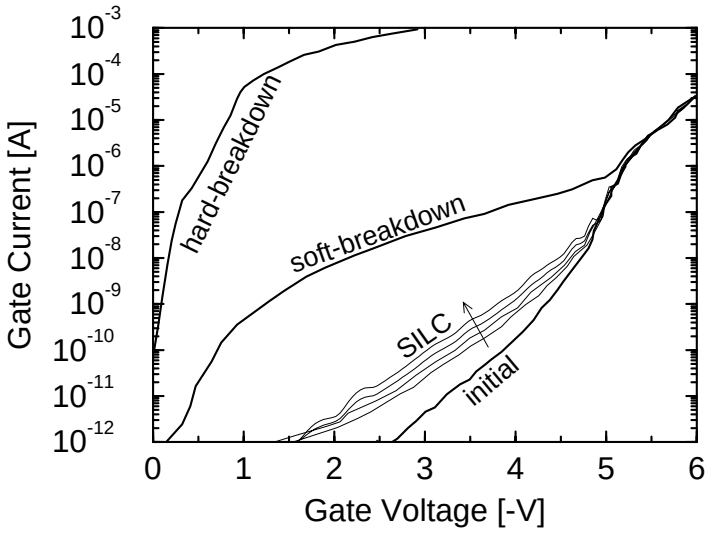

Fig. 4. Typical change of the $I-V$ characteristics during stress ($t_{ox} = 4nm$). Data from [7].

measurements; (d) the leakage current increases especially in the low voltage regime (see Fig. 4). This additional current is present only after stress and it is therefore called Stress Induced Leakage Current (SILC). It increases with stress, it still has an exponential dependence on the gate bias although with a smaller slope with respect to the initial curve, it becomes the main conduction mechanisms at low voltage, and it is generally attributed to trap-assisted tunneling [8,9] (see Par. 3.3). All these evidences suggest that defects (traps) are created within the oxide layer because of the electrical stress. This phase is therefore called "wear-out" of the oxide.

Continuing the stress, the gate current experiences sudden jumps and becomes much noisy. This phase is called quasi breakdown [10], or SILC B mode [11], or, more commonly, soft breakdown [12]. It is called soft breakdown because the leakage current increases significantly with respect to the initial $I-V$ characteristics or SILC (see Fig. 4), but is not as high as a complete or hard breakdown. Soft breakdown is characterized by a large noise of the leakage current [13,14] and an $I-V$ characteristic that follows a power law [15] as oppose to an exponential dependence of SILC [8,16–18] and an almost linear one of the hard breakdown [18]. Soft breakdown is generally attributed to carrier hopping between nearby traps as in a percolation phenomenon [12,19,20]. This is possible only if the the trap density is above some critical value.

Finally, the continuing creation of oxide traps leads eventually to hard breakdown that is characterized by a large increase of the leakage current with an almost ohmic conduction. Hard breakdown is also indicated with

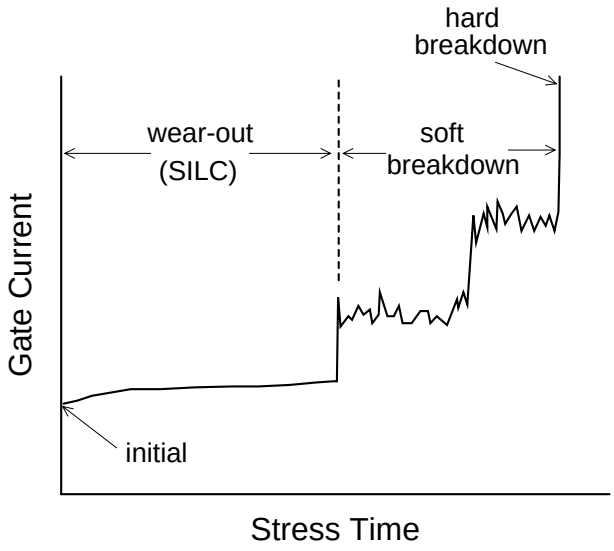

Fig. 5. Typical behavior of the gate current under constant voltage stress ($t_{ox} = 4nm$). Axis are not in scale to show the entire evolution of the current during stress. Data from [7].

"thermal" breakdown because is associated with an irreversible transformation of the oxide structure due to the discharge of the energy stored in the breakdown spot at the moment of th conductive path formation [20]. The probability to have an hard or soft breakdown depends on the conductive properties of the breakdown spot as well as the stress conditions (see Sec. 6). Because of this, soft breakdown is present only in thin oxides (such the one in Figs. 4,5).

Therefore, the general model of oxide breakdown suggested by these experimental facts is the following and it is schematically depicted in Fig. 6. The applied bias generates carrier with high energy. These energetic carriers create traps in the oxide. When the trap density reaches some critical value (that corresponds to the formation of a conductive path in a localized spot) there is breakdown. Breakdown is hard or soft depending on the stress condition, device parameters, etc.

Mathematically, this model can be expressed as [21]:

$$T_{BD} = k \, \frac{(N_{BD})^{1/m}}{R_G} \, e^{E_a/KT} \quad , \tag{9}$$

where N_{BD} is the critical trap density, R_G is the rate at which oxide trap are created by energetic carriers, m is the nonlinearity coefficient of the trap time evolution (e.g. $N_{trap}(t) \sim \alpha t^m$), k is a proportionality constant, and the exponential term represents the Arrhenius type dependence on temperature.

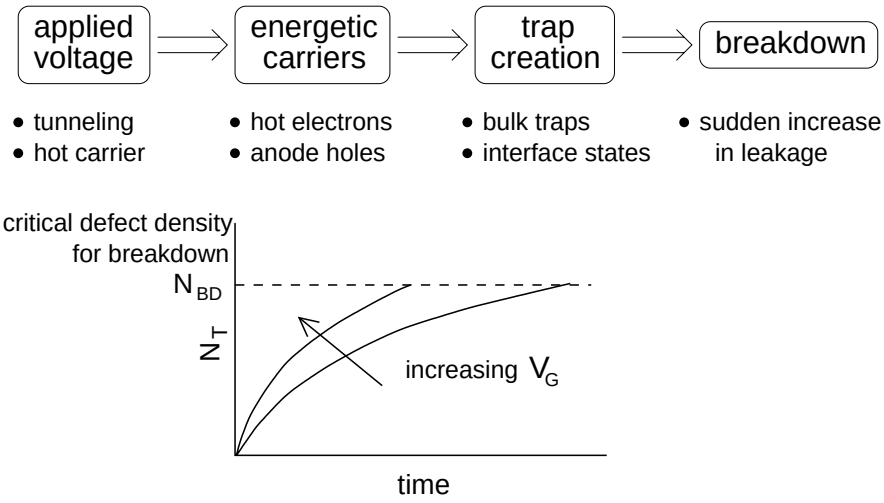

Fig. 6. Schematic illustration of the general framework of breakdown models.

A lot of models fit into this general framework. Obviously they differ for the physical mechanisms responsible for trap creation, and, therefore, for the numerical expressions of N_{BD}, R_G, m, E_a, and their dependence on the stress conditions. These models will be described in the following. In particular, Sect. 3 shows what kind of energetic carriers are generated depending on the stress conditions. Section 4 investigates the properties of N_{BD}. Section 5 describes the trap creation models and the corresponding reliability prediction they provide.

3 Electrical stress and carrier energy

3.1 Tunneling

One of the way to stress the oxide layer of a MOS devices is to apply an high voltage to the gate. In this way carrier can tunnel through the oxide and gain energy at the expenses of the high oxide field. Because of the exponential dependence of the tunneling current on the oxide thickness, tunneling current increases very rapidly with the continuous reduction of the oxide thickness. Therefore, tunneling is a very important phenomenon for oxide reliability and must be adequately described to enable accurate prediction of oxide lifetime.

Tunneling current has many components that are easily detectable in ultra thin oxides. They are schematically depicted in Fig. 7. DTE/DTH is the direct tunneling component for electrons/holes. VBE is the component due to valence band electron tunneling. Recently, also a tunneling component assisted by interface states (TEDit) has been reported [22,23]. ETAT and HTAT are the trap assisted components for electrons and holes, respectively

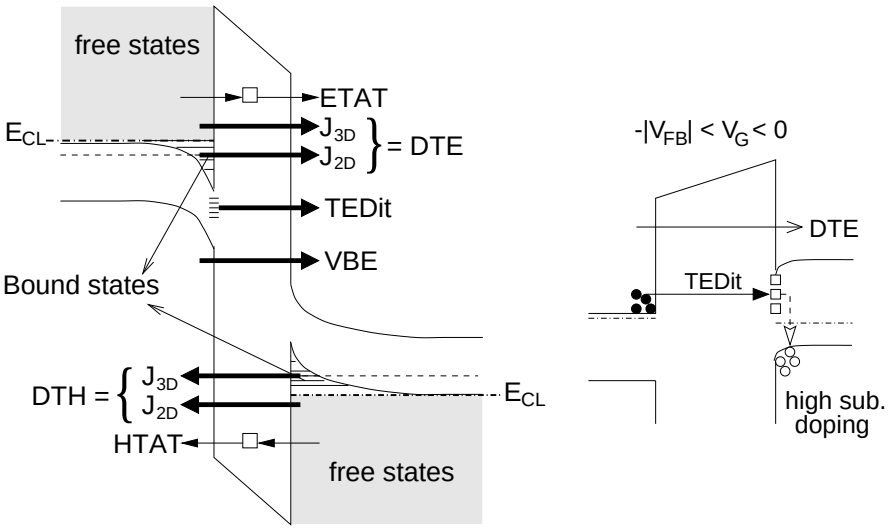

Fig. 7. Schematic representation of the potential profile and of the different components of the tunneling current. Electrons/holes with energy above/below E_{cl} form a free gas. Right: TEDit component for a n-MOSFET when $-|V_{FB}| < V_G < 0$.

(see Par. 3.3). All of these components are made of the contribution of both free and bound states (J_{3D} and J_{2D} respectively).

An accurate modeling of these tunneling current components in ultra thin oxide MOS devices requires the inclusion of important phenomena for nowday's MOS devices such as quantization effects of both electrons and holes in both inversion and accumulation regime, polysilicon depletion and several tunneling mechanisms. For this reason, these effects have been lately introduced in the tunneling current modeling in different ways, (see, for example, [24–30]), but not all of them in a single comprehensive model. In the following, we will outline an accurate model for tunneling current that includes all of the aforementioned effects. It has been coded in the simulation program QUASI [31] and proved to be accurate enough in many conditions [18]. This model is also one of the building block of the overall computational model for oxide breakdown [32,33] described later in Sec. 5.

Self-consistent Potential Profile

One of the key element for an accurate modeling of the tunneling current in thin oxides is the inclusion of quantization effects. There are many ways of different degree of accuracy to account for this effects. In our approach, charge quantization effects of both electrons and holes are accounted for by solving self-consistently the 1D Poisson and Schrödinger equations, providing best accuracy [34] and maximum flexibility. The adoption of the Fermi-Dirac

statistics is also needed for a precise determination of the potential profile. Since the tunneling current is generally too small to affect the potential profile, thermal equilibrium is assumed. In a limited number of cases involving minority carrier generation/recombination [35,23] an accurate simulation would require the inclusion of the continuity equation in the model.

Self-consistency is achieved through an iteration scheme. In order to speed up convergence, the Poisson equation is written in the following non linear form [36]:

$$
\begin{aligned}
&-\nabla \cdot (\epsilon \ \nabla V^{k+1}) = \\
&q\Bigg[N_V F_{\frac{1}{2}} \left(\frac{q}{k_B T}(V^k - V^{k+1}) + F_{\frac{1}{2}}^{-1}\left(\frac{p}{N_V}\right)\right) - \\
&\quad N_C F_{\frac{1}{2}} \left(\frac{q}{k_B T}(V^{k+1} - V^k) + F_{\frac{1}{2}}^{-1}\left(\frac{n}{N_C}\right)\right) + \\
&\quad N(z)\Bigg],
\end{aligned}
\tag{10}
$$

where $F_{\frac{1}{2}}(\eta)$ is the Fermi-Dirac integral of order $\frac{1}{2}$, $N(z)$ is the net doping, and n/p is the electron/hole concentration as provided by the Schrödinger equation:

$$
-\frac{\hbar^2}{2\ m_z}\ \frac{\partial^2 \zeta}{\partial z^2} - qV(z) = E_z\ \zeta\ . \tag{11}
$$

In order to compute the correct charge density in both accumulation and inversion layers using the minimum number of bound states possible, we take into account bound states up to a given threshold energy E_{cl}, above which carriers are thought to form a free gas [25] (see Fig. 7). Thus:

$$
n(z) = n_{3D}(z) + n_{2D}(z) \tag{12}
$$

$$
\begin{aligned}
n_{2D}(z) = \sum_{ij}^{E_{ij}<E_{cl}} & \frac{g_j m_{d_j} k_B T}{\pi \hbar^2} \times \\
& ln\left(\frac{1 + e^{\frac{E_F - E_{ij}}{k_B T}}}{1 + e^{\frac{E_F - E_{cl}}{k_B T}}}\right) |\zeta_{ij}(z)|^2
\end{aligned}
\tag{13}
$$

$$
E_M(z) = max(E_{cl}, E_C(z)) \tag{14}
$$

$$
n_{3D}(z) = N_C F_{\frac{1}{2}}^{(i)}\left(\frac{E_F - E_M}{k_B T}, \frac{E_{cl} - E_C(z)}{k_B T}\right) \tag{15}
$$

where $E_C(z)$ is the conduction band edge, g_j is the degeneracy of the $j-th$ valley, m_{d_j} its density of states effective mass, E_{ij} and ζ_{ij} are respectively

the energy level and the corresponding envelope function of the $i-th$ bound state in the $j-th$ valley, and $F^{(i)}_{\frac{1}{2}}(z,b)$ is the incomplete Fermi-Dirac integral as defined in [37]. Hole quantization is treated in a symmetric way.

Gate depletion effects are implicitly accounted for by solving for the potential also over the gate region. The poly is modeled as silicon, but considering the appropriate work function accounting for the correct doping level and band gap narrowing [38].

Transmission Probability

Another important element is the transmission probability P_T. A general, but approximated way to compute P_T is the WKB approximation:

$$\mathcal{P}_T = e^{-\int_a^b |k(x)|dx} \tag{16}$$

where the integral is extended to the forbidden gap and $k(x)$ is the (imaginary) wave vector.

A more accurate, yet general, way to compute P_T is based on the exact solution of the Schrödinger equation in term of Airy functions [39] and the transfer matrix method.

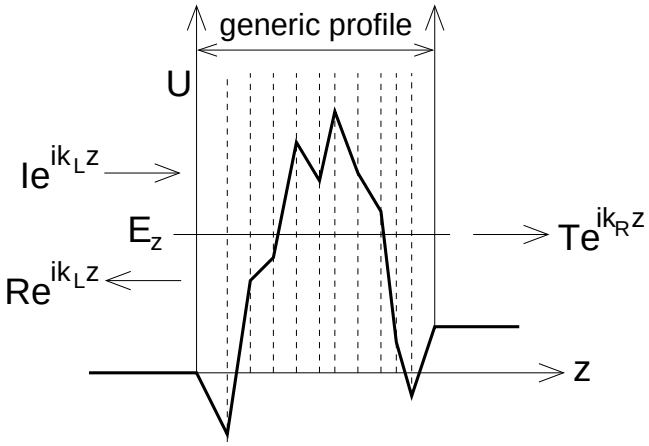

Fig. 8. Schematic representation of the potential profile for transmission probability computation.

Let's consider an incoming plane wave $\zeta_i = Ie^{ik_L z}$ which is partly reflected $\zeta_r = Re^{-ik_L z}$ and partly transmitted $\zeta_t = Te^{ik_R z}$ through a generic potential barrier (Fig. 8). The transmission probability is defined as the ratio of the transmitted and incoming current

$$\mathcal{P}_T = \frac{J_{\zeta_t}}{J_{\zeta_i}} = \frac{v_{gt}\ |\zeta_t|^2}{v_{gi}\ |\zeta_i|^2} = \frac{k_R\ m_L\ |T|^2}{k_L\ m_R\ |I|^2}. \tag{17}$$

where v_g is the group velocity. Following the example of [40], (11) can be solved analytically at all energies with open boundary conditions assuming the potential energy profile ($U = -qV$) be approximated with a piece linear function:

$$U^{(i)}(z) = U_i + s_i|F_i|(z - z_i); \quad z_i < z < z_{i+1} \tag{18}$$

where

$$F_i = \frac{U_{i+1} - U_i}{z_{i+1} - z_i}; \quad s_i = sign(F_i). \tag{19}$$

Using the variable

$$u^{(i)} = s_i \left(\frac{2m_i}{\hbar^2} |F_i| \right)^{\frac{1}{3}} \left[z - z_i + \frac{U_i - E_z}{s_i \ |F_i|} \right] \tag{20}$$

(11) becomes

$$\frac{d^2\zeta_i}{{du^{(i)}}^2} - u^{(i)} \ \zeta_i = 0, \tag{21}$$

whose solutions are, by definition, the Airy functions (Ai and Bi)

$$\zeta_i(z) = C_1^{(i)} \ Ai(u^{(i)}(z)) \ + \ C_2^{(i)} \ Bi(u^{(i)}(z)).$$

If $F_i = 0$ the solutions of (11) are:

$$\zeta_i(z) = C_1^{(i)} \ cos(k_i z) + C_2^{(i)} \ sin(k_i z)$$

where $k_i = \frac{\sqrt{2m_i(E_z - U_i)}}{\hbar}$ if $E_z > U_i$, or

$$\zeta_i(z) = C_1^{(i)} \ e^{-\alpha_i z} + C_2^{(i)} \ e^{\alpha_i z}$$

where $\alpha_i = \frac{\sqrt{2m_i(U_i - E_z)}}{\hbar}$ if $E_z < U_i$.

Summarizing, the solution of (11) in the range $z_i < z < z_{i+1}$ can be written as

$$\zeta_i(z) = C_1^{(i)} \ S_1^{(i)}(z) + C_2^{(i)} \ S_2^{(i)}(z) \ , \tag{22}$$

where

$$S_1^{(i)}(z) = \begin{cases} Ai(u^{(i)}(z)) & \text{if } F_i \neq 0 \\ cos(k_i z) & \text{if } F_i = 0 \text{ and } E_z > U_i \\ e^{-\alpha_i z} & \text{if } F_i = 0 \text{ and } E_z < U_i \end{cases}$$

$$S_2^{(i)}(z) = \begin{cases} Bi(u^{(i)}(z)) & \text{if } F_i \neq 0 \\ sin(k_i z) & \text{if } F_i = 0 \text{ and } E_z > U_i \\ e^{\alpha_i z} & \text{if } F_i = 0 \text{ and } E_z < U_i \end{cases}$$

In order to determine the envelope function over the whole potential barrier, boundary conditions must be enforced at each grid node [41]

$$\begin{aligned} \zeta_{i-1}(z_i) &= \zeta_i(z_i) \\ \frac{1}{m_{i-1}}\frac{d\zeta_{i-1}(z_i)}{dz} &= \frac{1}{m_i}\frac{d\zeta_i(z_i)}{dz} \end{aligned} \tag{23}$$

that can be written in a matrix form

$$M_{i-1}(z_i)\ C^{(i-1)} = M_i(z_i)\ C^{(i)} \tag{24}$$

where

$$M_i(z) = \begin{bmatrix} S_1^{(i)}(z) & S_2^{(i)}(z) \\ \frac{1}{m_i}\frac{dS_1^{(i)}(z)}{dz} & \frac{1}{m_i}\frac{dS_2^{(i)}(z)}{dz} \end{bmatrix};$$

$$C^{(i)} = \begin{bmatrix} C_1^{(i)} \\ C_2^{(i)} \end{bmatrix}.$$

Then, the transfer matrix is

$$M = \prod_{i=1}^{n} M_{i-1}^{-1}(z_i)\ M_i(z_i) \tag{25}$$

where

$$M_0(z) = \begin{bmatrix} e^{jk_L z} & e^{-jk_L z} \\ \frac{jk_L}{m_L}e^{jk_L z} & \frac{-jk_L}{m_L}e^{-jk_L z} \end{bmatrix}$$

$$M_n(z) = \begin{bmatrix} e^{jk_R z} & e^{-jk_R z} \\ \frac{jk_R}{m_R}e^{jk_R z} & \frac{-jk_R}{m_R}e^{-jk_R z} \end{bmatrix},$$

that relates the envelope functions at both sides of the potential barrier as

$$\begin{bmatrix} I \\ R \end{bmatrix} = M \begin{bmatrix} T \\ 0 \end{bmatrix}, \tag{26}$$

from which it derives immediately

$$\frac{|T|^2}{|I|^2} = \frac{1}{|M_{11}|^2}. \tag{27}$$

When compared with the standard WKB approximation, this method gives comparable results for energies well below the energy barrier, but provides a smoother transition between tunneling and thermionic emission and accounts for possible quantum reflection above the barrier and at the anode-oxide interface.

