Characterization of Random Telegraph Noises of MOSFET Subthreshold Currents for a 40 nm Process

Calvin Yi-Ping Chao, Meng-Hsu Wu, Shang-Fu Yeh, Chi-Lin Lee, Chin Yin, Kuo-Yu Chou, and Honyih Tu

Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan, ROC
Tel: (886) 3-5636688 ext 703-8243, email: calvin_chao@tsmc.com

INTRODUCTION

For stacked CMOS image sensors (CIS), the random telegraph noises (RTN) could degrade the performance of both the pixels on the top layer and the readout circuits on the bottom layer. In this paper, we investigate the MOSFET channel RTN in a 40 nm process, which is widely used for the CIS readout circuits in leading smartphone cameras.

Previous studies of RTN in CIS primarily focused on the pixel source follower (SF) [1-4]. However, the SF current is typically limited to a narrow range, roughly from 0.5 uA to 10 uA, due to the power consumption and circuit settling time requirements. In this work, we measured the RTN in the subthreshold region, which becomes increasingly important for ultra-low-power applications such as the Internet of Things (IoT). The objective was to identify the systematic trends of the RTN behavior over a much wider range of operation conditions.

CHIP DESIGN AND OPERATION

The test chip is fabricated in a 40 nm, 1P6M, low-power, mixed-mode process, consisting of an 800×1300 cell array. The signal-chain schematic is shown in Fig. 1(a). The chip is operated like an active-pixel CIS. The reset transistor (RST) is the device under test (DUT), acting as a current source controlled by the gate voltage (V_G) during the charge integration. Fig. 1(b) shows the in-line integration mode similar to the 4T pixel operation, used for measuring larger currents, from 1 pA to 500 pA, with relatively shorter integration time (t_{int}), from 0.5 us to 200 us. Fig. 1(c) shows the rolling-shutter integration mode similar to the 3T pixel readout, used for measuring smaller currents, from 1 fA to 1 pA, with longer t_{int}, from 0.52 ms to 516 ms. The key performance parameters of the chip are listed in Table I.

TABLE I
TEST CHIP KEY PERFORMANCE PARAMETERS

Parameter (at 1X analog gain)	Value
Conversion factor from ADC output to SF output	118.8 uV/DN
Conversion factor from ADC output to SF input (SN)	132.4 uV/DN
Source follower (SF) gain	0.90 V/V
Sense node (SN) capacitance C_{SN}	3.30 fF
Readout random noise (SF and readout circuits)	0.30 mV-rms

CORRELATION OF RTN DEVICES

The measured subthreshold current histograms are shown in Fig. 2, covering a range of almost 6 decades, from 1 fA to 1 nA. The exponential I-V characteristics is plotted in Fig. 3. The measured median subthreshold swing was 69.44

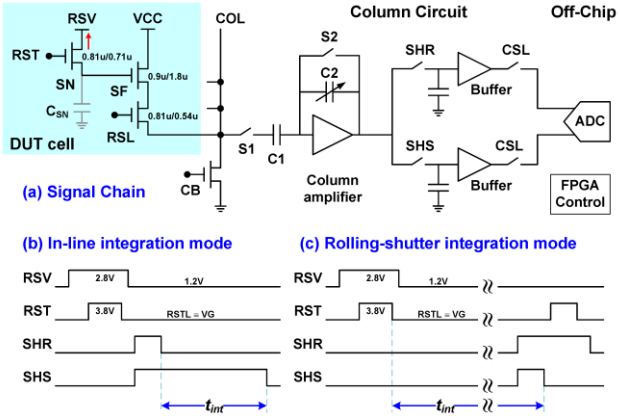

Fig. 1. (a) Schematic of the test chip signal chain. (b) In-line integration mode with 4T-like correlated double sampling for shorter integration time. (c) Rolling-shutter integration mode with 3T-like uncorrelated double sampling for longer integration time.

mV/dec at room temperature, with ±3 sigma's well within ±2% of the median, showing an excellent global matching.

In order to measure and compare the RTN of the currents over a wide range, different integration time was chosen for each current such that the average signal (the sense node voltage V_{SN}) was approximately kept at a constant level. As such, the measure random noise (RN) histograms corresponding to the cases in Fig. 2 had comparable medians and similar long tails. We verified that the noisiest 1,000 DUTs (the top 0.1% of the 1M array) predominately show RTN characteristics with clearly identifiable discrete signal levels.

