Citation	<p>Marko Simicic, Alexandre Subirats, Pieter Weckx, Ben Kaczer, Jacopo Franco, Philippe Roussel, Dimitri Linten, Aaron Thean, Guido Groeseneken, Georges Gielen (2016)</p> <p>Comparative experimental analysis of time-dependent variability using a transistor test array</p> <p>2016 IEEE International Reliability Physics Symposium (IRPS), pp. XT-10.1 - XT-10.6.</p>
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://ieeexplore.ieee.org/document/7574652/?reload=true
Journal homepage	http://irps.org/
Author contact	<p>marko.simicic@esat.kuleuven.be</p> <p>+ 32 (0)16 32 58 54</p>

(article begins on next page)

Comparative experimental analysis of time-dependent variability using a transistor test array

Marko Simicic ^{*(1,2)}, Alexandre Subirats ⁽²⁾, Pieter Weckx ⁽²⁾, Ben Kaczer ⁽²⁾, Jacopo Franco ⁽²⁾, Philippe Roussel ⁽²⁾, Dimitri Linten ⁽²⁾, Aaron Thean ⁽²⁾, Guido Groeseneken ^(1,2), Georges Gielen ⁽¹⁾

⁽¹⁾ Department of Electrical Engineering (ESAT)

KU Leuven

Leuven, Belgium

*email: Marko.Simicic@esat.kuleuven.be

⁽²⁾ imec

Leuven, Belgium

Abstract—As the minimum transistor length reaches the deca-nanometer scale, both time-zero and time-dependent variability, the latter including Random Telegraph Noise (RTN) and Bias Temperature Instability (BTI), become a great concern for IC design. Accurate statistical models describing these two variability sources are therefore necessary in order to design reliable circuits and systems. This paper gives insights in the geometric scaling of these variabilities and analyzes time-dependent variability through three different measurement techniques: 2-point Measure-Stress-Measure, Time-Dependent Defect Spectroscopy, and fine-step I_d - V_g . Advantages and downsides of each technique are discussed and compared.

Index Terms—transistor degradation, RTN, BTI, time-zero variability, time-dependent variability.

I. INTRODUCTION

Nowadays, transistor reliability studies demand a constantly growing sample population size in order to produce statistically relevant data and to enable extraction for time-zero and time-dependent variability. Statistical measurements are even more required for the study of Random Telegraph Noise (RTN) and Bias Temperature Instability (BTI), where the analysis of high percentiles becomes mandatory [1], [2]. For this reason, a large transistor test array has been designed and fabricated in a commercial 28-*nm* planar HKMG technology [3]. Measurements results will be discussed here with focus on the geometric scaling of the variability. For time-dependent variability, three different measurement techniques are compared: the 2-point Measure-Stress-Measure (2-point MSM) method [5] (Fig. 1a), the Time-Dependent Defect Spectroscopy (TDDS) method [6] (Fig. 1b) and the fine-step I_d - V_g method [7] (Fig. 1c). The various advantages and disadvantages of the techniques will be summarized and compared.

Figure 1. Techniques used to characterize the time-dependent variability: (a) 2-point MSM, (b) TDDS and (c) Fine-step I_d - V_g .

II. ARCHITECTURE OF THE TEST ARRAY AND TIME-ZERO VARIABILITY

All measurements in this work have been performed on the test array described in [3]. The geometries of the measured transistors are described in TABLE I and Fig. 2. The array contains 54,432 single devices, which are divided in 12 blocks (6 nMOS and 6 pMOS) of $42 \times 108 = 4,536$ nominally identical devices. A simplified schematic of the array is shown in Fig. 3 and the microphotograph of the chip in Fig. 4.

The measurements were performed with a pair of Keithley 2602 instruments (4 SMUs in total), combined with a custom-made matrix, a pulse generator, and two low-noise DC voltage sources, to provide power and signals to the array. All results presented in this work use integration times of 2 *ms* or more. With this setting, the repeatability of the measurements was confirmed by remeasuring the V_{th0} values. A difference of the

mean value virtually equal to zero and a standard deviation below 1 mV was found.

