From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation

M. Toledano-Luque[1,2], B. Kaczer[1], J. Franco[1,3], Ph.J. Roussel[1], T. Grasser[4], T.Y. Hoffmann[1], and G. Groeseneken[1,3].
[1]imec, Kapeldreef 75, B3001 Leuven, Belgium. [2]UCMadrid, Spain. [3]KULeuven, Belgium. [4]TUWien, Austria. (toleda@imec.be)

The large FET devices of the past were considered identical in terms of electrical performance. Consequently, the same workload resulted in an identical parameter shift in all devices. As downscaling of FET devices progressed, the gate dielectric was the first to reach nanometer dimensions, thus introducing the first stochastically distributed reliability mechanism—time dependent dielectric breakdown [1]. With the downscaling of *lateral* device dimensions to atomic levels, *variation* among devices appeared due to effects such as random dopant fluctuation [2-4] and line edge roughness [5]. Similarly, application of the identical workload in such devices results in *distributions of the parameter shifts* [6]. Since only a few, randomly behaving gate oxide defects can be expected in future nm-sized CMOS devices, the well-defined Bias Temperature Instability (BTI) lifetime of large devices (Fig. 1a) becomes drastically distributed (Fig. 1b) [7-9]. This paradigm shift is addressed here by demonstrating the methodology to predict the ΔV_{TH} *distributions* after BTI stress—the critical reliability issue in modern CMOS technologies. The methodology is based on a thorough understanding of the *atomistic* impact of *individual* traps.

Both pFETs and nFETs with drawn gate length L = 70nm (metallurgic length L ~ 35nm) and width W = 90nm were used in this study. The gate stack was formed by 0.8 nm-ISSG / 1.8 nm-HfSiO MOCVD (60%Hf) and 5nm-PEALD-TiN. 30 pFETs and 60 nFETs were tested at different stress conditions. The relaxation curves after different stress times and voltages were registered with the eMSM technique [10].

A representative set of typical quantized NBTI relaxation transients from pFETs is shown in Fig. 2. The *total* ΔV_{TH} (ΔV_{TH} at given t_{RELAX}) strongly varies from device to device and recovery occurs through a discrete number of quantized steps ΔV_{TH}. These steps correspond to individual discharge events [9-12]. Fig. 3 shows the complementary distribution (CCDF = 1-CDF) of the individual step heights ΔV_{TH} normalized to the number of tested devices. Step heights follow an exponential distribution [7] with the average step height $\langle \Delta V_{TH} \rangle = \eta$ equal to 3.4 mV, independent of stress conditions. Note that a single charged defect can cause up to tens of mV of ΔV_{TH}, which is much larger than the value predicted by the simple charge sheet approximation η_0 ~ 1.7 mV (η/η_0 = 2.0 in our case). This is due to the amplifying effect of the random dopants in the FET channel [13-14]. The number of detected steps increases with stress time (Fig. 3a) and stress voltage (Fig. 3b). The total number N_T can be obtained from a maximum likelihood fit of the data with eq. 4. Fig. 4 shows that N_T follows a power-law voltage dependence and can be fitted with both power law (eq. 9) and logarithmic (eq. 10) time dependences. The number of steps per device is Poisson distributed (Fig. 5, eq. 7). Fig. 6 gives the *total* ΔV_{TH} for pFETs for different (a) t_{STRESS} and (b) V_{STRESS}. The *total* ΔV_{TH} distribution $H_{\eta,NT}(\Delta V_{TH})$ (eq. 8) [15,16], a combination of the exponential discrete ΔV_{TH} step distribution and the Poisson distribution with average N_T, is traced in Fig. 6 for η = 3.4 mV and different values of N_T. Note that the lines in Fig. 6 follow the experimental *total* ΔV_{TH} data, and the N_T values given by eq. 8 excellently match those obtained *independently* in Figs. 4 and 5 (see symbols *, †, ‡, §, #), thus confirming the description derived in Table I.

