Characterizing the Impact of RTN on Logic and SRAM Operation Using a Dual Ring Oscillator Array Circuit

Qianying Tang, Student Member, IEEE, and Chris H. Kim, Senior Member, IEEE

Abstract—The impact of random telegraph noise (RTN) on circuit delay has been experimentally verified using a dual ring oscillator (ROSC) array test structure. The proposed on-chip monitor utilizes the tested-and-proven beat frequency detection technique to measure RTN-induced frequency shifts with high precision (>0.01%) and short sampling time (>1 μs). The main idea is to pair an ROSC in the first array with an ROSC having a similar frequency from a second array, so that the frequency measurement resolution is not compromised at sub-0.5-V supply voltages. RTN-induced frequency shifts at different supply voltages, temperatures, and stress conditions were measured from a 32-nm high-k metal-gate test chip. The impact of RTN on logic and SRAM performance was analyzed based on the measured RTN data. We also present the quantitative results of logic timing margin and SRAM noise margin, with and without RTN. According to this paper, RTN appears to have a modest 1% impact on circuit operating frequency in 32 nm, even under pessimistic conditions (i.e., $V_{dd} = 0.6$ V, multiple RTN traps in circuit path).

Index Terms—Logic timing margin, random telegraph noise (RTN), ring oscillator (ROSC), SRAM noise margin.

I. INTRODUCTION

Random telegraph noise (RTN) has become an increasing concern in scaled technologies affecting critical circuit parameters, such as delay and noise margin. RTN is attributed to the random capturing and emitting of charge carriers in gate dielectric traps as shown in Fig. 1. The traps in the oxide can either be defects created during the fabrication process or generated by voltage stress during normal operation. One direct impact of RTN on CMOS transistor is the $V_{th}$ fluctuation between capture and emission states, which resembles a random telegraph signal. Recent studies on RTN aided by new characterization methods have helped establish a better understanding of the underlying physics. This has also led to new fabrication techniques for minimizing the occurrences of RTN [1]–[4]. However, most of the data presented to date are from individual device probing, which provide limited insight into the circuit level RTN behavior.

Inferring circuit level parameters based on device $I$–$V$ data is prone to error due to the fast signal switching and complex circuit topology. Several circuit-based approaches have been demonstrated for RTN measurements. The metastable behavior of a counter circuit was used in [5]–[7] to extract RTN signatures, while an array of transistors was implemented in [7] for efficient $I$–$V$ sweeps. RTN-induced frequency fluctuation measured from an array of ROSC circuits was reported in [7]. On the modeling side, there has been a large body of work analyzing the impact of RTN on circuit parameters, such as logic gate delay and SRAM noise margins [9], [10]. For instance, a statistical timing estimation algorithm was proposed in [11] to calculate RTN-induced logic delay shift for a large circuit block. However, the lack of experimental data to verify the estimation results undermines the confidence of such work.

The main contribution of this paper is that we present detailed RTN-induced frequency fluctuation data collected from a 32-nm test chip operating at supply voltages as low as 0.45 V. Using the high-quality RTN data, we investigate the impact of RTN on logic timing and SRAM noise margin. The tested-and-proven beat frequency detection (BFD) technique was used to measure pico-second fluctuations in ROSC delay, by comparing two ROSC frequencies that are very close to each other [12]–[14]. One of the main shortcomings of this
Fig. 2. BFD circuit adopted in this paper for measuring RTN-induced delay shifts at sub-0.5-V supply voltages with high resolution. The output count $N$ represents the number of $f_B$ clock cycles that can fit within a single beat frequency (i.e., $f_A - f_B$) clock period.

Fig. 3. Limitation of prior art [8]. Due to the wide frequency spread, not all ROSCs under test can achieve high-measurement resolution at sub-0.5-V supply voltages.

Fig. 4. Measurement resolution comparison when pairing a 64 ROSCs with three reference ROSCs (left) and 64 reference ROSCs (right). A more precise waveform can be reconstructed using 64 reference ROSCs which is critical for collecting high-quality RTN statistics at low-supply voltages such as 0.5 V.

