Review on High-k Dielectrics Reliability Issues

G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo

Invited Paper

***Abstract*—High-k gate dielectrics, particularly Hf-based materials, are likely to be implemented in CMOS advanced technologies. One of the important challenges in integrating these materials is to achieve lifetimes equal or better than their SiO_2 counterparts. In this paper we review the status of reliability studies of high-k gate dielectrics and try to illustrate it with experimental results. High-k materials show novel reliability phenomena related to the asymmetric gate band structure and the presence of fast and reversible charge. Reliability of high-k structures is influenced both by the interfacial layer as well as the high-k layer. One of the main issues is to understand these new mechanisms in order to asses the lifetime accurately and reduce them.**

***Index Terms*—Breakdown, BTI, high-k dielectrics, hysteresis.**

I. Introduction

THE PHYSICAL limitations of the conventional silicon dioxide as gate dielectric has reached the point where films thickness are only a few atomic layers thick [1]. Below the physical thickness of 15 Å, the gate leakage current exceeds the specifications (1 A/cm^2). To get around this critical problem, high-k dielectrics have been introduced as hafnium-based [2], zirconium, aluminum oxides. In fact, while keeping the EOT constant high-k dielectrics allow us to increase the physical thickness of the gate stack. Hence, the gate leakage is found to be reduced by 2 to 3 orders of magnitude. Although a large amount of effort has been invested toward high-k gate dielectrics, many critical problems still remain. These problems include defects in the material which can lead to undesired transport through the dielectrics and trapping-induced instabilities. Furthermore, the asymmetric gate band structure induces polarity effects on the leakage and reliability. All of these stack properties lead to anomalous behavior with respect to the conventional SiO_2.

In this paper, we will discuss the main differences in terms of reliability between high-k dielectrics and the better known silicon dioxide. We will begin by discussing the high-k bulk properties and the defect leading to a fast reversible charge trapping never observed until now on silicon dioxide. A second section concerns a problem already met on silicon dioxide and called bias temperature instability (BTI). Section IV will deal with the breakdown or characteristics of the dielectrics.

Manuscript received September 30, 2004; revised November 16, 2004. This work was supported in part by the STmicroelectronics, LETI, IMEP.

G. Ribes, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, and E. Vincent are with STmicroelectronics, F-38926, Crolles France (e-mail:guillaume.ribes@st.com; michael.denais@st.com).

J. Mitard is with CEA-LETI/D2NT-I7, 28054 Grenoble, France (e-mail: mitardje@chartreuse.cea.fr).

G. Ghibaudo is with IMEP, Grenoble, France.

Digital Object Identifier 10.1109/TDMR.2005.845236

II. Hysteresis Phenomena

One of the challenges for the high-k integration is the threshold voltage stability during operation. In literature, the hysteresis which is a shift in $V_{\rm th}$ is attributed to a trapping of charges in the pre-existing traps without creation of additional traps. We present in this section several measurement techniques which allow a better assessment of the $V_{\rm th}$ instability. The hysteresis dependences are discussed allowing argument on trapping and detrapping mechanisms. Finally, some process solutions are presented in order to improve the electrical stability of Hf-based dielectrics.

A. Methodology of Characterization

1) DC Characterization: Charge trapping has been characterized until recently by quasi-static or dc measurement techniques on MOS devices. Two types of measure are possible: the first is a sweep of gate voltage and the drain current is sensed; the second is also a sweep of V_g, but this time, it is the capacitance which is measured. In both cases, a shift of characteristics (hysteresis) is observed and is related to a build-up or loss of charge in the high-k stack. Fig. 1 (a) and (b) obviously shows the effect of charge trapping on the MOSFET characteristics. Furthermore, the amount of hysteresis can be calculated by comparing the V_t shift between up and down traces at fixed drain current or gate capacitance.

Finally, this technique has the advantage to screen quickly the charge trapping effects on the characteristics of MOS device. However, given the weak ramp rate (10 V/s maximum), this static measurement technique is not adapted to characterize fast transient trapping which is responsible for the hysteresis effect. A time-resolved measurement technique down to microsecond time range is necessary so as to capture the fast transient component of the charge trapping observed in the HfO_2 stack.

2) Dynamic Characterization: To qualitatively assess the threshold voltage shift, alternative fast measurement techniques are set up. Two methods have been reported in the recent literature: single-pulsed [3] or multiple-pulsed [4] techniques. Both methods allow a better evaluation of the fast charging and discharging effects in SiO_2/HfO_2 gate stacks and deal with the impact of $V_{\rm th}$ instability on the device performance.

a) Single-Pulsed Technique: This technique, developed for the first time by Kerber *et al.* [3], minimizes the effect of charge trapping and detrapping on the characteristics of

1530-4388/$20.00 © 2005 IEEE

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

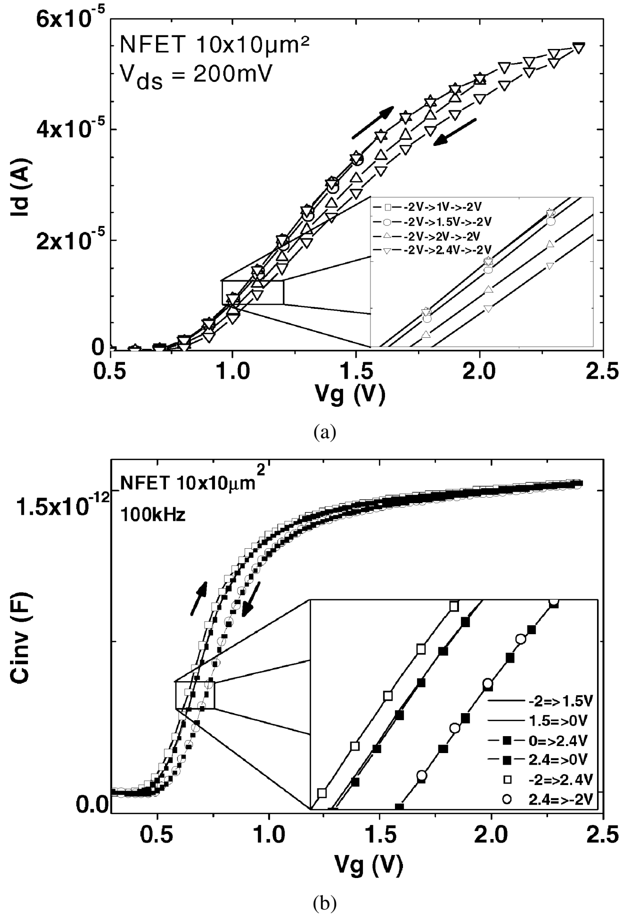

Fig. 1. (a) I_d–V_g characteristic of 1 nm SiO_2–4.5 nm HfO_2 NFET. (b) C–V_g characteristic of 1 nm SiO_2–4.5 nm HfO_2 NFET.

nMOS device. Fig. 2 illustrates the schematic measurement setup where the device under test is used in an inverter circuit. During measurement, a small static bias and a trapezoidal pulse is applied to the resistor and the gate, respectively. A digital oscilloscope is used to record simultaneously the drain and the gate voltage and I_d is extracted using the following relationship:

$$I_d = \alpha \cdot \left(\frac{V_{dd} - V_d}{R_L} \right) \quad \text{where } \alpha = \frac{V_{dd}}{V_d}. \tag{1}$$

Thus, from the gate leading edge, the up-trace of the I_d–V_g characteristic can be rapidly constructed. As for the down trace, it is determined by the down ramp of the sweep.

However, with this pulsed technique, we are obliged to use a normalization factor α since the drain voltage changes during measurement. This effect is due to the resistive load which forms a voltage divider with the channel of MOS transistor. This normalization to a constant V_d is justified only when the device under test operates in its linear regime, which is not always the case during the pulse. Moreover, when the pulse's rise or fall time becomes shorter, the charging current of parasitic capacitances such as C_{cables}, C_{gs}, or C_{gb} is more and more predominant and limits the speed of the measurement. Thus, it is more difficult to extract charge trapping characteristics for ultra-short time with this single technique (the resolution depends on the size of the tested device—for example, to characterize a $10 \times 10\ \mu\text{m}^2$ device, the minimum rise and fall time which can be used is about 2 μs; we use 50 μs).

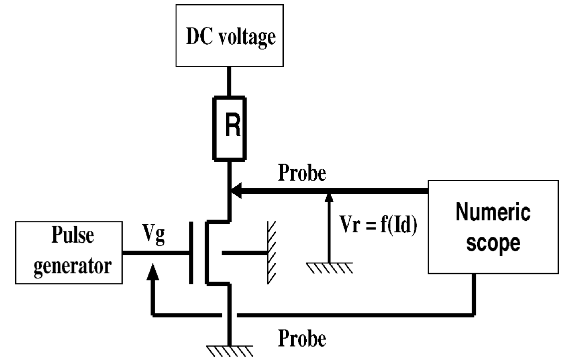

Fig. 2. Single-pulsed experimental setup to monitor the transient effects in high-k MOS structure.

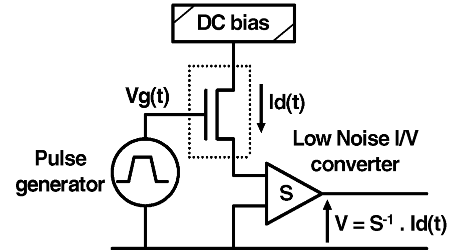

Fig. 3. Drawing of setup used for multiple-pulsed technique.

b) Multiple-Pulsed Technique: This technique, proposed by Leroux *et al.* [4], characterizes the trapping and detrapping transients and thereby the stability of the MOS transistor by monitoring the drain current evolution.

As shown in Fig. 3, a low noise and wide bandpass I/V converter is used to extract the $I_d(t)$ and a static bias generator is directly connected to the drain of the MOSFET. The advantages of this system as compared to the standard one [3] are threefold, namely: 1) a significant reduction of the noise level; 2) use of a constant drain voltage avoiding any drain current normalization as well as the charging current through the parasitic capacitances; and 3) accurate drain current transients.

Another originality of this fast measurement technique is the multiple $V_g(t)$ acquisitions. Indeed, typical $I_d(t)$ transients from microseconds up to seconds are represented in Fig. 4 using several gate voltage pulses with 2-ns leading edge. As a result, this complete set of transients in a large time-range enables an accurate evaluation of the trapping kinetics. In the same way, the detrapping kinetics of the fast traps can be deduced when the gate voltage goes from high to a lower level (for example: stress at 2 V and drop to 1.5 V for an NFET). Thus, by sampling these $I_d(t)$ and $V_g(t)$ curves *at specific times*, an $I_d(V_g)$ parametric plot can be reconstructed. As illustrated in Fig. 5, very short times for the $I_d(V_g)$ characteristics trace can be achieved (a few microseconds and only limited by the I/V converter speed).

Further, to convert the drain current transient into a threshold voltage shift, a reference characteristic is chosen. So, knowing the $I_d(V_g)$ parametric plot acquired at the shortest time (here 5 μs) and assumed without trapping, we can extract the V_{th} shift versus time for each drain current transient. Indeed, for a given

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

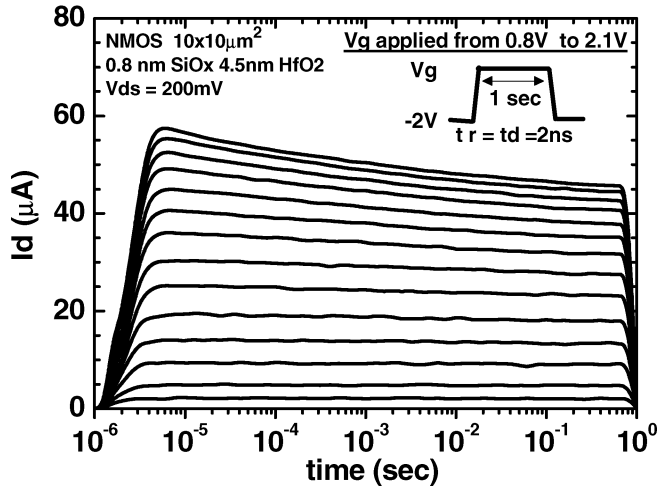

Fig. 4. Drain current degradation versus time and gate bias. Increasing drain current decay is measured with increasing gate bias.