A 10-year lifetime CDF prediction of the *total* ΔV_{TH} is obtained by combining eq. 8 with eq. 9 or 10 (N_T dependences on t_{STRESS} and V_{STRESS}). Fig. 7 shows the predicted lifetimes for different conditions. For failure criteria ΔV_{TH} = 30, 50, and 100 mV at t_{STRESS} = 10 years, Fig. 7a allows one to readily read off the fraction of devices expected to exceed a given failure criterion. As already alluded to in Fig. 1, the predicted ΔV_{TH} distribution gets steeper ("narrower") with increasing device area A (Fig. 7b). Since the *average total* ΔV_{TH} is given by $N_T \times \eta$ and $N_T \propto A$, the median *total* $\langle \Delta V_{TH} \rangle$ is constant with A if $\eta \propto 1/A$ [12]. Therefore, Probit($H_{\eta,N}$) = 0 determines the maximum overdrive for large, i.e., deterministic devices (vertical line in Fig. 7b). In contrast to that a considerable fraction of deeply scaled devices will exceed failure criteria even at lower overdrives (see e.g. circles in Fig. 7b). In Fig. 7c and 7d, the impact of N_T and η on the lifetime is explored. Fig. 7c shows that a reduction of the trap density N_T stretches out the overdrive (horizontal) axis, but the maximum fraction of working devices does not improve significantly at lower overdrives. On the other hand, *a reduction of the* η *value shifts the lifetime prediction vertically*, boosting the number of working devices to high percentages over the whole overdrive range. The largest gains in reliability can thus be achieved by moving to technologies with reduced dopant concentration N_A in the channel, see eq. 1 (Table I) [13,17].

An analogous analysis was performed on nFETs after positive gate bias stress. Fig. 8 shows the CCDF for the step heights obtained from 60 nFETs. Remarkably, CCDF follows a *bimodal* distribution with η_{IL} = 3.7 mV and η_{HK} = 0.85 mV. As shown in eq. 1, the η value is related to the charge centroid $\langle x \rangle$: distance of the trapped charges in the dielectric from the gate [14,18]. The lower value of η_{HK} corresponds to the defects in the high-k dielectric. The trap density for η_{HK} is significantly higher, but it has a reduced impact on the *total* ΔV_{TH} *shift* due to the lower η_{HK} value. Note that η_{IL} value is close to the η value obtained from pFETs (Fig. 3), suggesting that it is related to border traps, however, its trap density is 10 times lower w.r.t. pFETs under the same stress conditions. The lower *total* ΔV_{TH} CDF for pFETs w.r.t nFETs under identical stress conditions, Fig. 9a, and the ΔV_{TH} predictions for pFETs and nFETs (cf. Figs. 9b and 7a) confirm that PBTI is a less severe issue than NBTI in deeply scaled devices.

Conclusions: Based on detailed understanding of behavior and statistics of individual defects, we have presented a new methodology to predict the BTI lifetime distributions in deeply scaled FETs. Moreover, we have identified the sources of time dependent variability, some of which can be addressed technologically.

Acknowledgments: M.T.L. thanks the IberCaja Postdoctoral grant program and Spanish MEC (TEC2010-18051) for financial support.

References: [1] Degraeve *et al.*, Microelectron. Reliab. 39, 1445 (1999), [2] Declerck *et al.*, VLSI05, [3] Nagumo *et al.*, IEDM10, [3] Mizumo *et al.*, TED41, 2216 (1994), [5] Asenov *et al.*, TED 50, 1254 (2003), [6] Huard *et al.*, IRPS08, [7] Kaczer *et al.*, IRPS10, [8] Kaczer *et al.*, IRPS2011 (*accepted*), [9] Grasser *et al.*, IEDM10, [10] Kaczer *et al.*, IRPS08, [11] Kirton and Uren, APL48,1270 (1986), [12] Grasser *et al.*, IEDM09, [13] Bukhori *et al.*, TED57, 795 (2010), [14] Ghetti *et al.*, TED56, 1746 (2009), [15] Kaczer *et al.*, EDL31, 411 (2010), [16] Takeuchi *et al.*, VLSI09, [17] Franco *et al.*, IEDM10, [18] Toledano-Luque *et al.*, SISC10.

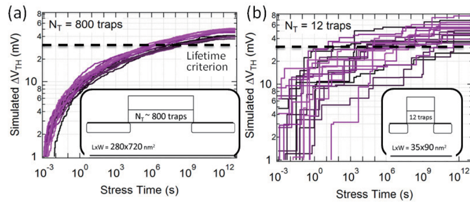

Fig. 1: Illustration of the BTI lifetime paradigm shift [9]. (a) The random properties of many defects in large devices average out, resulting in a well-defined lifetime while (b) the stochastic nature of a handful of defects in deeply-scaled devices becomes apparent, resulting in a large variation in lifetime. **pFET** parameters extracted hereafter were used.