Tunneling Current Components

The tunneling current can be generally computed summing up the contribution of all $\boldsymbol{k}$ states. In the case of direct tunneling electrons, and because of the silicon band structure, the sum over all free states reduces to:

$$J_{3D} = \frac{qgm_d}{4\pi^3\hbar^3} \int_{E_{cl}}^{\infty} dE_{\perp} \int_{0}^{\infty} dE_{||}\ \Delta f(E_{\perp} + E_{||}) \int_{0}^{2\pi} \mathcal{P}_T(E_{\perp}, E_{||}, \theta) d\theta, \tag{28}$$

where Δf is the difference of the Fermi-Dirac statistics at the two sides of the barrier.

The contribution to the tunneling current of a bound state depends on its lifetime (or decay time), that is the time spent by an electron in the bound state before tunneling away. It can be computed as [27]:

$$\tau_L = \frac{\tau(E_i)}{\mathcal{P}_T} = \frac{\int_a^b \sqrt{\frac{2m_z}{E_i - E_C(z)}}\, dz}{\mathcal{P}_T}, \tag{29}$$

where a and b are the classical turning points. Thus, the tunneling current from all bound states is given by

$$J_{2D} = \sum_{i}^{E_i < E_{cl}} \frac{qg_i m_{d_i}}{2\pi^2\hbar^2\tau(E_i)} \int_{0}^{\infty} dE_{||}\ \Delta f(E_i + E_{||}) \int_{0}^{2\pi} \mathcal{P}_T(E_i, E_{||}, \theta) d\theta. \tag{30}$$

The corresponding hole components are computed in a symmetric way accounting for the appropriate band structure and potential profile.

There is no conceptual difference between conduction band and valence band electrons, apart from their different band structure. Therefore the valence to conduction band tunneling current can be computed as in (28) taking into account the valence band structure and the appropriate potential profile. The trap-assisted components are discussed in Par. 3.3.

Figure 9 shows the accuracy of this model by comparing measurements and simulations of the gate current. Notice that simulations were performed adopting device parameters (t_{ox} and doping) from independent measurements (ellipsometry and $C - V$ curves).

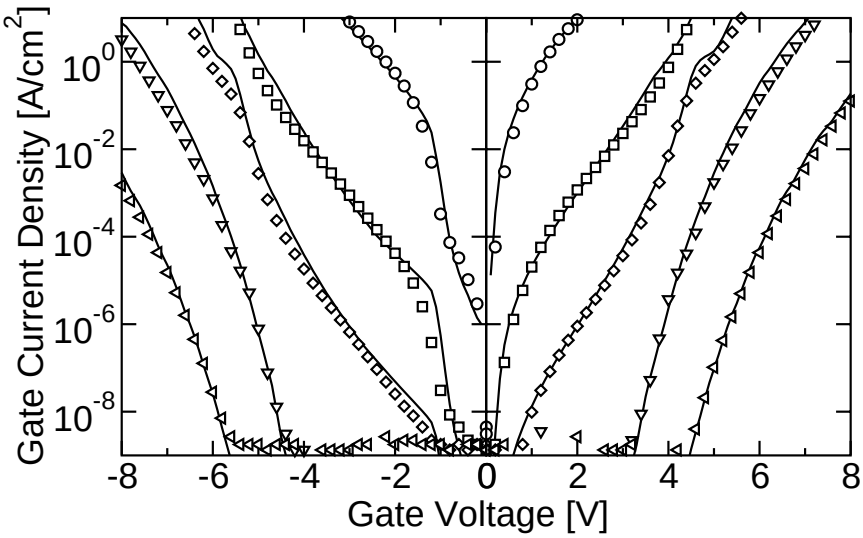

Fig. 9. Tunneling IV measurements (symbols) and simulations (lines) for n^+poly N-MOS transistors. Simulations were performed adopting the t_{ox} given by ellipsometry. Oxide thicknesses are (in nm): 1.56 (○), 2.47 (□), 3.27 (◇), 4.59 (△), 6.56 (◁). Devices with t_{ox} of 2.47nm and 3.27nm feature a TAT component at low gate voltage as explained in [42] (see Par. 3.3). From [31].

Fowler-Nordheim tunneling

If the voltage drop across the oxide $V_{ox} > 3V$, then tunneling electrons see a triangular energy barrier and also drift in the oxide conduction band. This situation is know as Fowler-Nordheim tunneling (FNT). While in the oxide conduction band, injected electrons can lose some of the energy gained by the oxide field (F_{ox}) through scattering. In this case, their energy distribution can accurately be computed with the Monte Carlo (MC) method [43–45]. Figure 10.a shows the average electron energy as a function of the distance from the point of injection in the oxide conduction band for different fields. After an initial transient, equilibrium between scattering and field is reached. In this condition, electron energy depends only on field. Figure 10.b shows the good agreement between MC simulations and experimental data of the final energy attained in the oxide. In the absence of MC simulation, electron energy in the oxide can be described analytically by [43]:

$$\frac{dE}{dx} = q\ F_{ox} - \frac{E}{\lambda} \tag{31}$$

whose solution is indicated by the dot-dashed line in Fig. 10. Notice the increase of the average energy near the emitting interface due to the lack of back scattered electrons. Because of this, slightly different λ are needed to fit the steady-state ($\lambda = 3nm$) or emission ($\lambda = 4.3nm$) energy.

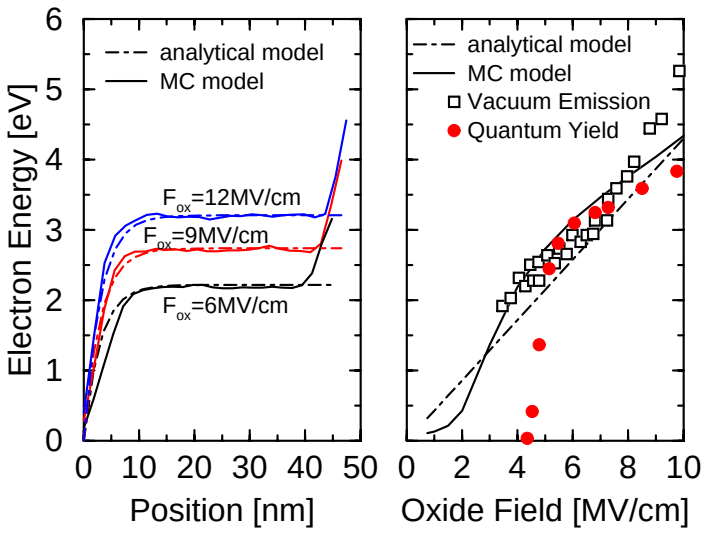

Fig. 10. Electron average energy in the oxide conduction band under uniform field. Left: steady-state average energy as a function of the distance from the injection point. Right: average emission energy at the interface with the silicon as a function of the oxide field. Symbols: experimental data from [43]. Solid lines: MC model. Dot-dashed lines: analytical model.

The WKB approximation provides the following expression for the Fowler-Nordheim tunneling current [46]:

$$J = A\ F_{ox}^2\ e^{-B/F_{ox}} \tag{32}$$

where F_{ox} is the electric field across the oxide, and A and $B \approx 265 MV/cm$ are known constants. Thus, not only the energy, but also the tunneling current depends only on F_{ox}. This suggested the practice of stressing an oxide by forcing a constant current (CCS). In this condition, same current implies same field, which implies same final energy. Thus a fair comparison among oxides of different thickness could be done. However, this is not true anymore in today's ultra thin oxides where direct tunneling takes place (see Fig. 11). In ultra thin oxides, carrier energy is mostly due to the applied voltage, thus constant voltage stress (CVS) is preferable.

Finally, Fig. 12 shows the energy distribution of injected electrons computed for different conditions. Direct tunneling and Fowler-Nordheim tunneling with no oxide scattering distributions feature a sharp peak around qV_{ox}, and have been computed with the tunneling model previously shown in details. Distribution accounting also for oxide scattering has a broader shape and has been computed with MC simulation [45].

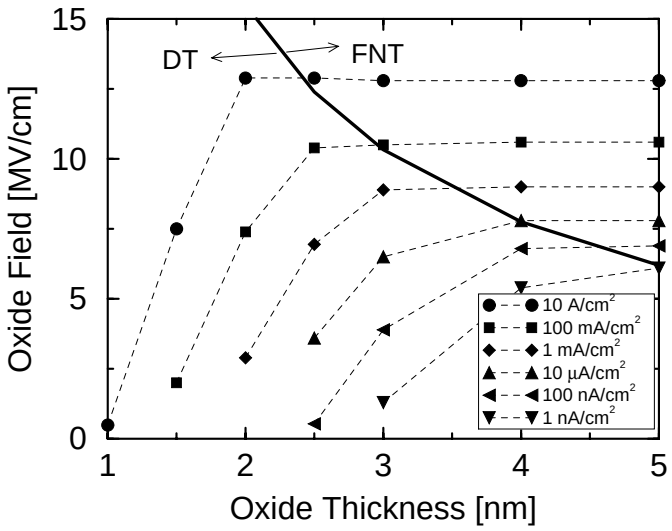

Fig. 11. Simulated relation between gate current and oxide field for different oxide thickness. Adapted from [47].

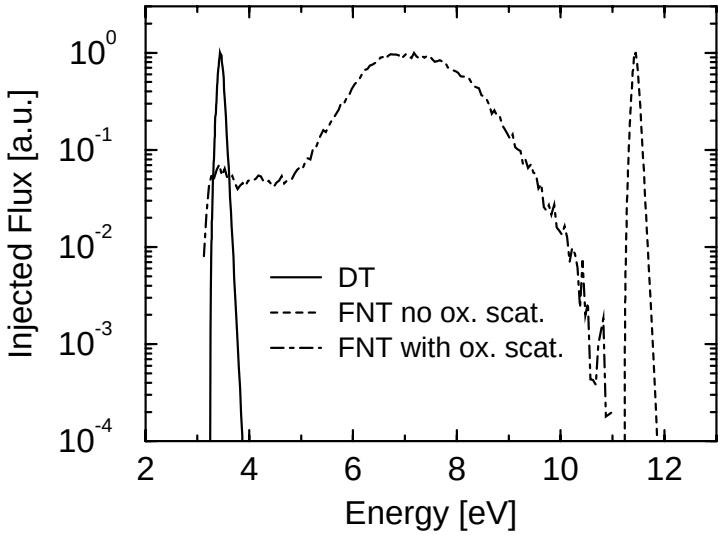

Fig. 12. Energy distribution of electrons injected in the silicon substrate for the same oxide field ($F_{ox} = 10MV/cm$), but different conditions. Solid line: $t_{ox} = 3.5nm$ and direct tunneling. Dashed line: $t_{ox} = 11.7nm$ and Fowler-Nordheim without oxide scattering. Dot-dashed line: $t_{ox} = 11.7nm$ and Fowler-Nordheim tunneling with oxide scattering.

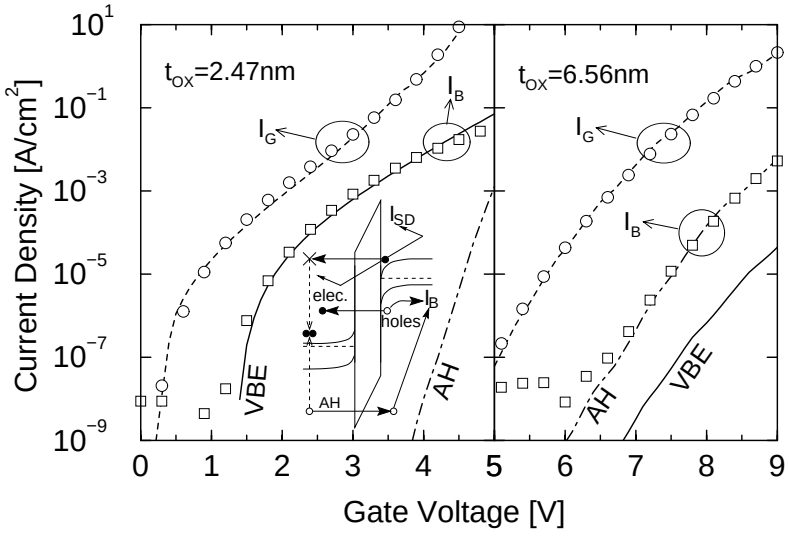

Fig. 13. Carrier separation experiment in n-MOS transistors with different oxide thickness. Symbols: measurements; lines: simulations. Adapted from [18].

3.2 Carrier separation experiments

An important type of measurements to investigate the physics of oxide breakdown are carrier separation experiments. Figure 13 reports the typical results of such kind of experiment for n-MOS transistors. For $V_G > 0$ the electrons tunneling from the channel are supplied by source/drain, while holes left behind by valence band tunneling electrons are collected by the substrate contact. Moreover, energetic electrons injected into the anode can generate secondary holes by impact ionization (II). These anode holes (AH) can tunnel back also contributing to the substrate current (I_B). This is a carrier separation experiment because electron current is supplied by source/drain electrodes, while hole current by the substrate contact.

This kind of experiments is important because it allows to separately measure electron and hole fluence through the oxide, thus allowing to correlate oxide breakdown with a particular type of carrier. As a matter of fact, one of the main models for oxide breakdown, the Anode Hole Injection model (AHI) (see Sec. 5), relates defect generation (hence breakdown) to these anode holes on the basis of the results of this kind of experiment. However, it must be noticed that the AH current is the dominant component of I_B only when t_{ox} is thick enough ($t_{ox} > \approx 4nm$ [48]), as demonstrated in Fig. 13.right. On the contrary, if t_{ox} is thin enough the AH component is negligible and I_B coincides with the VBE component as shown in Fig. 13.left.

The anode hole current (J_h) can be expressed as:

$$J_h = J_e \, \gamma \, T_H \tag{33}$$

where J_e is the initial electron current, γ is the average number of holes created by II by each injected electron (also known as Quantum Yield), and T_H is the average hole injection probability.

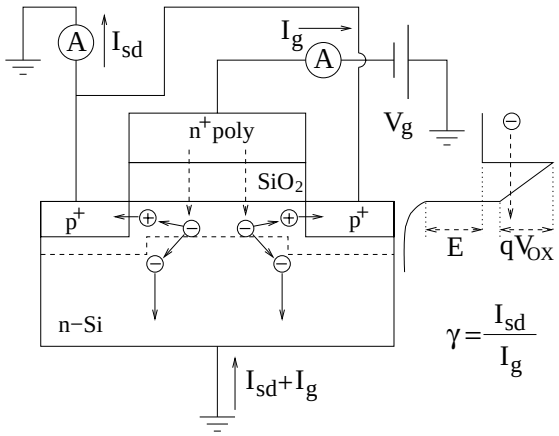

Fig. 14. Schematic illustration of quantum yield experiment. Basically, it is a carrier separation measurement on a p-MOS transistor.

The quantum yield γ can be measured with a carrier separation experiment in p-MOS transistors, which is schematically depicted in Fig. 14. Applying a negative bias to the gate of a p-MOSFET, electrons are injected by tunneling into the silicon bulk (anode) with a large kinetic energy. These energetic electrons create new electron-hole pairs by impact ionization. The bulk electric field pushes the electrons toward the substrate contact while holes are collected by source and drain. The ratio of source-drain current (I_{SD}) to the gate current (I_G) is the average number of hole created by each electron, i.e. the quantum yield $\gamma = I_{SD}/I_G$. γ is a function of the electron energy [49–51]. Neglecting the contribution of the substrate field (doping dependent) [52], this function can be considered the same in all devices. In any case, γ can be accurately computed in all conditions with Monte Carlo simulation as shown in Fig. 15.

Given the popularity of the AHI model, many attempts have been made to compute J_h. Analytical approaches such as [54,55] generally use (31) for the energy of injected electrons, the universal curve of γ, and some analytical function for the distribution of the anode hole needed to compute T_H. However, this have been shown not to be accurate in all conditions [56]. A more general and accurate way to compute J_h is to use coupled silicon/oxide Monte Carlo simulation of both electron and hole transport as done in Fig. 13, which also provides the correct distribution of hot anode holes.

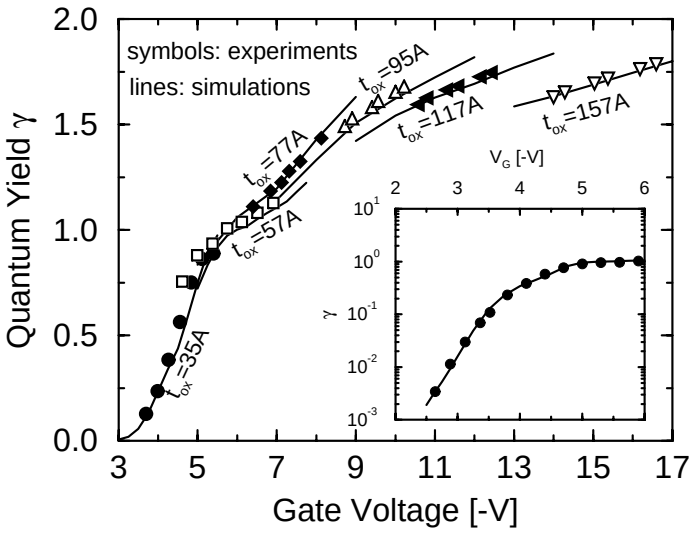

Fig. 15. Comparison of experimental data and Monte Carlo simulation of quantum yield. Experimental data for $t_{ox} = 35, 57\mathring{A}$ are from [49], the others from [51]. *Inset*: different voltage dependence of γ for high and low voltages. Data from [53].

3.3 Trap-Assisted tunneling

As the oxide layer wears out during stress, an additional leakage current components, called SILC, shows up, as pointed out in Par. 2.3. This current still retains an exponential dependence on the applied bias, but with about half the slope ($\mathrm{d}log(I)/\mathrm{d}V$) of the direct tunneling component (Fig. 4). For this reasons, when first detected, SILC was modeled with a Fowler-Nordheim type formula (32), but with unphysical parameters A and B to fit the experimental data [57].