Fig. 4 shows that the currents under 2 different V_G are well correlated. As a sharp contrast, the random noises (RN) under 2 different V_G are clearly not well correlated. Figs. 5(a)~(c) show the scatter plots of random noises for 3 pairs of voltages (V_{G1}, V_{G2}), corresponding to median currents (I_1, I_2). Evidently, the sets of noisiest devices under different conditions are not the same. The number of common devices (the blue dots) decreases as the difference between V_{G1} and V_{G2} increases, meaning that the correlation between two data sets becomes weaker.

To illustrate this observation quantitatively, we define S_1 and S_2 as the sets of noisiest 1,000 devices under the median currents I_1 and I_2, and define $C(I_1, I_2)$ as the percentage of common devices in sets S_1 and S_2. In a slightly more formal way:

$$C(I_1, I_2) \stackrel{\text{def}}{=} N(S_1 \cap S_2)/N(S_1), N(S_2) = N(S_1), \quad (1)$$

where $N(S)$ denotes the number of elements in the set S. We may interpret $C(I_1, I_2)$ as the "correlation coefficient"

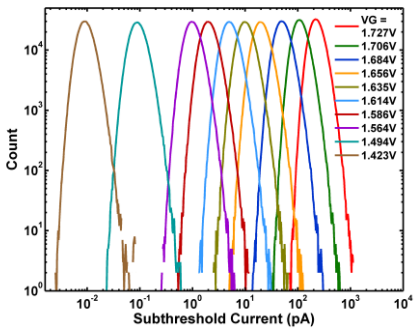

Fig. 2. Subthreshold currents histograms of the 1M devices at various gate voltages V_G, covering 6 decades from 1 fA to 1 nA.

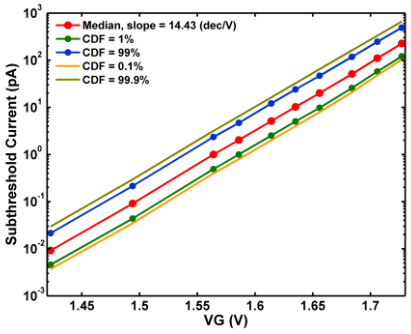

Fig. 3. Subthreshold currents versus V_G of the RST, biased at V_S = 1.2 V, V_B = 0 V, V_D = 1.65 V to 2.43 V.

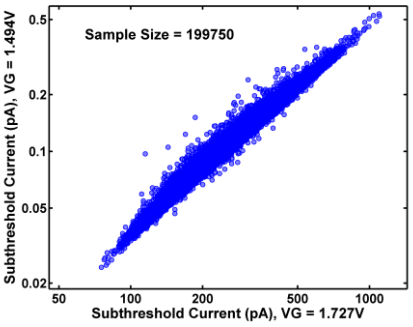

Fig. 4. Scatter plot of subthreshold currents measured at V_G = 1.727 V and V_G = 1.564 V showing strong correlation.

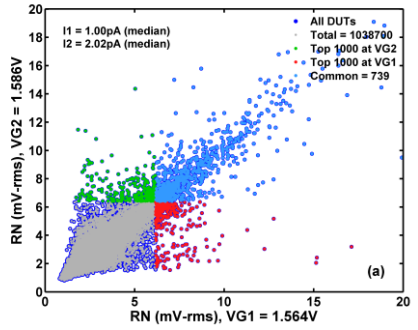

Fig. 5(a). Scatter plot of the random noises measured at V_{G1} = 1.564 V, V_{G2} = 1.586 V.

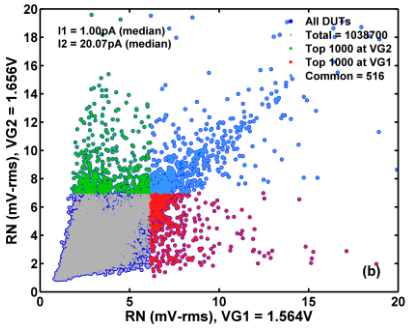

Fig. 5(b). Scatter plot of the random noises measured at V_{G1} = 1.564 V, V_{G2} = 1.656 V.

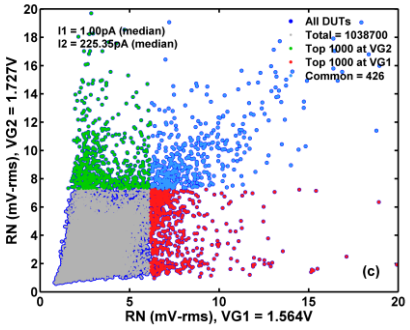

Fig. 5(c). Scatter plot of the random noises measured at V_{G1} = 1.564 V, V_{G2} = 1.727 V.