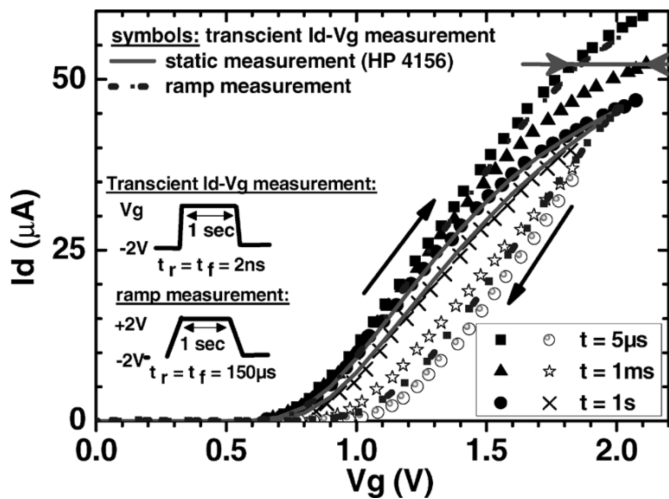

Fig. 5. Up and down I_d V_g trace at different times compared to static curves and classical dynamic measurement [1].

pulse, ΔV_t is obtained from the difference between applied gate voltage and corresponding gate bias taken at constant drain current on the reference curve (horizontal arrow in Fig. 5).

Finally, it is noted that all the results coming from this experiment are based on two assumptions. The first concerns the dielectric material polarization effects which, in the literature, were proposed as the cause for threshold voltage instability [5], [6]. However, a recent study [7] has shown that the origin of the dielectric relaxation current was the electron trapping and detrapping in the high-k dielectric stacks. Thus, it is supposed, in this experiment, the observed drain current variation is only due to a charge trapping effect. The second hypothesis is related to the carrier's mobility in the channel. In this section, the mobility alteration due to a scattering of trapped charge is considered as a second-order factor which does not affect the drain current degradation much.

c) Charge Pumping Technique: This technique, intensively investigated throughout the past decades, has been recently modified to characterize the charge trapping in Hf-based stacks, in particular HfO_2 [3].

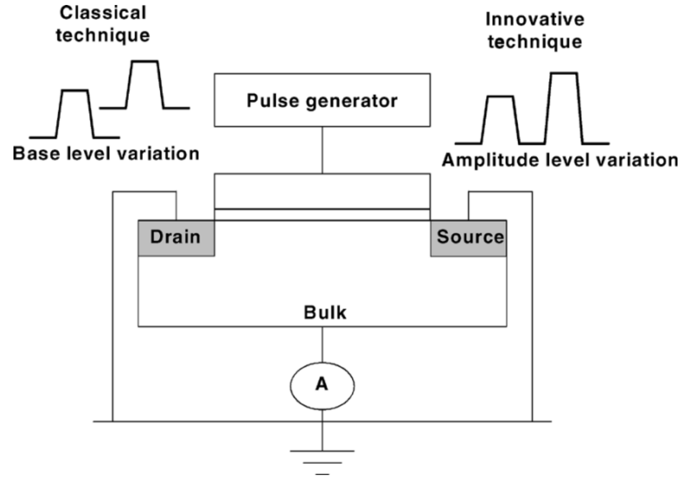

Fig. 6. Schematic diagram of the setup for charge pumping measurement.

At the beginning, charge pumping technique has been set up to study the interface properties between the substrate and gate dielectric (Si-SiO_2). Using this conventional technique, it has been possible to assess the interface states density (D_{it}) of MOSFETs only measuring the substrate current (elements of the theory are given in [8]). However, with the emergence of alternative gate dielectrics and the existence of "new" traps within SiO_2/high-k structure, a complementary method has been carried out.

As shown in Fig. 6, two types of signals can be applied on the gate in order to study different traps. The first, corresponding to the classical charge pumping (CP) method, has the amplitude of the pulse fixed while the base level is varied between two regimes, for instance from accumulation to inversion. With this conventional CP measurement, it has been demonstrated that it could be possible to isolate interface state densities from bulk traps. The second charge pumping procedure is quite different because the base level is fixed and a variable amplitude signal is applied on the gate. As mentioned for previous techniques, this method allows, once again, to underline the contribution of fast traps in the V_{th} instability phenomena and can help us to sense the defects in the high-k bulk. However, the originality of this last experiment lies in fast measurements which can be performed at different frequencies. So, the frequency response of traps can be obtained and information concerning their positions both in space and energy within the dual gate structure can be collected. Nevertheless, much care must be taken about geometric effects if a correct assessment of the defect density is desired [9].

B. Parameters Influencing the Hysteresis

In the previous section, fast measurement techniques allowing a correct assessment of the stability of high-k MOS devices have been outlined. These techniques are used to characterize charge trapping in SiO_2/HfO_2 dual stacks (details concerning the device fabrication are in [10]) and the dependence of charge trapping characteristics of these films with bias, time, stack thickness, and temperature are investigated in this section.

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

1) Polarity Dependence: Through the conventional measurement, I–V or C–V characteristics can be quickly measured and are useful monitors for charge trapping. Typical results obtained from multiple I_d–V_g and C–V_g traces are shown, respectively, in Fig. 1(a) and (b).

The first important point is that the superposition of all up traces contrary to the down traces which are increasingly shifted (inset of Fig. 1(a)). So, this fact suggests the charge trapping is a reversible phenomenon and a full $V_{\rm th}$ recovery is possible. However, it is noted that, if a starting voltage of 0 V rather than -2 V is used, a partial recovery of threshold voltage is obtained [Fig. 1(b)]. Thus, applying a sufficiently negative bias, a complete discharge of traps is assured. The second point is the raise of $V_{\rm th}$ shift when the maximum gate bias increasing. This shift of characteristics is consistent with a build-up of negative charge within the dual layer structure.

2) Time Dependence: Previously, the time dependence of hysteresis phenomena has been clearly evidenced from the short time measurement techniques (Figs. 4 and 5). Other research groups [13], [19] have also observed that the $V_{\rm th}$ shift due to the trapped charge follows logarithmic time dependence over several decades. The authors justified it in considering that the time constant t_0 is a function of the distance x from the dielectric/substrate interface. Indeed, the authors report that the trapping time constant $t_0 = q/J_G\sigma(x)$ increases from the interface due to a decrease in the transport probability of carriers from the substrate to the location of trap. Hence, the usual kinetics of charge trapping process can be written as

$$n_T(x,t) = N_T\left[1-\exp\left(\frac{-t}{t_0(x)}\right)\right] \tag{2}$$

where n_T is the density of trapped carrier, N_T is the trap density, and $t_0(x)$ is the time constant of the trap.

The time constant is a function of the capture cross section $\sigma(x)$ and the gate current density J_G. Since the transport probability of carriers changes exponentially with distance, the time constant $t_0(x)$ can be written as $t_o(x) = t_o(o)\exp(\alpha x)$ where α is the transport probability of carrier to reach the trap

$$t_0(x) = \frac{q}{J_G\sigma(x)} = \frac{q}{J_G\sigma_0\exp(-\alpha x)}. \tag{3}$$

Considering a uniform distribution of traps, the ΔV_t can be derived as

$$\Delta V_t = \frac{\mathrm{qN}_T}{C_{\rm ox}.\alpha}\ln\left(\frac{t}{t_0}\right). \tag{4}$$

Zafar *et al.* [13] have proposed a deferent V_t shift versus time dependence. The authors have used a model which assumes that the traps have a continuous distribution in capture cross section $\rho(\sigma)$ instead of single valued

$$\Delta Vt = \Delta V\max.\left[1-\exp\left(-\left(\frac{t}{\tau 0}\right)^{\beta}\right)\right]. \tag{5}$$

This assumption about $\rho(\sigma)$ allows the introduction of a constant β which is a measure of the distribution width. In other words, if $\beta = 1$, it implies that σ has discrete values with no distribution. However, for all ΔV_t curves, it has only found a good agreement with measurements when β is about 0.32. Thus, this result clearly shows the existence of a large time dependence of trapping phenomena.

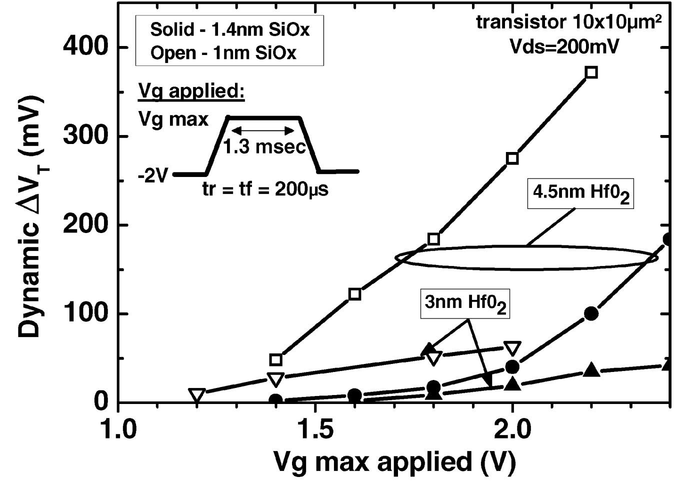

Fig. 7. Dynamic V_t shift for four samples with different gate architecture.

3) Stack Thickness Dependence: Based on the results of previous experiments, the threshold voltage instability ΔV_t in SiO_2/HfO_2 dual layer gate is due to the charging and discharging of pre-existing defects. To mention the impact of the gate stacks architecture on the hysteresis phenomena, results concerning four samples with different stacks are shown in Fig. 7. In this case, it is noted that the magnitude of $V_{\rm th}$ shift mainly depends on the HfO_2 thickness [11] whereas its bias dependence is mostly related to the interfacial oxide thickness. Hence, these results are consistent with the fact that reducing the thickness of the high-k material, the number of bulk defects available to be charged by electrons diminishes. Concerning the interfacial SiO_2 layer, it has been demonstrated in other studies [4], [12] that the thickness of this layer influences the mechanisms of charging and discharging of the HfO_2 defects. In all cases, at operational conditions for future transistors ($V_g = 1.2$ V), the $V_{\rm th}$ shift remains weak (< 40 mV) on three of the four tested splits.

4) Temperature Dependence: To study the influence of temperature on the $V_{\rm th}$ shift, single pulsed technique has been employed and the difference between the I_d-V_g curves generated by the up and the down swings of V_g has been measured. As illustrated in Fig. 8, a lowering of hysteresis is noticed when temperature increased. Up to now, this has been attributed to a temperature-enhanced detrapping [12], [14].

C. Mechanisms and Modeling

As shown previously, hysteresis phenomena are influenced by numerous parameters, namely, bias, time, architecture, and temperature. To understand these dependences, models have been put forward in the literature. In this section, a review of different models is proposed as well as possible mechanisms related to charge trapping and detrapping.

