Advanced MOSFET Variability and Reliability Characterization Array

M. Simicic[1,2,*], V. Putcha[1,2], B. Parvais[2], P. Weckx[1,2], B. Kaczer[2], G. Groeseneken[1,2], G. Gielen[1], D. Linten[2], A. Thean[2]
[1] ESAT, KU Leuven, 3000, Leuven, Belgium, [2] IMEC, Kapeldreef 75, 3001 Leuven, Belgium, *email: marko.simicic@imec.be

***Abstract*—Time-zero variability, bias temperature instability (BTI) and random telegraph noise (RTN) are issues that both analog and digital designers using scaled CMOS technologies have to face. In order to address them at design time, access to a sufficiently large number of individual devices is required for statistical technology characterization and modeling. In this paper we present a large MOSFET array designed and fabricated in an advanced 28nm technology, containing both nMOS and pMOS devices of different sizes, both single and stacked. Measurement data for time-zero and time-dependent variability are shown and discussed. Large scale transistor arrays are an indispensable tool to accurately capture the statistics of variability and reliability mechanisms in advanced technology nodes.**

Keywords—Bias temperature instability (BTI), Random telegraph noise (RTN), transistor array, variability, statistical distributions, defect-centric approach

I. INTRODUCTION

RTN and BTI are reliability issues which impact both analog and digital designs in highly scaled technologies [1]–[4]. Since both mechanisms are stochastic processes inducing high variability for minimum sized MOSFETs, a large number of devices need to be measured in order to obtain a sufficiently large sample size. Several array designs reported in literature have been proposed as solutions to obtain adequate sample sizes [5]–[11]. Compared to [5] and [11], the array described in this paper has the added advantage of withstanding higher than nominal stress voltages, needed for reliability characterization, thanks to large IO devices used in the periphery. These thicker gate oxide transistors are less sensitive to degradation and owing to their size, also to time-zero variability and RTN. The simplicity of this design also makes it more compact in size, compared to other more elaborate designs like [7]–[9]. Shift registers, which are used to program the array, make the selection of multiple transistors, Devices Under Test (DUTs), at the same time, possible. Also, this approach reduces the total parasitic resistance to the external measurement environment, compared to designs which use decoders, e.g. [10]. In the existing and upcoming multigate technologies gate length design will be restricted or quantized. This comes as a disadvantage for most analog designs.

Acknowledgments: The work has been in part supported by the European Commission under the FP7/ICT project 619234 MoRV.

TABLE I. GEOMETRIES OF THE TRANSISTORS IN THE ARRAY (UNITS ARE IN NM).

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

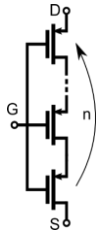

Fig. 1. The stacked transistor (** in table 1).

A stacked transistor, emulating a long gate, facilitating analog design in these technologies, is therefore also included in this design (Fig. 1). To summarize, the key points of this design include: (i) DUTs of several different sizes for area scaling, (ii) thick oxide (IO) switches capable of resisting stress voltages with minimum degradation, (iii) Kelvin sensing, (iv) calibration rows/columns, (v) leakage suppression structures and (vi) a shift register based control. In the following sections the design, simulations and measurements are discussed.

II. DESIGN OF THE ARRAY

The array was designed in a commercial CMOS 28 nm HK/MG bulk technology and includes 54,432 single FETs. The different transistor geometries are shown in TABLE I. each of which contains 108 x 42 = 4,536 single FETs.

All nMOS and pMOS transistors share the source and body terminals, the gates are shared among DUTs in rows, while the drains in columns. Thus, to measure a pMOS DUT, which requires the use of a negative gate V_{gs} and drain voltage V_{ds}, the source terminal has to be biased positively, e.g. 1.8 V, the nominal voltage for IO devices. Since the used technology has a p-type substrate, applying 1.8 V to the body terminal of an nMOS DUT would raise the body voltage of the whole chip to the same level, affecting the periphery. For this reason, a deep n-well is used to isolate the nMOS DUTs from the rest of the chip. Double transmission gates (switches) are used in order to select a single device out of the array while ensuring that the non-selected DUTs are never floating To reduce unwanted RTN signals and enable higher than nominal voltages on the DUTs, the switches are designed with sufficiently large IO devices. Other benefits of double switches are the leakage current steering and gate voltage control of the non-selected devices. Kelvin sensing on the gate