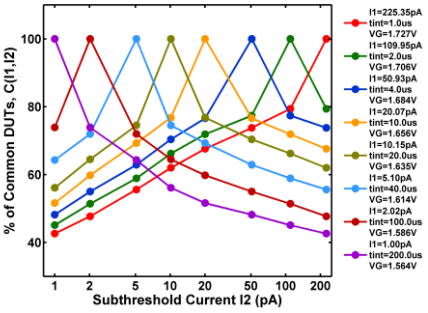

Fig. 6. RTN correlation coefficient $C(I_1, I_2)$ as the percentage of common devices in the sets of noisiest 1,000 devices under currents I_1 and I_2, defined in Eq. (1)

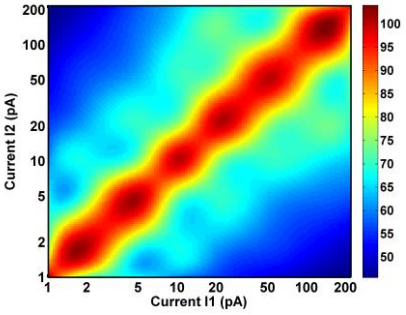

Fig. 7. 2D contour plot of the RTN correlation coefficient $C(I_1, I_2)$ as a function of I_1 and I_2, same data as that in Fig. 6 with some interpolation.

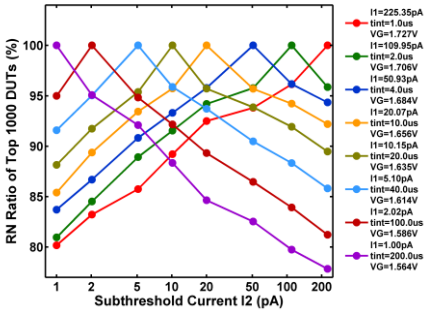

Fig. 8. Ratio of the RMS random noises of the noisiest 1,000 devices selected under I_1 and under I_2, both measured under I_2, the $T(I_1, I_2)$ defined in Eq. (2).

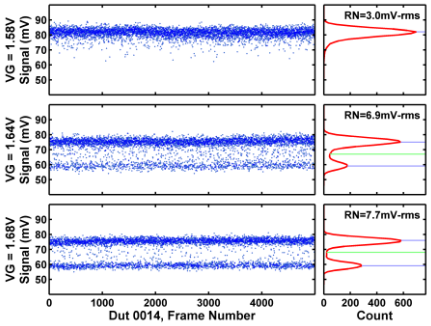

Fig 9(a). Waveform of DUT-0014 showing RTN activities under V_G = 1.68 V but no activities under V_G = 1.58 V, with time constants shorter than the sampling period.

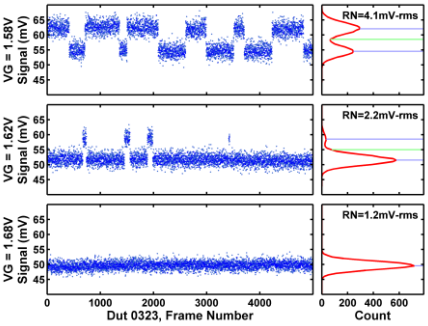

Fig. 9(b). Waveform of DUT-0323 showing RTN activities under V_G = 1.58 V but no activities under V_G = 1.68 V, with voltage dependent time constants.

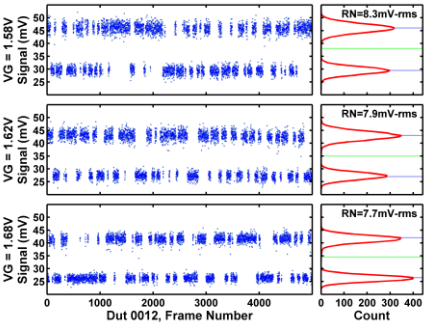

Fig. 9(c). Waveform of DUT-0012 showing RTN activities from V_G = 1.58 V to V_G = 1.68 V, with similar time constants independent of voltages.