SINGLE DEFECT: ΔV_{TH} exponentially distributed with $\eta = \langle \Delta V_{TH} \rangle = \frac{\langle x_0 \rangle \sqrt{N_A}}{WL}$ (eq. 1)

$$f_\eta(\Delta V_{TH}) = \frac{1}{\eta} e^{-\frac{\Delta V_{TH}}{\eta}} \quad \text{(eq. 2)} \qquad F_\eta(\Delta V_{TH}) = 1 - e^{-\frac{\Delta V_{TH}}{\eta}} \quad \text{(eq. 3)} \qquad \frac{1 - F_\eta(\Delta V_{TH})}{\#devices} = N_T e^{-\frac{\Delta V_{TH}}{\eta}} \quad \text{(eq. 4)}$$

DEVICE: *Total* ΔV_{TH} convolution of n individual exponential distributions = n traps

$$g_{\eta,n}(\Delta V_{TH}) = \frac{\Delta V_{TH}^{\,n-1}}{\eta^n (n-1)!} e^{-\frac{\Delta V_{TH}}{\eta}} \quad \text{(eq. 5)} \qquad G_{\eta,n}(\Delta V_{TH}) = 1 - \frac{n}{n!}\Gamma(n, \Delta V_{TH}/\eta) \quad \text{(eq. 6)}$$

CHIP: Traps Poisson distributed with $\langle n \rangle = N_T$

$$P_{N_T}(n) = \frac{e^{-N} N_T^{\,n}}{n!} \quad \text{(eq. 7)}$$

Total ΔV_{TH} cumulative distribution in a chip is the sum up of $G_{\eta,n}$ weighted by P_{N_T}

$$\boxed{H_{\eta,N_T}(\Delta V_{TH}) = \sum_{n=0}^{\infty} \frac{e^{-N} N_T^{\,n}}{n!} G_{\eta,n}(\Delta V_{TH})} \quad \text{(eq. 8)}$$

Table I: Flow to deduce the *total* ΔV_{TH} *shift distribution.*

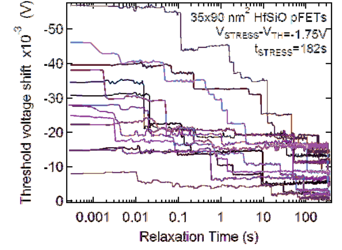

Fig. 2: **NBTI** relaxation transients obtained on 35×90 nm² HfSiO **pFETs**. Steps due to single-carrier discharge events are evident.

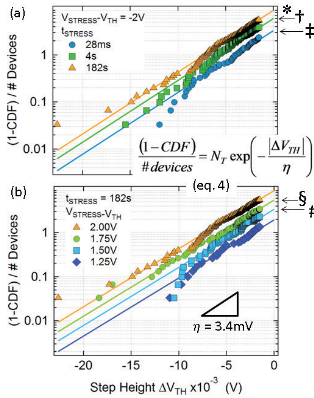

Fig. 3: Complementary cumulative distributions (CCDF=1-CDF) of step heights due to single oxide defects normalized to the number of tested **pFETs** after **negative** stress follow an exponential distribution (eq. 4) with the average step height η = 3.4mV. N_T values can be read from the intersection of the fit with the y-axis (ΔV_{TH} = 0).

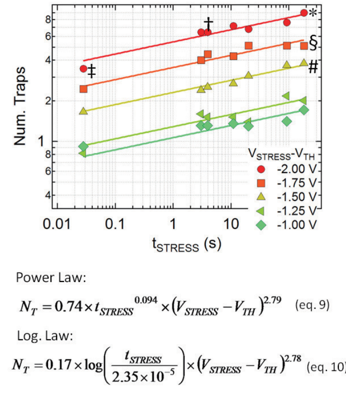

Fig. 4: The number of active traps per device N_T obtained from the fit of the CCDFs shown in Fig. 3 with eq. 4 (Intersection of CCDF with ΔV_{TH} = 0). Note that N_T increases with stress time and voltage. Data can be fitted with both a power-law and logarithmic time dependences.

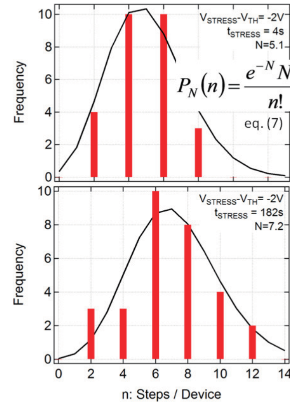

Fig. 5: Histogram of the number of steps per device detected (ΔV_{TH} > 1.5 mV) from the relaxation curves for 30 **pFETs** following different stress times. Steps per device n are Poisson distributed (eq. 7). The average value N increases with increasing stress time.

