On-Chip I-V Variability and Random Telegraph Noise Characterization in 28 nm CMOS

Amy Whitcombe, Scott Taylor*, Martin Denham*, Vladimir Milovanović, and Borivoje Nikolić
Dept. of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA
*Raytheon Vision Systems, Goleta, CA, USA
E-mail: amy.whitcombe@eecs.berkeley.edu

***Abstract*—Building reliable mixed-signal circuits in advanced process technologies requires an accurate understanding of device performance and variability. This work presents an on-chip transistor characterization platform built on a digital focal plane array readout circuit framework that enables highly parallel device measurements to be taken in the digital domain. This technique is used to quickly assess large-scale transistor characteristics and study the impact of random telegraph noise (RTN) in deeply scaled technologies. A 28 nm HKMG bulk LP CMOS test chip containing over 80,000 NFETs and PFETs of multiple sizes and threshold voltages was fabricated and tested to study device parameters and RTN performance down to cryogenic temperatures. Results support previous studies of RTN temperature dependence and suggest that threshold voltage has minimal impact on RTN relative to device type and dimension.**

I. Introduction

Increasing variability and random telegraph noise (RTN) in scaled CMOS processes pose a major challenge to scaling high-performance mixed-signal ICs. RTN in particular, which increases in magnitude as device sizes shrink, has been tied to issues ranging from imager variability [1] to SRAM bitcell failure [2]. The wide span of RTN time constants (microseconds to minutes) makes it difficult to fully characterize a statistically significant amount of RTN data. Most RTN studies use transistor arrays requiring complex and time-consuming off-chip data collection with high-performance test equipment. The test structure in this work is modeled after a digital focal plane array readout IC (ROIC) [3]. Applying readout methods for large pixel arrays to device characterization allows many RTN samples to be collected quickly in the digital domain.

As shown in Fig. 1, prior I-V measurement and RTN studies using custom ICs include on-chip measurement platforms with a single output analog-to-digital converter (ADC) [4] and analog CMOS image sensor-based structures [5], [6]. Ring oscillator-based methods, including array-based techniques [7], to study digital RTN have also been proposed [8]. However, these structures do not facilitate I-V measurements, providing less insight into mixed-signal RTN effects. In this work, a 96 $\times$ 18 array of compact, moderate-speed ADCs with 12 μm cell pitch enables parallel measurement of hundreds of test devices to quickly obtain meaningful statistical data for a wide range of RTN time constants. The chip, which contains devices with various channel doping (HVT/RVT/LVT) and lengths, was fabricated in a 28 nm HKMG bulk process and measured from room temperature to cryogenic temperatures (100 K) relevant to high-performance sensing applications.

II. On-Chip Characterization Platform

The test chip (Fig. 2(b)) resembles the architecture of a digital focal plane array ROIC (Fig. 2(a)). At the core of the design is a 96×18 array of addressable unit cell pairs containing both a device under test (DUT) cell to generate a test current and a measurement cell with a 16-bit current-mode single-slope ADC. The DUT cells and measurement cells are labeled 'A' and 'B,' respectively, in Fig. 2(a). The only analog inputs to the chip are the DUT gate voltage (V_G), drain voltage (V_D), and a chip-level reference current used for calibration.