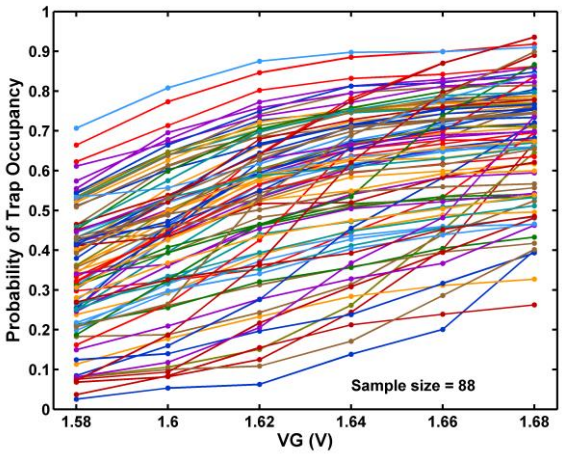

Fig. 10. The probability of trap occupancy (PTO) versus gate voltage V_G for DUTs with the well-behaved single-trap RTN behavior.

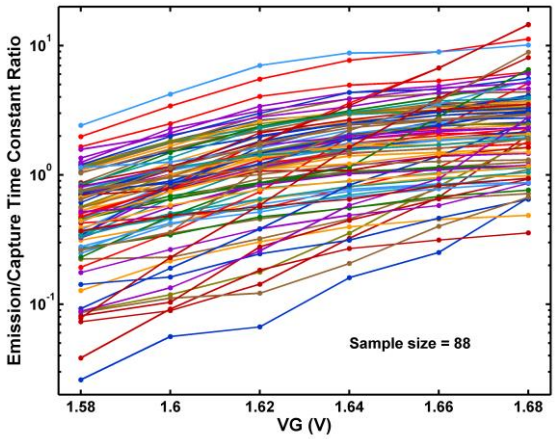

Fig. 11. The time constant ratio τ_e/τ_c versus gate voltage V_G for DUTs with the well-behaved single-trap RTN behavior.

between sets S_1 and S_2, where $C(I_1, I_2) = 1$ means 100% correlated and $C(I_1, I_2) = 0$ means no correlation.

The $C(I_1, I_2)$ family of curves are plotted in Fig. 6 as functions of I_2 indexed by I_1. The plot shows that $C(I_1, I_2)$ becomes smaller as the difference between I_1 and I_2 is larger. Adding a few interpolated points among the measured data, we plot $C(I_1, I_2)$ as 2D contours in Fig. 7.

This correlation can be viewed from a different perspective. Figs. 6~7 show that the set of the noisiest devices under I_1 are different from the set of the noisiest devices under I_2. It then makes sense to compare the RMS noises of these 2 different sets measured under the same bias current. We now define $R(I_1, I_2)$ as the RMS of the random noises of all the noisiest 1,000 devices selected under I_1 but measured under I_2. We then calculate the ratio of noises as:

$$T(I_1, I_2) \stackrel{\text{def}}{\equiv} R(I_1, I_2)/R(I_2, I_2). \quad (2)$$

The family of curves $T(I_1, I_2)$ as functions of I_2 indexed by I_1 are plotted in Fig. 8. It is interesting to notice that the curves $C(I_1, I_2)$ in Fig. 6 and the curves $T(I_1, I_2)$ in Fig. 8 look very much alike although their definitions are completely different.

Fig. 6 indicates that some members of the noisiest set of devices under current I_1 may not be among the noisiest devices under current I_2. Fig. 8 says that the RMS random noise of the noisiest devices selected under current I_1 is at the maximum under I_1. The RMS random noise of the same set of devices measured under any other current I_2 is always smaller.

Both results imply that different RTN devices have different operation windows where they are most active, or generate the highest random noises. One device may be RTN-active under one condition, but becomes less active or inactive under a different condition.

VOLTAGE DEPENDENT RTN BEHAVIOR

Figs. 6~8 show the collective behavior of the noisiest devices from the statistical point of view. We then examine the RTN waveforms of individual devices in time domain. The 1,000 noisiest devices under $V_G = 1.58$V and $V_G = 1.68$V were selected separately, and the signal (V_{SN}) waveforms of 5,000 consecutive frames were measured under various voltages. A low frame rate of 0.262 fps (frame time 3.815 s) was purposely chosen for data acquisition in order to capture the traps with long time constants anticipated for subthreshold operation.

In Figs. 9(a)~(c), the waveforms of 3 selected devices are shown on the left-hand side and the corresponding signal histograms are plotted on the right-hand side. The device in Fig. 9(a) shows clear single-trap RTN activities under V_G = 1.58 V but becomes inactive under V_G = 1.68 V. Another device in Fig. 9(b) shows an opposite behavior, active under V_G = 1.68 V but inactive under V_G = 1.58 V. A third device in Fig. 9(c) is RTN-active under all 6 voltages form V_G = 1.58 V to 1.68 V. These are concrete examples demonstrating that some RTN devices active under one bias condition may become inactive under another bias condition.