Today, SILC is generally attributed to trap-assisted tunneling (TAT) [8,9,58]: each tunneling electron tunnels in and out of only one trap (Fig. 16). The resulting steady-state current is given by the balance of these two components. In first approximation, if we assume that the direct tunneling (DT) component is proportional to the transmission probability T

$$J_{DT} \approx T(t_{ox}, V_G), \tag{34}$$

then the steady-state TAT component J_{TAT} is proportional to [59]:

$$J_{TAT} \approx \sigma N_{tr} \frac{T_1 T_2}{T_1 + T_2} \tag{35}$$

where σ is the trap capture cross section, N_{tr} is the trap density, and T_1 and T_2 are the in and out transmission probability, respectively. It is possible

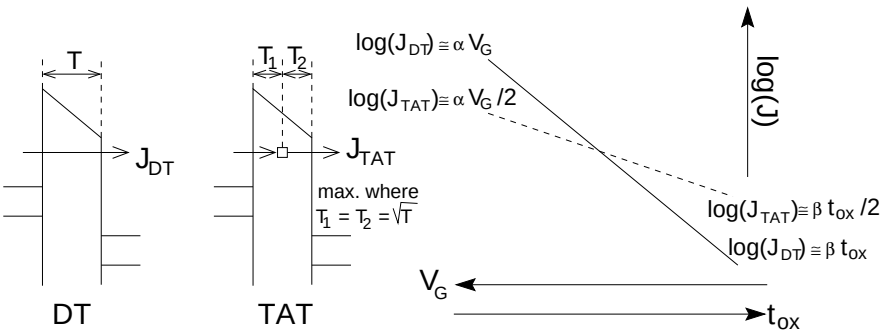

Fig. 16. Schematic representation of the voltage and oxide thickness dependence of direct (DT) and trap-assisted (TAT) tunneling.

demonstrate that (35) features a sharp peak where $T_1 = T_2$ [9]. Thus (35) is approximatively equivalent to

$$J_{TAT} \approx \sigma N_{tr} \frac{\sqrt{T(t_{ox}, V_G)}}{2} \ . \tag{36}$$

Equation (36) means that, if J_{DT} has a particular slope when plotted in semilog scale as a function of either V_G and t_{ox}, then J_{TAT} has half of that slope because the same T appears under square root (see Fig. 16). This is simply the mathematical representation of the fact that the tunneling distance is cut in half for the TAT component.

Another important feature of SILC for oxide breakdown modeling is the fact the SILC component is inelastic, i.e. electrons tunneling through a trap created by electrical stress lose some of their energy [8,50,53,60,61]. Although there are some uncertainties on the exact value of this energy loss [53], many experimental data reported in the literature converge to the range 1.2 - 1.5 eV [8,50,53]. This observation is extremely important for oxide breakdown because implies that, although SILC is the main current component at low voltage (i.e. at the nominal bias), it is not as effective as the direct tunneling current in damaging the oxide. Thus projection made from higher voltages where the direct tunneling current is the main component still make sense.

When taken all of this into account, (35) demonstrated to accurately describe the SILC component, as shown in Fig. 17. Notice that in Fig. 17 all the parameters of the TAT model were provided by independent measurements.

Equation (35) has several implications. First, it implies that SILC (i.e. TAT component) is proportional to the trap density [58,63]. Thus, SILC can be taken as a monitor of oxide degradation. Usually, oxide degradation is quantified by the relative increase of the leakage current

$$\Delta J/J_0 = \frac{J(V_P, t) - J(V_P, 0)}{J(V_P, 0)} \tag{37}$$

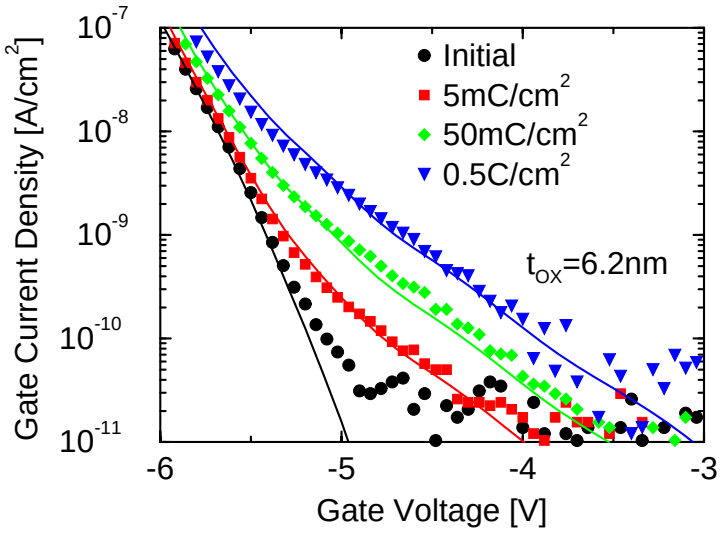

Fig. 17. Simulations based on the TAT model (lines) and experimental data (symbols) of SILC. All parameters of TAT model are from independent measurements: trap density from [17,62], trap cross section from percolation theory (see Sec. 4), energy loss from quantum yield experiment [50,53]. Experimental data from [17,62].

computed at a particular probe voltage V_P [64]. The numerator of (37) is the SILC component, whereas the denominator is simply the initial current. The normalization by J_0 aims to cancel out the effect of other parameters (such as t_{ox}), thus pointing out only the contribution of N_{tr}. However, because of (36), a $1/\sqrt{T(t_{ox}, V_G)}$ dependence still remains in $\Delta J/J_0$. Thus, comparisons of $\Delta J/J_0$ as a measure of N_{tr} are quantitatively correct only for oxide of similar thickness and for the same probe condition [65]. In addition, also charge trapping during stress can change the potential profile, hence the tunneling current, with different effect depending on t_{ox}, applied bias, stress time.

Second, (36) also explains why SILC is detectable only in a limited range of t_{ox}. It is not visible for very thick oxides because when it is the largest current component its contribution is still below the measurement threshold. And it is difficultly detectable for very thin oxides because it is always overcome by the direct tunneling component [18]. Thus other degradation monitors must be used for ultra-thin oxides.

Interface state assisted tunneling

Recently, it has been pointed out that in ultra-thin oxides ($t_{ox} < 3nm$) a current component due to tunneling into interface states is present for low applied voltages [22,23]. In particular, when the gate voltage (V_G) is between 0 and the flat-band voltage ($|V_{FB}|$) gate electrons face interface

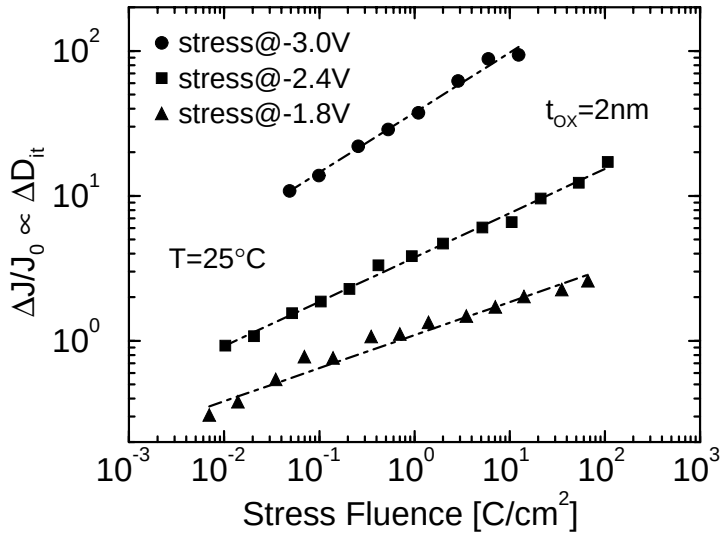

Fig. 18. Increase of $\Delta J/J_0$ due to TEDit component as a function of the stress fluence. Dot-dashed lines represent the least square best fit of the data with a power law. The sense voltage is $V_S = -0.4V$. Adapted from [23].

state of the substrate/oxide interface if the substrate doping is high enough [23], and can tunnel through them (TEDit component of Fig. 7). It has been shown that this component is very sensitive to electrical stress. Moreover, the conditions needed to detect this TEDit component are ultra-thin oxide and high substrate doping, i.e. the typical conditions of advanced devices. Thus, continuing the shrinking process, the TEDit component will become more and more relevant, and might be a good for monitoring oxide degradation.

In particular, Fig. 18 shows $\Delta J/J_0$ due to the TEDit component. Its value ($> 10^{-1}$) is at least two order of magnitude higher than conventional SILC ($10^{-3} - 10^{-1}$) [66,67]. Thus, TEDit appears to be more reliable and less sensitive to disturbing phenomena than conventional SILC detection in ultra-thin oxides. In particular, it allows the clear detection of oxide degradation in reasonable time even for stress at very low field.

It must be noticed that $\Delta J/J_0$ in Fig. 18 follows a power law:

$$\Delta J/J_0 = \alpha \; t^m \tag{38}$$

This relation is very important for reliability projection because indicates the rate at which defects are created in the oxide. As a matter of fact, the exponent m is the same that appears in the general formula for oxide lifetime (9). The majority of the data reported in the literature agree on similar power law evolution of oxide damage. However, some papers also reported linear behavior [68], or an initial linear relationship that later tends to saturate [58], or a sigmoidal shape [69] (a linear region bracketed by sub-linear portions at

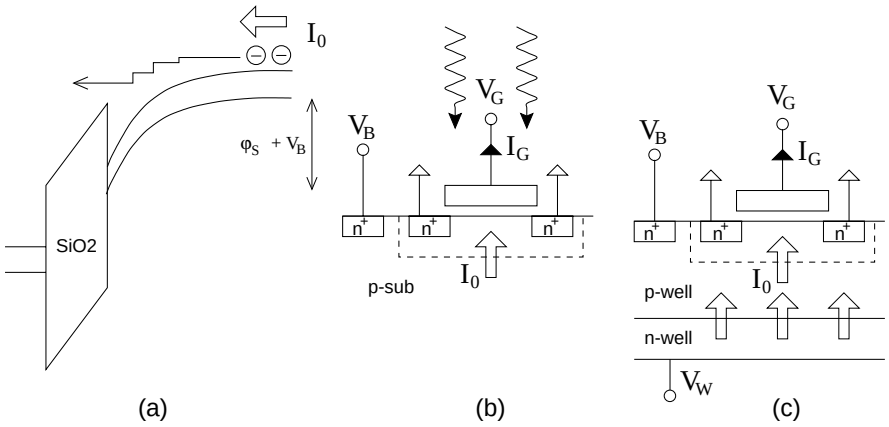

Fig. 19. Schematic representation of the substrate homogeneous hot electron injection experiment. a) band diagram; b) optical generation; c) electrical injection.

low and high fluence). The choice of the time evolution of oxide damage greatly impact lifetime extrapolation as explained in details in [65].

3.4 Hot carriers

Another type of energetic carriers that can be used to stress an oxide are the so-called "hot carriers". Hot carriers are particle that attain an high energy while drifting in the semiconductor under the action of an intense electric field. For example, electrons moving in the channel from source to drain see the entire V_{DS}, and can reach the drain with high energy. These are called channel hot electrons (CHE) [70]. CHE can generate holes by impact ionization. These holes, moving toward the substrate, can generate, again by impact ionization, additional electrons that are pushed toward the interface. These latter electrons are generally called Channel Initiated Secondary Electrons (CHISEL) [71]. They see V_{DB} that can be higher than V_{DS}, so they can be more energetic than CHE. However, CHE and CHISEL are not uniformly distributed along the interface, but their damage is localized near the drain.

For this reason, another configuration is usually adopted in reliability tests with hot carriers. It allows a uniform generation of energetic carriers at the interface exploiting the substrate field. For this reason, these energetic carriers are called substrate hot electrons/holes (SHE/SHH). Figure 19 schematically depicts this configuration in the case of SHE. A n-MOS transistor is biased in inversion, with source and drain grounded, and with a large substrate voltage. In these conditions, a large electric field that is essentially one-dimensional and perpendicular to the interface is present in the depletion region under the gate oxide. Since the surface potential (Φ_S) is pinned, the oxide field ($F_{ox} = (V_G - V_{FB} - \Phi_s)/t_{ox}$) is controlled only by the gate voltage V_G, while the substrate voltage (V_B) controls the potential drop and field in the

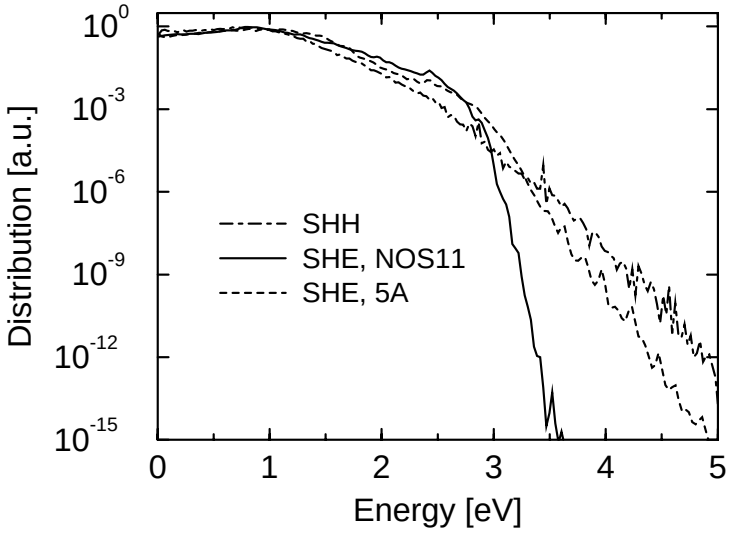

Fig. 20. Interface energy distribution of SHE/SHH computed for different devices. Dot-dashed line: SHH, device PPR1 of [73], $|V_B| = 14V$, $F_{ox} = 5.5MV/cm$. Solid line: SHE, device NOS11 of [74], $|V_B| = 1.9V$, $F_{ox} = 5MV/cm$. Dashed line: SHE, device 5A of [75], $|V_B| = 6V$, $F_{ox} = 2MV/cm$.

substrate. Electrons generated in the substrate at the edge of the depletion region gain energy at the expense of the substrate field while moving toward the interface. Here, some of them are injected in the gate, while the other are collected by source and drain. Electrons can be generated in the substrate by optical generation [72] (Fig. 19.b), or by electrical injection from a buried $p-n$ junction [73] (Fig. 19.c). This experiment, also called Ning's experiment [72] or homogeneous substrate injection, is very important because it allows to control independently oxide field, carrier energy, current intensity through the gate bias, the substrate bias and the light intensity (or forward bias of the buried junction), respectively. Substrate hot hole experiments can be performed similarly on p-MOSFETs [73].

Given its usefulness in studying many hot carrier related phenomena, this experiment has been widely investigated in the literature (see for example [75–77,74]). In particular, an effective tool to accurately compute the injection probability as well as the number of SHE/SHH is again the Monte Carlo method [77,74,78]. Figure 20 shows the energy distribution at the interface computed with the Monte Carlo method. The maximum energy available to SHE/SHH is $q(\Phi_S + V_{SB})$. In modern device with sufficiently high substrate doping such as those of [74] the distribution function features a "plateau" (i.e. it is sufficiently populated) up to that energy. On the contrary, for lower substrate doping or in the case of SHH, the distribution function is much

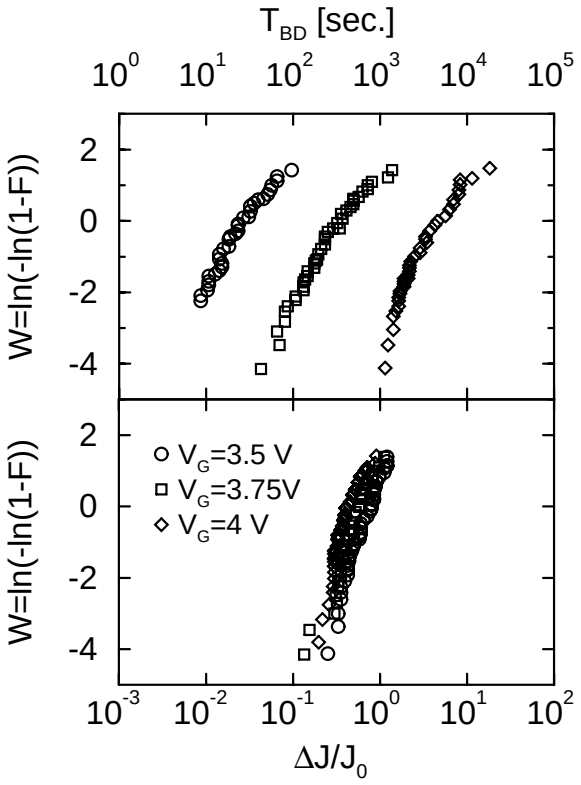

Fig. 21. Weibull distributions of time to breakdown T_{BD} and corresponding $\Delta J/J_0$ ($\propto N_{BD}$). Data from [66].

broader and smoother, so that a simple relation between applied substrate bias and average energy at the interface does not apply.

4 Critical trap density

4.1 Experimental evidence

One of the key insight into the physics of oxide breakdown enabling the development of a predictive model of oxide reliability has been the idea that damage generation up to a critical density would lead to a new conduction path resulting in oxide breakdown [79]. The existence of such critical density has been suggested by experimental data, as those reported in Fig. 21. Time to breakdown for different stress voltages shown in the upper frame feature a strong voltage dependence: T_{BD} changes of 3-4 orders of magnitude for a change of the stress voltage of only $0.5V$. On the other hand, $\Delta J/J_0$ (i.e. SILC) is relatively independent of the stress conditions [66]. Since $\Delta J/J_0$

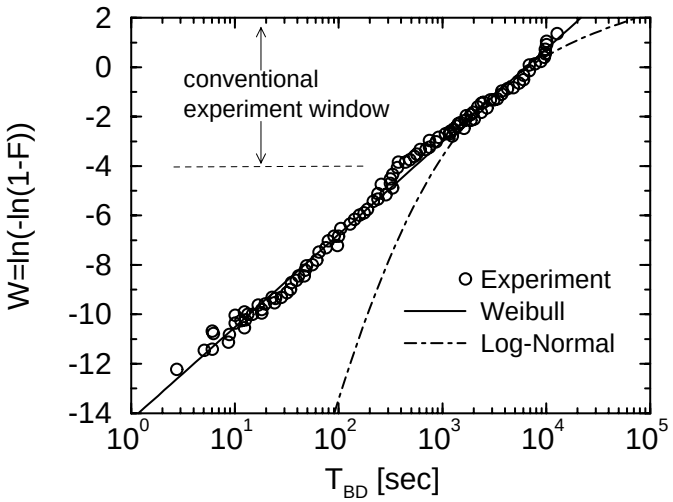

Fig. 22. Weibull an log-normal fits to breakdown distribution. Data from [84]. The Weibull distribution exhibits a much better fit, especially in the low failure percentile part used for reliability projection. Similar results have also been reported in [85].

is proportional to the defect density (see Par. 3.3), Fig. 21 indicates that a well define value of defect density is needed to trigger breakdown. Similar experimental evidence have also been reported in other papers (see for example [80,81]). However, a few papers have also criticized the use of $\Delta J/J_0$ as a probe of the trap density at breakdown [82,83]. But, it is likely that this criticisms stem out of an incorrect use of $\Delta J/J_0$ for quantitative comparisons due to the difficulties pointed out in Par. 3.3.

Moreover, data like those of Fig. 21 have also others implications. First, the critical trap density is, in first approximation, independent of the stress bias. Thus, it depends only on t_{ox}, and the large variation with the stress voltage of T_{BD} is mainly due to the voltage dependence of the trap generation rate. Second, the average trap density at breakdown (N_{BD}) is a stochastic variable. This implies that the breakdown spot (where the threshold density is reached) is much smaller than the device area. Thus, breakdown is a very localized phenomenon. The statistical nature of N_{BD} is due to the stochastic relationship between the local trap density and the total number of defect generated in all the device.