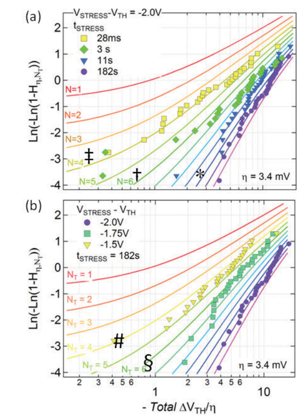

Fig. 6: (Symbols) Cumulative distributions of the *total* ΔV_{TH} normalized to η = 3.4 mV for 30 **pFETs** after stress (a) at different voltages and (b) for different times shown in Weibull plots. (Lines) *Total* ΔV_{TH} *CDFs* for different N_T values from eq. (8) match excellently the experimental data.

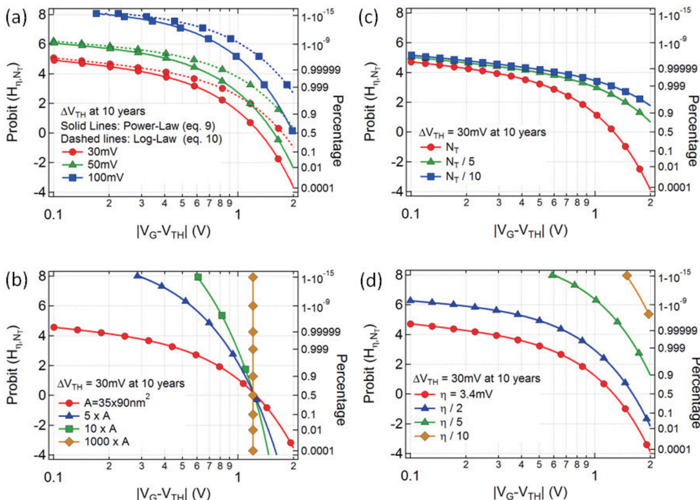

Fig. 7: Predicted 10-year lifetime cumulative distributions of the *total* **pFET** ΔV_{TH} at t_{RELAX} ~ 1 ms. For different failure criteria (a), a slightly more optimistic prediction is given by a logarithmic time dependent law. For different device areas (b), it is observed that the median *total* ΔV_{TH} is independent of area. Significant fraction of deeply scaled devices exceeds failure criteria at lower overdrives. For different trap densities (c), CDF stretches out. For different η values (d), a significant boost of working devices is obtained.

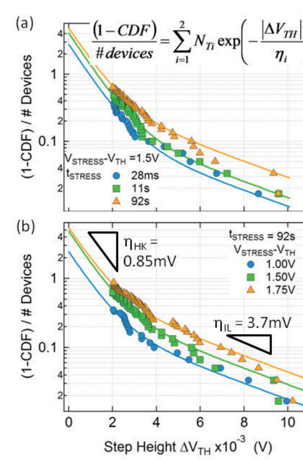

Fig. 8: (Symbols) CCDF of step heights normalized to the number of tested **nFETs** after **positive** stress. Data can be fitted with a *bimodal* distribution with η_{IL} = 3.7 mV and η_{HK} = 0.9 mV. Note that η_{IL} is similar to the η value obtained in pFETs.

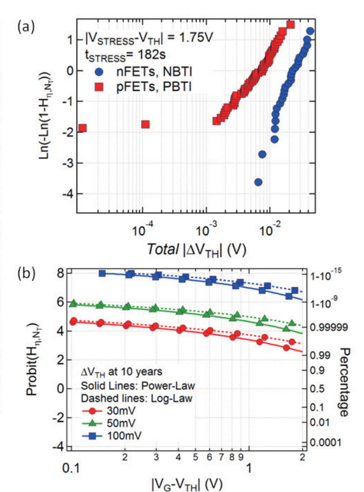

Fig. 9: (a) CDFs of the *total* ΔV_{TH} for **pFETs** and **nFETs** following identical stress conditions. (b) Predicted10-year lifetime CDF of the total **nFET** ΔV_{TH} indicates that **PBTI** is a less severe issue than NBTI (see Fig. 7a) in deeply scaled devices.