1) Charge Trapping: The magnitude of $V_{\rm th}$ instability in SiO_2/HfO_2 dual layer gate dielectrics is shown to depend strongly on experimental conditions as well. To take into account this polarity and time dependence, a team [16] has

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

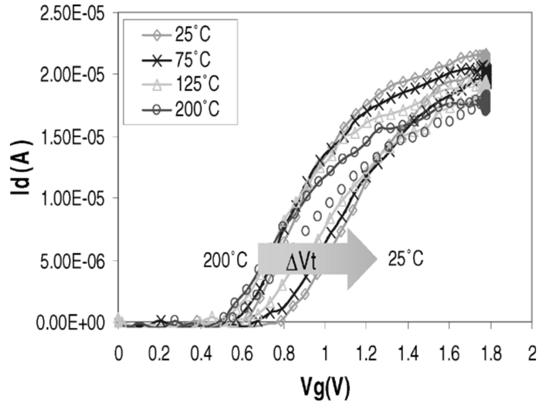

Fig. 8. Temperature dependence on hysteresis phenomena.

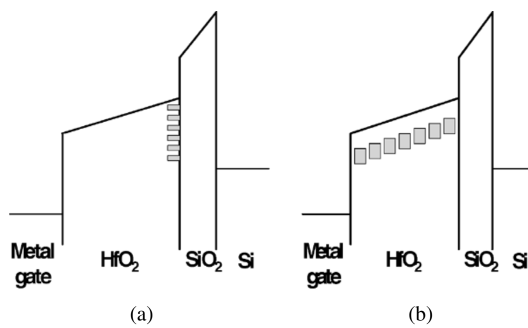

Fig. 9. Two types of model: defects are located (a) at the SiO_2/HfO_2 interface or (b) in the bulk of high-k film.

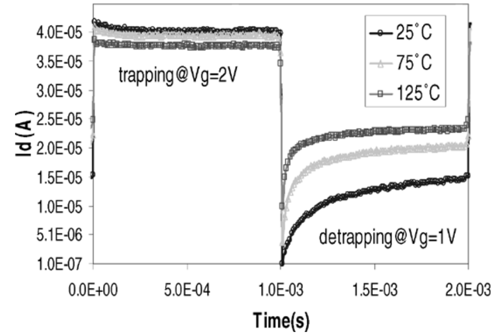

Fig. 10. $I_d(t)$ during gate pulse displays a charge trapping during pulse width at $V_g = 2$ V and charge recovery at $V_g = 1$ V.

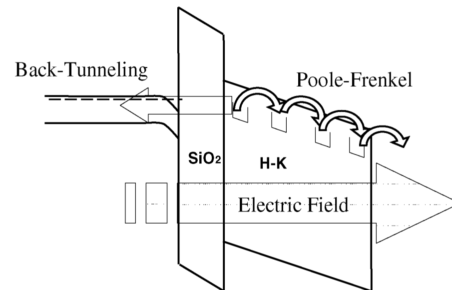

Fig. 11. Two mechanisms which may take part in the detrapping at positive bias: back tunnelling and Poole–Frenkel.

initially proposed a simple model with a defect band in HfO_2 layer located in energy between the silicon and hafnia conduction band [Fig. 9(a)]. This model has some attractive features, since it explains qualitatively, for negative and positive gate bias, the charging and discharging of defects by tunnelling through the interfacial SiO_2 layer. At the same time, studying the mobility degradation in high-k MOS devices, Morioka *et al.* [15] proposed to locate hysteresis traps at the HfSiO/SiO_2 interface as shown in Fig. 9(b). Thus, two types of models are encountered in the literature. To address the critical issue of the location of the trapped charge (bulk of the high-k film or the interface high-k/interface layer), Young *et al.* [11] have characterized the hysteresis phenomena on samples with different HfO_2 thicknesses. Their results suggest the trapping occurs mostly in the bulk of high-k material rather than only at the interface. This finding is consistent with data presented in Section II-C since less trapping in physically thinner high-k gate stacks is observed.

Considering the model with HfO_2 bulk traps, trapping mechanisms are finally investigated. In the previous section, we have observed the strong transient charging effects in particular with a thin SiO_2 bottom oxide. These transient effects are reasonably attributed to a tunnelling of carriers through the interfacial oxide. Depending on this interface, two mechanisms for electron trap filling can be mentioned: a filling by channel-to-defect tunnelling or by the capture of HfO_2 conduction band electrons. According to Pantisano *et al.* [12], for a bottom oxide intermediate thickness (~ 2 nm), trap filling is likely due to the second mechanism while for thin interfacial oxide (1 nm), a direct trap filling by electron tunnelling appears to contribute to the hysteresis phenomena. Recently, a study [4] has demonstrated that the trapping kinetic (V_{th} shift versus trapping time) can be modeled using a Shockley–Read–Hall statistic.

2) Charge Detrapping: It is commonly reported in the literature that applying a sufficient negative gate bias, a full recovery of trapped charge [3] is obtained. Nevertheless, even at positive bias, it has been observed that a partial recovery of charge can take place (Fig. 10). This partial recovery can originate from two possible detrapping mechanisms. The first may be due to a back tunnelling of electron from traps to substrate and the second, resulting from a Poole–Frenkel-like electrons conduction from traps to traps toward the gate electrode (Fig. 11).

Indeed, for the back tunnelling mechanism, a recent study [4] has shown that the detrapping kinetic can be modeled by a trap-like approach similar to a Shockley–Read–Hall model. Based on self-consistent solving of continuity potential relation and of the time evolution trap filling probability (6), simulations of threshold voltage shift with time account reasonably well for the obtained experimental results.

$$\frac{df_t}{dt} = c_n + e_p - (e_n + c_n + e_p + c_p) \cdot f_t \quad (6)$$

where c is capture, e is emission, n is electrons, and p is holes. Concerning the Poole–Frenkel mechanism, an experimental analysis [14] of recovery gate bias dependence of the detrapping characteristic time activation energy shows a field enhanced kinetics (activation energy decreases with recovery gate bias) (Fig. 12). This last result seems to be in agreement with a recovery assisted by Poole–Frenkel conduction. The temperature and gate bias dependence of detrapping characteristic time was also demonstrated to follow the Poole–Frenkel theory. In this

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

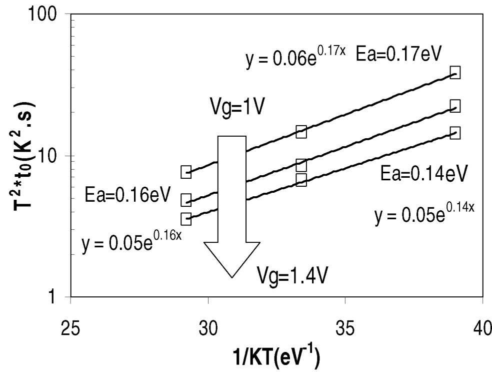

Fig. 12. Detrapping characteristic time t_o shows decreasing activation energy with respect to recovery gate voltage (1 V, 1.2 V, 1.4 V).

study [14], a good agreement is exhibited with experimental data and an extraction of the trap energy level distribution is given ($\langle Ec_{\mathrm{HfO_2}} - E_{\mathrm{trap}}\rangle = 0.35$ eV ± 0.034).

Thus, the detrapping kinetics seem to be dominated by a competition between two mechanisms. A detailed study of stack deposition processes (ALCVD, MOCVD, etc.) and stack architecture (chemical interfacial oxide, metal gate, or poly-Si gate) could distinguish the different mechanisms.

3) Origins of Defects: The physical and chemical nature of the electron traps in the HfO_2/SiO_2 structure remains unclear because the performance of high-k gate dielectrics is likely to be affected by various lattice defects. Among these defects, interstitial oxygen atoms and positively charged oxygen vacancies are suspected of playing a significant role in the observed V_{th} instability since they can trap electrons from the bottom of the hafnia conduction band and from silicon (results of *ab initio* calculations [22]). This result is consistent with investigations using electron spin resonance (ESR) measurements which have shown that O_2^- species, clearly a negatively charged defect, was the result of electron trapping in HfO_2 dielectric [23]. Additionally, chlorine impurities or water-related defects, such as OH (OH-) groups, are also possible contributors to the hysteresis phenomena.

D. Toward Minimizing Hysteresis

Hysteresis in scaled high-k dielectrics has been studied using both static and short time measurement techniques. The data shown previously indicates a significant amount of charge trapped especially for thick HfO_2 stacks. Although the exact origin of these traps still remains speculative, effective ways to improve the threshold voltage stability of alternative dielectrics are explored in this section.

1) Hafnium Silicate Integration: Among the candidates to replace SiO_2-based dielectrics, it is reported that nitrided silicate hafnium is also a promising material [18]. A publication has recently reported on the comparison of the electrical stability of nMOS devices using HfO_2 and HfSiON gate dielectrics [19]. A cross-sectional TEM image of HfSiON in a self-aligned planar CMOS transistor and an HRTEM image of the HfSiON layer show that the film is amorphous after a full CMOS thermal budget, including anneals at 1000 °C and 1050 °C.

This study clearly shows a hysteresis about ten times lower V_t shift for hafnium silicate compared to hafnium oxide.

A possible explanation of this result is the thermal stability of HfSiON films (it remains amorphous even after a high temperature annealing, contrary to an HfO_2 layer). Indeed, crystallization is, generally, an undesirable feature since it introduces grains into the gate stack. Thus, this nonuniformity of high-k stack can result in additional structural defects, contributing to the V_t instability.

2) Metal Gate Integration: In a previous study [19], it was also noted that a smaller hysteresis is observed on NFET using HfO_2 with TiN as metal gate rather than poly-Si gated HfO_2. This observation, confirmed by [20], indicates the gate process is also a contributor to the charge trapping. The reasons why the high-k dielectric is more stable with metal gate are still vague. Nevertheless, it is possible the top surface nitridation, a process essential to a compatibility of HfO_2 with polysilicon, leads to a degradation of high-k gate dielectric.

3) D_2O Annealing: Another way to effectively improve the hysteresis phenomena is the incorporation of deuterium during the ALD process. In fact, the D_2O device shows smaller charge trapping and thus a weaker V_{th} shift (about by a factor of 2) than H_2O ALD HfO_2 device [21]. It is likely this treatment passivates some dangling bonds responsible for the observed instability. As a result, the heavy water (D_2O) introduction into a conventional CMOS process allows appreciably reducing the number of traps within the high-k gate dielectric.

E. Conclusion on Hysteresis Phenomena

In this section, a presentation of hysteresis phenomena as well as the techniques to correctly quantify this effect has been made. Independent of the used high-k material, a few general remarks can be made.

To assess the severity of the charge trapping high-k gate dielectric, special fast measurement techniques are required. Furthermore, electron trapping in the high-k layer is the dominant charging mechanism at positive bias. Based on a model with bulk traps and depending on the interfacial layer thickness, two types of trapping mechanism are considered: a capture of HfO_2 conduction band electrons or an electron direct tunnelling from substrate to defect.

Concerning detrapping, it occurs when the positive bias applied to the gate is reduced. However, a total recovery of V_{th} shift can be obtained with a sufficient negative gate voltage. Two detrapping mechanisms seem to compete: a back tunnelling to the substrate and a Poole–Frenkel mechanism.

Finally, the V_{th} instability can be lowered by different ways. Indeed, the scaling of high-k material strongly reduces the number of defect sites to be charged. Also, the process improvements such as the D_2O use and the integration of new materials allow a significant reduction of charging effects. With these improvements, the problem of threshold stability during device operation can be prevented from being a major issue for the further integration of high-k gate dielectric.