II. DUAL RING OSCILLATOR ARRAY TECHNIQUE

Fig. 2 shows the basic principle of measuring RTN-induced frequency shift using the BFD technique [8], [15]. A standard DFF continuously monitors the frequency difference between two free running ROSCs, which is affected by RTN in either ROSCs. The period of the D-flip-flop (DFF) output signal is then digitalized by counting the number of ROSC cycles that fit within a single beat frequency cycle [i.e., $N = \text{floor}(f_B/(f_A - f_B))]$. The advantage of this technique is that the measurement resolution can be made very high by bringing the two frequencies $f_A$ and $f_B$ closer to each other. For example, when the initial frequency difference is calibrated to be 1%, an additional 1% frequency change due to RTN leads to an output count change from 100 to 50. Therefore, the minimum frequency measurement resolution, corresponding to a count change from 100 to 99, is 0.01%.

When a large number of ROSCs need to be measured at low supply voltages however, due to process variation between ROSCs, a small frequency difference (e.g., <1%) between the two ROSCs cannot always be guaranteed. This can be seen in Fig. 3 where the frequency variation of 64 ROSCs can be as high as ±15% at 0.45 V. In the previous design, the ROSC test array is paired with three reference ROSCs, the frequency difference can be as high as 8%, which limits the frequency measurement resolution to >0.6%, which is not sufficient for precise RTN measurements. Tuning the frequency of individual ROSCs using dedicated hardware is not desirable, since the tuning circuit itself may introduce additional RTN noise. Furthermore, adding tuning circuits will make the ROSCs less representative and increase the sensitivity to common-mode noise effects, such as temperature and voltage drifts.
Fig. 5. Block diagram of the proposed dual ROSC array-based RTN characterization circuit. By pairing ROSCs from two arrays, the BFD circuit can achieve a frequency measurement resolution less than 0.01%. The number of inverter stages can be configured from 9 to 15 using scan bits.

To overcome this limitation, in this paper, we propose a dual-array test structure, which guarantees that an ROSC from the main array can be paired with an ROSC from another array with a frequency difference less than 1%. This ensures a frequency measurement resolution of less than 0.01% even in the worst case. As shown in Fig. 4, as the number of reference ROSCs increases from 3 to 64, the worst-case measurement resolution is improved from 0.5% to 0.01% for the proposed dual ROSC array configuration. Test chip results in Section III indicate that a frequency resolution of 0.05% is attainable, which is significantly less than the frequency shift induced by a single RTN trap.

Figs. 5 and 6 show further details of the 32-nm test chip. It consists of two identical ROSC arrays, each comprising 64 ROSCs, along with two separate beat frequency detectors to determine which of the two input frequencies is higher. A 5-bit majority voter circuit is used to prevent functional errors caused by logic bubbles (e.g., lone 0 in a string of 1) or metastability issues, which are likely to occur when the two ROSC edges are about to cross each other. An ROSC in one array is sequentially paired with an ROSC in the other array until the BFD count falls within the desired range (e.g., > 100). A finite state machine sends out a “lock” signal to freeze the column and row selection signals, and then, the frequency difference is measured and scanned out. The

Fig. 6. 32-nm test chip microphotograph and feature summary table.
pairing process takes no more than 100 $\mu$s using our automated test setup. ROSCs are designed with programmable number of stages (i.e., 9, 11, 13, and 15) to study the impact of

Fig. 10. RTN trap location map measured at different supply voltages. Each cell represents a single ROSC.

the number of inverter stages on the amount of RTN-induced frequency shift. PMOS and NMOS transistors used in the ROSC circuit have a width of 624 nm and a length of 56 nm. The new test structure is well suited for bias temperature instability (BTI) stress experiments, since the ROSC can be configured as an open-loop inverter chain using tri-gate stages.

III. RTN-INDUCED FREQUENCY SHIFT MEASUREMENT

The proposed dual-array-based RTN monitor was fabricated in a 32-nm high-$k$ metal-gate process. The nominal supply voltage of this technology is 0.9 V. Fig. 7(a) shows frequency shift due to the same trap measured at different temperatures. (b) Capture and emission time constants both decrease at higher temperatures.