The uniformity of the array was checked by calculating V_{th0} differences δV_{th0} and their standard deviation $\sigma(\delta V_{th0})$, as a function of distance between the DUTs. As Fig. 5 suggests, no correlations or systematic variability effects as function of separation of the “matched” couples were noticed. Similar analysis of treating neighboring DUTs as matched pairs (MP) is also tested in Section III for the time-dependent variability.

The measured time-zero (as-fabricated) V_{th0} variability of the transistors is presented in the Pelgrom plots [8] of Fig. 6. These results show a rather low variability with respect to other planar non-SOI advanced technologies [9]. The pMOS DUTs exhibit extra variability in minimum-length devices, which can be attributed to the interplay of line edge roughness and the short channel effect, as discussed in [10] and [11].

TABLE I. GEOMETRIES OF THE DUTs IN THE TEST ARRAY (UNITS ARE IN nm).

L×n W	30	90	30×3	150
100	*	*	**	*
200	*			
300	*

Figure 2. The stacked transistor (** in TABLE I.).

Figure 3. The schematic of the array includes Kelvin sensing and thick-oxide (I/O) transistor transmission gates. The calibration structures, which are rows and columns containing no DUT, and shift registers which are used to control the switches, are not shown. Replotted from [3].

Figure 4. Microphotograph of the test chip containing the test array used for this work (top) [3], the individual devices (middle) and the pipelined array (bottom) [4].

Figure 5. The standard deviation of the V_{th0} difference δV_{th0} of neighboring transistors in columns as a function of distance. The leftmost point represents the $\sigma(\delta V_{th0})$ value of first neighbours, eg. column 1 and 2, 2 and 3 and so on.. For each point towards the right, the distance increases by one column, decreasing the sample size. The last point represents the $\sigma(\delta V_{th0})$ of the DUTs from the first and last column. An equivalent result was obtained for rows. No variability change is observed within a block of identical devices, concluding that systematic variability can be neglected. The dimensions in the legend represent the DUT length and width ($L \times W$) in nm $\times$ nm.

Figure 6. Pelgrom plots [8] for the nMOS and pMOS matched pairs. The blue triangles represent measurement data while the red lines represents the fit, which was forced to zero.

III. TIME-DEPENDENT VARIABILITY

In order to observe time-dependent variability caused by BTI and RTN, after initial characterization, the transistors have been stressed for 1000 s at room temperature with $|V_{gs_stress}| = 1.4$ V and $|V_{ds}| = 50$ mV. After stress, the drain current has been measured again, at $|V_{gs_relax}| = 0.55$ V, which is slightly above V_{th0} . Using the initial I_d - V_g characteristic measured on the fresh device, the ΔV_{th} caused by the stress has been extracted. This measurement scheme has been derived from the eMSM measurement [12], except that in this case only two points are measured (hence the name “2-point MSM”) [5].

The test array design allows for multiple devices to be stressed in parallel, but only one at a time can be measured, causing a delay from the first to the last device equal to 2 s. In order to make this delay more uniform across all measured devices, 5 s of additional relaxation were introduced before the start of the measurement. Thus, the first measured device had a total relaxation time of about 5 s while the last about 7 s. Since the relaxation proceeds on a logarithmic time scale, these 2 s of delay represent in this case a smaller error than it would be in the millisecond range.

The measured distributions of threshold voltage shifts ΔV_{th} after stress and the 5-7 s delay, as well as the fitting results using the Defect-Centric Model [12] and according to the fitting methodology described in [5], are shown in Fig. 7. Both the negative BTI (NBTI) tail of the distribution ($-\Delta V_{th} > 0$ V), the smaller RTN tail ($-\Delta V_{th} < 0$ V) and the Gaussian noise ($-\Delta V_{th} \sim 0$ V) are included in the fit. The Gaussian noise mean and spread parameters are 0 and 0.5 mV, respectively. The extracted average impact per gate oxide trap η and the average number of traps per transistor N_T are given in Fig. 7 and 8. The $N_{T,RTN}$ parameter describing the RTN part of the distribution is equal to 10 % of the N_T , while the η for BTI and RTN are taken equal.

