CSRO-Based Reconfigurable True Random Number Generator Using RRAM

Rekha Govindaraj, Swaroop Ghosh, *Senior Member, IEEE*, and Srinivas Katkoori, *Senior Member, IEEE*

***Abstract*— In this paper, we propose a high-speed (kilohertz–megahertz), reconfigurable current starved ring oscillator (CSRO)-based true random number generator (TRNG) design. The proposed TRNG exploits the intradevice stochastic variations in resistive RAM switching parameters and random telegraph noise (RTN). We demonstrate the effect of RTN on the jitter of CSRO oscillations. We also propose a methodology to reconfigure the TRNG to generate new random numbers. The proposed 10-bit TRNG is validated by NIST test suite for randomness in the data stream. Energy/bit is 22.8 fJ for generation, and the speed of random data generation is 6 MHz. Security vulnerabilities and countermeasures of the proposed TRNG are also investigated.**

***Index Terms*— Current starved ring oscillator (CSRO), hardware security, jitter, nonvolatile memory, random telegraph noise (RTN), resistive RAM (RRAM), true random number generator (TRNG).**

I. Introduction

INFORMATION security is one of the primary concerns with the growth of internet and cloud storage. Data encryption and cryptography are reliable techniques for protecting the data over communication channel (network and storage). Random number generator (RNG) is an integral part of cryptography algorithms in encryption engines [1]. Data and system security depends on the randomness of the bit stream generated by RNG [2], [3]. Entropy of the source is instrumental in ensuring the security of the encrypted data. RNGs also find numerous applications other than cryptography such as gaming, gambling, industrial testing and labeling, Monte Carlo simulations, and password generation. Software-based encryption engines depend on the random number generated by the computer which is only pseudo-random due to deterministic algorithms used for generating random number from an initial seed value. Hardware RNG exploits the randomness in physical processes such as electronic noise, quantum processes, and chaotic light emission to generate a continuous stream of random numbers. Although CMOS-based solutions [1], [4], [5] are promising they offer limited security-specific properties such as process variations, noise, and chaos. Emerging technologies such as spintronics [6], [7], memristor [8], and resistive RAM (RRAM) [9], [20] have demonstrated significant promise because in addition to low power, high density, and high speed, they also offer new sources of noise and randomness [40]. Furthermore, these technologies integrate with CMOS easily [12].

In this paper, we explore the RRAM technology and features such as cycle-to-cycle variations and random telegraph noise (RTN) for true RNG (TRNG) design. We make the following contributions in this paper.

1) We exploit inherent noise sources of RRAM to design a TRNG.
2) We propose a high-speed (kilohertz–megahertz) current starved ring oscillator (CSRO)-based TRNG using RRAM. We evaluate the proposed TRNG using NIST test suite.
3) We propose a methodology to reconfigure the TRNG when entropy reduces over time and to recover from noninvasive adversary attacks such as exploiting temperature sensitivity of RTN.
4) We discuss the security vulnerabilities of RRAM-based TRNGs and potential countermeasures in the proposed TRNG.

The remainder of this paper is organized as follows. Section II provides the background of TRNG, RRAM model with switching parameter variation and RTN, and RRAM-based TRNG. Sections III and IV describe the design and present simulation results of the proposed TRNG, respectively. Section V discusses potential adversary attacks on RRAM-based TRNG and countermeasures. Conclusions are drawn in Section VI.

II. Background

We discuss the details of TRNG, RRAM model, and RRAM as the source of randomness in the TRNG design.

A. Related Work on True Random Number Generator

RNGs are broadly categorized into two basic types based on the quality (in terms of randomness) and the method of bit stream generation, namely, pseudo-RNG (PRNG) and TRNG. In PRNGs, bit stream is not completely random as the algorithm is deterministic except the seed value [1]. More secure data encryption algorithms require fully random

Manuscript received September 2, 2017; revised February 14, 2018; accepted March 28, 2018. Date of publication May 3, 2018; date of current version November 30, 2018. This work was supported in part by the National Science Foundation under Grant CNS-1722557, Grant CCF-1718474, and Grant DGE-1723687, and in part by the DARPA Young Faculty Award under Grant D15AP00089. *(Corresponding author: Rekha Govindaraj.)*

R. Govindaraj and S. Katkoori are with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL 33620 USA (e-mail: rekhag@mail.usf.edu).