For reliability projection, it is very important to know the statistical distribution of N_{BD}, hence of T_{BD}. Initially, T_{BD} (Q_{BD}) distributions were described by either a Log-normal or Weibull distribution because both of them can fit equally well most sets of data collected on a limited number of samples [84] (see Fig. 22). However, it must be noticed that the Weibull and Log-normal distributions describe two different physical systems. The

Weibull distribution is based on extreme-value statistics and mathematically describes the so-called "weakest-link" problem: the first of many independent devices that fail determines the lifetime of the whole system. On the other hand, the Log-normal distribution derives from the multiplicative model in which the processes degrade over time [84]. From the physical point of view, the Weibull distribution would be more appropriate to describe oxide lifetime because gate oxide failure is a weakest-link type of problem: chip failure is defines by the failure of the first individual device, and a device fails in any small portion of the gate oxide area breaks down. In addition, when compared with a large sets of experimental data, the Weibull function fits much better the breakdown distribution, especially at low percentile (Fig. 22), and provides the correct area dependence on the contrary of the Log-normal distribution [84]. Therefore, the Weibull function has been generally accepted to describe oxide breakdown distribution. Notice that the Log-normal distribution would provide over-optimistic projection, especially for shallower breakdown distribution (thinner oxides).

4.2 Percolation

The concept of a critical defect density was numerically demonstrated to provide the correct statistical description of oxide breakdown in [86]. A device of area A was divided in smaller columnar cells with area S_0. Then, "defect" were generated randomly on the entire device. Breakdown was assumed to be triggered when a critical number of defects n_{bd} was reached in any of the cells. Assuming an uniform defect distribution with average density ρ, and considering each cell to be independent on each other, the probability to find a given number of defects in a cell is described by the Poisson distribution. Based on the properties of the Poisson distribution, it is possible to show that:

$$ln[-ln\{1-F(\rho)\}] = lnA + ln\left[\rho t_{ox} \; - \; \frac{1}{S_0} ln\left\{\sum_{n=0}^{n_{bd}-1} \frac{(S_0 t_{ox}\rho)^n}{n!}\right\}\right]. \quad (39)$$

It was demonstrated in [86] that with an appropriate choice of the two parameters S_0 (area of the breakdown spot) and n_{bd} (local critical number of traps) (39) reproduces satisfactory well the statistical properties of oxide breakdown. The obtained values of S_0 and n_{bd} in [86] are also in agreement with later determinations. Notice that (39) was derived only on the basis of the properties of the Poisson distribution without any hypothesis on the shape of the cumulative distribution failure. Furthermore, it must be pointed out that the area dependence is entirely given by the term lnA.

This approach it is not a predictive model since n_{bd} is treated as a fitting parameter. It was shown later in [32,87,88] that it is possible to compute the critical trap density N_{BD} and the slope of the Weibull distribution β with the percolation model. According to this model breakdown is envisioned as a conduction path of overlapping defects connecting the two oxide interfaces.

In the first implementation of this model [80,87] traps were represented by spheres. These spheres are placed randomly until a cluster of overlapping spheres connects the two interfaces (see Fig. 23). It was demonstrated in [80,87] that with an unique value of the sphere radius (i.e. trap size) the N_{BD} and β dependence on t_{ox} could be quantitatively reproduced.

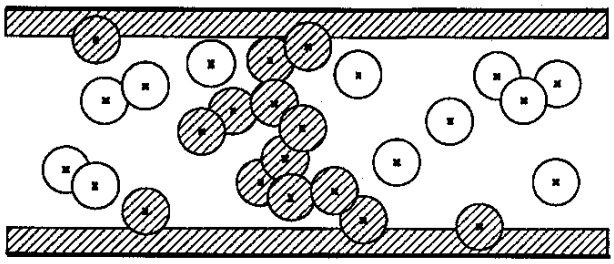

Fig. 23. Schematic illustration of the percolation model with spheres. From [80].

After that, many other implementations of the percolation method have been proposed in the literature (for a review see [89]). Another very popular implementation represents the oxide layer as a simple cubic lattice [32,88]. The random placement of defect is performed simply by choosing randomly one of the elementary cubes. Clusters are then formed with the selected cubes that have at least a face or edge in common with another cube of the cluster. The elementary cube can be of the same size of the trap (a_0) or smaller. In this latter case, a defect spans many lattice sites, and can overlap with another similarly the the previous implementation with spheres. In the limit case of elementary cubes much smaller than a_0, this approach tends to the sphere models. This implementation implies a discretization of the trap position. However, given its simple structure, the cubic lattice greatly simplifies cluster identification, with a significant reduction of the CPU time. This particular implementation of the percolation model has also been adopted in the overall computational model for oxide breakdown [32,33] described in more details later in Sec. 5.

Percolation results provided by this last method are shown in Fig. 24. For the same average defect density, the percolation model provides an increasing failure rate for increasing area, and the vertical shift of $ln(-ln(1-F))$ is $ln(A_2/A_1)$, precisely as indicated by (39). In addition, reducing t_{ox} it is more likely to find a defect cluster connecting the two interface for the same average defect density. In other words, as the oxide is made thinner a percolation path can form with some probability at a lower average defect density. The percolation model, thus, allowed to explain N_{BD} and β reduction for decreasing t_{ox} only on the basis of geometrical and statistical considerations. This also explains the experimental evidence that β is essentially independent of stress voltage and temperature.

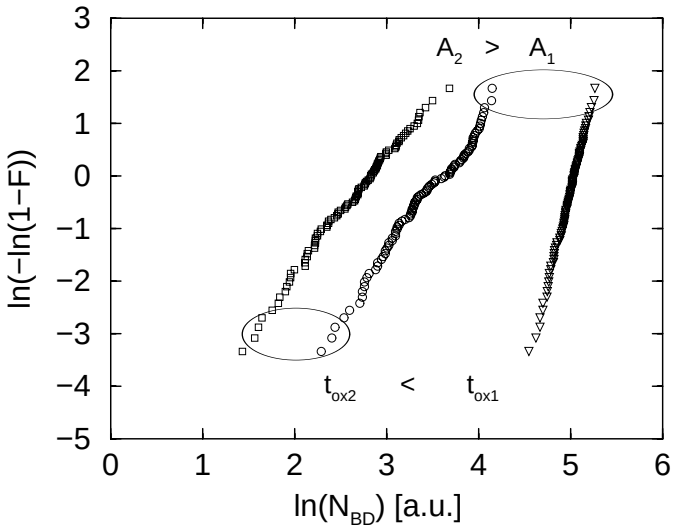

Fig. 24. Percolation simulation results for different oxide thickness ($t_{ox2} < t_{ox1}$) and device area ($A_2 > A_1$). □=(A_2, t_{ox2}); ○=(A_1, t_{ox2}); ∇=(A_1, t_{ox1}).

The trap size a_0 is the only free parameter of the percolation model that determines both N_{BD} and β. In particular, N_{BD} and β decrease for increasing a_0. In the limit case in which only one defect is enough to reach breakdown, β reaches its minimum theoretical value of one.

The value of a_0 can be found by comparing experimental and simulated Weibull distributions. However, the percolation model provides N_{BD} Weibull distribution, while experimentally only Q_{BD} or T_{BD} distributions are available. The translation of the slope of the N_{BD} Weibull distribution ($\beta_{N_{BD}}$) into the slope of the Q_{BD} Weibull distribution ($\beta_{Q_{BD}}$) is straightforward if the defect density obeys the power law relation of (38). In this case, from (9) it derives:

$$\beta_{Q_{BD}} = m\ \beta_{N_{BD}}\ . \tag{40}$$

Assuming $m = 0.33$ [23,83,90] $\beta_{N_{BD}}$ simulated with the cubic lattice approach is translated into $\beta_{Q_{BD}}$ and compared to experimental data in Fig. 25. The best fit was achieved adopting $a_0 = 8\mathring{A}$, which also well compares with the value given by sphere model and extracted from the distribution of soft-breakdown resistance (see Sec. 6).

A somehow larger value of a_0 ($3nm$) was used in [67,88]. However, this higher value might also be due to the restrictive assumption of $m = 1$ made in [67,88], that had to be balanced by a larger trap size in order to reproduce the experimental Weibull slope for $t_{ox} > 3nm$. This, in turn, lead to unrealistic low value of β for $t_{ox} < 2nm$, and thus to very pessimistic projection of oxide reliability for such thin oxides [67], that have been later shown incorrect

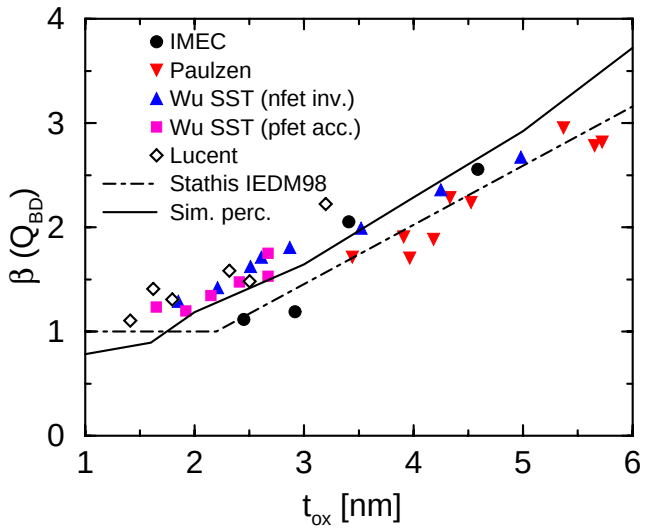

Fig. 25. Comparison of simulated (lines) and experimental (symbols) Q_{BD} Weibull slopes.

by experimental data reporting higher values of β (cfr. Fig. 25) and the compliance with the ITRS reliability specification for t_{ox} as thin as 1.6nm [6,91].

As a final remark about N_{BD}, it must be noticed that recent works have reported an increase of N_{BD} for very thin oxides stressed at very low voltage and for very long time [92]. Although this N_{BD} increase appears at different voltages (that are below $2.8V$ anyway) for different t_{ox}, it features an universal behavior as a function of the stress time. This observation may imply that defects undergo a slow relaxation process that reduces their ability to participate in breakdown [93].

4.3 Area and percentile scaling

For practical reasons, reliability tests are carried out on test structures with larger area than real devices, and on a limited number of samples. Thus, the results of these measurements must be scaled to the real device area and required failure rate. These operations are well defined when the Weibull distribution is known. If T_{test} is the measured time to breakdown corresponding to the measured failure rate F_{test}, and T_{life} is the lifetime at the requested failure rate F_{life}, then, from (7) we can obtain the following relation for failure rate scaling:

$$\frac{T_{life}}{T_{test}} = \left(\frac{ln(1 - F_{test})}{ln(1 - F_{life})} \right)^{1/\beta} . \tag{41}$$

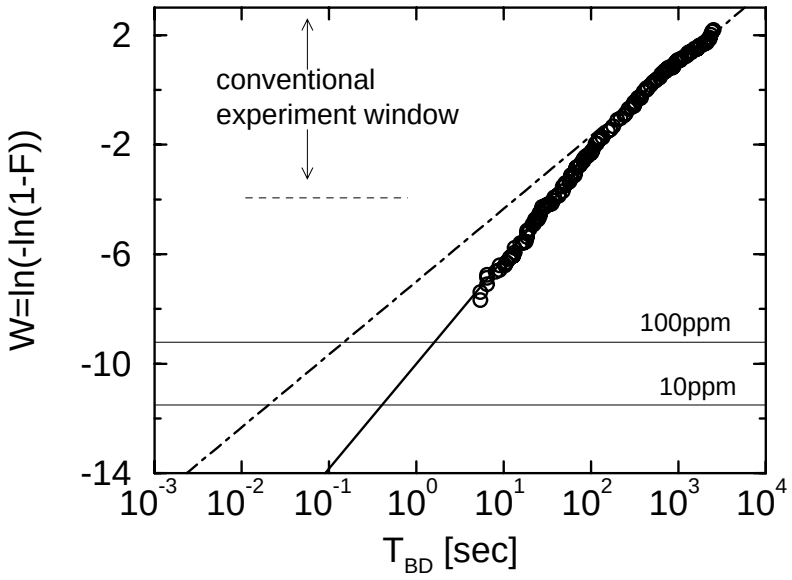

Fig. 26. Difference between percentile projections extrapolating from the higher or lower part of the breakdown distribution because of the non linearity of the Weibull distribution at higher percentile due to the non uniformity of the oxide thickness. Data from [94].

Similarly, combining (7) and (39) we obtain for area scaling

$$\frac{T_{life}}{T_{test}} = \left(\frac{A_{test}}{A_{ox}}\right)^{1/\beta} , \tag{42}$$

where A_{test} is the area of the test structure and A_{ox} is the area of real devices. It is clear from the two equations above that a smaller β means a greater sensitivity to the area and failure rate extrapolations, and implies a larger reduction of the lifetime with respect to the measured value when projecting at low failure rate. Thus β is an important parameter for accurate reliability projection.

A straightforward method to obtain β is an experimental determination of the breakdown distribution. In this case, there are two possible sources of error. The first is the statistical uncertainty due to the limited number of samples measurable. The second arises from possible non linearity of the Weibull distribution, as shown in Fig. 26. In this latter case, the high percentile part of the Weibull distribution features a smaller β than the low percentile part. Unfortunately, with conventional experiment on a limited number of samples, only the high percentile part is sensed, resulting in an underestimation of β and thus in a too pessimistic projection. It has been experimentally and theoretically shown that this non linearity is due to non uniformity of the oxide thickness [94,95].

Another method to obtain β is to exploit the area dependence of T_{BD}/Q_{BD}. because of (42), when plotted in a log-log graph versus oxide area, T_{BD}/Q_{BD} of all areas lay on a straight line with slope $1/\beta$ whose determination is less prone to errors. In the case of non uniform oxides, it is better to use Q_{BD} area dependence, because it is better described by the Poisson distribution at the basis of (42) [96].

Oxide non uniformity has also a detrimental impact on reliability. As a matter of fact, it determines a crowding of the stress current, and thus a localization of the defect generation in the thinnest area, resulting in an apparent reduction of N_{BD} with respect to an uniform oxide of the same average thickness [91]. The dependence of this current crowding on the stress bias can also explain the reported variation of N_{BD} with the stress conditions [32].

5 Defect generation and lifetime extrapolation

Time to breakdown has a strong dependence on the applied bias. It can vary of many orders of magnitude for each volt of variation of the stress voltage (see, for example, Fig. 21). Since in the general expression for T_{BD} (9), N_{BD} can be taken to be essentially independent of the stress voltage, as discussed in the previous Section, T_{BD} dependence on the applied bias is entirely due to the voltage dependence of the trap generation rate R_G. For this reason, R_G has been extensively investigated over the years by several groups. A lot of physical models have been proposed. In this Section, only the main ones can be briefly illustrated: first, the physically based models for which the expression of R_G stems out of a specific physical phenomenon. Then, the phenomenological models that adopts an empirical relation between T_{BD} and stress conditions suggested by experimental data, but without a well defined mechanism responsible for that relation. It must be pointed out that a comprehensive model is still lacking. This is because oxide breakdown is a very complex phenomenon not well understood yet in its microscopic aspects. A sound model for the trap creation process at the microscopic level it still missing.

One of the key element that a R_G model has to provide is the so-called "voltage acceleration factor" defined as:

$$AF = -\frac{\partial\ log(T_{BD})}{\partial V}\ . \tag{43}$$

It indicates how fast T_{BD} changes with the applied bias, thus it is of fundamental importance for reliability projection at the real operating conditions.

5.1 Anode Hole Injection Model

The basic idea of this model is that anode holes are responsible for oxide damage. However, the origin of these holes and the relation between hole

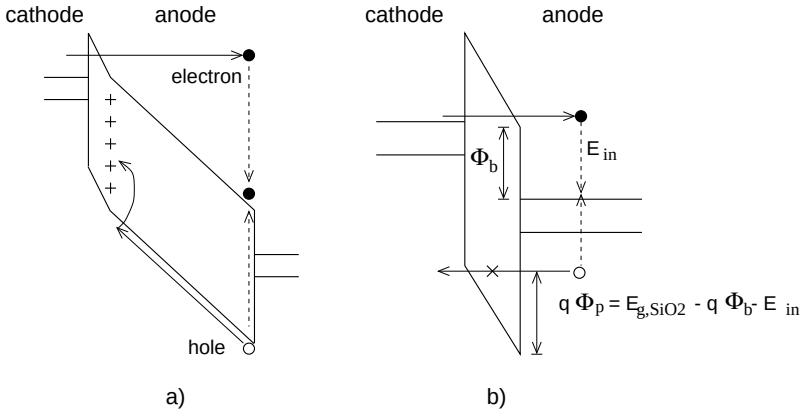

Fig. 27. Schematic illustration of the different versions of Berkeley anode hole injection model. a) for thick oxide/high voltages [97]; b) for thin oxide/low voltages [54] (B-AHI).

flux and breakdown is slightly different among the various versions of the model.

The first application of this concept to the modeling of oxide breakdown was proposed by the Berkeley University [97]. In this version, oxide breakdown is attributed to a positive feedback effect induced by hole generation and trapping at localized spots. At that time, only relatively thick oxides ($t_{ox} > 12nm$) were available that had to be stressed at voltages much larger than the oxide band gap ($V_G > 12 - 13V$). Under these conditions, hole generation was attributed to impact ionization inside the oxide. The resulting model is schematically depicted in Fig. 27.a. Electrons injected into the oxide by Fowler-Nordheim tunneling can gain enough energy to create holes by impact ionization inside the oxide. These holes are driven by the field back to the cathode where some of them get trapped, causing a local increase of the oxide field, and, thus, of the conduction, which further increases hole trapping and oxide conduction leading eventually to breakdown. In this case, the hole current can be evaluated as $J_h = \alpha\, J_e$, where J_e is the Fowler-Nordheim electron current (32), and α is the oxide impact ionization rate given by

$$\alpha(F_{ox}) = \alpha_0\, e^{-H/F_{ox}}, \tag{44}$$

where $B \approx 78MV/cm$ [98]. Thus

$$T_{BD} \sim \frac{1}{J_h} \sim e^{(B+H)/F_{ox}} = e^{G/F_{ox}}\ . \tag{45}$$

The anode hole injection model was then able to explain qualitatively and quantitatively the $1/E$ dependence of $log(T_{BD})$ experimentally found at that time. As a matter of fact, the pre-exponent of the model $G = B + H \approx 350MV/cm$ well compares with the experimental values [99]. For this reason, the anode hole injection model is generally referred to as the $1/E$ model.

If the stress voltage is below $12-13V$, injected electrons still drift in the oxide conduction band. However, they cannot attain the energy needed to impact ionize inside the oxide [44]. Thus, holes responsible for oxide damage are generated in the anode and then injected back through the oxide (Fig. 27.b). That is why the model is called Anode Hole Injection model (AHI). Holes can be generated in the anode by direct impact ionization [55,100,101] or through the excitation of surface plasmons that rapidly decay into hot electron-hole pairs [102]. Theoretical calculation of [102] showed that the latter mechanism is more efficient, but has an onset threshold of $\approx 7.5V$. Based on this observation, the AHI model was questioned because holes could not be present for $V_G < 7.5V$ [103]. However, taking into account also direct impact ionization in the anode the entire range of hole gate current can be consistently explained by AHI even below the $7.5V$ threshold, that is the normal condition for today's ultra thin oxides [21]. Note that while impact ionization inside the oxide depends exponentially on the electric field (44), the impact ionization in the anode depends only on the injected carrier energy, i.e. the applied voltage (cfr. Par. 3.2).