We also note that in Fig. 7 the long DUTs have a higher N_T value than the minimum-length DUTs. This can be explained by a lower V_{th0} in the longer devices. Since the applied V_g stress was equal for all DUTs, due to a lower V_{th0} , longer devices experienced a higher overdrive $V_{ov} = V_{gs} - V_{th0}$ and thus a higher effective stress. It is reasonable to expect that higher stress conditions cause more active traps per DUT and thus a higher N_T . However, note that the extracted values of η are not affected. The extracted η and N_T values plotted versus area are shown in Fig. 8a and b respectively.

Since the test array has been designed symmetrically and no spatial variability correlation has been observed (Fig. 5), neighboring devices are treated as matched pairs. This fact allows us to analyze the 2-point MSM ΔV_{th} results as if they come from matched pairs [14], therefore as

$$\delta\Delta V_{th} = \Delta V_{th,L} - \Delta V_{th,R}, \quad (1)$$

where L and R stand for the left and right device of the pair. In this case, the symmetrical $\delta\Delta V_{th}$ distribution was used instead of the asymmetrical ΔV_{th} distribution to compute the spread parameter

$$\sigma_{\Delta V_{th}}^2 = \sigma_{\delta\Delta V_{th}}^2 / 2 \quad (2)$$

Equations used for the η and N_T parameters extraction are:

$$\sigma_{\Delta V_{th}}^2 = 2 \eta < \Delta V_{th} > \quad (3)$$

$$< \Delta V_{th} > = N_T \eta \quad (4)$$

This approach is more robust towards systematic variability. The results are shown in Fig. 9. Again, Gaussian noise with spread equal 0.5 mV was added to the fitted curve.

Figure 7. The distributions of threshold voltage shifts ΔV_{th} fitted with the defect-centric model including RTN and Gaussian noise. The average impact per trap η , the average number of traps per transistor N_T , and the mean ΔV_{th} shift values are reported within the figures for each DUT geometry.

Figure 8. The extracted η (a) and N_T (b) parameters using three different measurement techniques plotted versus area: 2-point MSM for single devices and matched pairs (MP), TDDS and fine-step I_d - V_g .

Figure 9. The data points from Fig. 6 are analysed as matched pairs using the δV_{th} data points, as described in [14]. The average impact per trap η , the average number of traps per transistor N_T and the spread parameter $\sigma(\delta V_{th})$ are reported within the figures for each DUT geometry.

Figure 10. The maximum-likelihood fit of the bimodal defect-centric distribution on the minimum-size pMOS δV_{th} data. Note that almost all data points are beneath the fitting curve, so only a few points in the tail are visible on the plot.

Figure 11. a) The V_{th} standard deviation at time-zero $\sigma(V_{th0})$ (extracted from Fig. 6 by dividing the $\sigma(\delta V_{th0})$ values by $\sqrt{2}$) and the time-dependent variability $\sigma(\delta V_{th})$ for two different $\langle \Delta V_{th} \rangle$ values are compared similarly as in [15, 16]. b) The time-zero and the time-dependent variances are compared.

The extracted parameters (N_T , η) obtained with this method do not differ significantly from the ones in Fig. 7. This, again, confirms the negligible systematic variability of the array.

Because the η and N_T values extracted both in Fig. 7 and Fig. 9 could be in principle susceptible to outliers, a robust maximum-likelihood fit of the defect-centric distribution to the minimum-size device δV_{th} data was also designed, and is illustrated in Fig. 10.

The comparison of η and N_T is plotted in Fig. 8a and b respectively. Time-dependent variability shows a significant impact compared to time-zero variability [15, 16]. In Fig. 11, time-zero variability and two projections of time-dependent variability for a mean ΔV_{th} equal to 5 and 50 mV are compared. Equation (3) was used to calculate these projections. The results show a significant impact of the time-dependent variability which, at the end-of-life criterion ($\Delta V_{th} = 50$ mV), is expected to be comparable with the time-zero variability.

IV. COMPARISON OF THE EXTRACTION OF (N_T , η) PARAMETERS WITH THE DIFFERENT TECHNIQUES

In current reliability studies, different techniques exist to extract the two key parameters describing the number of gate oxide traps and their impacts (N_T , η). Here, we focus on the TDDS technique (Fig. 1b) [6] and the fine-step I_d - V_g technique (Fig. 1c) [7]. The N_T and η values extracted with the 2-point MSM method are compared to those obtained with these two other techniques. In both cases, smaller sample sizes are used due to the longer measurement times required, 108 samples per DUT geometry for the fine-step I_d - V_g method and 42 samples for the TDDS method. A full comparison of the three measurement techniques (both advantages and drawbacks) is given in TABLE II.