Fig. 9. RTN-induced frequency shift versus the number of ROSC stages. The frequency shift caused by the same RTN trap is reduced as the number of stages increases.

Fig. 10. RTN trap location map measured at different supply voltages. Each cell represents a single ROSC.
shift traces of a nine stage ROSC from 0.45 to 0.6 V. Measurements show the signature RTN behavior caused by trapping and de-trapping events. The measured RTN-induced frequency shift decreases from 0.38% to 0.15% as the supply voltage is increased from 0.45 to 0.6 V. The telegraph-shaped RTN waveform was only observable at supply voltages below 0.6 V. This suggests that RTN is not a major issue at nominal supply voltages, but will become more significant when the supply is lowered to near-threshold voltages. The magnitude of frequency shift due to RTN measured from six different ROSCs is shown in Fig. 7(b). Variation in RTN-induced frequency shift can be attributed to the different trap locations in the gate oxide [16]. The frequency shift monotonically decreases at higher supply voltages. One possible reason for this is that ROSC frequency is more sensitive to the same $V_{th}$ change at lower supply voltages due to the smaller overdrive voltage $|V_{gs} - V_{th}|$. Fig. 8(a) shows the frequency shift waveforms at 27 °C, 55 °C, and 85 °C. The magnitude of the frequency shift shows little dependence on temperature; however, trapping and de-trapping occur more frequently at higher temperatures, which is in line with the previous studies. RTN time constants are defined as the average time a trap site stays in the occupied state or in the unoccupied state. The capture ($\tau_c$) and emission ($\tau_e$) time constants can be extracted using an exponential model fit to the measured distribution, as shown in Fig. 8(b) (top). Theoretically, the proposed BFD can measure time constants shorter than a microsecond. However, due to the slow data scan out, the minimum time constant measurable by our design is a few microseconds. The maximum time constant we can measure is limited only by the measurement time. To study the impact of logic depth on frequency shift, we first selected an ROSC with an RTN trap and, then, varied the number of stages using scan signals. Experimental data in Fig. 9 show that as the number of stages increases from 9 to 15, the frequency fluctuation reduces from 0.38% to 0.24% for the same RTN trap due to the lower sensitivity.

Fig. 10 shows the occurrence and location of RTN traps across a single test chip from 0.45 to 0.6 V. RTN traps may appear or disappear as the supply voltage is varied which we suspect is due to the Fermi level shift [17]. That is, the RTN trap is more likely to be detected if the trap energy level and the Fermi level are closely aligned. The number of ROSCs affected by RTN remained relatively constant under different supply voltages.

Both RTN and BTI have been reported to originate from the same defect sources [1], [2]. To understand the interplay between RTN and BTI better, we measured the location and occurrence of RTN while applying a voltage stress to the ROSC array. The ROSC frequencies were sampled periodically at 0.45 V while the test chip was subject to a 1.8 V voltage stress ($=2 \times$ the nominal $V_{dd}$) for 14 h. Stress results in Fig. 11 reveal several newly generated RTN traps as well
as few annealed traps. The former can be attributed to defects created during BTI stress, while the latter may be related to the BTI recovery phenomenon [18]. The higher occurrence rate with longer stress time implies that RTN along with BTI further degrades the circuit long-term performance. The percentage of ROSCs affected by RTN measured from six different chips is shown in Fig. 12.

IV. RTN IMPACT ON LOGIC TIMING

To estimate RTN-induced delay shift in circuits other than simple inverters, we first translated the frequency shift measured from the 32-nm test chip to $V_{th}$ shift using the frequency versus $V_{th}$ relationship simulated in SPICE. Then, we apply the $V_{th}$ shift to various logic gates and DFFs. DFFs are typically implemented with minimum or near-minimum sized transistors, so to keep the simulations representative, RTN-induced $V_{th}$ shifts have been estimated accordingly based on the area scaling equation $\Delta V_{th} \propto 1/(W \cdot L)$ widely used in the previous literature [19]–[21]. Fig. 13 shows three possible RTN-induced timing violations in a typical pipeline circuit.