S. Ghosh is with the College of Electrical and Computer Engineering, Pennsylvania State University, State College, PA 16802-1503 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2018.2823274

and nondeterministic method of generation. Such streams are generated using TRNGs. Several TRNGs have been proposed in [1] and [4]–[8] based on randomness in electrical noise, thermal noise, and oscillator-based RNGs such as free-running oscillator, Fibonacci RO, and Galois RO. Noise-based RNGs postprocess the noise from the analog source (resistance, voltage source, and temperature) to generate random numbers for a digital system. Amplifying tiny noise voltage or converting noise from physical environment to a digital signal often requires multiple stages of processing [1], [4] which depreciates the randomness from the source. Furthermore, the TRNGs which employ the analog parts are weak due to their vulnerability to various adversary attacks.

Emerging technologies such as spintronics and RRAM [6]–[10], which are compatible with the CMOS technology [12] and provide rich sources of entropy on-chip, are attractive in such scenario [40]. However, the resistance range of spintronic device is limited and the speed of RRAM-based TRNGs is as low as few kilohertz due to their dependency on programing speed of RRAM. A high-speed TRNG is proposed in [20], which employs the RRAM RTN noise. The principle is to utilize the differential change in the bias voltage to modify the sampling frequency. The distinction between [20] and the proposed work is as follows.

1) RTN of RRAM in the cell of a memory array modulates the bias voltage of a voltage-controlled oscillator (VCO). However, the bitline interconnect noise could be large enough to suppress the effect of RTN on voltage differential eventually affecting the available entropy. Furthermore, on-chip noise from pseudorandom source such as power supply, temperature, and crosstalk [4] can overpower RTN noise of RRAM which is as small as in the range of nanoampere when bias voltage is generated from a cell in large memory array. The proposed design incorporates a dedicated RRAM in TRNG circuit which preserves the entropy of RRAM RTN.
2) Once the adversary can predict stable frequency of the faster clock (by the method of frequency injection when memory and digital supply rails are accessible) [21], random number samples could be predicted for various sampling frequencies under such weak frequency modulation method. In the proposed design, power supply of the TRNG can be isolated and placed such that it is not accessible externally. This prevents from the possibility of frequency injection attack.
3) TRNG in [20] employs multiplexers to drive each of the inverters in the VCO for frequency trimming which adds considerable area overhead.
4) The peak-to-peak (p–p) amplitude of current variations due to RTN is a figure of merit (FoM) for an RRAM in storage application. Considering the FoM of RRAM to be used in memory array, it is not feasible to use the RTN of RRAM from a memory cell as source of entropy [22] because of their contrary FoM requirements. RO is placed as boundary circuit in the memory architecture. RTN being a noise voltage of less than100 nA, the method uses a bias current of greater than 50 μA to generate bias voltage differential of $\sim$200 mV. The proposed TRNG can operate with bias voltage differential as low as 0.7 mV without any current source for biasing. Furthermore, having a microampere range of current source in the bias circuit also increases the power dissipated in the bias circuit compared to the proposed TRNG. Reconfiguration of the faster clock oscillator and using a dedicated RRAM cell within bias voltage circuit is essential for a robust design under these circumstances. The proposed method provides two levels of recovery from external adversary attacks by configuring the TRNG through programing RRAM (SET/RESET), and by tuning the sampling frequency to obtain good statistical properties of the generated bit stream.

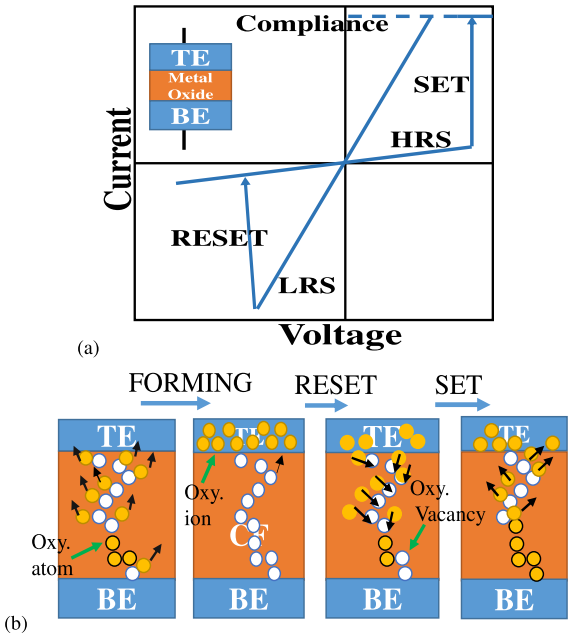

Fig. 1. (a) RRAM memory device and resistance transfer characteristics. (b) Forming, SET, and RESET mechanisms.