A. The TDDS method

It is important to note that in this work only a single relaxation trace was measured on each device, while a full TDDS measurement normally requires multiple relaxation curves. In any case, when using this technique all delays are reduced to a minimum, 1 ms in this work. The total measurement time was 5 s and the stress time 1,000 s. This is graphically illustrated in Fig. 12. A drawback is that only 1 device per array row was measured while the remaining 107 were stressed, but not measured. That was to avoid measuring prestressed devices. One of the advantages of this technique is that the impact of the traps is monitored directly, but at a cost of having a complex data analysis involving step detection from noisy relaxation traces and avoiding double-counting of the defects causing RTN [14]. TDDS measurements performed on 42 devices allowed to extract the (N_T , η) parameters for all available transistor geometries. The results are given in Fig. 8.

The N_T and η parameters extracted by the TDDS technique are different from those obtained by the 2-point MSM method. Indeed, the N_T values extracted with the TDDS technique are lower. The reason for that is the short measurement time window of only 5 s, thus only the active traps which emitted during that window were observed (Fig. 12.) The difference in η is explained by the statistical error related to the small DUT sample size (42) that could be measured with this characterization-intensive technique.

B. The fine-step I_d - V_g method

This is the third technique used to extract the N_T and η parameters. The highlight of this technique is the absence of measurement delay dependency, as the DUT is never pre-stressed. Furthermore, it also provides capture and emission time constants (τ_c and τ_e) with the voltage dependence included. It is also compatible with full transistor arrays. On the other hand, compared to the 2-point MSM method, the analysis more complex and the throughput is lower by about 5 times for the measurement setup used in this work.

The η and N_T parameters extracted with this technique are shown in Fig. 8. The N_T extracted by the model are extremely low. This is expected as this technique is “passive” (same as standard RTN measurement), i.e., it does not activate (charge) additional traps by pre-stress and relies on the chance of a trap appearing in the measurement window (Fig. 12). Furthermore we notice an opposite trend of the variation of N_T with the transistor size (decreasing with the area, which is incorrect). This is explained by the low sensitivity of this technique for the traps with low impact that are commonly present in larger devices. Since the low-impact traps are screened, this also explains the higher extracted η values.

C. Considerations on the time dependency of the observed traps using the different measurement techniques

When measuring the effect of oxide traps on the V_{th} , it is important to understand the time dependence of the traps. This time dependence is well summarized by the capture-emission time map (CET map) [17]. Fast measurements, like the TDDS method in this paper, are able to observe traps which emit captured charges already in the millisecond range or even faster. This is possible because the delay between the stress and the measurement phase is only limited by the measurement equipment. In this work the TDDS measurement was stopped

Figure 12. Graphical representation of the measurement time windows for the three methods used in this work. The x-axis represents the relaxation time of the charged traps while the y-axis represents the stress time. The traps with longer or shorter capture and/or emission times (τ_c and/or τ_e) that fall outside of the measurement windows cannot be observed.

at 5 seconds. In Fig. 12 the block named TDDS depicts all the traps this measurement could have observed.

The 2-point MSM method could achieve the same short delay as the TDDS method, but not if several stressed devices are measured sequentially, like in our test array. This is a simple throughput versus accuracy tradeoff. Introducing an additional delay, 5 s in this work, the measured data have more uniform delays, but the fast-relaxing traps cannot be observed anymore and thus do not appear in the results. Timewise, the advantage of this method is that it indirectly accounts for all charged traps with emission times slower than the applied delay.

If the fine-step I_d - V_g method is used to extract the (N_T , η) couples, the timing is not a real concern since the DUT is never stressed like in the 2-point MSM or the TDDS method. The ΔV_{th} information is extracted directly from the I_d - V_g sweep. Timing becomes a concern only if the capture and emission time constants of the active traps (τ_c , τ_e) need to be measured. A measurement time window shown in Fig. 12 was estimated from the integration time of 20 ms as described in [18].