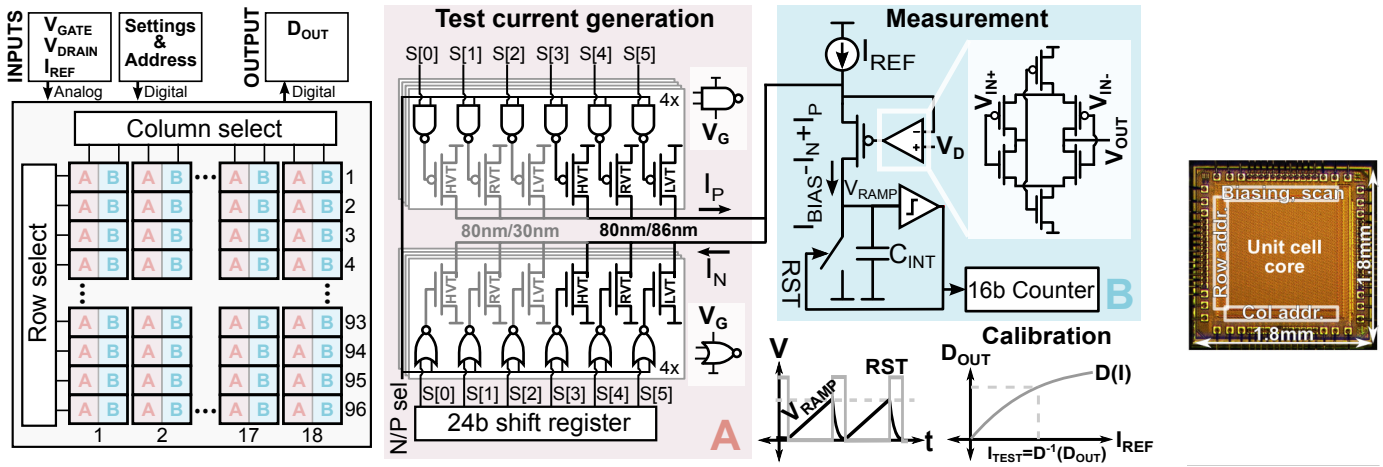
	Concept	t_{test}	$f_{s,ADC}$	Ref.	Tech, DUTs
Standard	CHIP, SMU	$\frac{N \cdot k}{f_{samp}}$	f_{samp}	-	-
On-chip DAC/ADC	CHIP, ADC	$\frac{N \cdot k}{f_{samp}}$	f_{samp}	[4]	45 nm CMOS ~4000 N/PFETs, 6 W/L ratios
Analog array	CHIP, SHIFT REG, ADC, (Fast), SHIFT REG	$\frac{k}{f_{samp}}$	$N \cdot f_{samp}$	[5] / [6]	0.18µm / 0.22µm CMOS 10^6 / 10^5 NFETs, 14 / 1 W/L ratio(s)
This work	CHIP, ADC	$\frac{k}{f_{samp}}$	f_{samp}	-	28 nm CMOS ~10^5 N/PFETs, 2 W/L ratios, 3 V_{TH}

Fig. 1. Prior I-V characterization approaches. This work minimizes the testing time (t_{test}) and ADC/source-measure unit (SMU) sample rate ($f_{s,ADC}$) needed to take k current readings at a sample rate f_{samp} for N devices.

Each DUT cell includes 24 NFETs and 24 PFETs, split evenly between HVT, LVT, and RVT devices with W=80 nm and L=30 nm or 86 nm. Four of each DUT type are included per cell. Specific devices are selected for testing out of the 48 total DUTs using a 24-bit shift register and a binary select signal that toggles between NFET and PFET sampling.

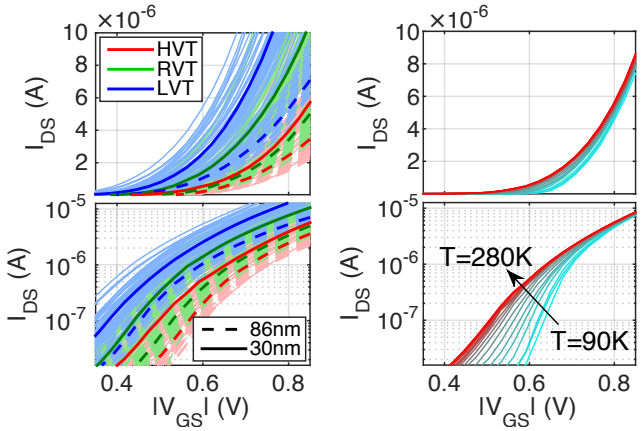
The measurement cell is analogous to a digital ROIC pixel [3], using a single-slope ADC that integrates the test current onto a capacitor to generate a variable-slope voltage ramp that is reset once it exceeds a threshold (Fig. 2(a)). A gain-boosted cascode provides high current source output resistance and enables control of the DUT drain voltage. A compact (12 μm^2) self-biased amplifier [9] is used in the gain-boosting stage to minimize analog area and eliminate the need for external amplifier biasing. The single chip-level reference current is used to calibrate cell measurements. This accounts for both nonlinearity in the I_{in} to D_{out} transfer characteristic caused by the finite switch reset time (see inset in Fig. 2(a)) and cell-to-cell variation in the integration capacitance and comparator offset. Calibrating for cell variation and nonlinearity allows