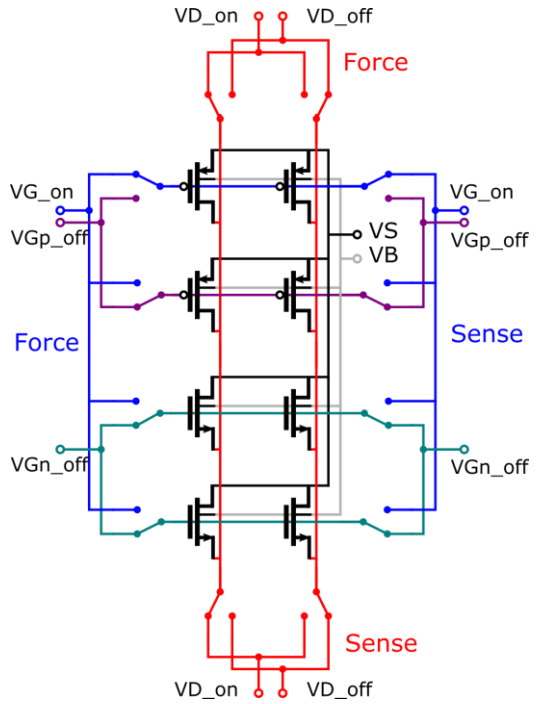

Fig. 2. Reduced schematic of the array including both nFETs and pFETs, double transmission gates (shown as switches), Kelvin sensing for the drain and gate terminals, the body and the source connections. The suffix "on" represents the pins where the useful currents are measured while the "off" represents the pins used to suppress the leakage from unselected FETs by both current steering (VD_off) and by applying a deep sub-threshold gate voltage (VG_off).

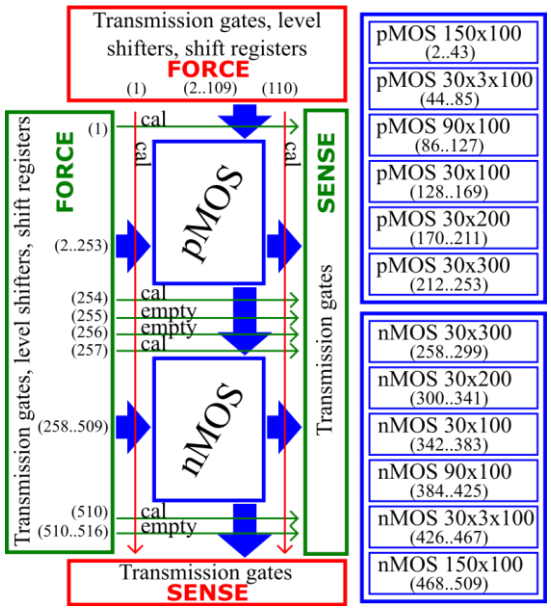

Fig. 3. Placing plan of the array. The pMOS DUT block is on the top, while the nMOS is on the bottom. Transmission gates are placed on both sides, to enable Kelvin sensing. The shift registers ans level shifters are placed on the top and left side of the design.

Fig. 4. Microphotograph of the chip with (top) the large array, (center) single FETs and (bottom) the pipelined array [6]-[7].

and drain DUT terminals compensate for the parasitic resistance in the metal lines and the transmission gates. A schematic of the switches and the Kelvin sensing is shown in Fig. 2.

Two shift registers are used to address the transmission gates, with the added possibility of measuring multiple DUTs in parallel. This solution is also area efficient, since it requires a total of four pins. If the shift register was substituted with a simple decoder, it would require ten pins to control the switches.

The array is organized in 110 columns and 516 rows. Calibration and open lines are included to characterize the leakage from the periphery, the parasitic resistances and as a sanity check for the programming of the chip during measurements. The placement plan of the array is depicted in Fig. 3. The drawn area of the design is 0.18 mm^2 (top part of Fig. 4, highlighted in red).

III. Measurements

Measurements were performed at room temperature using Keithley K26xx series source meters, an HP81150A pulse generator and standard power supplies. Both simulations and measurements show that the leakage current from the non-selected DUTs and the periphery can affect the measurements. This issue can be dealt with by subtracting the leakage current from the measured current. Comparing the measurements to the simulations (Fig. 5), shows that currents below the leakage limit can be measured to a certain extent, mostly depending on the noise level and the integration time of the measurement setup.