TABLE II. SUMMARIZED COMPARISON OF THE THREE RTN/BTI MEASUREMENT TECHNIQUES.

	2-point MSM	TDDS	Fine-step I_d - V_g
Throughput	High	Low; can be increased with specialized array designs ^a	Intermediate; depends on measurement conditions
Analysis complexity	Low	High	High
Parallelization	High, but with inherent delay	Intermediate; can be increased with specialized array designs ^a	High
Accuracy	Intermediate ^b	High	Dependent on measurement conditions
Extracted parameters	η	Yes	Yes
	N_t	Yes	Yes
	τ_c ^c	No	Yes, if multiple measurements
	τ_e ^c	No	Yes, if multiple measurements
	τ_{nV} ^c	No	Yes, if multiple measurements

^a All arrays where every device is completely isolated from the rest, e.g. [4, 15].

^b Depends on the measurement equipment. Normally, the longer the integration time, the higher the accuracy.

^c The τ_c , τ_e and τ_{nV} parameters are the trap mean capture time, mean emission time and time-constant gate voltage scaling.

In the case no trap is active, the measured I_d - V_g values are like those of a trap-free transistor. If increasing the V_g activates the trap, which is often the case, a shift in the I_d - V_g curve will be measured. Most of the capture and emission times scale with V_g , so in this kind of measurement it is only important to find the V_g range where the trap is active. In case a trap is in the metastable state ($\tau_c \sim \tau_e$), the measured values depend only on the integration time and not the actual delay. In this state, it is possible to measure 1) the captured state, 2) the relaxed state or 3) the average. According to [18], using an integration time equal to 20 ms this method can observe traps in the metastable state with τ_c and τ_e in the range from 100 ms to 5 s. Refining the V_g step size helps to distinguish to which state the measured values belong by providing more data points, but also prolonging the total measurement time.

V. CONCLUSION

In this paper both the time-zero and the time-dependent variability of an advanced HKMG CMOS technology have been studied on an extensive number of transistors in a test array environment. The two key parameters, the average number of oxide traps per transistor N_T and the trap average impact η , used in the description of the ΔV_{th} distribution due to BTI degradation and RTN, have been extracted using three different techniques. Compared to the time-zero variability, the time-dependent variability has been shown to be relevant. Discussing the three techniques, the fine-step I_d - V_g has been shown to be more comprehensive, but inaccurate for transistors with multiple traps or traps with a low impact. The TDDS method has been proven to be the most accurate, but laborious compared to the 2-point MSM method. The latter, combined with the defect-centric model, gives good estimates for both N_T and η , provided that the measurement delay can be minimized.