The probability of trap occupancy ($p \stackrel{\text{def}}{\equiv}$ PTO) [5] and the emission-capture time constant ratio (τ_e/τ_c) can be calculated from the signal waveforms for all devices showing clear single-trap RTN behaviors such as those in Fig. 9, where the lower signal level corresponds to the occupied state and the higher signal level corresponds to the empty state. The results are plotted in Figs. 10 and 11. A consistent trend of increasing PTO and τ_e/τ_c versus increasing V_G is evident. These behaviors can be described by the well-known theoretical formula [6]:

$$\tau_e/\tau_c = \exp\left(\frac{E_F - E_T}{k_B T}\right) = p/(1-p), \text{ and} \quad (3)$$

$$p = \tau_e/(\tau_e + \tau_c) = 1/\left(1 + \exp\left(\frac{E_T - E_F}{k_B T}\right)\right). \quad (4)$$

where E_T is the trap energy; E_F is the electron quasi-Fermi level in the p-type substrate; k_B is the Boltzmann constant; T is the absolute temperature.

Several selected devices from $V_G = 1.68$ V and $V_G = 1.58$ V were fit to the theoretical formula using 2 fitting parameters c_1 and c_2,

$$(E_F - E_T)/k_B T = c_1 V_G + c_2. \quad (5)$$

The fitting results in Fig. 12 show that the selected devices match with the theory reasonably well. But we note that not all curves in Fig. 11 follow the simple exponential form in Eq. (3).

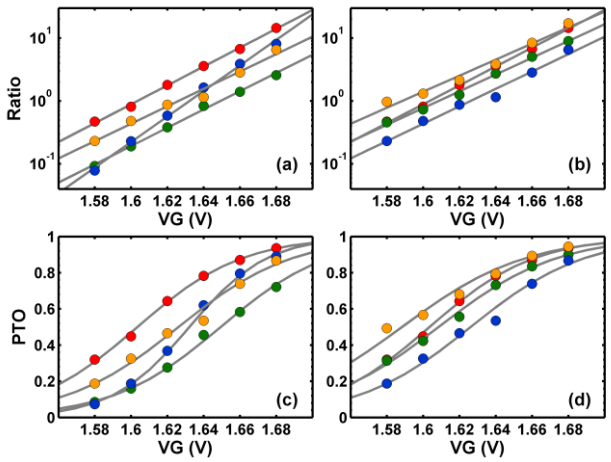

Fig. 12. Fitting of time constant ratio τ_e/τ_c and PTO as functions of V_G to the theoretical formula in Eqs. (3) and (4) for a few selected devices. (a) and (c): samples from the noisiest 1,000 devices under $V_G = 1.68$ V. (b) and (d): samples from the noisiest 1,000 devices under $V_G = 1.58$ V.

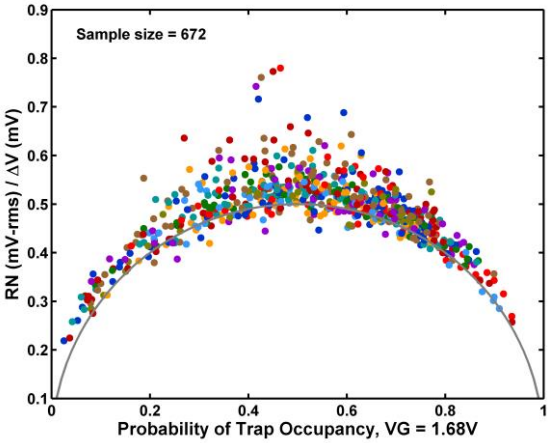

Fig. 13. The ratio of the random noise $\sigma(V)$ versus the RTN amplitude ΔV for DUTs with well-behaved single-trap RTN behavior. The gray curve is the theoretical prediction described by Eq. (3).