In order to reflect this new physical scenario, the University of Berkeley proposed a modified version of its AHI model in [54] (B-AHI). According to this second version of the AHI model

$$T_{BD} = \frac{Q_p}{J_h} \tag{46}$$

where $Q_p \approx 0.1 C/cm^2$ is a constant value of hole fluence as a function of stress voltage, believed to correspond to the threshold for the onset of the positive feedback [48]. The existence of the critical Q_p was supported by carrier separation experiments like the one reported in Fig. 28 showing a correlation of breakdown with Q_p. The general expression for J_h (33) was evaluated in an analytical way. From WKB approximation

$$T_H = exp\left(-B_h \frac{\Phi_p^{3/2}}{F_{ox}}\right) \tag{47}$$

where Φ_b is defined in Fig. 27.b and E_{in} is given by the solution of (31). γ was assumed constant to 0.08, while J_e was given by a Fowler-Nordheim type of expression, modified to account also for direct tunneling

$$J_e = A\ F_{ox}^2 e^{-B\frac{\left[1-(1-\frac{V_{ox}}{\Phi_b})^{3/2}\right]}{F_{ox}}}\ . \tag{48}$$

When taken all of this into consideration, the model still gives a dependence on $1/E$ of T_{BD}.

This variant of the AHI model suffers of a few limitations. Q_p was found not to be constant as a function of the stress voltage for temperature below 300K [104]. Even at 300K, it decreases if $t_{ox} < 5nm$ [54] making Q_p an additional fitting parameter. Since Q_p cannot be experimentally detected

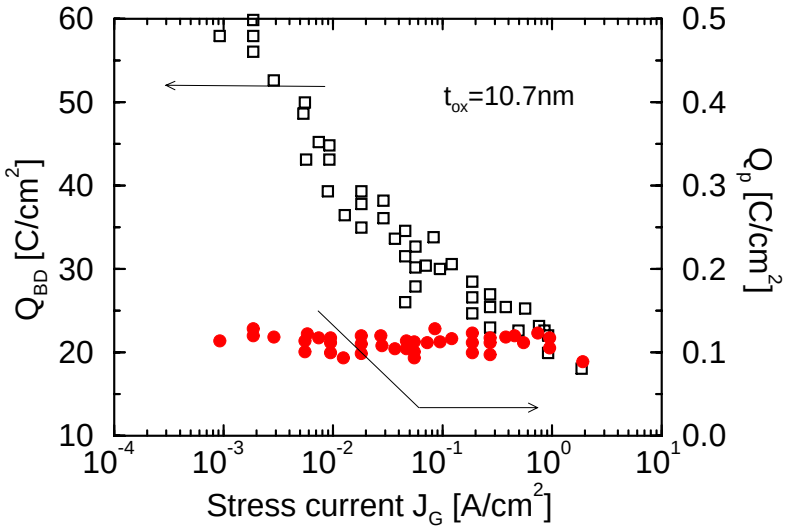

Fig. 28. Electron charge to breakdown (Q_{BD}) and hole fluence to breakdown (Q_p) for a 107$\mathring{A}$ oxide. Q_{BD} increases for decreasing stress current, while Q_p remains constant. Adapted from [48].

for $t_{ox} < 4nm$ because, in this case, the substrate current is dominated by valence band electron tunneling [48], the model looses its predictivity for ultra thin oxides. Moreover γ is assumed constant, while it is a strong function of voltage, especially in the low voltage regime [50,53]. But, above all, it cannot explain the E-dependence of T_{BD} experimentally found at low voltages [105,106].

All of these limitations have been overcome by the latest version of the AHI model proposed by people at Lucent Technologies (L-AHI) [21,32,33]. This variant is based on an accurate numerical implementation of the three basic phenomena involved in the anode hole model: electron tunneling and transport, hole generation and injection, and percolation (Fig. 29). The L-AHI retains only the basic idea of the anode hole concept, i.e. holes created in the anode travel back to the cathode through the oxide, interact with the lattice, and create defects in the process. However, no critical Q_p is used, but T_{BD} is given directly by (9) with $R_G = J_h$. N_{BD} and its statistical properties are simulated with the percolation algorithm (Par. 4.2). Simulated oxide layer is divided is columnar "super-cells" with their own thickness in order to account for thickness variation among devices and due to surface roughness. Traps are placed in the oxide according to the local J_h. The simulation of each element concurring to form J_h (33) has been independently validated. In each super-cell, J_e is computed with the self-consistent Schrödinger-Poisson model shown in details in Par. 3.1, including all tunneling components, that well compares with experiments (Fig. 9). Electron transport in silicon and oxide

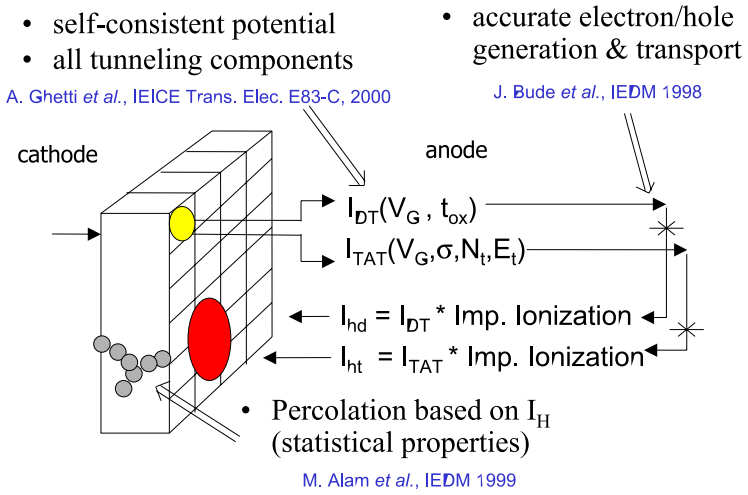

Fig. 29. Schematic illustration of the accurate implementation of the anode hole injection model (L-AHI) [21,32].

(if needed), and hole generation and transport in the anode are simulated with Full-Band Monte Carlo simulation, with independently calibrated rates of phonon scattering and impact ionization. The most significant aspect of the new impact ionization model is the possibility of minority ionization (see later on p. 39). This allows to take into account also the effect of the back gate bias, anode field and doping, and provides an accurate estimate of γ (Fig. 15). Finally, J_h is computed convolving the non equilibrium distribution of anode hole hitting the interface with energy resolved T_H computed with the general and accurate method illustrated on p. 12.

It has been shown that this physically and numerically accurate implementation of the anode hole injection model can quantitatively explain many of the existing experimental data. In particular, it is able to explain the passage from $1/E$ to E dependence of T_{BD} for decreasing voltages (Fig. 30). At high voltage, $J_e \propto exp(-B/E)$ and $T_H \propto exp(-H/E)$ because of Fowler-Nordheim tunneling (32), while γ is approximatively constant (Fig. 15). Thus $log(T_{BD}) \sim log(1/J_h) \sim 1/E$ and the $1/E$ dependence is recovered. On the contrary, for low voltages, J_e and T_H have a weaker dependence on the applied bias because of direct tunneling, while $\gamma \approx exp(B\ V)$ has a strong dependence on the applied voltage (inset of Fig. 15), and dominates the conduction. Thus $log(T_{BD}) \sim log(1/J_h) \sim V \sim E$, and the E dependence is found.

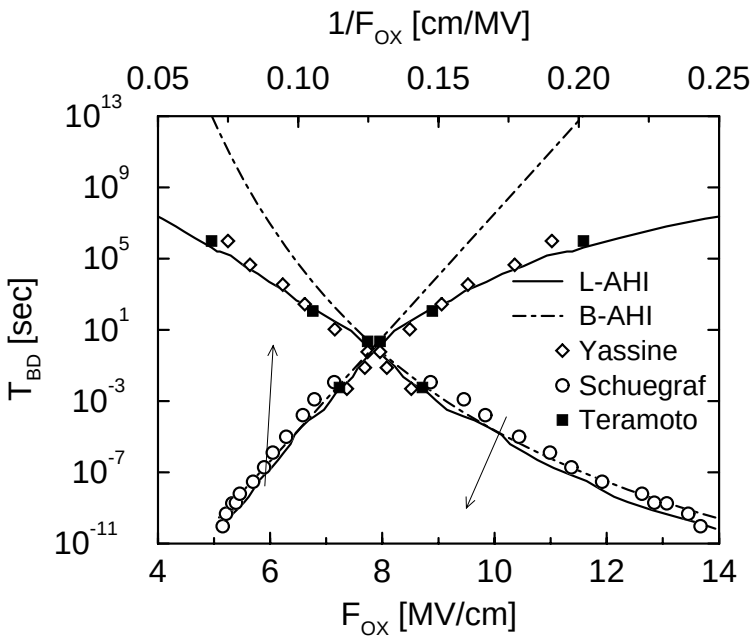

Fig. 30. Simulated (lines) and measured (symbols) field dependence of T_{BD}. The simple Berkeley AHI (B-AHI) always provides a $1/E$ type of dependence, while the accurate implementation of L-AHI is able to catch the passage from $1/E$ to E dependence. Adapted from [33].

The L-AHI is also able to reproduce the thickness dependence of T_{BD} as shown in Fig. 31. For the same F_{ox}, thinner oxides feature higher projected reliability than thicker oxides (Fig. 31.left). This puzzling result can be explained in the framework of L-AHI model remembering that breakdown is energy driven. Same F_{ox} implies a lower applied voltage for the thinner oxides. At lower voltage, the tunneling electrons can produce fewer holes. Since T_{BD} is inversely proportional to J_h, thinner oxide show better reliability. As a matter of fact, when T_{BD} is plotted against the voltage (that is proportional to energy for thin oxides) this apparent inconsistency disappears (Fig. 31.right). This indicates that breakdown for ultra thin oxides is voltage driven [5,107,108].

It is known that ultra thin oxide reliability features a gate voltage polarity asymmetry. That is, for the same F_{ox} and t_{ox}, devices stressed with negative gate voltage (gate injection) exhibit a smaller T_{BD} than devices stressed with a positive gate voltage (substrate injection), as shown in Fig. 32.left [5,107]. In the framework of L-AHI voltage polarity asymmetry is explain by the newly introduced minority ionization mechanisms, schematically depicted in Fig. 33. In the conventional mechanism both recoil and secondary electrons end up in the conduction band. The maximum secondary hole energy (E_H)

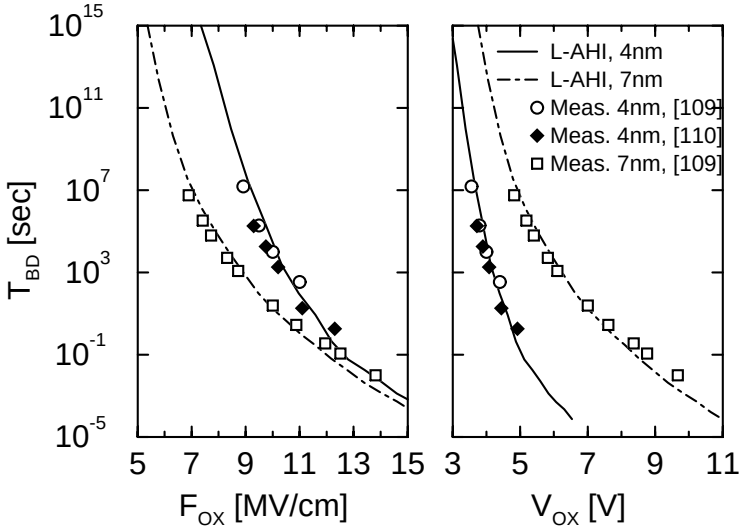

Fig. 31. Simulated (lines) and measured (symbols) T_{BD} for different oxide thickness as a function of the oxide field (left) or the voltage drop across the oxide (right). Experimental data from [109,110]. Adapted from [33].

is $E_H = E_{in} - E_{Gap}$, where E_{in} is the initial electron energy and E_{Gap} is the silicon band gap. However, if the Fermi level is inside the valence band other two mechanisms (called "minority") are possible. In the type 1 mechanism, secondary electron ends up in an unoccupied state of the valence band. In this case $E_H = E_{in} + \Delta E_F$, where ΔE_F is the distance between the bottom of the valence band and the Fermi level. In the type 2 mechanism, both recoil and secondary electrons end up in the valence band and $E_H = E_{in} + E_{Gap} + 2\Delta E_F$. Thus, minority events generate holes with a larger energy than conventional mechanism [100], and are possible only for electron injection toward an hole accumulation or inversion layer, i.e. $V_G < 0$. This causes the asymmetry of the anode hole generation ($G_h(E) = J_h/J_e$) shown in the graph of Fig. 33. If $E_{in} < 5-6eV$ (i.e. $V_G < 5-6V$), more anode hole current is generated under negative voltage stress than under positive stress at a given E_{in} (i.e. F_{ox}), thus explaining the polarity asymmetry when T_{BD} is plotted as a function of F_{ox}. Notice that, in this voltage regime, hole generation rates for the two polarities are shifted approximatively of $1.2V$ (i.e. $\approx V_{FB}$). So, if $G_h(E)$ is plotted as a function of V_G, hole generation rates for the two opposite polarities nearly coincide, explaining why there is no polarity asymmetry if T_{BD} is plotted as a function of V_G [5,107] (Fig. 32.right). On the contrary, for high energy ($V_G > 5 - 6V$) $G_h(E)$ does not feature any polarity asymmetry. Since less energetic electrons are injected for $V_G < 0$ (because the smaller field due to the compensation of V_{FB}), T_{BD} at a given $|V_G|$ is larger for negative V_G than for positive V_G [5].

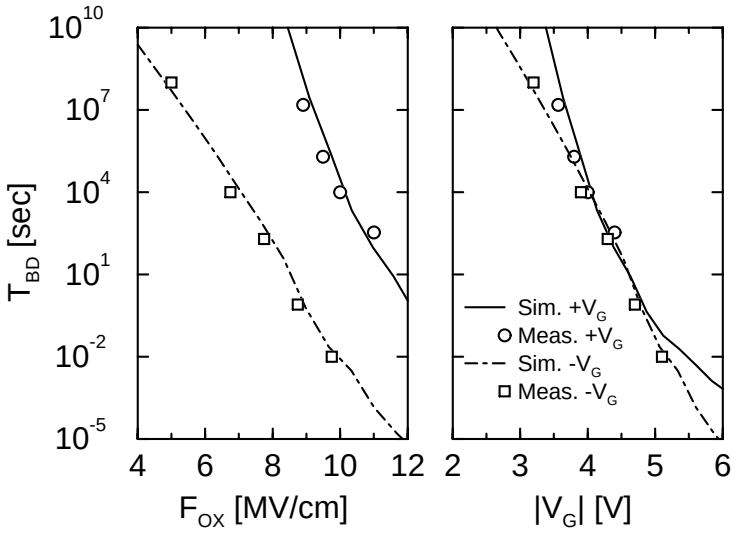

Fig. 32. Simulated (lines) and measured (symbols) T_{BD} for different stress polarity as a function of the oxide field (left) or the gate voltage (right). Experimental data from [109]. Adapted from [33].

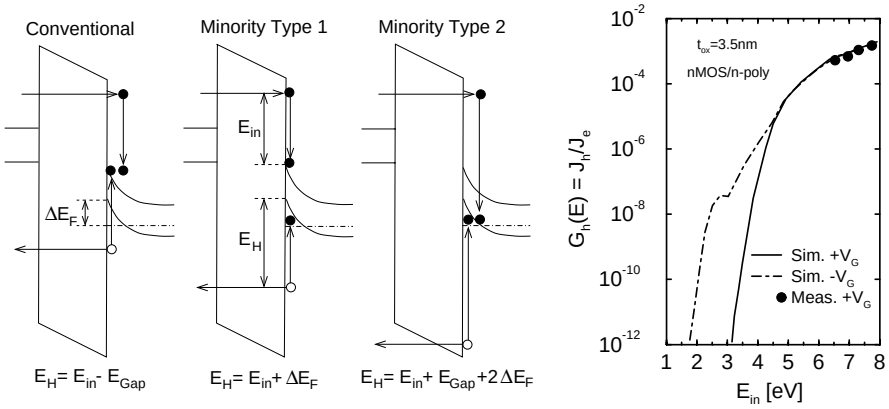

Fig. 33. Schematic illustration of the conventional and minority impact ionization mechanisms and their effect on the voltage polarity of the anode hole generation. Adapted from [100].

The L-AHI model is also consistent with the anode doping dependence of T_{BD}. It has experimentally found that reducing the anode doping T_{BD} of N-MOSFET in inversion decreases [110], featuring a sharp reduction when the poly becomes inverted [108]. Thin is explained, within the L-AHI, with the increase of T_H due to voltage drop in poly depletion region that results in a smaller energy barrier for hot holes. The sharp reduction when the poly gets inverted is due to the onset of minority ionization.

Moreover, L-AHI model also provide a voltage acceleration factor in agreement with experimental data (see discussion on p. 46), that further validates the model.

Recently, the AHI model, although very successful in explaining experimental data, has been criticized based on the observation that hole defect generation rate is much smaller than that required by the model. Since the hole current is much smaller than the initial electron current (Fig. 33), hole defect generation rate should be much larger than electron defect generation rate to account for the amount of damage needed to break down the oxide. Direct measurement of this hole defect generation rate carried out on p-MOSFET in inversion [69] gives comparable values for electron and holes. However, it must be pointed out that in this experiment p-MOSFETs were stressed with "cold" thermal holes, while anode holes are "hot" carrier. It is reasonable to assume that only energetic holes above a critical energy threshold participate in the damage creation process [111]. Therefore, much fewer holes than those injected from the inverted channel of p-MOSFET effectively damage the oxide, thus with a larger generation rate, resolving this apparent inconsistency.

5.2 Anode Hydrogen Release Model

The anode hydrogen release (AHR) model is schematically depicted in Fig. 34. Electrons injected into the anode with enough energy interact with hydrogen ions present at the silicon/oxide interface releasing some of them. Hydrogen ions (positively charged) diffuse under the effect of the oxide field through the oxide layer creating traps in the process.

It is well known that hydrogen is involved in the generation of defects [112]. For example, intentional exposure to hydrogen generates a number of defects in silicon dioxide films, even without field [113]. The degradation of MOSFET performance under channel hot electron stress is due to the creation of interface traps by hot electrons breaking silicon-hydrogen bonds [70]. It was also shown that hydrogen can create oxide bulk traps [114]. This latter process has a threshold energy of approximatively $5eV$ [114]. However, it has been recently reported that it continues down to voltages as low as 1.2V [115].

There are two main arguments against the AHR model. The first is the absence of the isotope effect on breakdown. It is known that MOS devices annealed in deuterium have an improved immunity to CHE stress because

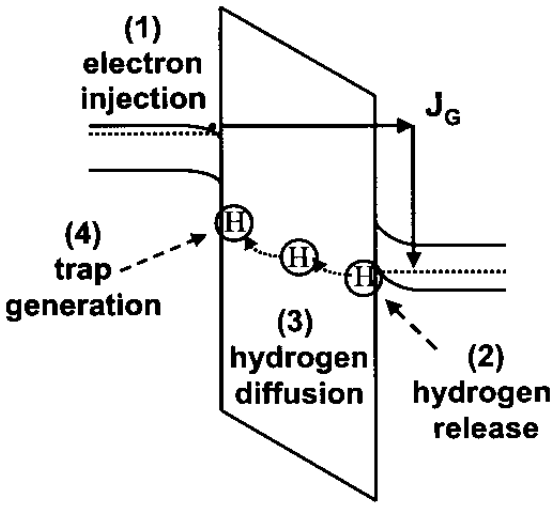

Fig. 34. Schematic illustration of the Anode Hydrogen Release model.

of the larger energy needed to desorb deuterium [116]. However, the same isotope effect has not been found on T_{BD}/Q_{BD} [117], although opposite results have been reported on this issue [118]. The second objection is the effect of the substrate bias. Since the hydrogen is present only at the silicon/oxide interface, the AHR cannot account for the substrate bias dependence of SILC and breakdown.

