Statistics of Multiple Trapped Charges in the Gate Oxide of Deeply Scaled MOSFET Devices—Application to NBTI

B. Kaczer, Ph. J. Roussel, T. Grasser, *Senior Member, IEEE*, and G. Groeseneken, *Fellow, IEEE*

***Abstract*—The statistical distribution of negative bias temperature instability (NBTI) in deca-nanometer p-channel FETs is discussed. An exponential distribution of threshold voltage shifts due to a single charge trapped in the gate oxide is observed, resulting in single-charge shifts exceeding 30 mV in some of the studied 35-nm-long and 90-nm-wide devices. The exponential distribution is justified with a simple channel percolation model. Combined with the assumption of the Poisson-distributed number of trapped gate oxide charges, an analytical description of the total NBTI threshold voltage shift distribution is derived. This allows, among other things, linking its first two moments with the average number of defects per device, which is found < 10 in the studied devices.**

***Index Terms*—MOSFETs, negative bias temperature instability (NBTI), reliability, variability.**

I. Introduction

As CMOS devices scale toward atomic dimensions, device parameters become statistically distributed. Similarly, parameter *shifts* during device operation, once studied in terms of the average value only, will have to be described in terms of their probability density function (PDF) or cumulative density function (CDF).

The statistical aspects of the negative bias temperature instability (NBTI), which is a critical reliability issue in modern CMOS technologies, have been already investigated in downscaled devices [1]–[4]. The threshold voltage shift $\Delta V_{\rm th}$ variation following NBT stress has been described analytically by Skellam's distribution [1], [2] or calculated in computationally intensive atomistic device simulations [3]. In this letter, we show that the NBTI $\Delta V_{\rm th}$ distribution can be described analytically in an intuitive and compact manner if two simple assumptions are considered: First, much like in the case of random telegraph noise (RTN) [5]–[7], we observe the distribution of $\Delta V_{\rm th}$ due to *individual* discharging events to be exponentially distributed. This distribution is confirmed by a simplified channel percolation model [8]. Second, the number of defects in each device is Poisson distributed [1]–[4]. We find that, on average, less than ten oxide-trapped charges are responsible for NBTI in our devices and stress conditions.

Manuscript received January 17, 2010. Date of publication April 5, 2010; date of current version April 23, 2010. This work was supported by the IMEC Industrial Affiliation Program, funded by IMEC's core partners: Intel, Texas Instruments, Micron, Infineon, NXP, ST, Panasonic, TSMC, Samsung, and Elpida. The review of this letter was arranged by Editor J. Cai.

B. Kaczer and Ph. J. Roussel are with IMEC, 3001 Leuven, Belgium (e-mail: kaczer@imec.be; roussel@imec.be).

T. Grasser is with TU Wien, 1040 Vienna, Austria (e-mail: grasser@iue.tuwien.ac.at).

G. Groeseneken is with IMEC, 3001 Leuven, Belgium, and also with ESAT, KU Leuven, 3001 Leuven, Belgium (e-mail: groes@imec.be).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LED.2010.2044014

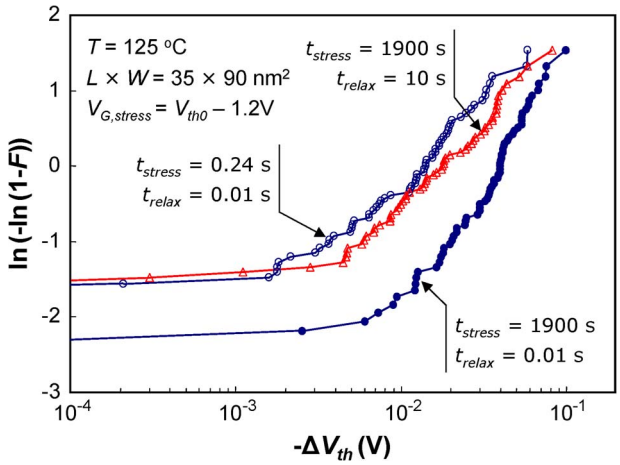

Fig. 1. Cumulative distribution of $\Delta V_{\rm th}$ for 72 pFETs following stress and relaxation at indicated times, shown in a Weibull plot. With ($\circ \rightarrow \bullet$) increasing $t_{\rm stress}$, the average $\langle \Delta V_{\rm th} \rangle$ increases, while the fraction of devices with negligible $\Delta V_{\rm th}$ decreases. The opposite trends are observed during ($\bullet \rightarrow \Delta$) relaxation following the longest stress.