Another interesting observation related to PTO is that the ratio of the measured random noise $\sigma(V)$ versus the RTN amplitude ΔV is dependent on PTO. Consider an ideal RTN trap with the occupied-state signal level at V_o and the empty-state signal level at $V_o + \Delta V$, the random noise for a large number (N) of repetitive measurements ($V_n, n = 1$ to N) can be calculate as:

$$\sigma(V)^2 = (\sum_{n=1}^{N} V_n^2)/N - (\sum_{n=1}^{N} V_n/N)^2$$
$$= [pV_o^2 + (1-p)(V_0 + \Delta V)^2]$$
$$-[pV_o + (1-p)(V_0 + \Delta V)]^2 = p(1-p)(\Delta V)^2.$$
$$\therefore\ \sigma(V)/\Delta V = \sqrt{p(1-p)} = 1/\left[2\cosh\left(\frac{E_F - E_T}{2k_BT}\right)\right]. \quad (6)$$

In Fig. 13, the $\sigma(V)/\Delta V$ ratio of the measured data for the well-behaved RTN devices are plotted against PTO, and compared to the curve by Eq. (6). Apparently, the equation describes the data reasonably well. The reason why most of the data points are above the theoretical semi-circle is because the simplified noise calculation in Eq. (6) only includes RTN but no other noises such as the thermal noises, the flicker noises, the readout circuit noises, and the current shot noises. To our knowledge, this simple relation was not verified or reported in literature before.

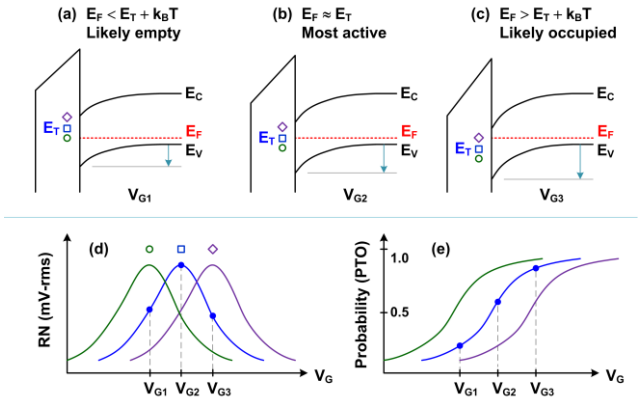

Fig. 14. (a), (b), and (c): band diagrams of a NMOS with 3 hypothetical traps labelled by the purple diamond, the blue square, and the green circle, under 3 gate voltages, V_{G1}, V_{G2}, and V_{G3}. The trap energy for the blue-square trap is E_T. (d): RN as function of V_G for 3 traps. (e): PTO as function of V_G for 3 traps.

MODEL INTERPRETATION

All above data may be understood using a conceptual RTN model. Figs. 14(a)~(c) represent the simplified band diagrams of a NMOS device with 3 hypothetical traps with different trap energies under 3 gate voltages V_{G1}, V_{G2}, and V_{G3}. The trap energy for the blue-square trap is labelled as E_T. Under voltage V_{G1}, the quasi-Fermi level E_F is lower than E_T, and the trap is likely to be empty. Under voltage V_{G2}, E_F aligns with E_T, the PTO is close to 0.5, the trap is most active, accordingly the RN reaches the maximum. Under voltage V_{G3}, E_F is higher than E_T, and the trap is likely to be occupied. Fig. 14(d) shows 3 RN curves as functions of V_G for 3 traps, which explains qualitatively the measured data in Fig. 13 and described by Eq. (6). The RN curves peak at different voltages according to the trap energies (E_T) and the physical location of the trap inside the gate dielectric. Fig. 14(e) shows the corresponding PTO curves. When PTO ≈ 0.5, the trap is most active, and the RN is the highest. When PTO ≈ 0 or PTO ≈ 1, the trap is least active and the RN is the lowest, which explains the data in Fig. 10 and described by Eq. (4). The question why some devices match the theoretical formula better than the others needs to be answered by more study in the future.

CONCLUSION

In summary, we designed a test chip and demonstrated a method to characterize the MOSFET-channel RTN in sub-threshold region. The key finding was that each RTN trap was mostly active within a specific window of operating voltages, outside which the trap became less active or inactive. We found that at different gate biases, the sets of the noisiest RTN devices were not the same. In addition, we derived and verified a relation between the random noise, the RTN amplitude, and the PTO.

REFERENCE

[1] C. Y.-P. Chao *et al.*, J-EDS, vol. 5, pp. 79–89, 2017.
[2] C. Y.-P. Chao *et al.*, Sensors 17, no. 12, 2017, Art. no. 2704.
[3] C. Y.-P. Chao *et al.*, J-EDS, vol. 7, pp. 227–238, 2019.
[4] C. Y.-P. Chao *et al.*, Sensors 19, no. 24, 2019, Art. no. 5447.
[5] X. Wang *et al.*, IEDM 2006, pp. 115–118.
[6] M. J. Kirton *et al.*, Adv. Phys. 38(4), pp. 367–468, 1989.