III. Bias Temperature Instabilities (BTI)

Bias temperature instability (BTI) is a degradation phenomenon in MOS field effect transistors (MOSFETs), known since the late 1960s on SiO_2 dielectrics [24], [25]. Even though the

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

root causes of the degradation are not yet well understood, it is now commonly admitted that under a constant gate voltage and an elevated temperature, a build-up of charges occurs either at the interface Si/SiO_2 or in the oxide layer leading to the reduction of MOSFET performances. As a consequence of both the nitridation process step and the use of surface-channel devices, many researchers ascribed an accelerated BTI-like degradation of pMOSFETs under negative bias and elevated temperatures, the so-called negative bias temperature instabilities (NBTI) effect [26], [27].

Unlike SiO_2, the high-k dielectrics such as Hf-based dielectrics present serious instabilities for negative and positive bias, after NBT and positive bias temperature (PBT) stresses. The trapped charges are sufficiently high to represent one of the high-k integration's most critical showstopper. The instability is worrying, especially in the case of nMOS PBTI. In this section, we present a review of process optimizations found in the literature. A new experimental methodology is also presented in order to asses with accuracy the real degradation and allows us to argue on possible NBTI and PBTI mechanisms.

A. *Process Optimizations*

Recently, authors have reported sub-15-Å EOT pMOS NBTI exhibiting sufficient lifetime [28] but until now it is not the case for nMOS PBTI. Fortunately, even if nMOS PBTI lifetime remains strongly upsetting ($<$1000 s), some important process improvements were reported. It has been found that deuterium annealing at 600 °C improves significantly the V_t shift during PBTI stress in limiting both, the generation of oxide traps and interface traps [21]. Recently, Shanware *et al.* [19] has presented that nitride hafnium silicate (HfSiON) exhibits about ten times lower V_t shift than HfO_2 for identical drive current. The same authors compare TiN and polysilicon gate with HfO_2 dielectrics and found a factor 5 of V_t shift reduction during PBTI stress with the metal gate. Such process improvements allow us to expect a possible future integration.

Nevertheless, the physics of this mechanism is not yet understood and the reasons for the process improvements not completely found. In order to succeed in the integration of the hafnium-based gate dielectrics, it is indispensable to deeper understand the trapping mechanism. Following this viewpoint, we will propose in the following section an analysis of trapping mechanism and try to find the possible origins of the traps acting in the BTI stress.

B. *BTI Mechanisms*

1) Experimental Methodology: First, it was discussed in the first section that the high-k dielectrics present fast reversible charge trapping. This recovery induces underestimation of the V_t instability since the stress is stopped (Fig. 13). Hence, during a constant voltage stress, conventional methodology of degradation measurement is unable to estimate the total instability of the MOSFET because of the stress break during measurements. In order to correctly characterize it in the high-k, we need to measure the threshold voltage without interrupting the stress. An identical issue is already being dealt with for NBTI degradation in nitride-silicon dioxide, where it was observed [29] that hole trapping and detrapping introduces a significant error on the V_t shift evaluation. A neat measurement methodology resolving this problem has been proposed by Denais *et al.* [29], [30]. Using small gate voltage pulses during the stress, the authors measure the V_t shift without any interruption of the stress. Using this methodology, we obtain the degradation characteristics displayed in Fig. 14 which help us to gain better understanding of the high-k dielectrics instabilities.

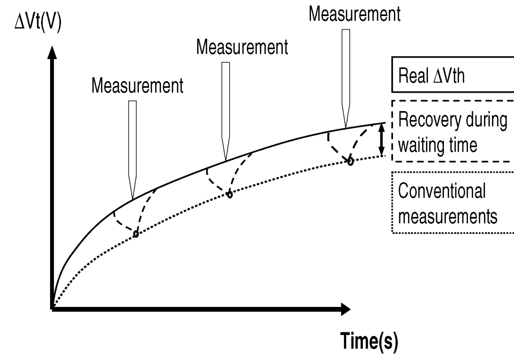

Fig. 13. Schema of the conventional V_t shift measurement methodology. During each measurement the stress is stopped and a part of charges recover.

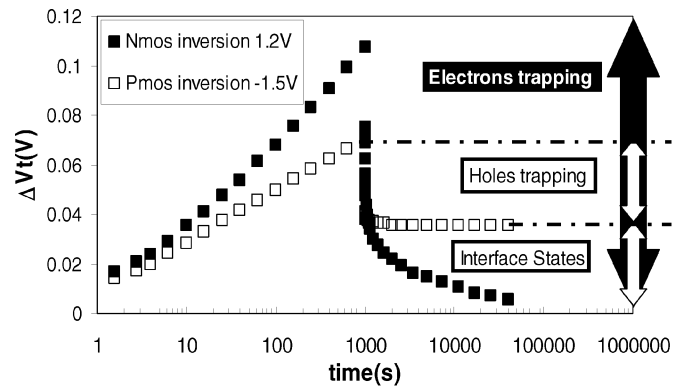

Fig. 14. PMOS and nMOS BTI stresses. The nMOS PBTI presents a total charge recovery and the pMOS NBTI a partial one.

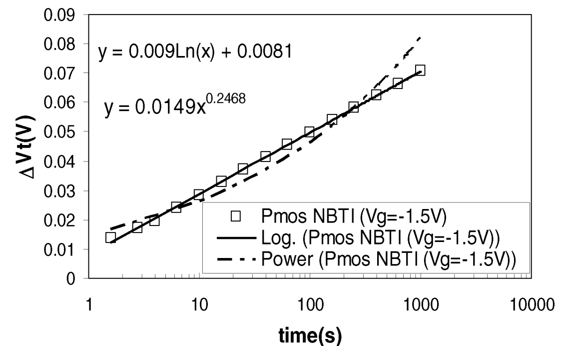

Fig. 15. PMOS NBTI V_t shift measurement without stress break. The V_t shift dynamic follows a logarithmic rather than a power law.

2) PMOS NBTI Mechanisms: The pMOS NBTI stress ($V_g = -1.5$ V, 125 °C) presented in Fig. 14 displays a positive charge ($\Delta V_t < 0$) trapping dynamic. As for hole trapping on nitride-silicon dioxide [30], a best fit is obtained for logarithmic law (Fig. 15) and a charge recovery is observed.

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

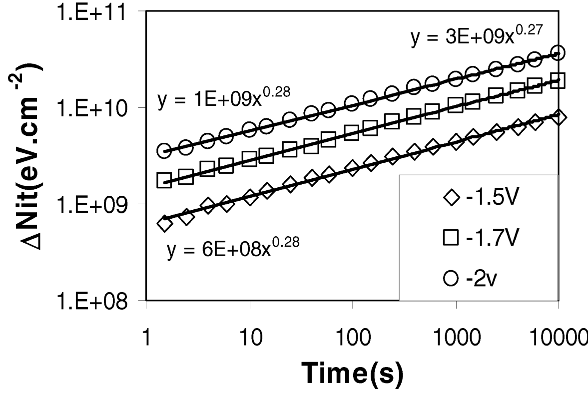

Fig. 16. Interface states generation during NBTI stress measured by charge pumping techniques on TiN/HfO_2 stack. ($HfO_2 = 30$ Å;$SiO_2 = 10$ Å).

The trapping dynamic can be well explained by the model proposed by Shanware *et al.* [31]. This is explained as follows:

$$\Delta V_t = \frac{qN_T}{C_{\text{ox}}.\alpha} \ln\left(\frac{t}{t_0}\right). \quad (7)$$

Hence, the V_t shift during the stress can be well explained by a hole tunnelling from the interfacial layer/substrate interface to the hole traps.

After applying a zero gate bias a recovery is observed until saturation level (Fig. 14). The recovery of ΔV_t corresponds to holes detrapping and the saturation level corresponds to interface traps. A similar behavior was found by Huard *et al.* [32] on nitride-silicon dioxide. The authors report that the V_t shift due to NBTI stress is composed of two traps: first, hole traps (metastable states) which recover, and interface states (Pb0) which are stable (no recovery). This is in agreement with the interface states generation measured using charge pumping measurement during NBTI stress (Fig. 16). The device used is TiN metal gate with 30 Å HfO_2 layer and 10 Å interfacial layer.

This is not the only common point between SiO_2 and HfO_2. Indeed Onishi *et al.* report that as SiON the NBTI lifetime of HfO_2 is deteriorated by the introduction of an NH_3 surface nitridation technique [28]. The observed similarities between high-k dielectrics and nitride silicon dioxide in gate injection mode (NBTI) are certainly due to the degradation prevalence of the interfacial layer [33], [34]. In fact, this layer being very close to amorphous silicon dioxide, we can easily understand that the mechanisms taking part in the degradation (Fig. 17) are similar to these observed in the NBTI silicon dioxide. Hence, understanding the NBTI mechanisms in SiO_2 gate oxide undoubtedly will help us to tackle instabilities associated with the High-k dielectrics

3) NMOS PBTI Mechanisms: In contrast to NBTI, the nMOS PBTI displayed in Fig. 2 shows an electron trapping ($\Delta V_t > 0$). The main difference with pMOS NBTI is that the whole V_t shift is recovered. That means that no interface traps are generated at this gate bias stress. As for the NBTI the PBTI characteristics displays a logarithmic law (Fig. 18) and it can be well explained by the direct tunneling electron trapping. As exhibited In Fig. 19, the electron trapping increases with temperature. This increase is very slight and characterized by

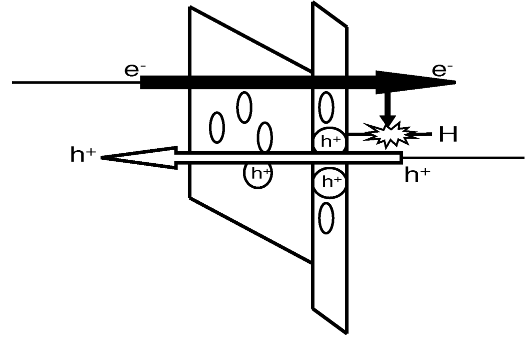

Fig. 17. Schema summing up the mechanisms taking part in the NBTI degradation. Hydrogen release induces interface states and oxide traps capture holes.

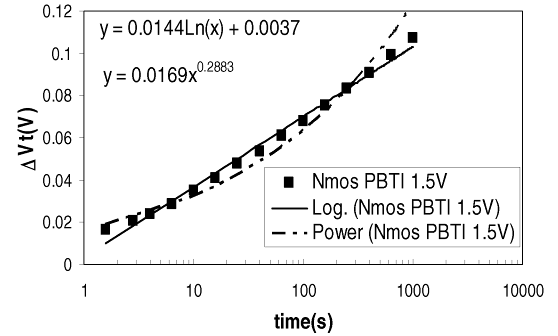

Fig. 18. NMOS PBTI V_t shift measurement without stress break. The V_t shift dynamic follows a logarithmic law.

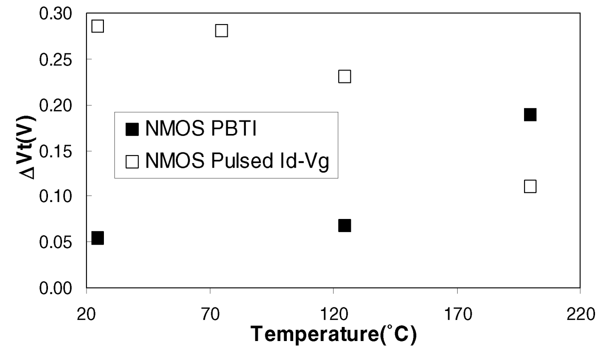

Fig. 19. Temperature dynamic for nMOS PBTI V_t shift for the fast electron trapping (open squares: V_t shift measured with pulsed I_d-V_g) and the long electron trapping (filled squares: V_t shift measured by the Denais method).

activation energy of 0.075 eV (Fig. 20). The weak value is in agreement with the low activation energy of the direct tunnel current. It confirms the assumption of carriers direct tunneling from interface to traps toward the dielectrics [3], [4], [11], [12], [14].