Modeling hydrogen desorption, diffusion, and defect generation is a very complex task to tackle. Only recently, an attempt to give a quantitative description of the AHR model based on the physics of the microscopic processes involving hydrogen has been proposed [119]. But, when this model was first quantitatively used to make reliability projection [67] the defect generation rate (P_g) was extracted from experimental data on the basis of a simplified version of (9):

$$Q_{BD} = q\ N_{BD}/P_g \tag{49}$$

where, beside the temperature factor that has been dropped only for clarity, a linear function has been assumed for the time evolution of the trap density ($m = 1$). On the basis of this equation, P_g is simply evaluated as the damage (assessed in any way) divided the electron fluence needed to generate it. P_g computed in this way from a number of different types of experiment is shown in Fig. 35. It is reported that P_g is almost insensitive to t_{ox}, oxidation type and voltage polarity, showing an universal behavior as a function of V_G [93]. P_g decreases exponentially for $V_G < 6V$ with a steeper decrease for $V_G < 3V$ (not shown in Fig. 35, but see [93]). However, the maximum slope is approximatively $5dec/V$.

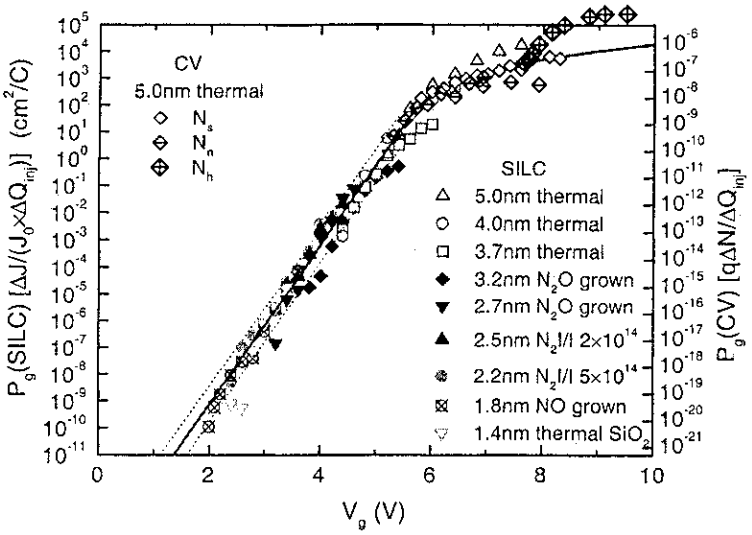

Fig. 35. Defect generation rate measured from SILC ($P_g(SILC)$) and C-V stretch-out ($P_g(CV)$ and N_s), and electron and hole trapping rates (N_n and N_h, respectively) as a function of stress voltage for different t_{ox}. Adapted from [93].

5.3 Thermochemical Model

The thermochemical model attributes the generation of oxide trap to the interaction of the electric field with weak bonds [120]. These weak bonds are believed to be oxygen vacancies that generate polarized Si-Si bonds (Fig. 36.b). When the external oxide field (E) is applied, these bonds acquire a dipole energy that lowers the activation energy needed to break the Si-Si bond, leading to the formation of a trap (Fig. 36.c). Thus, if ΔH_0 is the energy needed to break a bond in absence the field, then, in presence of field the activation energy becomes $E_a = \Delta H_0/KT - \gamma(T)\ E$ resulting in the following expression for oxide lifetime

$$T_{BD} = \tau_0\ e^{-\gamma(T)\ E}\ e^{\Delta H_0/KT}\ . \tag{50}$$

The thermochemical model became very popular because it provided a theoretical foundation to the linear dependence of T_{BD} on field previously experimentally found at low voltages. That's why it is also referred to as the E model.

However, it must be pointed out that fitting of experimental data is not a proof of the validity of a particular model. As a matter of fact, the thermochemical model indicates that oxide breakdown is a field driven phenomenon, where current flowing through the oxide play a minor role. This is in contrast with experimental data reporting different T_{BD} for the same oxide field depending on t_{ox} (Fig. 31), voltage polarity (Fig. 32), and substrate bias [100,121].

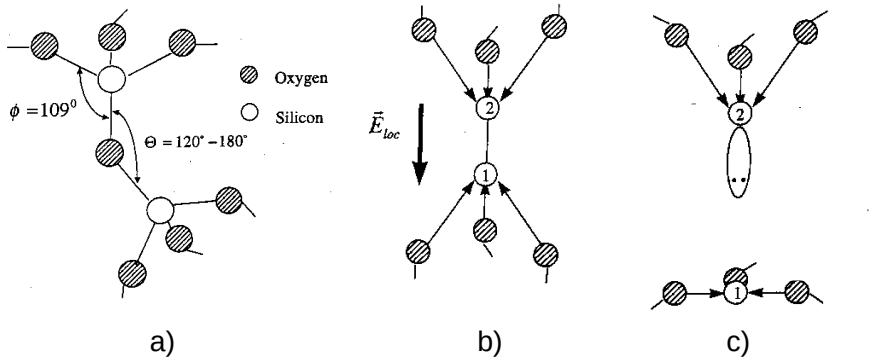

Fig. 36. Schematic illustration of SiO_2 bonds. a) normal structure; b) oxygen vacancy with polarized dipole; c) broken bond (i.e. trap). Adapted from [120].

5.4 Phenomenological Models

Beside the main physical models discussed so far, a lot of other models have been proposed in the literature. Many of them are based on empirical observations of the T_{BD} voltage dependence on a limited number of experimental conditions. The two most popular, E-model and $1/E$-model, have already been presented. In the framework of this Chapter, they are not used as synonymous of the thermochemical and AHI model respectively, but indicate a broader class of models postulating that kind a dependence.

In order to explain the entire voltage/field dependence of T_{BD}, a few attempts to unify the E and $1/E$ models by treating them as competing mechanisms have been proposed, on the basis also of theoretical considerations [122–124].

Following the experimental findings that breakdown of ultra thin oxides is voltage driven, a few models relating directly the applied voltage to T_{BD} have been presented [5,125]. One of the most interesting is the model proposed in [125], that is based on a power law:

$$T_{BD} \sim V^{-n(T)} \ . \tag{51}$$

This relation has been suggested by experimental data like those in Fig. 37. The same set of T_{BD} data relative to different t_{ox} features an increasing slope for decreasing t_{ox} in a linear-log graph (Fig. 37.left), while the slope is constant if plotted in a log-log graph (Fig. 37.right), suggesting that a power law might be a better fit. This has also been directly confirmed by long stress spanning a larger voltage window. The exponent of the power law depends only on the temperature. At room temperature the best fit was attained with 44 [125]. The voltage acceleration factor is thus $AF = n/V$, that increases as V_G^{-1} for decreasing voltage.

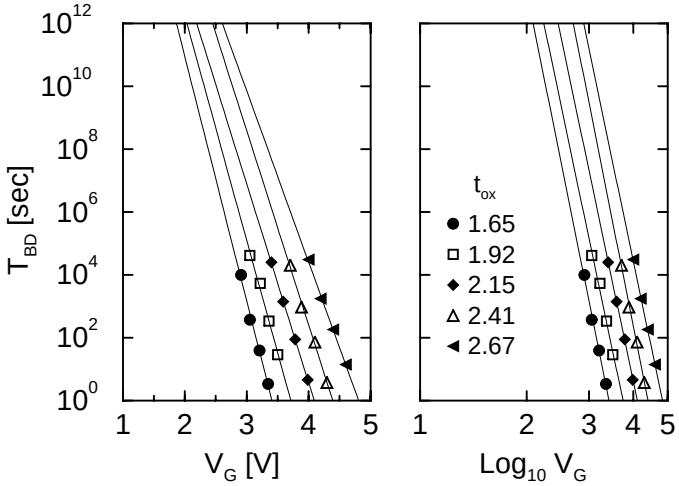

Fig. 37. T_{BD} as a function of the gate voltage. Left: linear-log scale. Right: log-log scale. Adapted from [125].

5.5 Reliability projection

Figure 38 shows the reliability projections provided by different models. Using the P_g shown in Fig. 35, pessimistic reliability projections were made in [67] forecasting that the minimum t_{ox} respecting the ITRS specifications would had been approximatively $2.2nm$. On the contrary, the L-AHI and power law models provided larger reliability margins. Later on, these margins have been also validated experimentally [6]. The power law model projection better fits experimental data simply because it is an extrapolation to thinner oxide of real data. In contrast, the L-AHI projections in Fig. 38 represents the theoretical limit in case all the nonidealities associated with thickness uniformity and determination of voltage acceleration factor could be eliminated [91]. In practice, this limit will be difficult to achieve. However, the large spread in the experimental data indicates that there is still room for improvement. Data of Fig. 38 also indicate that silicon dioxide will meet ITRS reliability roadmap yet for some time providing more time to find a replacement [126].

There are many factors that could have contributed to underestimate oxide lifetime in [67]. From (9) it is clear that an overestimation of m ($m = 1$ instead of a value smaller than one) provides a smaller T_{BD}. In addition, the percolation algorithm used to compute the Weibull slope β to be used in the projection provided a smaller value than what experimentally found later on (see Fig. 25 and discussion on p. 31) determining more pessimistic projection.

However, the most important factor was the low acceleration factor provided by the AHR model. According to (49), the voltage dependence is entirely due to P_g. Thus, the acceleration factor is simply the slope (in log scale) of P_g. Figure 39 compares experimental values of the voltage acceleration fac-

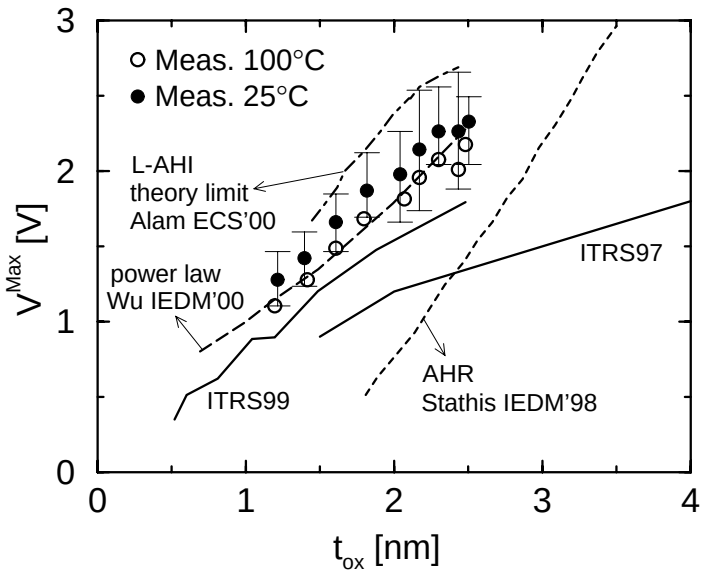

Fig. 38. Reliability projections provided by L-AHI in the ideal case, power-law and AHR models in comparison with experimental data. The ITRS roadmap is also shown for reference.

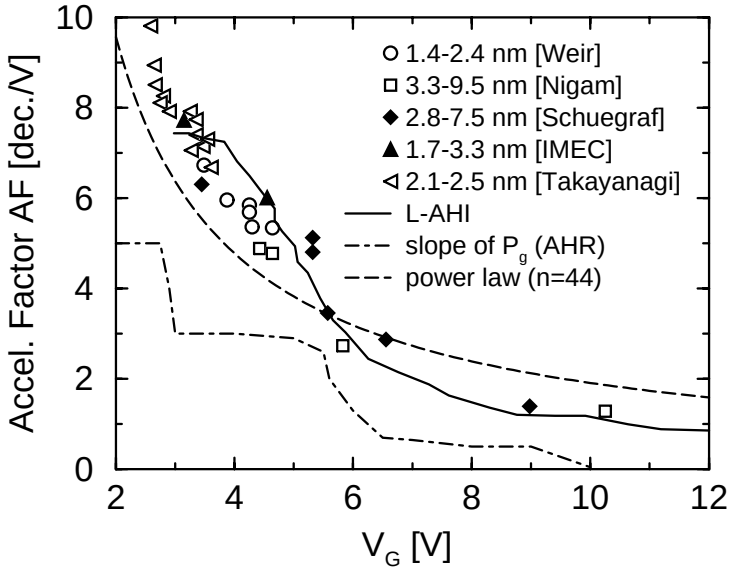

Fig. 39. Comparison of measured (symbols) and simulated (lines) acceleration factor.

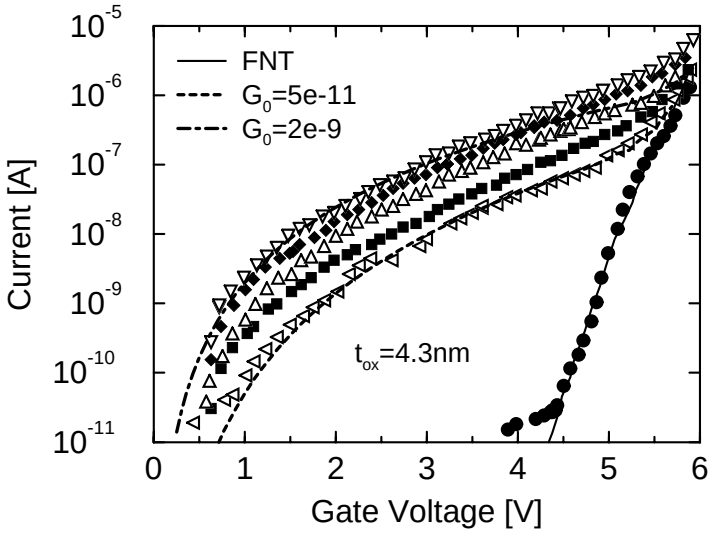

Fig. 40. Fresh and soft breakdown $I-V$ characteristics. Adapted from [15].

tor with predictions of different models. The slope of P_g underestimates AF, especially in the low voltage regime that is of most interest today, even with the latest correction [93]. On the other hands, the power law and L-AHI model better agree with experiments. In particular, they indicate that the voltage acceleration factor increases for decreasing voltages. It is precisely this increasing immunity to defect generation that allows silicon dioxide to meet ITRS requirements from the point of view of intrinsic reliability.

6 Breakdown mode

Silicon dioxide can feature many breakdown modes, characterized by very different conduction properties. Initially, when only thick oxides were available, only hard breakdown was observed. Hard breakdown is characterized by a large current with roughly linear (i.e. ohmic) $I-V$ characteristics with resistance $\sim 10k\Omega$. Successively, while investigating breakdown of thin oxides, another breakdown mode, called soft breakdown, was found [10,12]. As shown in Fig. 5, stress current (voltage) becomes much noisy [13,14], featuring also several small jumps before the final runaway. The characteristic feature of soft breakdown is the intermediate (between SILC and hard breakdown) $I-V$ curve, that is best described by a power law $I = G_0 V^{\delta}$. The parameters G_0 and δ feature a statistical distribution, i.e. similar devices stressed under the same conditions exhibit different soft breakdown $I-V$ characteristics as shown in Fig. 40. It was also found that they are statistically correlated [15]. For a given δ, G_0 varies approximatively of 1 order of magnitude, while δ is in the range 3-6 [15,19].

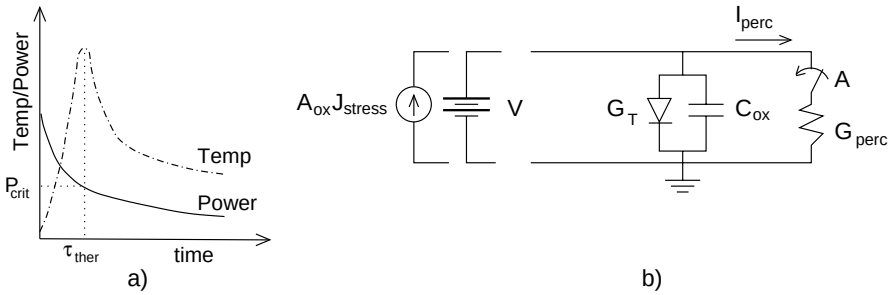

Fig. 41. a) Schematic illustration of the power and temperature evolution during breakdown. b) Oxide equivalent circuit for the calculation of the dissipated power. Adapted from [20].

Soft breakdown was seen only for $t_{ox} < 5nm$. It is the most common failure mode for CCS stress, while hard breakdown is more likely to occur in CVS. Extensive characterization of soft breakdown [127] showed that soft breakdown conduction is essentially independent of gate area, oxide thickness, injection polarity, sense polarity, and substrate type. This indicates that soft breakdown is a phenomenon localized in a small spot, whose size is 10^{-14} to $10^{-12}cm^2$, and that can be viewed with optical emission microscopy [10,128]. On the contrary, a strong dependence of soft breakdown $I-V$ characteristics on stress conditions has been seen.

Under the same stress conditions devices can feature directly hard breakdown or several soft breakdown events before the final hard breakdown. This generated some controversy about the relation between soft and hard breakdown, whether soft breakdown is a precursor or not of hard breakdown, if they result from different physical phenomena as the different voltage and temperature acceleration seem to indicate [128,129], or they share a common origin. However, accurate statistical analysis showed that soft and hard breakdown have similar statistical properties and occur, in general, in different locations [130], and that they feature consistent acceleration parameters [131], pointing out that they are actually two different manifestations of the same failure mechanism. Only the severity of breakdown is different.

In order to explain (and predict) the occurrence of the different breakdown modes, a few models have been proposed [20,132–134]. The basic idea of all these models is that if during the discharge transient the dissipated power on the breakdown spot (P_{perc}) exceeds some critical value (P_{crit}) at least for the duration of the thermal response of the system (τ_{ther}), then the local temperature will become high enough to melt the silicon near the breakdown spot and allow it to flow through the oxide creating a short-circuit, i.e. an hard breakdown [20]. On the contrary, if the dissipated power is below P_{crit} then breakdown will be soft.

The dissipated power can be computed using the simple oxide equivalent circuit shown in Fig. 41, where G_T is the conductance due to leakage

$J_{tun} = \alpha(t_{ox})exp(-\beta(t_{ox}/V)$, and G_{perc} is the conductance of the breakdown spot ($I_{perc}(t) = G_0V(t)^{\delta}$) [20]. After the breakdown spot has been created (represented by the closure of switch A in Fig. 41), the system is described by the following non linear differential equation for the case of constant current stress [132]:

$$A_{ox}J_{stress} = A_{ox}C_{ox}dV/dt + A_{ox}\alpha(t_{ox})e^{-\beta(V)/t_{ox}} + G_0V^{\delta} \quad . \tag{52}$$

Once solved, the dissipated power is simply $P_{perc}(t) = V(t > 0)I_{perc}(t > 0)$. For the case of constant voltage stress $P_{perc} = G_0V^{\delta+1}$. In any case the dissipated power depends on G_0 and δ that have a statistical distribution and so does P_{perc}. Thus the simulation of G_0 is important to determine the severity of breakdown. In addition, it is also useful to simulate post soft-breakdown $I - V$ characteristics in view of a possible use of soft broken devices if the leakage current is not too high [13].