II. Experimental

pFETs with metallurgic length $L = 35$ nm, width $W = 90$ nm, and HfO_2 dielectrics with $EOT = 0.8$ nm were used. To account for the considerable variability in these aggressively scaled devices, each DUT was stressed at V_G 1.2 V below its initial threshold voltage $V_{\rm th0}$ using the extended measure–stress–measure sequence ($T = 125$ °C) [4]. The resulting $\Delta V_{\rm th}$ was found to be uncorrelated with the initial $V_{\rm th0}$, confirming that the NBTI mechanism is decoupled from the sources of the $V_{\rm th0}$ variation and can be studied separately.

III. Results and Discussion

The distribution of $\Delta V_{\rm th}$ following NBTI stress at several stress and relaxation times is shown in Fig. 1. As expected, the mean of the distribution increases with the stress. Perhaps surprisingly, Fig. 1 also shows that a fraction of devices exists with negligible $\Delta V_{\rm th} \leq\sim 1$ mV even after the longest stress. As will become apparent, this is a direct consequence of the two assumptions given before. The opposite trends are observed in Fig. 1 when the devices are left to relax after the longest stress.

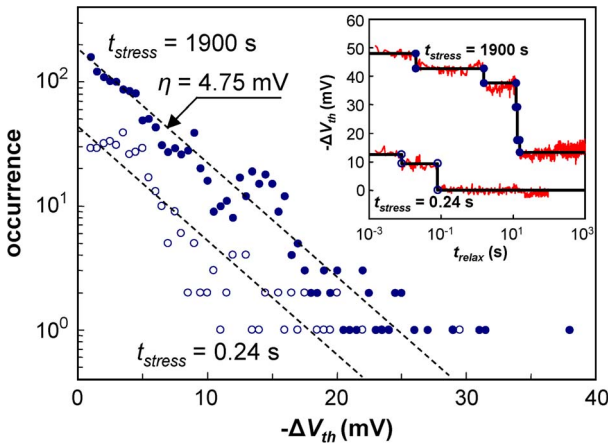

Fig. 2. Histograms of NBTI transient step heights following two stress times for 72 pFETs show an exponential distribution (1). (Inset) Example of two NBTI relaxation transients corresponding to the same stress times. Individual steps, as determined by the step-detection algorithm, are delineated. Only down steps, overwhelmingly prevalent at $t_{relax} < 10^2$ s, are counted.

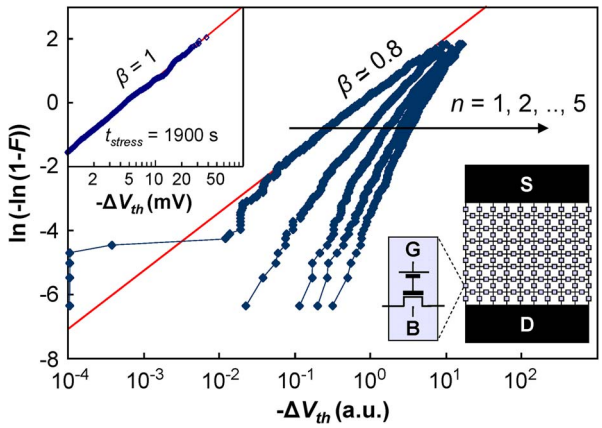

Fig. 3. Cumulative $\Delta V_{\rm th}$ distributions generated for an increasing number of gate oxide defects n by the simplified channel percolation model shown schematically in the inset. For a single-charged defect ($n = 1$), the model well reproduces the observed exponential distribution from Fig. 2, now plotted as a CDF in the inset.