978-1-5090-2969-3/16/$31.00 ©2016 IEEE



(a) (b)

Fig. 2. (a) Device characterization platform consisting of a 96×18 array of 12μm× 12μm unit cell pairs. Current from one of 48 DUTs in cell (A) is measured by the single-slope ADC in (B), using a gain-boosted cascode with a self-biased amplifier to maintain the DUT drain voltage. An external reference current I_{REF} is used to measure the nonlinear current to digital output code (D_{OUT}) transfer function. (b) Chip die photo. Unit cell core area is 1.44 mm^2.



(a) (b)

Fig. 3. Sample PFET I-V traces on linear and log scales, $|V_{DS}| = 500$ mV. (a) All types, T=290K. (b) RVT L=30 nm PFET vs. T.

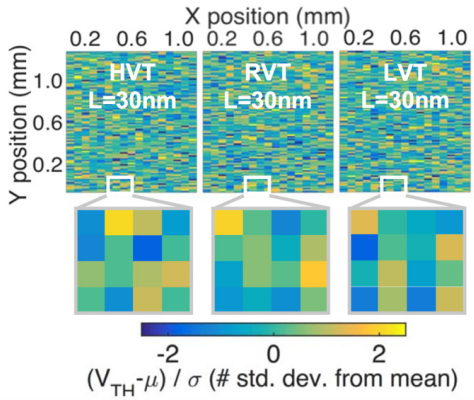
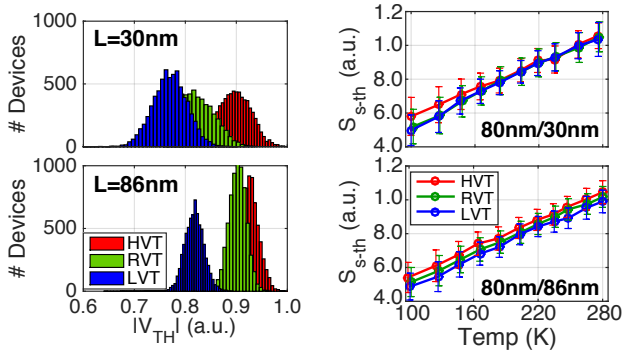


Fig. 4. Map of PFET V_{TH} variation, with no clear spatial dependence. Inset shows large variation between cells, and between DUTs in a cell.



(a) (b)

Fig. 5. Threshold voltage and subthreshold slope performance for PFETs. (a) $|V_{TH}|$ distribution by device type at room temperature (T=290 K). (b) Subthreshold slope (S_{s-th}) vs. temperature. Error bars mark standard deviation of distribution.

the integration capacitor to be built with high-density MOS capacitance and allows the comparator to be implemented with an inverter chain, so that the ADC can be built in a 12μm ×12μm area and require no external biases other than the DUT gate and drain voltages and the reference current.