Many models about soft breakdown conduction have been presented: local thinning of the damaged region [10], variable range hopping [135], quantum point contact [136], and percolation of non linear resistors [19,132]. Percolation was proved to provide the correct power law $I - V$ characteristics, and, from the practical point of view, it allows to compute the statistical distribution of G_0 as shown in [20] and here briefly summarized.

Percolation simulation is performed according to the cubic-lattice algorithm introduced at Par. 4.2. Each trap is represented by 6^3 smaller cubes. This allows overlapping of traps that is important for thin oxides. The different degree of overlapping and the topology of the percolation path determine the percolation conductance. This conductance is computed replacing each side of the smaller cubes with a non linear resistor $i = gv^{\delta}$, where g is a parameter that must be determined empirically by comparing experiments and simulation. Once g is determined, its value it is kept constant for all other calculation. In this way it is possible to quantitatively simulate soft-breakdown conductance as shown in Fig. 42.

Therefore, this model is able to explain why soft breakdown is more likely under CCS than CVS (the reduction of of applied voltage during the transient reduces the dissipated power), the existence of bimodal failure distribution (G_0 may or may not be enough to dissipate P_{crit} because of its statistical distribution), the effect of current limitation (less maximum current less dissipated power), the dependence on stress voltage, oxide area, oxide thickness, and back gate bias [132,138]. On the basis of this theory, it is believed that breakdown at very low voltage ($\sim 1V$) will always be soft allowing continued operation of many circuits after breakdown.

7 Summary and Conclusions

In this Chapter, a review of the physical models about silicon dioxide wear-out and breakdown has been presented with the purpose to address their

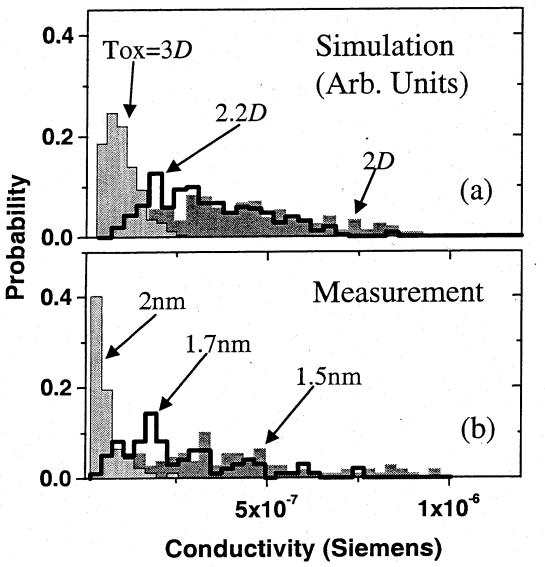

Fig. 42. Probability distribution of percolation conductance for oxides with different thicknesses. a) percolation results; b) experimental data. From [137].

numerical implementations and the reliability projections they provide. Oxide breakdown is a very complex phenomenon, and that is why many models have been proposed. However, some aspects of these models are not in accord with certain experimental data, and so more experimental and theoretical work is still required for a complete understanding.

Here, only the main models that attained the largest widespread acceptance have been illustrated. They all share a common background. Oxide wear-out is due to energetic carriers, thus it is an energy driven process. In thin oxides/low voltages it is the applied voltage to control carrier energy, while for thick oxide/high voltages it is the oxide field to determine carrier energy. We showed that carrier energy distribution can be satisfactory simulated with different techniques (Sec. 3).

Statistical properties can be well reproduced by the percolation algorithm (Sec. 4), that also provides the correct post breakdown conductance (Sec. 6).

The physical mechanism responsible for the generation of defect is still controversial. This is most investigated phenomenon among those involved in oxide breakdown because it determines the voltage dependence that is of fundamental importance to project reliability to the real operating conditions. Projection made with the most accurate models available today suggest that silicon dioxide still meet ITRS requirements down to $1nm$ at 1V (Sec. 5) if the leakage current can be tolerated.

Nevertheless there are still many things to be understood. Temperature dependence has to be investigated in more details. This also requires a mi-

croscopic model for the oxide defect formation, which is still missing. As a matter of fact, it is generally accepted that temperature dependence of oxide breakdown is due to the defect generation process, rather than to the temperature dependence of the driving force.

Another important issue to be investigated more in depth is how the functionality of a complex circuit is affected by the failure of single devices. It was proposed in [13] that a soft-broken device is still acceptable for logic operation if the leakage or noise are not too high. As a matter of fact, it was later shown that simple circuit such as a ring oscillator continued to function even after several devices experienced breakdown [139]. Frequency, standby and dynamic current were changed, however, the logic operation was still correct. From this point of view, it is important to investigate post soft breakdown conduction and how device characteristics are affected. In addition, even if a circuit survives the first breakdown, its functionality depends on the stability of the leakage path. Thus, it is also interesting to study subsequent soft breakdown events [140,141].

Finally, the impact of limiting the maximum current of the stress source must also be addressed in more details. In real circuit, MOS devices are driven by other MOS devices. Thus, the maximum current is limited by the saturation current of a complementary transistor in series. Recent studies showed that the post breakdown leakage current has a strong dependence of the current available during the breakdown runaway, suggesting that breakdown would be much softer under actual circuit operating conditions [142].

References

1. W. Weibull, "A Statistical Distribution Function of Wide Application," *J. Appl. Mech.*, vol. 18, pp. 293–297, 1951.
2. R. Degraeve, J. Ogier, R. Bellens, P. Roussel, G. Groeseneken, and H. Maes, "A New Model for the Field Dependence of Intrinsic and Extrinsic Time-Dependent Dielectric Breakdown," *IEEE Trans. on Electron Devices*, vol. 45, no. 2, pp. 472–481, 1998.
3. "International technology roadmap for semiconductors. 2001 edition." `http://public.itrs.net`, 2001.
4. D. DiMaria and J. Stathis, "Non-Arrhenius temperature dependence of reliability in utlrathin silicon dioxide films," *Applied Physics Letters*, vol. 74, no. 12, p. 1752, 1999.
5. R. Degraeve, N. Pangon, B. Kaczer, T. Nigam, G. Groeseneken, and A. Naem, "Temperature acceleration of oxide breakdown and its impact on ultra-thin gate oxide reliability," in *Proc. VLSI Technology Symposium*, pp. 59–60, 1999.
6. B. Weir, M. Alam, J. Bude, P. Silverman, A. Ghetti, F. Baumann, P. Diodato, D. Monroe, Y. Ma, M. Brown, A. Hamad, D. Hwang, and P. Manson, "Gate Oxide Reliability Projection to the Sub 2nm Regime," *Semicond. Sci. Technol.*, vol. 15, no. 5, pp. 455–461, 2000.
7. K. Okada, "Extended Time Dependent Dielectric Breakdown Model Based on Anomalous Gate Area Dependence of Lifetime in Ultra Thin Silicon Dioxide," *Japanese Journal of Applied Physics*, vol. 36, no. 1-3B, pp. 1443–1447, 1997.

8. J. Wu, L. Register, and E. Rosembaum, "Trap-Assisted Tunneling Current Throgh Ultra-Thin Oxide," in *Proc. IRPS Symposium*, pp. 389–395, 1999.
9. S. Takagi, N. Yasuda, and A. Toriumi, "A New $I-V$ Model for Stress-Induced Leakage Current Including Inelastic Tunneling," *IEEE Trans. on Electron Devices*, vol. 46, no. 2, pp. 348–354, 1999.
10. S.-H. Lee, B.-J. Cho, J.-C. Kim, and S.-H. Choi, "Quasi-Breakdown of ultra-thin gate oxide under high field stress," in *IEDM Technical Digest*, pp. 605–608, 1994.
11. K. Okada, "The gate oxide lifetime limited by 'B-mode' stress induced leakage current and the scaling limit of silicon dioxide in the direct tunnelling regime," *Semicond. Sci. Technol.*, vol. 15, no. 5, pp. 478–484, 2000.
12. M. Depas, T. Nigam, and M. Heyns, "Soft Breakdown of Ultra-Thin Gate Oxide Layers," *IEEE Trans. on Electron Devices*, vol. 43, no. 9, pp. 1499–1504, 1996.
13. B. Weir, P. J. Silverman, D. Monroe, K. Krisch, M. Alam, G. Alers, T. Sorsch, G. Timp, F. Baumann, C. Liu, Y. Ma, and D. Huang, "Ultrathin gate dielectrics: They break down, but do they fail?," in *IEDM Technical Digest*, pp. 73–76, 1997.
14. A. Cester, L. Bandiera, G. Ghidini, I. Bloom, and A. Paccagnella, "Soft breakdown current noise in ultra-thin gate oxides," *Solid-State Electronics*, vol. 46, no. 7, pp. 1019–1025, 2002.
15. E. Miranda, J. Suñé, R. Rodriguez, M. Nafria, and X. Aymerich, "A Functional-Fit Model for the Soft Breakdown Failure Mode," *IEEE Electron Device Letters*, vol. 20, no. 6, pp. 265–267, 1999.
16. P. Olivo, T. Nguyen, and B. Riccò, "High-field-induced degradation in ultra-thin sio$_2$ films," *IEEE Trans. on Electron Devices*, vol. 35, no. 12, p. 2259, 1988.
17. J. D. Blauwe, J. V. Houdt, D. Wellekens, G. Groeseneken, and H. Maes, "SILC-Related Effects in Flash E^2PROM's-Part I: A Quantitative Model for Steady-State SILC," *IEEE Trans. on Electron Devices*, vol. 45, no. 8, pp. 1745–1750, 1998.
18. A. Ghetti, "Characterization and Modeling of the Tunneling Current in Si/SiO_2 Structures with Ultra-Thin Oxide Layer," *Microelectronic Engineering*, vol. 59, no. 1-4, pp. 127–136, 2001.
19. M. Houssa, T. Nigam, P. Mertens, and M. Heyns, "Model for the current-voltage characteristics of ultra-thin gate oxides after sof-breakdown," *Journal of Applied Physics*, vol. 84, no. 8, pp. 4351–4355, 1998.
20. M. Alam, B. Weir, and P. Silverman, "A Study of Soft and Hard Breakdown - Part I: Analysis of Statistical Percolation Conductance," *IEEE Trans. on Electron Devices*, vol. 49, no. 2, pp. 232–238, 2002.
21. M. Alam, B. Weir, J. Bude, P. Silverman, and A. Ghetti, "A Computational Model for Oxide Breakdown: Theory and Experiments," *Microelectronic Engineering*, vol. 59, no. 1-4, pp. 137–147, 2001.
22. P. Nicollian, M. Rodder, D. Grider, P. Chen, R. Wallace, and S. Hattangady, "Low Voltage Stress-Induced-Leakage-Current in Ultrathin Gate Oxides," in *Proc. IRPS Symposium*, pp. 400–404, 1999.
23. A. Ghetti, E. Sangiorgi, J. Bude, T. Sorsch, and G. Weber, "Tunneling into Interface States as Reliability Monitor for Ultra-Thin Oxides," *IEEE Trans. on Electron Devices*, vol. 47, no. 12, pp. 2358–2365, 2000.

24. S.-L. Lo, D. Buchanan, Y. Taur, and W. Wang, "Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's," *IEEE Electron Device Letters*, vol. 18, pp. 209–211, 1997.
25. C. Bowen, C. Fernando, G. Klimeck, A. Chatterjee, D. Blanks, R. Lake, J. Hu, J. Davis, M. Kulkarni, S. Hattangady, and I. Chen, "Physical Oxide Thickness Extraction and Verification using Quantum Mechanical Simulation," in *IEDM Technical Digest*, p. 869, 1997.
26. J. Suñé, P. Olivo, and B. Riccò, "Self-Consistent Solution of Poisson and Schrödinger Equations in Accumulated Semiconductor-Insulator Interfaces," *Journal of Applied Physics*, vol. 70, pp. 337–345, 1991.
27. F. Rana, S. Tiwari, and D. Buchanan, "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides," *Applied Physics Letters*, vol. 69, no. 8, p. 1104, 1996.
28. N. Yang, W. Henson, J. Hauser, and J. Wortman, "Modeling Study of Ultrathin Gate Oxides Using Direct Tunneling Current and Capacitance-Voltage Measurements in MOS Devices," *IEEE Trans. on Electron Devices*, vol. 46, no. 7, p. 1464, 1999.
29. L. Register, E. Rosembaum, and K. Yang, "Analytic model for direct tunneling current in polycrystalline silicon-gate metal-oxide-semiconductor devices," *Applied Physics Letters*, vol. 74, no. 3, pp. 457–459, 1999.
30. K.-N. Yang, H.-T. Huang, M.-C. Chang, C.-M. Chu, Y.-S. Chen, M.-J. Chen, Y.-M. Lin, M.-C. Yu, S. Jang, D. Yu, and M. Liang, "A Physical Model for Hole Direct Tunneling Current in P^+ Poly-Gate PMOSFETs with Ultrathin gate Oxides," *IEEE Trans. on Electron Devices*, vol. 47, no. 11, pp. 2161–2166, 2000.
31. A. Ghetti, J. Bude, P. Silverman, A. Hamad, and H. Vaidya, "Modeling and Simulation of Tunneling Current in MOS Devices including Quantum Mechanical Effects," *IEICE Trans. on Electronics*, vol. E83-C, no. 8, pp. 1175–1182, 2000.
32. M. Alam, J. Bude, B. Weir, P. Silverman, A. Ghetti, D. Monroe, K. Cheung, and S. Moccio, "An Anode Hole Injection Percolation Model for Oxide Breakdown - The "Doom's Day" Scenario Revisited," in *IEDM Technical Digest*, pp. 715–719, 1999.
33. M. 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?," in *Proc. IRPS Symposium*, pp. 21–26, 2000.
34. C. Moglestue, "Self-consistent calculation of electron and hole inversion charges at silicon-silicon dioxide interfaces," *Journal of Applied Physics*, vol. 59, no. 9, p. 3175, 1986.
35. H. Massoud and J. Shiely, "The Role of Substrate Carrier Generation in Determining the Electric Field in the Oxide of MOS Capacitors Biased in the Fowler-Nordheim Tunneling Regime," *Microelectronic Engineering*, vol. 36, pp. 263–266, 1997.
36. A. Pacelli, "Self-consistent Solution of the Schrödinger equation in Semiconductor Devices by Implicit Iteration," *IEEE Trans. on Electron Devices*, vol. 44, no. 7, pp. 1169–1171, 1997.
37. M. Goano, "Series Expansion of the Fermi-Dirac Integral $\mathcal{F}_j(x)$ over the Entire Domain of Real j and x," *Solid-State Electronics*, vol. 36, no. 2, p. 217, 1993.

38. N. Lifshitz, "Dependence of the Work-Function Difference between the Polysilicon Gate and Silicon Substrate on the Doping Level in Polysilicon," *IEEE Trans. on Electron Devices*, vol. 32, no. 3, pp. 617–621, 1985.
39. K. Gundlach, "Zur berechnung des tunnelstroms durch eine trapezformige potentialstufe," *Solid-State Electronics*, vol. 9, pp. 949–957, 1966.
40. W. Lui and M. Kukuma, "Exact solution of the Schrödinger equation across an arbitrary one-dimensional piecewise-linear potential barrier," *Journal of Applied Physics*, vol. 60, no. 5, p. 1555, 1986.
41. B. Nag, "Boundary condition for tunneling through potential barriers in non-parabolic semiconductors," *Applied Physics Letters*, vol. 59, pp. 1620–1622, 1991.
42. A. Ghetti, E. Sangiorgi, T. Sorsch, and I. Kizilyalli, "The Role of Native Traps on the Tunneling Characteristics of Ultra-Thin (1.5-3 nm) Oxides," *Microelectronic Engineering*, vol. 48, pp. 31–34, 1999.
43. M. Fischetti, D. DiMaria, S. Bronson, T. Theis, and J. Kirtley, "Theory of high-field electron transport in silicon dioxide," *Phys. Rev. B*, vol. 31, no. 12, pp. 8124–8142, 1985.
44. D. Arnold, E. Cartier, and D. DiMaria, "Theory of high-field electron transport and impact ionization in silicon dioxide," *Phys. Rev. B*, vol. 49, no. 15, pp. 10278–10297, 1994.
45. A. Ghetti, "MOSFET Hot-Carrier Induced Gate Current Simulation by Self-Consistent Silicon/Oxide Monte Carlo Device Simulation," in *Proc. SISPAD Conference*, pp. 231–234, 2002.
46. M. Lenzlinger and E. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2," *Journal of Applied Physics*, vol. 40, pp. 278–283, 1969.
47. A. Ghetti, C. Liu, M. Mastrapasqua, and E. Sangiorgi, "Characterization of Tunneling Current in Ultra-Thin Gate Oxide," *Solid-State Electronics*, vol. 44, no. 9, pp. 1523–1531, 2000.
48. I. Chen, S. Holland, K. Young, C. Chang, and C. Hu, "Substrate hole current and oxide breakdown," *Applied Physics Letters*, vol. 49, no. 11, p. 669, 1986.
49. C. Chang, C. Hu, and R. Brodersen, "Quantum yield of electron impact ionization in silicon," *Journal of Applied Physics*, vol. 57, pp. 302–309, 1985.
50. S. Takagi, N. Yasuda, and A. Toriumi, "Experimental Evidence of Inelastic Tunneling in Stress-Induced Leakage Current," *IEEE Trans. on Electron Devices*, vol. 46, no. 2, pp. 335–341, 1999.
51. Y. Kamakura, I. Kawashima, K. Deguchi, and K. Taniguchi, "Verification of hot hole scattering rates in silicon by quantum-yield experiment," *Journal of Applied Physics*, vol. 88, no. 10, pp. 5802–5809, 2000.
52. A. Ghetti, M. Alam, J. Bude, and F. Venturi, "Assessment of Quantum Yield Experiments Via Full Band Monte Carlo Simulations," in *IEDM Technical Digest*, pp. 873–876, 1997.
53. A. Ghetti, M. Alam, J. Bude, D. Monroe, E. Sangiorgi, and H. Vaidya, "Stress Induced Leakage Current Analysis via Quantum Yield Experiments," *IEEE Trans. on Electron Devices*, vol. 47, no. 7, pp. 1341–1348, 2000.
54. K. Schuegraf and C. Hu, "Hole Injection SiO_2 Breakdown Model for Very Low Voltage Lifetime Extrapolation," *IEEE Trans. on Electron Devices*, vol. 41, no. 5, pp. 761–767, 1994.
55. K. Kobayashi, A. Teramoto, and M. Hirayama, "Model for the substrate hole current based on thermionic hole emission from the anode during Fowler-