A. Time zero variability

By representing the I_{ds} currents measured at a certain voltage as a color-coded matrix (Fig. 6), no obvious variability correlation within a block of DUTs of the same geometry can be revealed. The threshold voltage values (V_{th}) for both the nMOS and pMOS DUTs are extracted at a constant current. The V_{th} distributions are shown in Fig. 7. All data, for both nMOS and pMOS DUTs, show a linear behavior on a standard normal quantile (Probit) plot, meaning that they are normally distributed, as typically observed. The mean V_{th} values vary with geometry, but not sufficiently to make any decisive conclusions. The standard deviation values, on the other hand, give more substantial results.

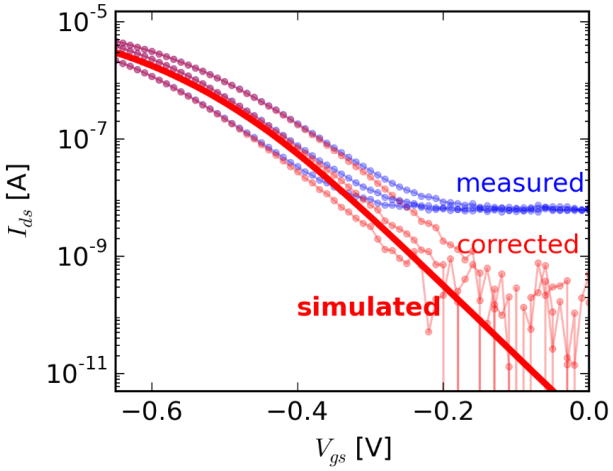

Fig. 5. Three pMOS IdVg measurements, corrected for leakage, plotted against simulations, proving that information below the leakage level can be extracted and that no significant IR drop exists owing to the use of Kelvin sensing.

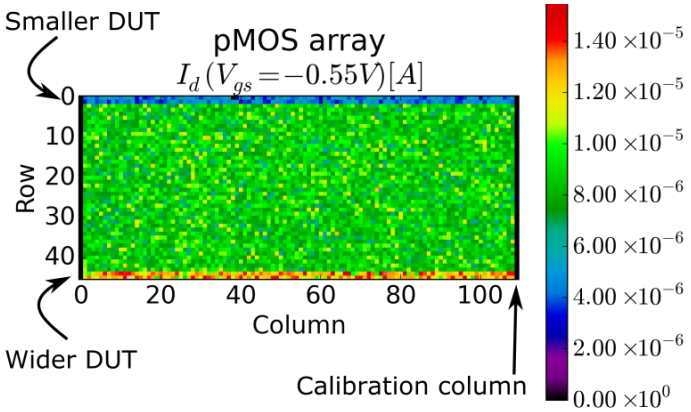

Fig. 6. Color coded matrix of a block of FETs. Calibration columns are found on the left and right edge, while devices of a different geometry from adjacent arrays are on the two top and bottom rows.

The standard deviation values are analyzed using the Pelgrom plot [12]. For the pMOS DUTs, as shown in Fig. 8, the fit trends towards zero, as expected, but also reveals a higher variability of minimum length devices. This can be attributed to the nonlinear interplay of device length, the line edge roughness and the short channel effect, explained in more detail by Reid et al. in [13] and [14]. The nMOS data shows a different trend. The fitting line on the Pelgrom plot matches the data well even when forcing a zero intercept (Fig. 8), but compared to the pMOS, it does not show an increased variability for minimum length DUTs. However, the variability of the largest device is higher than expected. These two observations imply that a source of variability other than the line edge roughness, has a higher impact on the nMOS DUTs. The cause for this could simply come from the inherently different fabrication steps of the n- and pMOS devices. An in-depth analysis would be needed for a more definite conclusion.