An example of the NBTI relaxation transients collected on one device is shown in the inset of Fig. 2. As already reported [2], [4], abrupt $\Delta V_{\rm th}$ steps caused by *single-hole* discharge events are visible in the transients. The down steps larger than 1 mV were detected in relaxation traces in all measured pFETs, and a histogram of the step heights was constructed. The histogram in Fig. 2 demonstrates that these NBTI relaxation steps are exponentially distributed, with their PDF being

$$f_1(\Delta V_{\rm th}, \eta) = \frac{1}{\eta} e^{-\frac{\Delta V_{\rm th}}{\eta}} \tag{1}$$

where η is the mean $\Delta V_{\rm th}$ value for a single charge. The variance of this distribution is $\sigma^2 = \eta^2$. The CDF corresponding to (1) is then

$$F_1(\Delta V_{\rm th}, \eta) = 1 - e^{-\frac{\Delta V_{\rm th}}{\eta}}. \tag{2}$$

The average $V_{\rm th}$ shift η corresponding to a single carrier discharge is 4.75 $\pm$ 0.3 mV in our devices, with several steps in Fig. 2 exceeding 30 mV(!). Note that a simple charge sheet approximation predicts $\Delta V_{\rm th}$ of less than 2 mV. The observed $\Delta V_{\rm th}$ values are significantly larger than those reported for NBTI earlier [2], [4]. As will be discussed hereinafter, the cause lies in the deeply scaled dimensions of the pFETs used.

The exponential distribution of single-charge $\Delta V_{\rm th}$ can be understood if the nonuniformities in the pFET channel due to random dopant fluctuations (RDFs) are considered [5]–[8]. The threshold voltage of such a FET corresponds to carrier energy sufficient to generate a conduction (percolation) path in the random dopant potential between source and drain. To zeroth order, depending on the position of the NBTI-stress-generated oxide charge, the conduction path could be either unaffected or obstructed by the new charged defect. In the latter case, the drop in the current has to be compensated by an increase of the gate voltage, resulting in the observed $\Delta V_{\rm th}$.

The essence of this process can be *qualitatively* captured in a simplified channel percolation model without the need of a full device simulation with RDF [3], [7]. In the model (inset of Fig. 3), a mesh of "elementary" FETs with random $V_{\rm th}$'s, representing variations in the local potential, is set up to represent the channel of our pFETs. For the sake of simplicity, a uniform distribution of the random "elementary" $V_{\rm th}$'s (voltage source in series with gate) is used, and short-channel effects are not considered. A script is used to generate 400 instances of the randomized mesh, to call SPICE to solve them, and to extract the $V_{\rm th0}$'s of the simulated pFETs. The resulting $V_{\rm th0}$'s are normally distributed, and their variance scales reciprocally with the FET area [8].

A number n of "charged defects" are then inserted, each represented by an additional $V_{\rm th}$ shift of one random "elementary" FET in the netlist, and a new $V_{\rm th}$ is calculated, resulting in $\Delta V_{\rm th}$ for each instance. For a single additional charged defect, the simplified model shows the $\Delta V_{\rm th}$ distribution to be Weibull distributed with $\beta = 0.8$ for a range of dimensions of the channel mesh (Fig. 3). This confirms that the previously described process can be responsible for the observed exponential distribution (i.e., Weibull with $\beta = 1$; inset of Fig. 3) of step heights. We also note that the exponential distribution of the step heights has been repeatedly reported for RTN amplitudes [5]–[7].

The simplified model predicts [8] that η scales inversely with both W and L, i.e., the smaller the device, the larger the steps, thus explaining the large observed value of η. A more thorough discussion of the dependence of η on W, L, EOT, and channel doping in the framework of RTN is given in [7], which infers $\eta \sim L^{-1/2}$ for short devices.