REFERENCES

- [1] M. Luo, R. Wang, S. Guo, J. Wang, J. Zou and R. Huang, "Impacts of Random Telegraph Noise (RTN) on Digital Circuits," in IEEE Transactions on Electron Devices, vol. 62, no. 6, pp. 1725-1732, June 2015.
- [2] P. Weckx, B. Kaczer, M. Toledano-Luque, P. Raghavan, J. Franco, P.J. Roussel, G. Groeseneken, F. Catthoor, "Implications of BTI-Induced Time-Dependent Statistics on Yield Estimation of Digital Circuits," in IEEE Transactions on Electron Devices, vol. 61, no. 3, pp. 666-673, March 2014.
- [3] M. Simicic, V. Putcha, B. Parvais, P. Weckx, B. Kaczer, G. Groeseneken, G. Gielen, D. Linten, and A. Thean, "Advanced MOSFET Variability and Reliability Characterization Array", presented at IEEE International Integrated Reliability Workshop (IIRW), (2015).
- [4] V. Putcha, M. Simicic, P. Weckx, B. Parvais, J. Franco, B. Kaczer, D. Linten, D. Verkest, A. Thean, and Guido Groeseneken, "Smart-array for pipelined BTI characterization", presented at IEEE International Integrated Reliability Workshop (IIRW), (2015).
- [5] P. Weckx, B. Kaczer, C. Chen, J. Franco, E. Bury, K. Chanda, J. Watt, Ph. J. Roussel, F. Catthoor, and G. Groeseneken, "Characterization of time-dependent variability using 32k transistor arrays in an advanced HK/MG technology", Reliability Physics Symposium (IRPS), 2015 IEEE International, Monterey, CA, 2015, pp. 3B.1.1-3B.1.6.
- [6] T. Grasser, H. Reisinger, P. J. Wagner, F. Schanovsky, W. Goes and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," Reliability Physics Symposium (IRPS), 2010 IEEE International, Anaheim, CA, 2010, pp. 16-25.
- [7] M. Toledano-Luque, R. Degraeve, P.J. Roussel, V.S. Luong, B.H. Tang, J.G. Lisoni, C.L. Tan, A. Arreghini, G. Groeseneken and J. Van Houdt, "Statistical spectroscopy of switching traps in deeply scaled vertical poly-Si channel for 3D memories," Electron Devices Meeting (IEDM), 2013 IEEE International, Washington, DC, 2013, pp. 21.3.1-21.3.4.
- [8] M.J.M. Pelgrom, H.P. Tuinhout, M. Vertregt, "Transistor matching in analog CMOS applications," in Technical Digest, International Electron Devices Meeting (IEDM), Dec. 1998.
- [9] R. Ritzenthaler, T. Schram, A. Spessot, C. Caillat, M. Aoulaiche, M.J. Cho, K.B. Noh, Y. Son, H.J. Na, T. Kauerauf and B. Douhard, "A Low-Power HKMG CMOS Platform Compatible With Dram Node 2x and Beyond," in IEEE Transactions on Electron Devices, vol. 61, no. 8, pp. 2935-2943, Aug. 2014.
- [10] D. Reid, C. Millar, S. Roy, A. Asenov, "Understanding LER-Induced MOSFET Vt Variability-Part I: Three-Dimensional Simulation of Large Statistical Samples," IEEE Transactions on Electron Devices (TED), Nov. 2010.
- [11] D. Reid, C. Millar, S. Roy, A. Asenov, "Understanding LER-Induced MOSFET Vt Variability-Part II: Reconstructing the Distribution," in IEEE Transactions on Electron Devices (TED), Nov. 2010.
- [12] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," Reliability Physics Symposium, 2005. Proceedings. 43rd Annual. 2005 IEEE International, 2005, pp. 381-387.
- [13] B. Kaczer, P. J. Roussel, T. Grasser and G. Groeseneken, "Statistics of Multiple Trapped Charges in the Gate Oxide of Deeply Scaled MOSFET Devices—Application to NBTI," in IEEE Electron Device Letters, vol. 31, no. 5, pp. 411-413, May 2010.
- [14] B. Kaczer et al., "Extraction of the Random Component of Time-Dependent Variability Using Matched Pairs," in IEEE Electron Device Letters, vol. 36, no. 4, pp. 300-302, April 2015.
- [15] C. Schlunder, F. Proebster, J. Berthold, W. Gustin, H. Reisinger, Influence of MOSFET geometry on the statistical distribution of NBTI induced parameter degradation, presented at IEEE International Integrated Reliability Workshop (IIRW), (2015).
- [16] P. Weckx, B. Kaczer, P. Raghavan, J. Franco, M. Simicic, P.J. Roussel, D. Linten, A. Thean, D. Verkest, F. Catthoor and G. Groeseneken, "Characterization and simulation methodology for time-dependent variability in advanced technologies," Custom Integrated Circuits Conference (CICC), 2015 IEEE, San Jose, CA, 2015, pp. 1-8.
- [17] T. Grasser, P.J. Wagner, H. Reisinger, T. Aichinger, G. Pobegen, M. Nelhiebel and B. Kaczer, "Analytic modeling of the bias temperature instability using capture/emission time maps," Electron Devices Meeting (IEDM), 2011 IEEE International, Washington, DC, 2011, pp. 27.4.1-27.4.4.
- [18] M. Toledano-Luque, R. Degraeve, P. J. Roussel, L. A. Ragnarsson, T. Chiarella, N. Horiguchi, A. Mocuta and A. Thean, "Fast Ramped Voltage Characterization of Single Trap Bias and Temperature Impact on Time-Dependent V_{th} Variability," in IEEE Transactions on Electron Devices, vol. 61, no. 9, pp. 3139-3144, Sept. 2014.