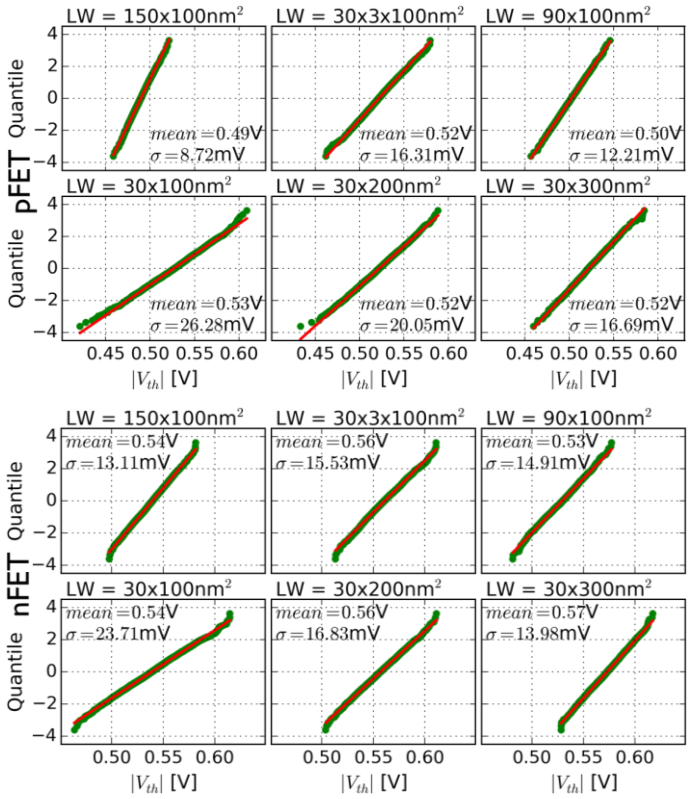

Fig. 7. Probit plots of the V_{th} distributions across all n- and pFET geometries. All plots fit to the normal distribution well at least up to the three sigma point.

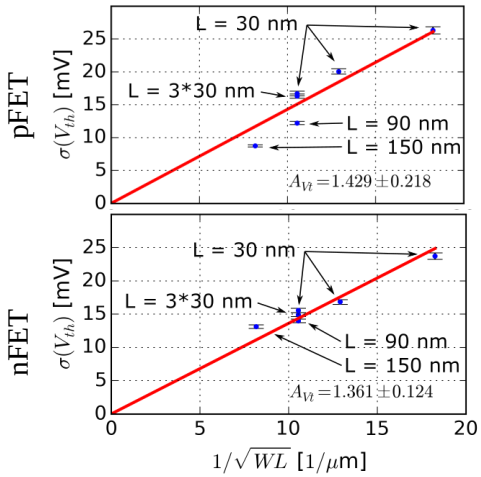

Fig. 8. Pelgrom plots for the nMOS and pMOS DUTs with error bars for a confidence level of 99%.

B. *RTN and BTI measurements*

After initial characterization, a whole row of DUTs (nMOS or pMOS) is stressed with $|V_{gs}|$ = 1.4 V and $|V_{ds}|$ = 50 mV at room temperature for 1000 s to characterize BTI. Once the stress is removed, the currents are measured again at $V_{gs} \sim V_{th}$ and $|V_{ds}|$ = 50 mV. Extracting the V_{th} from the fresh and stressed I_{ds},

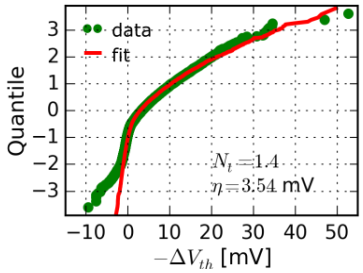

Fig. 9. The current at approximately V_{th} is measured before and after stress and converted to the ΔV_{th} value using the initial I_dV_g sweep. The converted values are plotted on the probit graph and compared to the defect centric model, showing a good fit. The positive ΔV_{th} datapoints are not taken into consideretion for the fit.

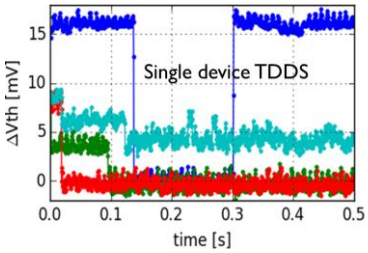

Fig. 10. TDDS measurement. The measured current is transformed into ΔVth values using the initial I_dV_g curve measured on the same DUT.

gives information about BTI and RTN for all DUTs in the row as explained in [15] and plotted in Fig. 9. Preliminary results show that the defect centric model [16] fits the data well, as expected. From this fit it is possible to extract the average number of traps per transistor N_t and the average impact of a trap, η, on the V_{th} value.

As a sanity check, single DUT current relaxation transients are observed after stress. This measurement is also known as the time dependent defect spectroscopy (TDDS) [17]. Fig. 10 depicts relaxation traces from four different DUTs. As expected, immediately after the stress is released, the traps are still charged. In time, they can emit charges producing distinct steps in the transients. A steady sate condition where charge is emitted and recaptured is attributed to RTN. It is important to note that these measurements are not affected by BTI or RTN in the peripheral circuitry as they were never observed when measuring the calibration lines.