Both long-term stress and the fast transient charging characterized by pulsed I_d-V_g technique correspond to electron trapping. Nevertheless both mechanisms present opposite temperature behavior. Indeed we have already shown that electron trapping characterized by pulsed I_d-V_g decreases versus temperature (Fig. 9). The opposite evolutions versus temperature mean that the trapping mechanisms are different: One fast and decreasing with temperature and the others slow and slightly increasing. The differences, between fast and long-term stresses,

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

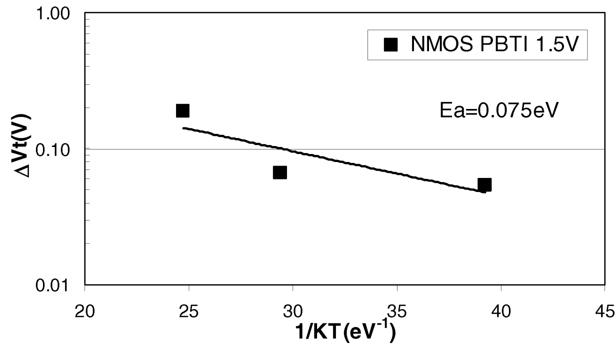

Fig. 20. Activation energy of the PBTI long-term stress V_t shift.

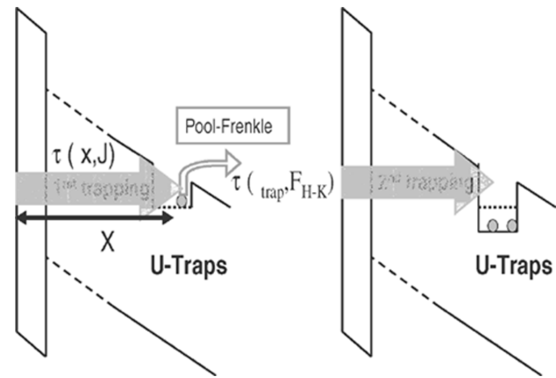

Fig. 21. Schema illustrating the trapping mechanism of U-traps for Hf-based dielectrics. The first electron trapping is a meta-stable state and the electron can be emitted by Pool-Frenkle detrapping. The second electron trapping transforms the meta-stable state into a stable one (the trapped electrons emission is strongly reduced).

can help us to identify the defect responsible of the PBTI instability. In the first section, it has been reported that the decrease of I_d-V_g hysteresis with temperature (Pulsed I_d-V_g) is explained by the following assumption. During positive bias the V_t shift is equilibrium between electron trapping assisted by direct tunneling and electron-detrapping assisted by Pool-Frenkel mechanism. A possible explanation of the different temperature behaviors can be found in only one defect type: the U-traps [35]. Very recently Kang *et al.* [36] have reported that hydrogen in monoclinic HfO_2 is negative U traps capturing two electrons with $U = -1.6$ eV. The process could be hypothesized as follows: when the electrons are injected into HfO_2, the neutral vacancy V^0 will first capture one electron, which is meta-stable state. Therefore it subsequently captures the second and the defect become stable for positive bias (due to a large lattice relaxation). Now considering that the fast traps correspond to the first electron trapping which is meta-stable state, a weak energy is necessary to detrap the electron ($\mathrm{Ec}_{\mathrm{HfO}_2} - \mathrm{E}_{\mathrm{traps}} = 0.35$ eV [14]). The second electron trapping induces a stabilization of the meta-stable traps into a stable one and so the energy necessary to release the second electron is higher ($\mathrm{E}_V{}^- - \mathrm{E}_V{}^0 > 0$). This increase of energy can explain a limited Pool-Frenkel detrapping for slow traps (Fig. 21).

Hence, the equilibrium between direct tunneling electrons trapping and Pool-Frenkel detrapping moves to a strong prevalence of the direct tunneling electron trapped. By this way, the well known direct tunneling weak temperature activation can be the root cause of the slight increase of the trapped charge with higher temperature of the long-term stress.

C. Conclusion on BTI

In this section we have seen that both mechanisms present serious issues in terms of high-k reliability and need to be reduced by process optimization. The principal improvement is found for hafnium silicate coupled with a metal gate and deuterium annealing. It has been discussed that defects called U-traps with energy distributed and localized in the high-k layer could explain the very efficient electron trapping observed on high-k dielectric for PBTI in long and short term stress (hysteresis).

In the case of NBTI the charge is rather composed by two defects as in nitride-silicon dioxide. The first is a stable state called the Pb_0 center located at the Si-substrate/interfacial layer interface and a meta-stable one: the hole traps.

IV. Dielectric Breakdown

In the last few years, competition has driven the pace of decrease the oxide thickness and even leads us to change the nature of the gate dielectrics (H-k dielectrics). Nowadays the margin in term of intrinsic lifetime, as a result of downscaling and dielectrics changes, has disappeared. Hence, an accurate projection of ultrathin gate-oxide and high-k dielectrics become crucial in order to assess the dielectrics breakdown issue. Area scaling consistent with Weibull statistics has been shown to apply in high-k stacks with both metal gate and polysilicon + nitride capping layer, demonstrating that intrinsic effects rather than manufacturing induced defects, dominate TDDB [34], [37]. Some controversy remains in substrate injection [33], [34] but for gate injection, a consistent picture has begun to emerge recently indicating that the reliability of high-k dielectrics is dominated by the breakdown of the interfacial layer rather than the high-k layer itself [33], [34]. Hence, the large experience on the silicon dioxide breakdown is a precious background to understand the High-k stack breakdown.

The goal of this section is to give an overview of the High-k dielectrics breakdown literature and try to link it with the silicon dioxide breakdown knowledge.

A. Breakdown Occurrence

In order to asses an accurate lifetime it is crucial to detect the first occurrence of the dielectric breakdown. The different occurrence of the wear out was already analyzed and defined in literature. Furthermore, since some MOS digital circuits could remain functional after gate oxide breakdown (BD) provided that the post-BD resistance is high enough [38], the separate consideration of soft (SBD), progressive (PBD) and hard (HBD) breakdown events is necessary to set up an adequate application-specific reliability assessment methodology. The link between gate oxide failure and device failure is now a major concern for the reliability community. So it is important to well identify the different breakdown occurrence and give an accurate definition.

1) Definitions of the Different Occurrences: The SBD and the HBD are localized and randomly distributed all over the device area [39]. It is now commonly admitted that SBD is not a HBD precursor. Both do coexist, soft and hard breakdowns

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

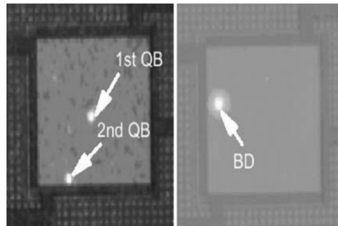

Fig. 22. Emission microscope measurements displaying SBD(QB) and HBD(BD) on SiO_2 dielectric shows that both breakdown occurrence are spatially uncorrelated. [39].

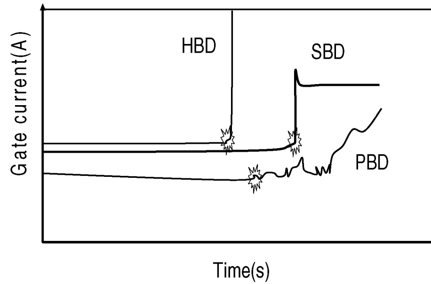

Fig. 23. Schema illustrating the three different occurrences of the breakdown (HBD, SBD, PBD).

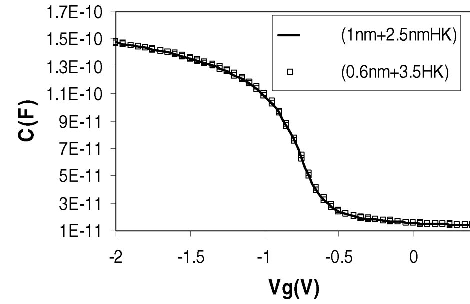

Fig. 24. C-V characteristics measurements on two stacks displaying identical characteristic.

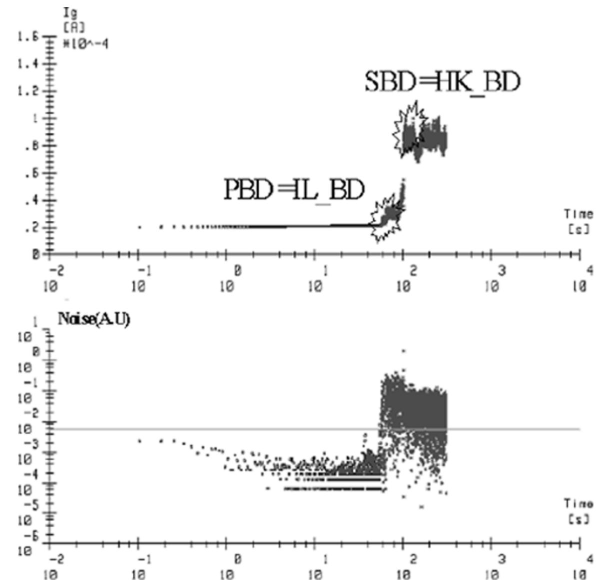

Fig. 25. $\mathrm{Ig}(t)$ characteristics under CVS stress on the stack 1 nm of interfacial layer and 2.5 nm of silicate.

being spatially uncorrelated (Fig. 22). The prevalence ratio that is the fraction of each failure mode as a first failure varies with stress conditions, oxide thickness and gate oxide area [39], [40]. When a hard or a soft breakdown occurs, the current rapidly reaches a steady level that can last hundreds of seconds. This state is then interrupted by another failure occurrence (that can be either soft or hard). Finally, the magnitude of this current level (low for SBD and high for HBD) is attributed to a difference of the size or resistance of the conduction path created in the oxide by defect percolation [41], [42]. Regardless, these (HBD and SBD) current levels are oxide thickness, device area and stress conditions independent [39], [42].

For oxide thinner than 25 Å, a new failure manifestation appears as in increase in the noise level called progressive breakdown PBD [43]. It has been shown that only the noise occurrence is randomly distributed all over the device area (area scaling and Weibull statistics are followed) [43]–[45]. As a result, only the noise detection enables to extrapolate the time to failure occurrence with respect to the device area and the cumulative failure percentage. The "noisy breakdown" prevalence ratio reaches 100% for oxide thinner than 20 Å. After the failure occurrence, the current increases continuously till the hard or soft breakdown occurs showing a direct link with the hard or soft breakdown failure mode [46]. Indeed the PBD was reported to be the initiator of the hard or soft [46].

To summarize, the PBD is the occurrence which announces the percolation path creation toward the layer. It is the initiator of a SBD or a HBD. The different occurrence of these two breakdown types results from the breakdown spot size or resistance. A figure illustrating the different breakdown occurrences is given Fig. 23. In the case of high-k dielectrics the presence of high-k layer implies that the origin of the PBD, SBD and HBD occurrence may be changed and then the definition may have to be modified. Hence, it is crucial to analyze the breakdown occurrences and perform an accurate phenomenological and statistical study.