Nordheim electron tunneling in n-channel metal-oxide-semiconductor field-effect transistors," *Journal of Applied Physics*, vol. 77, no. 7, pp. 3277–3282, 1995.
56. P. Palestri, M. Pavesi, P. Rigoli, L. Selmi, A. D. Serra, A. Abramo, F. Widdershoven, and E. Sangiorgi, "Impact ionization and photon emission in MOS capacitors and FETs," in *IEDM Technical Digest*, pp. 97–100, 2000.
57. A. Halimaoui, O. Briere, and G. Ghibaudo, "Quasi-Breakdown in Ultrathin Gate Dielectrics," *Microelectronic Engineering*, vol. 36, pp. 157–160, 1997.
58. D. DiMaria and E. Cartier, "Mechanism of stress-induced leakage current in thin silicon dioxide films," *Journal of Applied Physics*, vol. 78, no. 6, p. 3883, 1995.
59. E. Suzuki, D. Schroder, and Y. Hayashi, "Carrier conduction in ultrathin nitrided oxide films," *Journal of Applied Physics*, vol. 60, no. 10, pp. 3616–3621, 1986.
60. A. Ghetti, M. Alam, J. Bude, E. Sangiorgi, G. Timp, and G. Weber, "Native and Stress-Induced Traps in SiO_2 Films," in *Proc. of the 4th Symposium on the Physics and Chemistry of SiO_2 and the Si-SiO_2 Interface*, pp. 419–428, Toronto, Canada; 5/14-19/2000.
61. T.-K. Kang, M.-J. Chen, C.-H. Liu, Y. Chang, and S.-K. Fan, "Numerical Confirmation of Inelastic Trap-Assisted Tunneling (TAT) as SILC Mechanism," *IEEE Trans. on Electron Devices*, vol. 48, no. 10, pp. 2317–2322, 2001.
62. J. Deblauwe, "Stress induced leakage current in thin dielectrics." Ph.D. Thesis - IMEC, 1998.
63. D. Dumin and J. Maddux, "Correlation of Stress-Induced Leakage Current in Thin Oxides with Trap Generation Inside the Oxides," *IEEE Trans. on Electron Devices*, vol. 40, no. 5, p. 986, 1993.
64. D. DiMaria, D. Arnold, and E. Cartier, "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon," *Journal of Applied Physics*, vol. 73, no. 7, pp. 3367–3384, 1993.
65. M. Alam, "SILC as a Measure of Trap Generation and Predictor of T_{BD} in Ultrathin Oxides," *IEEE Trans. on Electron Devices*, vol. 49, no. 2, pp. 226–231, 2002.
66. D. Buchanan, J. Stathis, E. Cartier, and D. DiMaria, "On the Relationship between Stress Induced Leakage Currents and Catastrophic Breakdown in Ultra-Thin SiO_2 Based Dielectrics," *Microelectronic Engineering*, vol. 36, pp. 329–332, 1997.
67. J. Stathis and D. DiMaria, "Realiability Projection for Ultra-Thin Oxides at Low Voltage," in *IEDM Technical Digest*, pp. 167–170, 1998.
68. J. Suñé, M. Nafria, E. Miranda, X. Oriols, R. Rodriguez, and X. Aymerich, "Failure physics of ultra-thin SiO_2 gate oxides near their scaling limit," *Semicond. Sci. Technol.*, vol. 15, no. 5, pp. 445–454, 2000.
69. D. DiMaria and J. Stathis, "Anode hole injection, defect generation, and breakdown in ultrathin silicon dioxide films," *Journal of Applied Physics*, vol. 89, no. 9, pp. 5015–5024, 2001.
70. C. Hu, S. Tam, F. Hsu, P. Ko, T. Chan, and K. Terrill, "Hot-electron-induced MOSFET degradation: Model, Monitor, and Improvement," *IEEE Trans. on Electron Devices*, vol. 32, pp. 375–385, 1985.
71. J. D. Bude, A. Frommer, M. R. Pinto, and G. R. Weber, "EEPROM/Flash Sub 3.0 V Drain-Source Bias Hot Carrier Writing," in *IEDM Technical Digest*, p. 989, 1995.

72. T. Ning and H. Yu, "Optically induced injection of hot electrons into silicon dioxide," *Journal of Applied Physics*, vol. 45, pp. 5373–5378, 1974.
73. L. Selmi, E. Sangiorgi, R. Bez, and B. Riccò, "Measurement of the Hot Hole Injection Probability from *Si* Into SiO_2 in p-MOSFET's," in *IEDM Technical Digest*, p. 333, Dec. 1993.
74. B. Fischer, A. Ghetti, L. Selmi, R. Bez, and E. Sangiorgi, "Bias and Temperature Dependence of Homogeneous Hot-Electron Injection from Silicon into Silicon Dioxide at Low Voltages," *IEEE Trans. on Electron Devices*, vol. 44, pp. 288–296, Feb. 1997.
75. T. Ning, C. Osburn, and H. Yu, "Emission probability of hot electrons from silicon into silicon dioxide," *Journal of Applied Physics*, vol. 48, pp. 286–293, 1977.
76. J. Tang and K. Hess, "Theory of hot electron emission from silicon into silicon dioxide," *Journal of Applied Physics*, vol. 54, pp. 5145–5151, 1983.
77. M. Fischetti, D. DiMaria, S. Bronson, and T. Theis, "Understanding hot-electron transport in silicon devices: is there a short cut?," *Journal of Applied Physics*, vol. 78, pp. 1058–1087, 1995.
78. A. Ghetti, L. Selmi, E. Sangiorgi, A. Abramo, and F. Venturi, "A Combined Transport-Injection Model for Hot-Electron and Hot-Hole Injection in the Gate Oxide of MOS Structures," in *IEDM Technical Digest*, pp. 363–366, 1994.
79. Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "Trap generation and occupation dynamics in SiO_2 under charge injection stress," *Journal of Applied Physics*, vol. 60, pp. 2024–2035, 1986.
80. R. Degraeve, G. Groeseneken, R. Bellens, J.-L. Ogier, M. Depas, P. Roussel, and H. Maes, "New Insights in the Relation Between Electron Trap Generation and the Statistical Properties of Oxide Breakdown," *IEEE Trans. on Electron Devices*, vol. 45, no. 4, pp. 904–911, 1998.
81. A. Ghetti, J. Bude, and G. Weber, "T_{BD} Prediction from Low Voltage Near-Interface Trap-Assisted Tunneling Current Measurements," *IEEE Trans. on Electron Devices*, vol. 48, no. 7, pp. 1354–1359, 2001.
82. S. Bruyére, D. Roy, E. Robilliart, E. Vincent, and G. Ghibaudo, "Body effect induced wear-out acceleration in ultra-thin oxides," *Microelectronics Reliability*, vol. 41, pp. 1031–1034, 2001.
83. E. Wu, J. Suñé, E. Nowak, W. Lai, and J. McKenna, "Weibull Slopes, Critical Defect Density, and The Validity of Stress-Induced-Leakage Current (SILC) Meassurements," in *IEDM Technical Digest*, pp. 125–129, 2001.
84. E. Wu, W. Abadeer, L.-H. Han, S.-H. Lo, and G. Hueckel, "Challeges for Accurate Reliability Projections in the Ultra-Thin Oxide Regime," in *Proc. IRPS Symposium*, pp. 57–65, 1999.
85. R. Degraeve, B. Kaczer, and G. Groeseneken, "Reliability: a possible show-stopper for oxide thickness scaling," *Semicond. Sci. Technol.*, vol. 15, no. 5, pp. 436–444, 2000.
86. J. Suñé, I. Placencia, N. Barniol, E. Farres, F. Martin, and X. Aymerich, "On the breakdown statistics of very thin SiO_2 films," *Thin Solid Films*, vol. 185, pp. 347–362, 1990.
87. R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. Maes, "A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides," in *IEDM Technical Digest*, pp. 863–866, 1995.

88. J. Stathis, "Quantitative model of the thickness dependence of breakdown in ultra-thin oxides," *Microelectronic Engineering*, vol. 36, pp. 325–328, 1997.
89. J. Stathis, "Percolation models for oxide breakdown," *Journal of Applied Physics*, vol. 86, no. 10, pp. 5757–5766, 1999.
90. T. Nigam, R. Degraeve, G. Groeseneken, M. Heyns, and H. Maes, "A Fast and Simple Methodology for Lifetime Prediction of Ultra-thin Oxides," in *Proc. IRPS Symposium*, pp. 381–388, 1999.
91. M. Alam, B. Weir, P. Silverman, J. Bude, G. Timp, and A. Ghetti, "Physics and Prospects of Sub 2nm Oxides," in *Proc. of the 4th Symposium on the Physics and Chemistry of SiO_2 and the Si-SiO_2 Interface*, pp. 365–376, Toronto, Canada; 5/14-19/2000.
92. J. Stathis, A. Vayshenker, P. Varekamp, E. Wu, C. Montrose, L. McKenna, D. DiMaria, L.-K. Han, E. Cartier, R. Wachnik, and B. Linder, "Breakdown Measurements of Ultra-Thin SiO_2 at Low Voltage," in *Proc. VLSI Technology Symposium*, pp. 94–95, 2000.
93. J. Stathis, "Physical and Predictive Models of Ultra Thin Oxide Reliability in CMOS Devices and Circuits," in *Proc. IRPS Symposium*, pp. 132–149, 2001.
94. E. Wu, E. Novak, L. Kan, D. Dunfresne, and W. Abadeer, "Nonlinear Characteristics of Weibull breakdown distributions and its impact on reliability projection for ultra-thin oxides," in *IEDM Technical Digest*, pp. 441–444, 1999.
95. B. Weir, P. Silverman, M. Alam, A. Hamad, T. Oberdick, N. Zhao, F. Baumann, G. Timp, A. Ghetti, Y. Ma, M. Brown, and T. Sorsch, "Gate Oxides in 50nm Devices: Thickness Uniformity Improves Projected Reliability," in *IEDM Technical Digest*, pp. 437–450, 1999.
96. E. Wu, J. Stathis, and L.-K. Han, "Ultra-thin oxide reliability for ULSI applications," *Semicond. Sci. Technol.*, vol. 15, no. 5, pp. 425–435, 2000.
97. I.-C. Chen, S. Holland, and C. Hu, "Electrical Breakdown in Thin Gate and Tunneling Oxides," *IEEE Trans. on Electron Devices*, vol. 33, no. 2, pp. 413–422, 1985.
98. I. Chen, S. Holland, and C. Hu, "Hole Trapping and Breakdown in Thin SiO_2," *IEEE Electron Device Letters*, vol. 7, no. 3, p. 164, 1986.
99. K. Schuegraf and C. Hu, "Reliability of thin SiO_2," *Semicond. Sci. Technol.*, vol. 9, pp. 989–1004, 1994.
100. J. Bude, B. Weir, and P. Silverman, "Explanation of Stress-Induced Damage in Thin Oxides," in *IEDM Technical Digest*, p. 179, 1998.
101. A. Ghetti, M. Alam, and J. Bude, "Anode hole generation mechanisms," *Microelectronics Reliability*, vol. 41, no. 9-10, pp. 1347–1354, 2001.
102. M. Fischetti, "Model for the generation of positive charge at the Si-SiO2 interface based on hot-hole injection from the anode," *Phys. Rev. B*, vol. 31, no. 4, pp. 2099–2113, 1985.
103. D. DiMaria, E. Cartier, and D. Buchanan, "Anode hole injection and trapping in silicon dioxide," *Journal of Applied Physics*, vol. 80, no. 1, pp. 304–317, 1996.
104. H. Satake and A. Toriumi, "Temperature dependent hole fluence to breakdown in thin gate oxides under Fowler-Nordheim electron tunneling injection," *Applied Physics Letters*, vol. 66, no. 25, pp. 3516–3517, 1995.
105. J. McPherson, V. Reddy, K. Banerjee, and H. Le, "Comparison of the E and 1/E T_{DDB} Models for SiO_2 under long-term/low-field test conditions," in *IEDM Technical Digest*, pp. 171–174, 1998.

106. A. Yassine, H. Narim, and K. Olasupo, "Field and Temperature Dependence of TDDB of Ultrathin Gate Oxide," *IEEE Electron Device Letters*, vol. 20, no. 8, pp. 390–392, 1999.
107. D. DiMaria, "Explanation for the polarity dependence of breakdown in ultrathin silicon dioxide films," *Applied Physics Letters*, vol. 68, no. 21, p. 3004, 1996.
108. P. Nicollian, W. Hunter, and J. Hu, "Experimental Evidence for Voltage Driven Breakdown Models in Ultrathin Gate Oxides," in *Proc. IRPS Symposium*, pp. 7–15, 2000.
109. A. Teramoto, H. Umeda, K. Azamawari, K. Kobayashi, K. Shinga, J. Komori, Y. Ohno, and H. Miyoshi, "Study of Oxide Breakdown under Very Low Electric Field," in *Proc. IRPS Symposium*, pp. 66–71, 1999.
110. T. Nigam, "Growth Kinetics, Electrical characterization and Reliability study of sub 5-nm Gate Dielectrics." Ph.D. Thesis, Dept. Electrotechniek, Katholieke Universiteit Lueuven, May 1999.
111. M. Alam. Private Communication.
112. E. Nicollian and J. Brews, *MOS Physics and Technology.* New York, NJ: Wiley, 1983.
113. R. Stahlbush and E. Cartier, "Interface defect formation in MOSFETs by atomic hydrogen exposure," *IEEE Trans. Nucl. Sci.*, vol. 41, pp. 1844–1853, 1994.
114. D. DiMaria and J. Stasiak, "Trap Creation in Silicon Dioxide Produced by Hot Electrons," *Journal of Applied Physics*, vol. 65, pp. 2342–2356, 1989.
115. D. DiMaria, "Electron energy dependence of metal-oxide-semiconductor degradation," *Applied Physics Letters*, vol. 75, no. 16, pp. 2427–2428, 1999.
116. K. Hess, I. Kizilyalli, and J. Lyding, "Giant Isotope Effect in Hot Electron Degradation of Metal Oxide Silicon Devices," *IEEE Trans. on Electron Devices*, vol. 45, no. 2, p. 406, 1998.
117. J. Wu, E. Rosenbaum, B. MacDonald, E. Li, B. Tracy, and P. Fang, "Anode Hole Injection versus Hydrogen Release: The Mechanism for Gate Oxide Breakdown," in *Proc. IRPS Symposium*, pp. 27–32, 2000.
118. Y. Mitani, S. H, H. Itoh, and A. Toriumi, "Highly reliable gate oxide under Fowler-Nordheim electron injection by deuterium pyrogenic oxidation and deuterated poly-Si deposition," in *IEDM Technical Digest*, pp. 343–346, 2001.
119. J. Suñé and E. Wu, "Modeling the breakdown and breakdown statistics of ultra-thin SiO_2 gate oxides," *Microelectronic Engineering*, vol. 59, no. 1-4, pp. 149–153, 2001.
120. J. McPherson and H. 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*, vol. 84, no. 3, pp. 1513–1523, 1998.
121. E. Vogel, J. Suehle, M. Edelstein, B. Wang, Y. Chen, and J. Bernstein, "Reliability of Ultrathin Silicon Dioxide Under Combined Substrate Hot-Electron and Constant Voltage Tunneling Stress," *IEEE Trans. on Electron Devices*, vol. 47, no. 6, pp. 1183–1191, 2000.
122. C. Hu and Q. Lu, "A Unified Gate Oxide Reliability Model," in *Proc. IRPS Symposium*, pp. 47–51, 1999.
123. K. Cheung, "Unifying the thermal-chemical and anode-hole-injection gate-oxide breakdown models," *Microelectronics Reliability*, vol. 41, pp. 193–199, 2001.

124. J. McPherson and R. Khamankar, "Molecular model for intrinsic time-dependent dielectric breakdown in SiO_2 dielectrics and the reliability implications for hypher-thin gate oxide," *Semicond. Sci. Technol.*, vol. 15, no. 5, pp. 462–470, 2000.
125. E. Wu, J. Aitken, E. Novak, A. Vayshenker, P. Varekamp, G. Hueckel, and J. McKenna, "Voltage-Dependent Voltage-Acceleration of Oxide Breakdown for Ultra-Thin Oxides," in *IEDM Technical Digest*, pp. 541–544, 2000.
126. J. Suehle, "Ultrathin Gate Oxide Reliability: Physical Models, Statistics, and Characterization," *IEEE Trans. on Electron Devices*, vol. 49, no. 6, pp. 958–971, 2002.
127. E. Miranda, J. Suñé, R. Rodriguez, M. Nafria, X. Aymerich, L. Fonseca, and F. Campabadal, "Soft Breakdown Conduction in Ultrathin (3-5 nm) gate Dielectrics," *IEEE Trans. on Electron Devices*, vol. 47, no. 1, pp. 82–89, 2000.
128. S. Bruyére, E. Vincent, and G. Ghibaudo, "Quasi-breakdown in Ultra-thin SiO_2 Films: Occurrence Characterization and Reliability Assessment Methodology," *Proc. IRPS Symposium*, pp. 48–54, 2000.
129. T. Pompl, H. Wurzer, M. Kerber, and I. Eisele, "Investigation of ultra-thin gate oxide reliability behavior by separate characterization of soft breakdown and hard breakdown," *Proc. IRPS Symposium*, pp. 40–47, 2000.
130. J. Suñé, G. Murra, and E. Miranda, "Are Soft Breadown and Hard Breakdown of Ultrathin Gate Oxides Actually Different Failure Mechanisms?," *IEEE Electron Device Letters*, vol. 21, no. 4, pp. 167–169, 2000.
131. J. Suehle, E. Vogel, B. Wang, and B. Bernstein, "Temperature Dependence of Soft Breakdown and Wear-Out in Sub-3nm SiO_2 Films," in *Proc. IRPS Symposium*, pp. 33–39, 2000.
132. M. Alam, B. Weir, J. Bude, P. Silverman, and D. Monroe, "Explanation of Soft and Hard Breakdown and its Consequences for Area Scaling," in *IEDM Technical Digest*, pp. 449–452, 1999.
133. H. Satake and A. Toriumi, "SiO_2 dielectric breakdown mechanism studied by the post-breakdown resistence statistics," *Semicond. Sci. Technol.*, vol. 15, no. 5, pp. 471–477, 2000.
134. J. Suñé, E. Wu, D. Jiménez, R. Vollertsen, and E. Miranda, "Understanding Soft and Hard breakdown statistics, prevalence ratios and energy dissipation during breakdown runaway," in *IEDM Technical Digest*, pp. 117–120, 2001.
135. K. Okada and K. Taniguchi, "Electrical stress-induced variable range hopping conduction in ultrathin silicon dioxide," *Applied Physics Letters*, vol. 70, no. 3, pp. 351–353, 1997.
136. J. Suñé and E. Miranda, "Post Soft Breakdown Conduction in SiO_2 Gate Oxides," in *IEDM Technical Digest*, pp. 533–536, 2000.
137. M. Alam, B. Weir, P. Silverman, Y. Ma, and D. Hwang, "The Statistical Distribution of Percolation Resistence as a Probe into the Mechanics of Ultra-Thin Oxide Breakdown," in *IEDM Technical Digest*, pp. 529–532, 2000.
138. M. Alam, B. Weir, and P. Silverman, "A Study of Soft and Hard Breakdown - Part II: Principles of Area, Thickness, and Voltage Scaling," *IEEE Trans. on Electron Devices*, vol. 49, no. 2, pp. 239–246, 2002.
139. B. Kaczer, R. Degraeve, M. Rasras, K. V. de Mieroop, P. Roussel, and G. Groeseneken, "Impact of MOSFET Gate Oxide Breakdown on Digital Circuit Operation and Reliability," *IEEE Trans. on Electron Devices*, vol. 49, no. 3, pp. 500–506, 2002.

140. M. Alam, R. Smith, B. Weir, and P. Silverman, "Statistically indipendent soft breakdown redefine oxide reliability specifications," in *IEDM Technical Digest*, pp. 151–154, 2002.
141. J. Suñé and E. Wu, "Statistics of successive breakdown events for ultra-thin gate oxides," in *IEDM Technical Digest*, pp. 147–150, 2002.
142. B. Linder, J. Stathis, R. Wachnik, E. Wu, S. Cohen, A. Ray, and A. Vayshenker, "Gate Oxide Breakdown under Current Limited Constant Voltage Stress," in *Proc. VLSI Technology Symposium*, pp. 214–215, 2000.