Setup time violation is illustrated in Fig. 13(a) in the presence of RTN traps in the clock tree. In the worst case, the launching clock CLK1 arrives late and the sampling clock CLK2 arrives early due to RTN. This introduces a skew between CLK1 and CLK2, which reduces the available time for logic computation. The second scenario is shown in Fig. 13(b) where the combinational logic delay increases due to RTN. Finally, as shown in Fig. 13(c), RTN in the DFF can affect setup and hold times.

For a better understanding, Fig. 14(a) shows that the worst case DFF setup time occurs when traps appear in alternating pMOS and nMOS devices on the signal path from D to Q. Fig. 14(b) displays the D-to-CLK and CLK-to-Q delays with and without RTN. Since RTN becomes more significant at low supply voltages, our simulations are performed at 0.5 V. It can be seen that in the presence of RTN, the setup time and hold time curves shift either to the right or left depending on the location of the RTN trap. The fluctuation in DFF setup time ranges from $-0.08$ to $0.18$ FO4 inverter delays.

The following discussion will focus on setup time violation. A similar analysis can be performed for hold time violation,
which is not included in this paper. As shown in Fig. 15(a),
to operate without any logic errors, the clock period \( t_{\text{clk}} \) must
be greater than \( t_{\text{clk-to-q}} + t_{\text{setup}} + t_{\text{clk-skew}} \). Fig. 15(b)
compares the max-delay time under different RTN scenarios.
In the worst case, traps may be present in the input and output
DFFs as well as the clock tree and logic path. Clock buffers
and combinational path are implemented with \( 4 \times \) and \( 2 \times \)
sized inverters, respectively. Note that the actual RTN-induced
frequency shift of a circuit depends on the specific sizing
which is different from design to design. The max-delay time
allowed for correct operation is reduced by 0.21 FO4 inverter
delays under this worst-case condition.

The following two factors have been incorporated for estimating RTN-induced timing errors of a large circuit: 1) the
frequency shift magnitude of an individual trap; and 2) the
spatial distribution of traps. The probability of RTN-induced
timing errors for a given timing guard band \( x \) can be expressed as

\[
\Pr(\text{Timing error} | \text{guard band} = x) = \frac{1}{2} \left[ 1 - \sum_{i,j,k} \Pr(N_{\text{clk}} = i) \cdot \Pr(N_{\text{data}} = j) \cdot \Pr(N_{\text{DFF}} = k) \right]
\]

\[
\forall i, j, k : \Delta t_{\text{skew}}(N_{\text{clk}} = i) + \Delta t_{\text{data}}(N_{\text{data}} = j) + \Delta t_{\text{DFF}}(N_{\text{DFF}} = k) \leq x
\]  

(1)

which compares the max-delay time under different RTN scenarios.
Fig. 15(b) shows the RTN impact on SRAM write margin. In this
simulation, RTN-induced \( V_{\text{th}} \) shift of each SRAM cell trans-
sistor was estimated using the area scaling equation \( \Delta V_{\text{th}} \propto 1/(W \cdot L) \). Read SNM and write margin move in opposite
directions for the same RTN trap. The worst case for write
happens when RTN trapping occurs in PG. For better illustration,
we ran Monte Carlo simulations on SRAM read and
read SNM under a 0.6 V supply voltage assuming random trap
locations. As shown in Fig. 18, with RTN, the 99.9 percentile
read SNM and write margin are reduced by 12% and 3.9%,
respectively.