The simplified model also predicts the *total* $\Delta V_{\rm th}$ distribution for the number of defects $n > 1$ (Fig. 3). Since the subsequent charge lateral locations are uncorrelated, the overall $\Delta V_{\rm th}$ distribution can be readily expressed as a convolution of individual exponential distributions (1), and the PDF and CDF are respectively described by

$$f_n(\Delta V_{\rm th}, \eta) = \frac{e^{-\frac{\Delta V_{\rm th}}{\eta}}}{(n-1)!} \frac{\Delta V_{\rm th}^{n-1}}{\eta^n} \tag{3}$$

$$F_n(\Delta V_{\rm th}, \eta) = 1 - \frac{\Gamma(n, \Delta V_{\rm th}/\eta)}{(n-1)!}. \tag{4}$$

The CDF in (4) well describes the result in Fig. 3 for $\beta = 1$.

An actual population of stressed devices will consist of devices with a *different* number n of oxide defects in each device. That number will be Poisson distributed [1]–[4]. The

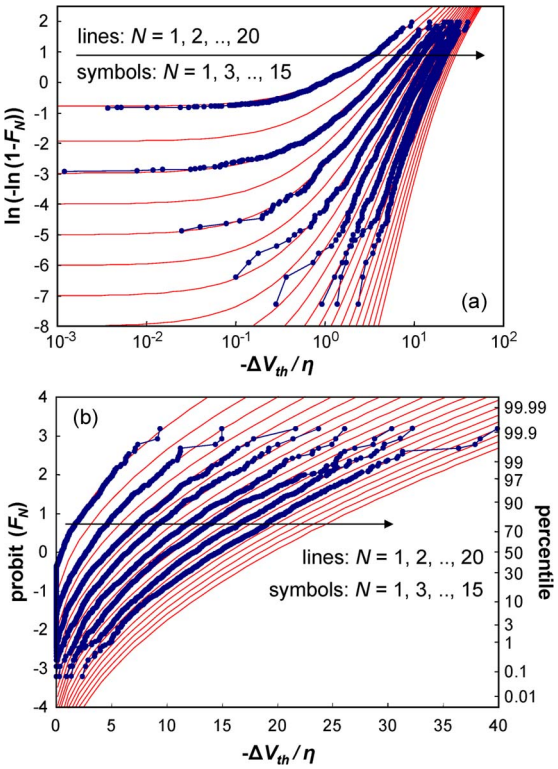

Fig. 4. Equation (6) for several values of the (lines) average number defects N compared with (symbols) a Monte Carlo calculation with 1000 samples. (a) Weibull plot emphasizes the fraction of devices with $\Delta V_{\rm th} \sim 0$ V; cf. Fig. 1. (b) Same results in a probit plot, as shown, for example, in [1]–[3].

total $\Delta V_{\rm th}$ distribution can be therefore obtained by summing the distributions F_n weighted by the Poisson probability

$$P_N(n) = \frac{e^{-N}N^n}{n!}. \tag{5}$$

In (5), N is the mean number of defects in the FET gate oxide and is related to the oxide trap (surface) density $N_{\rm ot}$ as $N = WLN_{\rm ot}$ (note that N is not an integer). This results in a CDF given by

$$F_N(\Delta V_{\rm th}, \eta) = \sum_{n=1}^{\infty} \frac{e^{-N}N^n}{n!} F_n(\Delta V_{\rm th}, \eta). \tag{6}$$

The corresponding PDF is

$$f_N(\Delta V_{\rm th}, \eta) = e^{-N}\left[\delta(\Delta V_{\rm th}) + N\frac{e^{-\frac{\Delta V_{\rm th}}{\eta}}}{\eta} {}_0\mathcal{F}_1\left(2; N\frac{\Delta V_{\rm th}}{\eta}\right)\right] \tag{7}$$

where ${}_0\mathcal{F}_1(2;x)$ is the hypergeometric function. The Dirac $\delta(\Delta V_{\rm th})$ term represents the fraction of devices with 0-V shift [6], which decreases with increasing N.