IV. Conclusion

An advanced array to measure variability, RTN and BTI is described, simulated, fabricated in silicon and measured. Shift registers, leakage suppression, calibration lines and the Kelvin sensing structures have been successfully implemented in a single design making it a flexible but compact solution for statistical FET data acquisition. Preliminary results of time zero variability, RTN and BTI are presented concluding that the design is a valuable asset for both variability and reliability studies.

References

[1] M. Luo, R. Wang, S. Guo, J. Wang, J. Zou, R. Huang, "Impacts of Random Telegraph Noise (RTN) on Digital Circuits," in *IEEE Transactions on Electron Devices (TED)*, 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 (TED)*, March 2014.

[3] C. Leyris, F. Martinez, M. Valenza, A. Hoffmann, J.C. Vildeuil, F. Roy, "Impact of Random Telegraph Signal in CMOS Image Sensors for Low-Light Levels," *Eu. Solid-State Circuits Conference (ESSCIRC)*, 2006.

[4] J.P. Campbell, L.C. Yu, K.P. Cheung, J. Qin, J.S. Suehle, A. Oates, K. Sheng, "Large random telegraph noise in sub-threshold operation of nano-scale nMOSFETs," *IEEE International Conference on IC Design and Technology (ICICDT)*, May 2009.

[5] Y. Shimizu, M. Nakamura, T. Matsuoka, K. Taniguchi, "Test structure for precise statistical characteristics measurement of MOSFETs," *Int. Conference on Microelectronic Test Structures (ICMTS)*, 2002.

[6] V. Putcha, E. Bury, P. Weckx, J. Franco, B. Kaczer, G. Groeseneken, "Design and simulation of on-chip circuits for parallel characterization of ultrascaled transistors for BTI reliability," *IEEE International Integrated Reliability Workshop Final Report (IIRW)*, Oct. 2014.

[7] V. Putcha, M. Simicic, P. Weckx, B. Parvais, B. Kaczer, J. Franco, D. Linten, D Verkest, A. Thean, G. Groeseneken, "Smart-array for pipelined BTI characterization," *IEEE International Integrated Reliability Workshop Final Report (IIRW)*, in press.

[8] M.B. da Silva, B. Kaczer, G. Van der Plas, G.I. Wirth, G. Groeseneken, "On-chip circuit for massively parallel BTI characterization," *IEEE International Integrated Reliability Workshop Final Report (IIRW)*, 2011.

[9] C. Schlunder, J.M. Berthold, M. Hoffmann, J.-M. Weigmann, W. Gustin, H. Reisinger, "A new smart device array structure for statistical investigations of BTI degradation and recovery," *IEEE International Reliability Physics Symposium (IRPS)*, April 2011.

[10] C.S. Chen, L. Liping, Q. Lim, Hong Hai Teh, N.F. Binti Omar, Chun Lee Ler, J.T. Watt, "A compact array for characterizing 32k transistors in wafer scribe lanes," *International Conference on Microelectronic Test Structures (ICMTS)*, March 2014.

[11] K. Abe, S. Sugawa, S. Watabe, N. Miyamoto, A. Teramoto, Y. Kamata, K. Shibusawa, M. Toita, T. Ohmi, "Random Telegraph Signal Statistical Analysis using a Very Large-scale Array TEG with 1M MOSFETs," *IEEE Symposium on VLSI Technology*, June 2007.

[12] 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.

[13] D. Reid, C. Millar, S. Roy, A. Asenov, "Understanding LER-Induced MOSFET V_t Variability-Part I: Three-Dimensional Simulation of Large Statistical Samples," *IEEE Transactions on Electron Devices (TED)*, 2010.

[14] D. Reid, C. Millar, S. Roy, A. Asenov, "Understanding LER-Induced MOSFET V_t Variability-Part II: Reconstructing the Distribution," in *IEEE Transactions on Electron Devices (TED)*, Nov. 2010.

[15] P. Weckx, B. Kaczer, C. Chen, J. Franco, E. Bury, K. Chanda, J. Watt, P.J. Roussel, F. Catthoor, G. Groeseneken, "Characterization of time-dependent variability using 32k transistor arrays in an advanced HK/MG technology," *IEEE International Reliability Physics Symposium (IRPS)*, April 2015.

[16] B. Kaczer, T. Grasser, P.J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," *IEEE International Reliability Physics Symposium (IRPS)*, May 2010.

[17] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," *IEEE International Reliability Physics Symposium (IRPS)*, May 2010.