III. Measurement Results

A. I-V Characterization

Device parameters were extracted by measuring I_D vs. V_{GS} for all devices at room temperature, and for a subset of 512 devices of each type down to cryogenic temperatures. Figs. 3(a) and 3(b) present examples of measured PFET I-V characteristics, demonstrating the expected linear above-threshold and exponential subthreshold behavior, as well as the expected distinction between HVT, RVT, and LVT devices. Devices are measured in saturation (V_{DS}=500 mV) to model analog performance relevant to mixed-signal and imaging purposes. The sample I-V characteristic measured from 280 K to 90 K in Fig. 3(b) demonstrates a clear improvement in subthreshold slope across temperature, with little reduction in above-threshold current. The x-intercept of the above-threshold fit to the I_{DS}-$|V_{GS}|$ curve is used to estimate threshold voltage (V_{TH}). Mapping mean V_{TH} for all four devices of each type within a cell suggests that random variation surpasses spatial dependence in the 1.44 mm^2 core area (Fig. 4). The room temperature V_{TH} distribution (Fig. 5(a)) shows the expected difference between HVT, RVT, and LVT devices, in addition to higher variation in minimum-length devices. Subthreshold slope varies linearly with temperature, as expected, with more relative variability evident at cold temperatures (Fig. 5(b)).

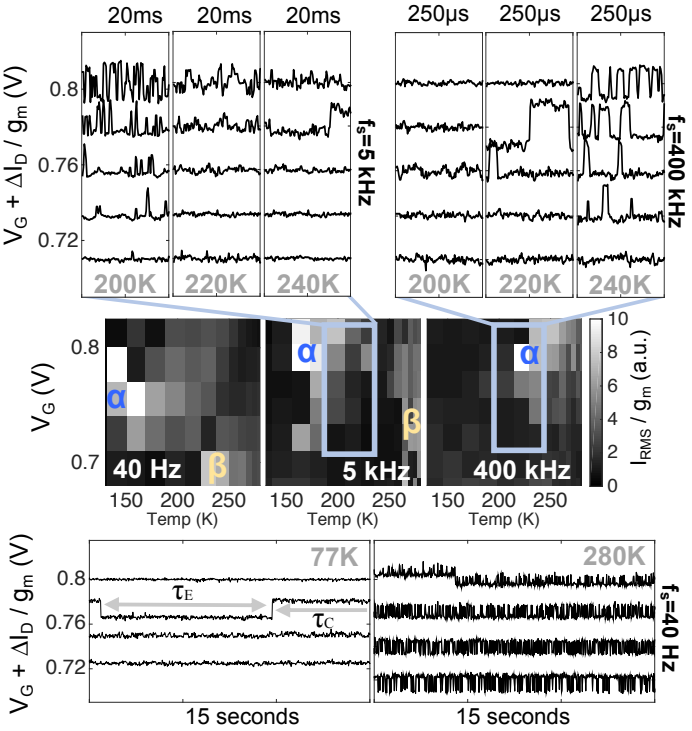


Fig. 6. Sample NFET $\Delta I_D/g_m$ for f_s=40 Hz, 5 kHz, and 400 kHz vs. temperature and V_G (V_{DS} = 500 mV). Clusters in map of I_{RMS}/g_m (with peaks marked α, β) show RTN traps.

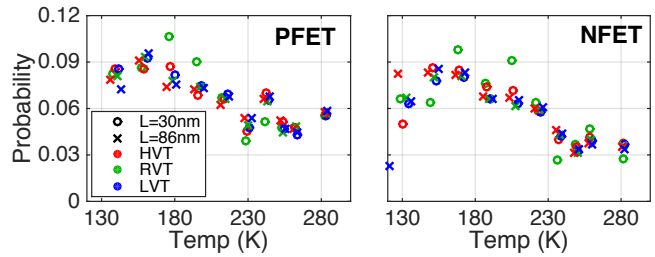


Fig. 7. Distribution of temperatures at which cluster peaks were detected (f_s=40 Hz).

B. RTN Characterization

RTN was measured under static bias conditions for a subset of 6,144 total DUTs at temperatures ranging from 100 K to 290 K. The test chip can capture a wide range of RTN time constants (τ) by varying the ADC sample rate (f_s), as discussed in [10]. Reducing f_s averages the DUT current over a longer sampling period, which improves the minimum measurable current level but restricts the range of detectable time constants to $\tau > 1/f_s$. While long-duration, high-speed measurements can accurately characterize both large and small τ, they also generate orders of magnitude more data than necessary to extract RTN parameters. When many DUTs are tested across bias voltage and temperature, this can easily amount to terabytes of additional measurement data.