The CDF of (6) is shown in Fig. 4(a) for several values of N. We can see that it has the same properties as the distributions obtained on the limited population in Fig. 1; specifically, the fraction of devices showing negligible $\Delta V_{\rm th}$ varies with $\langle \Delta V_{\rm th} \rangle$. For completeness, the CDF is also compared with a simple 1000-sample Monte Carlo calculation assuming the following: 1) exponentially distributed individual $\Delta V_{\rm th}$ steps and 2) Poisson distributed in number. Fig. 4(b) then shows the same data in a probit plot, such as that used in [1]–[3].

The mean of the previously derived distribution is

$$\langle \Delta V_{\rm th} \rangle = N\eta \tag{8}$$

i.e., it should be independent of FET gate area WL provided that N and η are directly and inversely proportional to WL, respectively. The variance of the distribution is then

$$\sigma^2 = 2N\eta^2 \tag{9}$$

i.e., it increases with a decreasing gate area.

With the value of η extracted in Fig. 2, we can use (8) to convert $\langle \Delta V_{\rm th} \rangle$ to the average number of trapped defects N. In our devices, N increases from 2.6 ($t_{\rm stress} = 0.24$ s; Fig. 1) to 6.9 ($t_{\rm stress} = 1900$ s) and then decreases to 3.4 ($t_{\rm relax} = 10$ s). These values correspond to the effective trap densities of $1 - 2 \times 10^{11}$ cm^{-2}, typically observed for NBTI in large devices.

Finally, we note that obtaining the value of η as in Fig. 2 could be rather laborious. However, (8) and (9) allow us to express both N and η in terms of $\langle \Delta V_{\rm th} \rangle$ and σ^2, i.e., the first two moments of the measured NBTI distributions. This way, we *independently* obtain N, increasing from 1.9 to 4.6 with stress, and η, varying between 7 and 9 mV. For the limited population of devices measured, these values are very close to those obtained directly by counting individual $\Delta V_{\rm th}$ step heights in Fig. 2.

IV. Conclusion

We have laid out the two main causes of the statistical distribution of NBTI in deeply scaled MOSFETs. The analytical description derived for this distribution should prove useful for both reliability data analysis and simulations of deeply scaled CMOS circuitry.

Acknowledgment

The authors would like to thank E. Simoen, R. Degraeve, J. Franco, L.-Å. Ragnarsson, and H. Reisinger for their input.

References

[1] S. E. Rauch, "Review and reexamination of reliability effects related to NBTI statistical variations," *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 4, pp. 524–530, Dec. 2007.

[2] V. Huard, C. Parthasarathy, C. Guerin, T. Valentin, E. Pion, M. Mammasse, N. Planes, and L. Camus, "NBTI degradation: From transistor to SRAM arrays," in *Proc. Int. Rel. Phys. Symp.*, 2008, pp. 289–300.

[3] M. F. Bukhori, S. Roy, and A. Asenov, "Simulation of statistical aspects of reliability in nano CMOS," in *Proc. Int. Integr. Rel. Workshop*, 2009, pp. 82–85.

[4] B. Kaczer, T. Grasser, J. Martin-Martinez, E. Simoen, M. Aoulaiche, P. J. Roussel, and G. Groeseneken, "NBTI from the perspective of defect states with widely distributed times," in *Proc. Int. Rel. Phys. Symp.*, 2009, pp. 55–60.

[5] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decananometer MOSFETs: 3-D simulation study," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 839–845, Mar. 2003.

[6] K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai, and Y. Hayashi, "Single-charge-based modeling of transistor fluctuations based on statistical measurement of RTN amplitude," in *VLSI Symp. Tech. Dig.*, 2009, pp. 54–55.

[7] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive analysis of random telegraph noise instability and its scaling in deca-nanometer Flash memories," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1746–1752, Aug. 2009.

[8] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," in *Proc. Int. Rel. Phys. Symp.*, 2010, to be published.