2) Phenomenological Study: Authors have published that the three breakdown occurrences appearing in the SiO_2 appears in high-k dielectrics too [37], [47]. The first is a noise increase so called the PBD and the second, a fast increase of gate current which can be hard or soft. Using a carrier separation analysis Mizubayashi *et al.* [48] have reported that in gate injection for p-MOSFET the dominant transport after the soft breakdown is an electron current. They report that the defect site in the high-k stack after soft breakdown are located at energies near the conduction band edge of n+ Poly-Si. But the question remains as to which layer triggers the breakdown.

In order to determine which layer trigger the breakdown we have reproduce these experiments on two stacks one with 0,6 nm of interfacial layer and 3.5 nm of hafnium silicate and the other with 1 nm of interfacial layer and 2.5 nm of silicate. As it is shown by the C-V characteristics displayed on the Fig. 24, the two stacks have the same EOT and the same V_{fb}.

We stress the two stacks of same area at the same gate voltage. Typical $\mathrm{Ig}(t)$ observed are depicted Figs. 25 and 26. PBD is observed on both but the PBD appears sooner on the physical thicker stack. It is probably linked to the fact that the physical

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

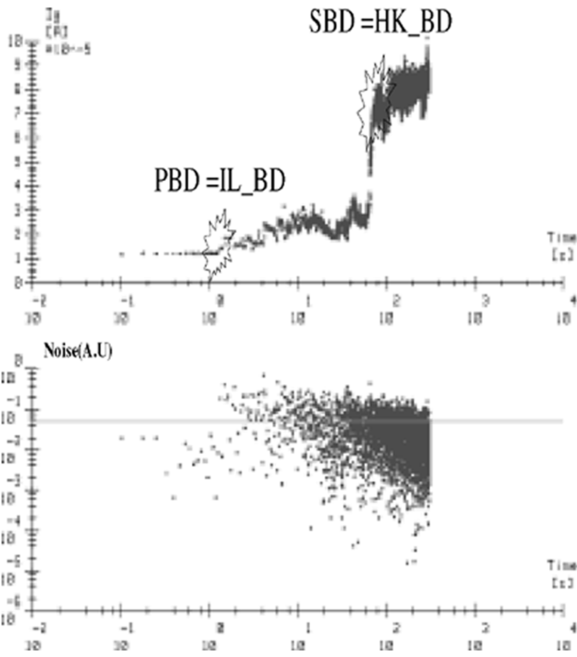

Fig. 26. $\mathrm{Ig}(t)$ characteristics under CVS stress on the stack 0.6 nm of interfacial layer and 3.5 nm of silicate.

thicker stack is the one with the thinnest interfacial layer. Hence, this result could be an indication of the correlation between PBD and the creation of a percolation path toward the interfacial layer. After the PBD occurrence the SBD appears (Figs. 25 and 26). Surprisingly, in spite of a PBD appearing sooner the progressiveness is longer for the stack (0.6 nmIL + 3.5 nmHK). This is probably due to the higher High-k thickness (3.5 nm) of this stack. Hence, the PBD seems to be dependent of the interfacial layer breakdown and the SBD of the high-k layer.

3) Statistical Analysis: Authors report that the PBD and SBD have different Weibull slope and that both events are area scalable [47], [49].

That means that the PBD and the SBD or HBD verify the Weibull statistic:

– breakdown appears on one site among a lot,
– weakest site determines the breakdown
– breakdown occurrence of one site is independent of others breakdown sites.

The different Weibull slope between PBD and SBD means that they correspond with two statistical different events. So the thickness of the broken layer and, or the defects are not identical. The area scaling of the two events means that the PBD and the SBD are uncorrelated.

In order to enhance these observations, we have proceeded to a statistical analysis of the PBD and SBD observed on the two stacks described above (0.6 nmIL + 3.5 nmHK) and (1 nmIL + 2.5 nmHK). First we verify that the PBD and the SBD occurrence are area scalable (Figs. 27 and 28).

Plotting the distributions of the two stacks PBD and SBD (Fig. 29), we observe that the two PBD distributions have the same Weibull slope. This result is in agreement with the constant Weibull slope measured on SiO_2 layer thinner than 1.2 nm. Indeed the defect cross section in the SiO_2 is supposed to be close to this thickness [50]. In all likelihood the PBD is the electrical

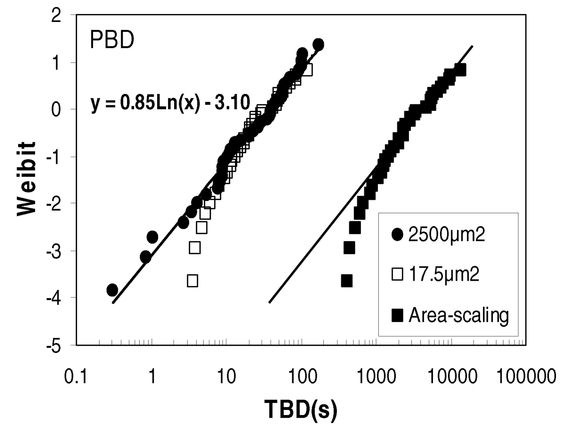

Fig. 27. Time to breakdown Weibull distribution detected at PBD occurrence displays area scaling from 2500 $\mu\mathrm{m}^2$ to 17.5 $\mu\mathrm{m}^2$.

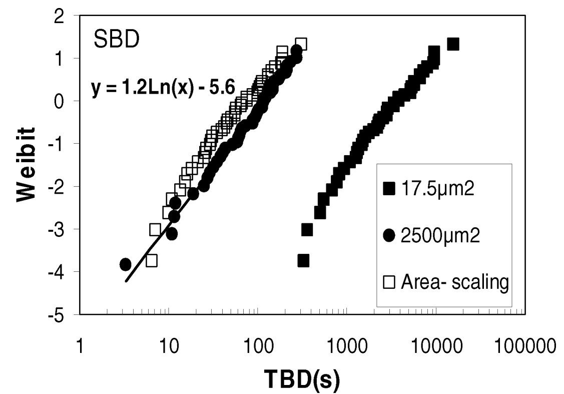

Fig. 28. Time to breakdown Weibull distribution detected at SBD occurrence displays area scaling from 2500 $\mu\mathrm{m}^2$ to 17.5 $\mu\mathrm{m}^2$.

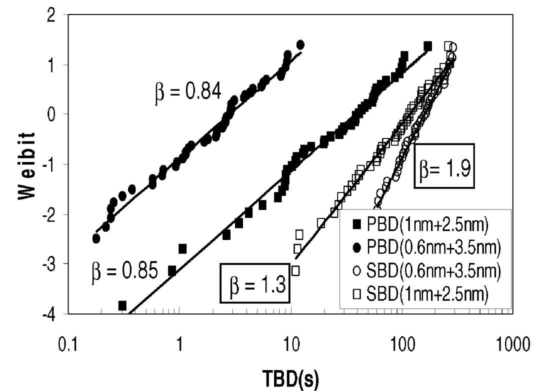

Fig. 29. Weibull distribution on (1 nm SiO_2 and 2.5 nm HK) and (0.6 nm SiO_2 and 3.5 nm HK) stacks detected for both PBD and SBD.

occurrence of percolation path appearance toward the interfacial layer.

In contrast the soft breakdown (SBD) displays different Weibull slope ($\beta_{2.5\ \mathrm{nmHK}} = 1.3$ and $\beta_{3.5\ \mathrm{nmHK}} = 1.9$). This Weibull slope increase, with respect to high-k layer thickness, is in agreement with the link between SBD and high-k layer.

The percolation model ($\beta = t_{\mathrm{ox}}/\mathrm{a}_0$) [51]–[53] predicts that the cross section of the defect is equal to 1.9 nm (Fig. 30). By this way, below a thickness of 1.9 nm, the Weibull slope of the SBD should remain constant as SiO_2 below 1.2 nm. Hence, a ratio close to 1.5 between the silicate defect cross section and the

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

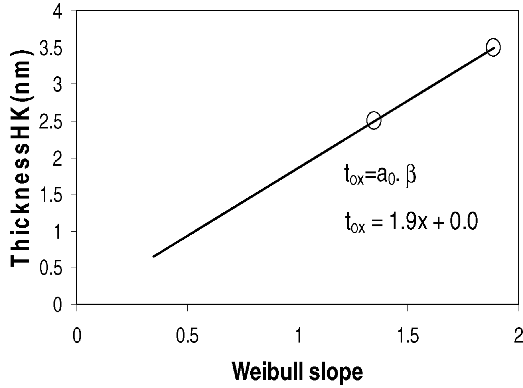

Fig. 30. Plot of high-k thickness versus Weibull slope detected at SBD displays a 1.9 nm defect cross section.

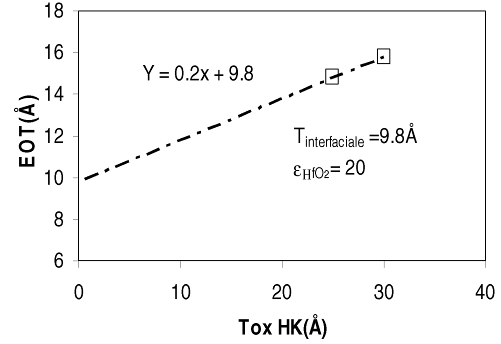

Fig. 31. EOT extracted from NCSU versus physical oxide thickness of high-k dielectrics exhibits a consistent interfacial layer thickness with the targeted one.

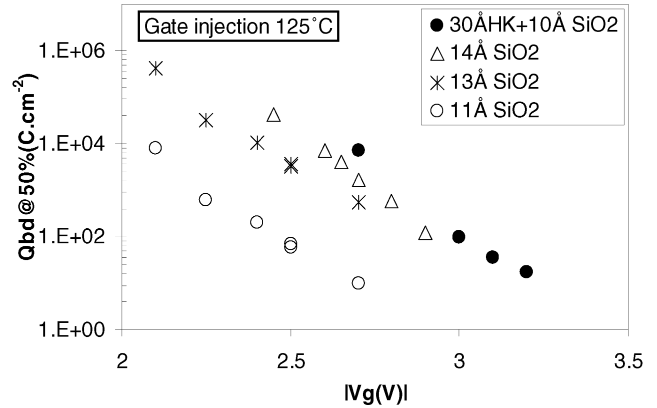

Fig. 32. Charge to breakdown versus gate voltage for thin SiO_2 (from 14 Å to 11 Å) and (30 ÅHK+10 ÅIL) stack.

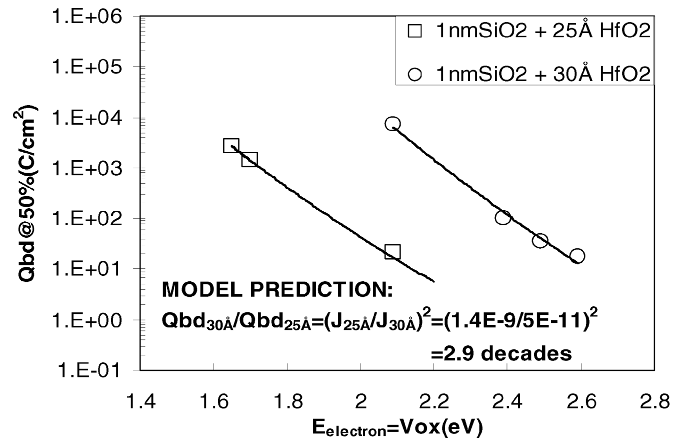

Fig. 33. Charge to breakdown versus Vox measured on two stacks with 25 Å HfO_2 and 30 Å HfO_2 thickness and same interfacial layer thickness (1 nm). A shift of around 3 decades between the two stacks is observed which is in agreement with the MVHR prediction.