Fig. 6 illustrates how f_s, temperature (T), and gate bias (V_G) impact the detectability of RTN in a sample NFET. Change in current (either root mean square current, I_{RMS}, or difference from mean, ΔI_D) divided by device transconductance (g_m) is used as a normalized measure of RTN magnitude. Under a single set of operating conditions (fixed V_G and T), even a multi-trap device will appear trap-free if τ is smaller

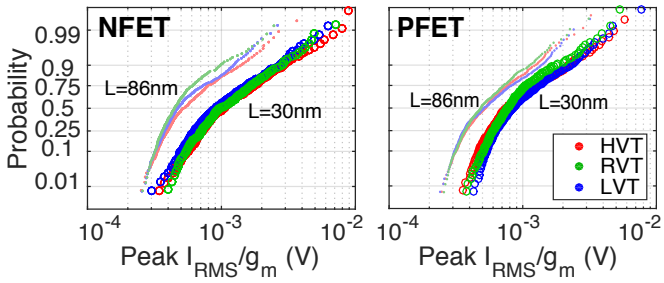


Fig. 8. Lognormal probability of peak I_{RMS}/g_m in each cluster.

than $1/f_s$ or larger than the sample duration. Because RTN time constants are bias-dependent and inversely proportional to temperature [11], most traps will be observable even at a fixed f_s for some set of operating conditions. The center plot in Fig. 6 maps I_{RMS}/g_m measured from 100-point samples of DUT current taken at a range of gate voltages and temperatures for f_s=40 Hz, 5 kHz, and 400 kHz.

In Fig. 6, two traps introduce distinct clusters of measurements, marked α and β, with high I_{RMS}/g_m for a specific bias and temperature range. The peak I_{RMS}/g_m in a cluster occurs when emission (τ_E) and capture (τ_C) time constants are equal and contained in the sample bandwidth. The trap associated with cluster α is first detectable at T=130 K when f_s=40 Hz. As temperature increases, the trap becomes undetectable (filtered out) at f_s=40 Hz, but is observable at higher f_s and V_G. Time-domain samples of the input-referred gate voltage shift ($V_G + \Delta I_D/g_m$) at the top of Fig. 6 illustrate this filtering effect. Because RTN time constants (τ) decrease with temperature, the trap seen at 200 K when f_s=5 kHz is only evident at 240 K when f_s=400 kHz. Similarly, the trap associated with cluster β is first detectable at 230 K when f_s=40 Hz, but cannot be seen even at room temperature when f_s=400 kHz. Nevertheless, a 15 s sample at 280 K (f_s=40 Hz), shown at the bottom of Fig. 6, reveals an additional slow trap just becoming detectable at room temperature. While the DUT may exhibit single-trap or trap-free behavior when a single gate bias and temperature are considered, more comprehensive measurements reveal the presence of at least three traps.

Because most traps will be detectable at some temperature and gate bias for a fixed f_s, RTN amplitude and number of traps are estimated by identifying high-I_{RMS}/g_m clusters in 1000-point, 40 Hz samples of 6,144 DUTs (512 of each type) for 6 near-V_{TH} gate biases and 9 temperatures from 100 K to 290 K. The 512 measurements per DUT type were taken simultaneously, for a total testing time 4.5 hours instead of the 96 days required to measure each DUT sequentially. At this f_s, the peak I_{RMS}/g_m of most clusters are found at lower temperatures for all device types (Fig. 7). The magnitude of RTN, however, is device-dependent. Fig. 8 shows the cumulative probability distribution of peak I_{RMS}/g_m values for each cluster as a function of device type. The distribution tail at high I_{RMS}/g_m levels, where RTN dominates, fits a lognormal distribution (Fig. 8). The 30 nm channel length FETs have RTN amplitudes roughly twice as large as the 86 nm channel length FETs, while channel doping has a relatively small impact on RTN magnitude. In general, NFETs exhibit larger RTN amplitudes than PFETs. The number of RTN traps can be estimated by the number of clusters per DUT with large ($>75^{\text{th}}$ percentile) peak I_{RMS}/g_m values. As shown in Fig. 9, this quantity follows a Poisson distribution. The results indicate