V. RTN IMPACT ON SRAM STABILITY AND TIMING

A 6T SRAM cell is shown in Fig. 17. RTN either improves
or worsens the read margin depending on the trap location
inside the SRAM cell [22]. The read margin is determined primarily by the relative strength between the pull down NMOS
transistor (PD) and the pass gate (PG). In the worst case,
the diagonal PD and PU transistor pair becomes weaker while
PG becomes stronger due to multiple RTN traps. Fig. 17(b)
shows the RTN impact on SRAM write margin. In this
simulation, RTN-induced \( V_{\text{th}} \) shift of each SRAM cell tran-
sistor was estimated using the area scaling equation \( \Delta V_{\text{th}} \propto 1/(W \cdot L) \). Read SNM and write margin move in opposite
directions for the same RTN trap. The worst case for write
happens when RTN trapping occurs in PG. For better illustration,
we ran Monte Carlo simulations on SRAM read and
read SNM under a 0.6 V supply voltage assuming random trap
locations. As shown in Fig. 18, with RTN, the 99.9 percentile
read SNM and write margin are reduced by 12% and 3.9%,
respectively.
Next, we analyze how the SRAM read path delay, namely the CLK to DOUT delay, is affected by RTN. Fig. 19 shows the schematic and timing diagram of a 128-kbit SRAM subarray used in this discussion. First, when RTN traps are present in the row decoder, the CLK to WL delay increases causing the read delay to increase. Similarly, read delay may increase due to RTN traps in the sense amplifier enable signal (SAE) generation path. The worst case read delay occurs when the trap is located in the PG transistors because SRAM read speed is determined by the read current. Finally, RTN in the sense amplifier may degrade the resolving time. Fig. 20 shows a typical latch based sense amplifier. When BitLine is discharged, traps on transistors 2 and 3 increase the SAE to DOUT delay while traps on transistors 4 and 5 decrease the delay. RTN has a stronger impact on sense amplifier delay for smaller bitline voltage differences.

To capture the above discussion, we simulated the CLK to DOUT delay of a realistic 128-kbit SRAM sub-array in 32-nm technology assuming traps in different locations. Considering that transistors on the critical path have different sizing and fan-out, the RTN-induced $V_{th}$ shift was applied to the most sensitive stage while the $V_{th}$ shift was estimated using the area scaling equation. For the sense amplifier, we assumed the input transistors have a W/L of 624 nm/56 nm which is comparable to the sizing used in an industrial design [23]. As shown in Fig. 21, traps located in the row decoder show negligible impact on the overall read delay while traps in the sense amplifier have a greater impact. Assuming a scenario in which a single RTN trap exists in each block (i.e., row decoder, SRAM access transistor, sense amplifier, and SR latch), the read path delay increases by 0.51% at 0.6 V and by 0.90% at 0.55 V.

VI. CONCLUSION

In this paper, we present an array-based circuit for detailed characterization of RTN-induced frequency shift, fabricated in a 32-nm technology. A novel dual ROSC array structure based on the tested-and-proven BFD technique enables fully automated collection of RTN statistics with high measurement accuracy at supply voltages as low as 0.45 V. The magnitude and occurrences of RTN-induced ROSC frequency shift were measured under different supply voltages, temperatures, and voltage stress conditions. Based on the measured frequency shift data, we estimated the RTN impact on logic timing margins and SRAM performance.
REFERENCES


Qianying Tang (S’13) received the B.E. degree in electrical engineering from the University of Electronic and Information Technology and China, Chengdu, China, in 2011, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2017. She joined the VLSI Research Laboratory, University of Minnesota, in 2012, with a focus on design for characterizing circuit reliability, including random telegraph noise and radiation induced soft errors, and circuit design for hardware security, such as true random number generators and physical unclonable functions. She was an Intern with IBM, Fishkill, NY, USA, in 2015. Dr. Tang was a recipient of the Fellowship Award from the China Scholarship Council from 2011 to 2015.

Chris H. Kim (M’04–SM’10) received the B.S. and M.S. degrees from Seoul National University, Seoul, South Korea, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA. In 2004, he joined the Electrical and Computer Engineering Faculty, University of Minnesota, Minneapolis, MN, USA, where he is currently a Professor. He has authored or co-authored over 200 journal and conference papers. His current research interests include digital, mixed-signal, and memory circuit design in silicon and non-silicon (organic TFT and spin) technologies. Prof. Kim has served as a Technical Program Committee Member of several circuit design and semiconductor device conferences. He was a recipient of the SRC Technical Excellence Award, the Council of Graduate Students Outstanding Faculty Award, the National Science Foundation CAREER Award, McKnight Foundation Land-Grant Professorship, the 3M Non-Tenured Faculty Award, the DAC/ISSCC Student Design Contest Awards, the IBM Faculty Partnership Awards, the IEEE Circuits and Systems Society Outstanding Young Author Award, and the ISLPED Low Power Design Contest Awards.