SiO_2 one is found. So the Weibull slope of the High-k at same thickness must decrease by the same ratio. This is in agreement with low Weibull slope observed in substrate injection [33] with respect to those measured on SiO_2 layers.

B. Breakdown Mechanisms

1) Gate Injection:

a) Thick High-k Stack: Further indications that the interfacial layer plays a strong role in the breakdown in gate injection have already been published [33], [34]. It has been found that, for stack with high-k layer thicker than 8 nm and carrier energy higher than 5 eV, the Qbd for ZrO_2/SiO_2 stacks are equivalent with those for SiO_2 with a thickness identical to the interfacial layer [33]. This result is in agreement with the similarity between NBTI SiO_2 degradation and NBTI high-k stack degradation discussed in the previous section. Moreover, the Weibull slope of the time to breakdown exhibits a good correlation with SiO_2 interfacial-layer thickness. The authors explain this mechanism by the Anode Hole Injection model, considering that hot tunnelling electron can give its energy to an electron-hole pair and release a hot hole which is then able to generate a defect by inelastic heat. The hydrogen release model could also explain the breakdown of the SiO_2 interfacial layer. Indeed the process is similar to the Anode Hole Injection apart from the fact that the excited element is not the electron-hole pair but interfacial Si-H bond.

b) Thin High-k Stack:

Experimental Investigations: Below 5 eV electron energy, the interpretation of the results is more complicated. The Fig. 31 displays the equivalent oxide thickness extracted from CV measurement (NCSU) for two MOS transistors one with 30 Å high-k thickness and 10 Å interfacial layer, the other 20 Å and 10 Å interfacial layer. As is shown, the EOT's measured confirm the targeted thicknesses. Knowing the stack layers' thicknesses, we can see that contrary to thick high-k stressed at carrier energy higher than 5 eV [33], the Qbd is no longer scalable with interfacial SiO_2 layer thickness. Seen either as electron energy, the gate voltage (Fig. 32), or as the voltage drop toward the dielectric, Vox (Fig. 33), the nonscalability is verified in the both cases.

As it is shown on Fig. 31, the high-k stack charge to breakdown is higher than the corresponding SiO_2 interfacial layer thickness one (10 Å). That means that probably the High-k layer plays a role in the defect generation probability (Pg) or in the critical defect density (N_{BD}).

However, we have to note that the Weibull slopes remain identical close to $\beta = 1.1$ for the two stacks, which is in agreement with the breakdown of the interfacial layer. The Weibull slope being directly linked to the N_{BD} [53], this result seems to show that the N_{BD} is independent of the high-k layer thickness.

The temperature behavior of the breakdown can help us to determine the first layer broken. On Fig. 34, we present results

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

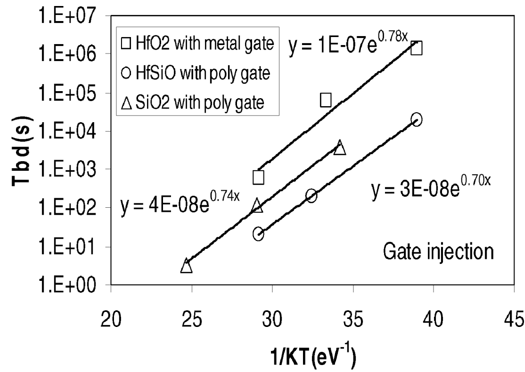

Fig. 34. Time to breakdown at 50% displays similar activation energy on three different dielectrics (HfO_2 with metal gate, HfSiO with poly gate and SiO_2 with poly gate).

on HfO_2 with metal gate and HfSiO with polysilicon gate. The times to breakdown display similar activation energy for both stacks (0.7 eV $<$ Ea $<$ 0.8 eV). Furthermore, the activation energies are in agreement with those measured on SiO_2. The temperature behavior similarity between these completely different stacks is an additional indication about the interfacial layer breakdown. This way the high-k thickness dependence of the Weibull slope and the temperature behavior are in agreement with a critical defect density ($N_{\rm BD}$) independent with the high-k thickness.

Hence, the $\rm Qbd$ shift between both stacks seems to originate rather from a high-k thickness dependence of the defect generation probability ($\rm Pg$).

Discussions: It has been reported that, below 5 eV, only one electron is not sufficient to break the Si-H bond by direct excitation [54]. Hence, below this carrier energy, the defect generation probability ($\rm Pg$) depends on gate current density ($\rm Qbd\% \ 1/Pg\% \ J^{-2}$) [55]. This dependence is linked to the multivibrational excitation of the Si-H bond [54], [56]. It was reported that cooperation between electrons is needed in order to break the bond [54], [56]. This process allows electrons to further excite the bond until an energy relaxation time. Hence, following this process, the defect generation probability depends on gate current density.

This theory was verified on SiO_2 MOS transistors with SHE and SHH stresses which enable to decorrelate all the stress parameters (electric field, gate current, electron energy) [57]. With this model, we explain all CVS stress polarities for pMOS and nMOS SiO_2 dielectrics [55]. Hence, according to the multivibrational hydrogen release model (MVHR), below 5 eV the $\rm Qbd$ should decrease with decreasing high-k layer thickness.

That way, the current dependence of the defect generation probability predicted by the MVHR model may be an explanation. Indeed, considering that the electron energy, in accumulation regime, is the oxide voltage drop toward the stack ($\rm Vox$) [55] and that the defect generation probability follows a power law of the gate current density ($\rm Pg\% J^2$) as predicted by the MVHR model, the $\rm Qbd$ shift between the two stacks depicted on Fig. 33 is explained and quantified with accuracy.

Hence, these results confirm that in spite of the $\rm Qbd$ nonscalability, below 5 eV, with respect to the interfacial layer thickness, the breakdown remains driven by the interfacial layer.

Finally, other degradation mechanism models are favored by different groups. For reliability projection, the voltage dependence of the failure time must be clarified. Both charge injection model (HR, MVHR, AHI) predicts a rapidly increasing life time for high-k stacks, as gate voltage is reduced, due to the current limiting effect of the high-k layer. Recent measurements of the voltage acceleration with decreasing gate voltage ([58]) are consistent with a charge injection model. The thermo-chemical model has also been applied to explain high-k breakdown phenomena ([59]), but this model results in more conservative and pessimistic reliability projection. Experimental investigation in order to determine which degradation mode gives the appropriate voltage extrapolation appears an important challenge for the future. All these results seem to show that answers can be found in the good understanding of the SiO_2 breakdown.

2) Substrate Injection: Because of the good agreement between high-k stack and interfacial SiO_2 layer thickness charge to breakdown, [33] and [60] have first reported that even in substrate injection, the stack breakdown is driven by the interfacial oxide layer. This proposition was supported by the low Weibull slope observed with respect to the high-k layer thickness. Nevertheless, study and especially degradation analysis such as SILC [33] or trap density in the high-k layer characterize by charge pumping [60] are in agreement and seems to show that as for Al_2O_3 in substrate injection the layer which drives the stack breakdown is the high-k dielectrics. Since it has been proved that the trapped charge ($\rm D_{HfO}$) [60] or SILC [33] remains constant at breakdown for different stress conditions, the authors report that electron trapping into the high-k dielectrics would be responsible of the gate stack breakdown. Hence, the dielectric breakdown and the PBTI could be directly correlated. The breakdown assisted by electron trapping in the high-k layer seems to be coherent with degradation analyses but the low Weibull slope (β) with respect to the high-k layer thickness remains surprising. An explanation could be found with a percolation model [51]–[53]

$$\beta = \frac{t_{\rm ox}}{\rm a_0} \tag{8}$$

with a defect cross section ($\rm a_0$) higher for HfO_2 than for SiO_2 resulting from the different nature of the two dielectrics: covalent and amorphous for SiO_2 and ionic crystal or poly-crystal for HfO_2.

Further investigations are required in order to understand the high-k dielectrics breakdown complexity. Notably, all degradation mechanisms analysis like interface states, oxide traps generation and temperature behavior appears to be important issues which must be understand before trying to explain the creation of the breakdown path in the high-k dielectrics. Comprehension of these mechanisms is crucial in order to make extrapolation models trustworthy and to accurately assess the high-k dielectrics reliability in the substrate injection regime

C. Conclusion on Dielectrics Breakdown

The degradation, the breakdown occurrence phenomenological and statistical analyses are found to be consistent.

In gate injection, the first stack wear-out is the interfacial layer breakdown. It is characterized by a progressive breakdown

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

(PBD). According to the MVHR model the charge to breakdown of this event is dependent of the gate current and so to the high-k layer thickness. The gate current increases, until second breakdown which can be soft (SBD) or hard (HBD) depending of the spot size. Both breakdowns were identified as the breakdown of the high-k layer. This event is area scalable and then independent of the interfacial layer breakdown. So probably, it appears at others breakdown site than the interfacial layer breakdown path. Hence, in order to assess the lifetime of the whole high-k stack an accurate detection must be performed, especially to determine the circuit lifetime.

In the case of substrate injection further research is required. In all likelihood, the layer responsible of the breakdown event seems to be the high-k layer. Accurate statistical analyses, with thickness layers discriminations, remain crucial in order to understand the breakdown in this stress regime.

V. Conclusion

The reliability of high-k gate dielectric stacks is influenced both by the interfacial layer as well as the high-k layer. TDDB results indicate these mechanisms do not pose fundamental limitations but need to be understood deeply. Charge trapping, however, presents a serious reliability challenge, primarily as a result of V_t instabilities. Data is available to show that required levels of long-term V_t stability of both nMOS and pMOS devices can be attained, but will require careful process optimizations. However as the basic understanding of hot carrier aging mechanisms and plasma-induced damage (PID) begins to be published, a deeper understanding is urgently required. It is clear that this there is a link between the major performance and reliability factors, so that reliability improvement can be expected to result in continued process improvements.

References

[1] D. A. Buchanan, *IBM J. Res. Dev.*, vol. 43, p. 245, 1999.

[2] E. P. Gusev, "The physics and chemistry of SiO_2 and the Si-SiO_2 interface," *Electrochem. Soc.*, pp. 477–485, 2000.

[3] A. Kerber *et al.*, "Characterization of VT instability in SiO_2/HfO_2 gate stacks dielectrics," in *Proc. Int. Reliability Physics Symp.*, 2003, pp. 41–45.

[4] C. Leroux *et al.*, "Characterization and modeling of hysteresis phenomena in high-k dielectrics," in *IEDM Tech. Dig.*, 2004, pp. 737–740.

[5] H. Reisinger *et al.*, "A comparative study of dielectric relaxation of losses in alternative dielectrics," in *IEDM Tech. Dig.*, 2001, pp. 740–743.

[6] J. R. Jameson *et al.*, "Problems with metal-oxide high-k dielectrics due to 1/t dielectric relaxation current in amorphous materials," in *IEDM Tech. Dig.*, 2003, pp. 431–434.

[7] Z. Xu *et al.*, "A study of relaxation current in high-k dielectric stacks," *IEEE Trans. Electron Devices*, vol. 51, no. 3, pp. 402–408, Mar. 2004.

[8] G. Groeseneken *et al.*, "A reliable approach to charge pumping measurements in MOS transistors," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 42–53, 1984.

[9] G. Van den Bosch, G. Groeseneken, and H. E. Maes, "On the geometric component of charge-pumping current in MOSFETs," *IEEE Electron Device Lett.*, vol. 14 , no. 3, pp. 107–109, Mar. 1993.