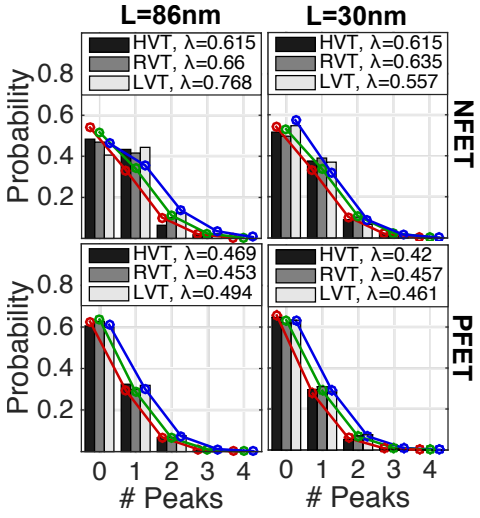


Fig. 9. Measured probability (bar) and Poisson fit (line) of # clusters per DUT with peak levels above 75th percentile (f_s=40 Hz, T=100 K to 290 K, 512 DUTs per type).

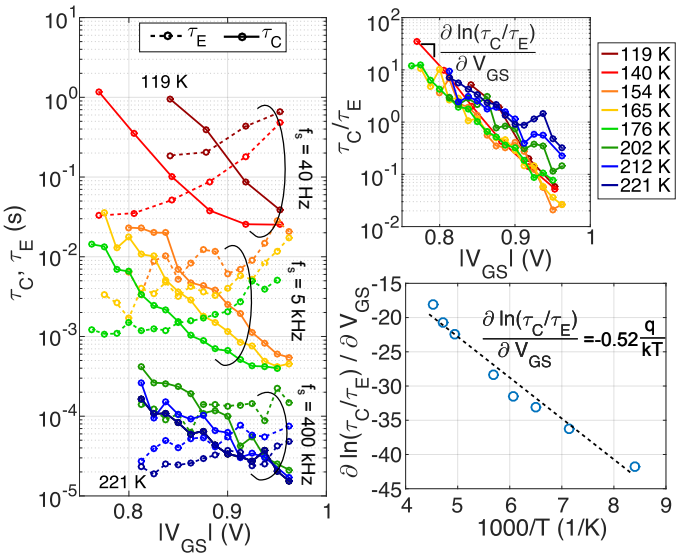


Fig. 10. Example voltage and temperature dependence of mean capture (τ_C) and emission (τ_E) times and mark-space ratio τ_C/τ_E in a sample PFET.

that NFETs are slightly more likely to exhibit RTN and that increasing device area slightly increases RTN probability, but changing doping level has little consistent effect.

The temperature dependence of RTN time constants was also considered. Fig. 10 shows the mean capture (τ_C) and emission (τ_E) times measured in 1000-point samples of a single PFET for all three sampling frequencies. The majority of observed traps exhibit either the type-I behavior shown, in which τ_C decreases with $|V_{GS}|$ and τ_E increases with $|V_{GS}|$ [12], or neutral behavior in which τ_E changes little with $|V_{GS}|$ while τ_C decreases. The mark-space ratio (τ_C/τ_E) dependence on V_{GS} also fits the model presented in [12] (where X_T/T_{OX} is the trap depth relative to the oxide thickness):

$$\frac{X_T}{T_{OX}} = \frac{-kT}{q}\frac{\partial \ln(\tau_C/\tau_E)}{\partial V_G} \quad (1)$$

Fig. 10 shows that $\frac{\partial \ln(\tau_C/\tau_E)}{\partial V_G}$ varies directly with -1/T, as predicted assuming a fixed relative trap depth of 0.52.

IV. Conclusion

Integrating high-performance digital processors directly with sensors requires moving analog designs to scaled process nodes that are increasingly susceptible to variability and random telegraph noise (RTN). To help understand these challenges, this work presents a compact device characterization platform modeled after a digital focal plane array readout circuit capable of measuring hundreds of devices simultaneously in the digital domain. This technique eliminates the need to use high-performance bench supplies in device characterization, and enables a wide range of random telegraph noise (RTN) time constants to be characterized quickly and efficiently by adopting a variable sampling rate. Results indicate that device doping level (threshold voltage) has a much smaller impact on RTN magnitude and frequency than device size, and that low-frequency noise is less pronounced in PFETs than NFETs.

Acknowledgment

This work was supported in part by the Intelligence Advanced Research Projects Activity (IARPA) via the Space and Naval Warfare Systems Command (SPAWAR) contract number N66001-12-C-2010, the NDSEG fellowship, and the BWRC. The authors also thank Bruce Bozovich, Mike Jack, Dave Madajian, Eileen Herrin, Chris Tracy, Henry Lee, Roya Mokhtari, Adam Kennedy, Eric Beuville, Tsu-Jae Liu, Dan Connelly, Ana Arias' group, and the UCB Nanolab for support.

The U.S. Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, SPAWAR, or the U.S. Government.

References

[1] R. Kuroda et al., "A statistical evaluation of random telegraph noise of in-pixel source follower equivalent surface and buried channel transistors," *IEEE Trans. on Elec. Devices*, vol. 60, no. 10, pp. 3555-3561, Aug. 2013.

[2] H. Qiu et al., "Impact of random telegraph noise on write stability in Silicon-on-Thin-BOX (SOTB) SRAM cells at low supply voltage in sub-0.4V regime," *VLSI Technology*, June 2015.

[3] B. Tyrrell et al., "Time delay integration and in-pixel spatiotemporal filtering using a nanoscale digital CMOS focal plane readout," *IEEE Trans. on Elec. Devices*, vol. 56, no. 11, pp. 2516-2523, Sept. 2009.

[4] S. Realov and K. Shepard, "Analysis of random telegraph noise in 45-nm CMOS using on-chip characterization system," *IEEE Trans. on Elec. Devices*, vol. 60, no. 5, pp. 1716-1722, May 2013.

[5] K. Abe et al., "Random telegraph signal statistical analysis using a very large-scale array TEG with 1M MOSFETs," *VLSI Technology*, June 2007.

[6] A. Yonezawa et al., "The study of time constant analysis in random telegraph noise at the subthreshold voltage region," *IRPS*, April 2013.

[7] M. Giles et al., "High sigma measurement of random threshold vatiation in 14nm logic FinFET technology," *VLSI Technology*, June 2015.

[8] Q. Tang and C. Kim, "Assessing the impact of RTN on logic timing margin using a 32 nm dual ring oscillator array," *IEDM*, Dec. 2015.

[9] V. Milovanovic and H. Zimmermann, "A 40 nm LP CMOS self-biased continuous-time comparator with sub-100ps delay at 1.1V & 1.2 mW," *ESSCIRC*, Sept. 2013.

[10] C. Theodorou et al., "New LFN and RTN analysis methodology in 28 and 14nm FD-SOI MOSFETs," *IRPS*, April 2015.

[11] H. Miki et al., "Voltage and temperature dependence of random telegraph noise in highly scaled HKMG ETSOI nFETs and its impact on logic delay uncertainty," *VLSI Technology*, June 2012.

[12] T. Nagumo, K. Takeuchi, T. Hase, and Y. Hayashi, "Statistical characterization of trap position, energy, amplitude, and time constants by RTN measurement of multiple individual traps," *IEDM*, Dec. 2010.