[10] B. Guillaumot *et al.*, *IEDM Tech. Dig.*, 2002, pp. 355–358.

[11] C. D. Young *et al.*, "Charge trapping in MOCVD Hafnium-based gate dielectric stack structures and its impact on device performance ," in *Integrated Reliability Workshop Final Report*, 2003, pp. 28–35.

[12] L. Pantisano *et al.*, "Dynamics of threshold voltage instability in stacked high-k dielectrics: Role of the interfacial oxide," in *Symp. VLSI Technology Dig. Tech. Papers*, 2003, pp. 163–164.

[13] S. Zafar, A. Callegari, E. Gusev, and M. V. Fischetti, "Charge trapping in high-k gate dielectric stacks," in *IEDM Tech. Dig.*, 2002, pp. 517–520.

[14] G. Ribes *et al.*, "Characterization of Vt instability in hafnium based dielectrics by pulse gate voltage techniques," in *Proc. ESSDERC*, 2004, pp. 89–92.

[15] A. Morioka *et al.*, "High mobility MISFET with low trapped charge in HfSiO films," in *Symp. VLSI Technology Dig. Tech. Papers*, 2003, pp. 165–166.

[16] A. Kerber *et al.*, "Origin of the threshold voltage instability in SiO_2/HfO_2 dual layer gate dielectrics," *IEEE Electron Device Lett.*, vol. 24, no. 2, pp. 87–89, Feb. 2003.

[17] T. H. Hou *et al.*, "Direct determination of interface and bulk traps in scaled HfO_2 dielectrics using charge pumping method," in *Proc. Int. Reliability Physics Symp.*, 2004, pp. 581–582.

[18] M. R. Visokay *et al.*, "Application of HfSiON as a gate dielectric material," *Appl. Phys. Lett.*, vol. 80, p. 3183, 2002.

[19] A. Shanware *et al.*, "Characterization and comparison of the charge trapping in HfSiON and HfO_2 gate dielectrics," in *IEDM Tech. Dig.*, 2003, pp. 939–942.

[20] E. P. Gusev *et al.*, "Charge trapping in aggressively scaled metal gate/high-*K* stacks," in *IEDM Tech. Dig.*, 2004, pp. 403–405.

[21] H.-H. Tseng *et al.*, "ALD HfO_2 using heavy water (D_2O) for improved MOSFET stability," in *IEDM Tech. Dig.*, 2003, pp. 411–414.

[22] A. S. Foster, F. L. Gejo, A. L. Shluger, and R. M. Nieminen, "Vacancy and interstitial defects in hafnia," *Phys. Rev. B*, vol. 65, p. 174117, 2002.

[23] A. Y. Kang, P. M. Lenahan, and J. F. Conley, "Electron spin resonance observation of trapped electron centers in atomic-layer-deposited hafnium oxide on Si," *Appl. Physics Lett.*, p. 8316, Oct. 2003.

[24] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, *J. Electrochem. Soc.*, vol. 114, p. 266, 1967.

[25] D. Frohman-Bentchkowsky, "A fully decoded 2048-bit electrically programmable FAMOS read-only memory," *IEEE J. Solid-State Circuits*, vol. 6, no. 5, pp. 301–306, Oct. 1971.

[26] K. Uwasawa, T. Yamamoto, and T. Mogami, "A new degradation mode of scaled p+ polysilicon gate pMOSFETs induced by bias temperature (BT) instability," in *IEDM Tech. Dig.*, 1995, pp. 871–874.

[27] G. La Rosa *et al.*, "NBTI-channel hot carrier effects in PMOSFETs in advanced CMOS technologies," in *Proc. Int. Reliability Physics Symp.*, 1997, p. 282.

[28] K. Onishi *et al.*, *IEDM Tech. Dig.*, 2001, p. 659.

[29] M. Denais *et al.*, "New methodologies of NBTI characterization eliminating recovery effects," in *Proc. ESSDERC*, 2004, p. 265.

[30] M. Denais *et al.*, "On-the-fly characterization of NBTI in ultrathin gate-oxide PMOSFETs," in *IEDM Tech. Dig.*, 2004, pp. 109–111.

[31] A. Shanware *et al.*, "Evaluation of positive bias stress stability in HfSiON gate dielectrics," in *Proc. Int. Reliability Physics Symp.*, 2003, pp. 208–213.

[32] V. Huard, F. Monsieur, G. Ribes, and S. Bruyère, "Evidence for hydrogen-related defects during NBTI stress in p-MOSFETs," in *Proc. Int. Reliability Physics Symp.*, 2003, pp. 178–182.

[33] T. Kauerauf *et al.*, "Toward understanding degradation and breakdown of SiO_2/high-k stacks," in *IEDM Tech. Dig.*, 2002, pp. 521–524.

[34] X. Garros *et al.*, "Reliability assessment of ultrathin HFO_2 oxides with TiN gate and polysilicon-N+ gate," in *Proc. Int. Reliability Physics Symp.*, 2004, pp. 176–180.

[35] P. W. Anderson *et al.*, *Phys. Rev. Lett.*, vol. 34, p. 953, 1975.

[36] A. Y. Kang *et al.*, "An electron spin resonance study of interface defects in atomic layer deposited hafnium oxide on Si,," *Appl. Phys. Lett.*, vol. 84, p. 3894, 2004.

[37] S. J. Lee, C. H. Lee, C. H. Choi, and D. L. Kwong, "Time-dependent dielectric breakdown in poly-Si CVD Hf02 gate stack," in *Proc. Int. Reliability Physics Symp.*, 2002, pp. 402–414.

[38] B. Kaczer *et al.*, "Impact of MOSFET oxide breakdown on digital circuit operation and reliability," in *IEDM Tech. Dig.*, 2000, pp. 553–556.

[39] S. Bruyère, E. Vincent, and G. Ghibaudo, "Quasibreakdown in ultrathin SiO_2 films: Occurrence characterization and reliability assessment methodology," in *Proc. Int. Reliability Physics Symp.*, 2000, pp. 48–54.

[40] J. Suñé, E. Y. Wu, D. Jiménez, R. P. Vollertsen, and E. Miranda, "Understanding soft and hard breakdown statistics, prevalence ratios and energy dissipation during breakdownrunaway," in *IEDM Tech. Dig.*, 2001, pp. 611–614.

[41] M. A. Alam, B. E. Weir, P. J. Silverman, Y. Ma, and D. Hwang, "The statistical distribution of percolation resistance as a probe into the mechanics of ultrathin oxide breakdown," in *IEDM Tech. Dig.*, 2000, p. 529.

[42] E. Miranda and J. Suñé, "Analytic modeling ofleakage current through multiple breakdown paths in SiO_2 films," in *Proc. Int. Reliability Physics Symp.*, 2001, pp. 367–379.

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.

[43] F. Monsieur, E. Vincent, G. Pananakakis, and G. Ghibaudo, "Wear-out, breakdown occurrence and failure detection in 18–25 Å ultrathin oxides," *Microelectron. Reliab.*, vol. 41, pp. 1035–1039, 2001.

[44] J. S. Suehle, E. M. Vogel, B. Wang, and J. B. Bernstein, "Temperature dependence of soft breakdown and wear-out in sub-3 nm SiO_2 films," in *Proc. Int. Reliability Physics Symp.*, 2000, pp. 33–39.

[45] F. Monsieur, E. Vincent, D. Roy, S. Bruyere, J. C. Vildeuil, G. Pananakakis, and G. Ghibaudo, "A thorough investigation of progressive breakdown in ultrathin oxides. Physical understanding and application for industrial reliability assessment," in *Proc. Int. Reliability Physics Symp.*, 2002, pp. 45–54.

[46] F. Monsieur, E. Vincent, G. Ribes, V. Huard, S. Bruyere, G. Pananakakis, and G. Ghibaudo, "Evidence for defect-generation-driven wear-out of breakdown conduction path in ultra thin oxides," in *Proc. Int. Reliability Physics Symp.*, 2003, pp. 424–431.

[47] Y. H. Kim *et al.*, "Hard and soft-breakdown characteristics of ultrathin HfO_2 under dynamic and constant voltage stress," in *IEDM Tech. Dig.*, 2002, pp. 629–632.

[48] W. Mizubayashi *et al.*, "Carrier separation analysis for clarifying leakage mechanism in unstressed and stressed $HfAlO/SiO_2$ stack dielectrics layers," in *Proc. Int. Reliability Physics Symp.*, 2004, pp. 188–193.

[49] M. Koyama *et al.*, "Degradation mechanism of HfSiON gate insulator and effect of nitrogen composition on the statistical distribution of the breakdown," in *IEDM Tech. Dig.*, 2003, pp. 849–852.

[50] J. Sune, "New physics-based analytic approach to the thin-oxide breakdown statistics," *IEEE Electron Devices Lett.*, vol. 22, no. 6, pp. 296–298, Jun. 2001.

[51] R. Degraeve *et al.*, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," *IEEE Trans. Electron Devices*, vol. 45, no. 4, pp. 904–911, Apr. 1998.

[52] J. H. Stathis, "Percolation models for gate oxide breakdown," *J. Appl. Phys.*, vol. 86, pp. 5757–5766, 1999.

[53] D. J. DiMaria and J. H. Stathis, "Explanation for the oxide thickness dependence of breakdown characteristics of metal-oxide-semiconductor structures," *Appl. Phys. Lett.*, vol. 70, pp. 2708–2710, 1997.

[54] P. Avouris *et al.*, "Breaking individual chemical bonds via STM-induced excitations," *Surf. Sci.*, vol. 363, pp. 368–377, 1996.

[55] G. Ribes *et al.*, "Evidence and modeling current dependence of defect generation probability an dits impact on charge to breakdown," in *WODIM Proc.*, 2004, pp. 40–42.

[56] K. Stokbro *et al.*, "First-principles theory of inelastic currents in a scanning tunneling microscope," *Phys. Rev. B*, vol. 58, no. 12, 1998.

[57] G. Ribes *et al.*, "Breakdown mechanisms in ultrathin oxides: Impact of carrier energy and current through substrate hot carrier stress study," in *Proc. INFOS*, 2003, pp. 40–41.

[58] Y. Harada *et al.*, "Specific structural factors influencing on reliability of CVD-HfO_2," in *Symp. VLSI Technology Dig. Tech. Papers*, 2002, pp. 26–27.

[59] J. W. McPherson *et al.*, "Trends in the ultimate breakdown strength of high dielectric-constant materials," *IEEE Trans. Electron Devices*, vol. 50, no. 8, pp. 1771–1778, Aug. 2003.

[60] R. Degraeve, A. Kerber, P. Roussel, E. Cartier, T. Kauerauf, L. Pantisano, and G. Groeseneken, "Effect of bulk trap density on HfO_2 reliability and yield," in *IEDM Tech. Dig.*, 2003, pp. 3851–3854.

G. Ribes received the engineer degree in electrical engineering from the Institut Supérieur d'Electronique et du Numérique (ISEN), Toulon, France, in 2002. He received the M.S.E.E degree in micro- and nanoelectronics from Provence University, Marseille, France. He is now pursuing the Ph.D degree in the Reliability group of the Central R&D, STMicroelectronics, Crolles, France.

He works in the areas of ultrathin silicon dioxide and high-k dielectrics layers reliability in collaboration with IMEP.

Photographs and biographies of the co-authors not available at the time of publication.

Authorized licensed use limited to: PORTLAND STATE UNIVERSITY. Downloaded on May 22, 2009 at 01:02 from IEEE Xplore. Restrictions apply.