Therefore, the proposed design is more effective in exploiting the entropy of the RRAM device for TRNG application. We discuss various potential adversary attacks on TRNG in Section V.

B. Resistive RAM

RRAM is a promising candidate for future nonvolatile memory applications. It is designed by sandwiching an oxide material between two metal electrodes i.e., top electrode (TE) and bottom electrode. RRAM resistive switching is primarily due to the mechanism of oxide breakdown and reoxidation which modifies a conduction filament (CF) in the oxide. Fig. 1(a) shows the voltage and current transfer characteristics during the SET and RESET process cycles. The minimum resistance of the filament depends on the current compliance used in the process of forming. The two states of the RRAM in low resistance and high resistance are termed as low-resistance state (LRS) and high-resistance state (HRS). We have used the expressions from [11], [18], and [23] as the basis to model

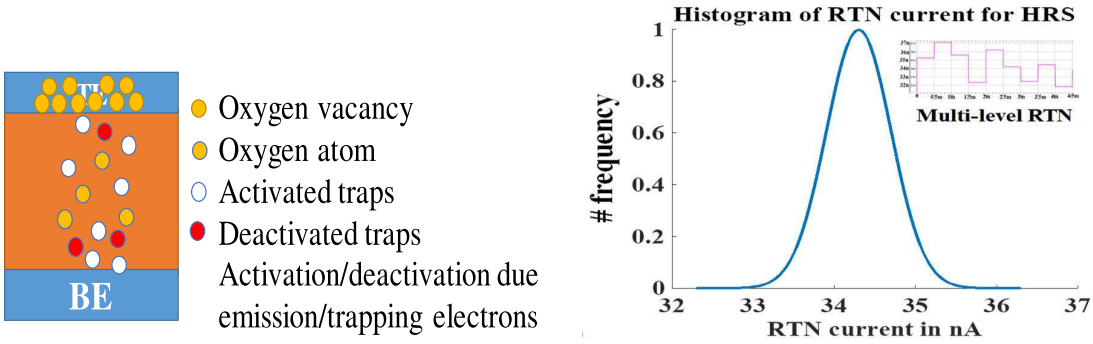

Fig. 2. (a) Mechanism of RTN due to trapping/emission of electrons in the vicinity of CF. (b) RTN current distribution in HRS state of RRAM based on measurement data [22]. I_{RTN} with the frequency of 5 kHz is shown in the inset.

the resistance of Hafnium oxide-based RRAM at different voltages applied at the TE. The resistance switching of RRAM involves three elementary processes such as formation, SET, and RESET.

The forming voltage is applied across the electrodes to create an electric field in the oxide material. Oxygen atoms are knocked out of oxide material forming oxygen vacancies under the influence of high electric field, typically as high as 10 MV/cm [Fig. 1(b)]. The conduction through the CF is primarily due to the transportation mechanism of electrons in these oxygen vacancies termed as trap-assisted tunneling (TAT). After the process of forming, the resistance of the RRAM is at the lowest (LRS). The resistance in LRS depends on the current compliance as shown in characteristic plot in Fig. 1(a). The SET process is the same as forming except that only a part of CF is recovered as compared to forming process [Fig. 1(b)]. Also, SET is performed following a RESET process and SET voltage depends linearly on the RESET voltage [11], [15]. The process of setting state to HRS state is called RESET process. During RESET, the oxygen ions drifted to the anode return to the bulk to combine with the oxygen vacancies or oxidize the metal precipitates. The rate of reoxidation depends on the magnitude of the RESET voltage [11].

C. Source of Randomness in RRAM and Model

RRAM shows the intradevice temporal variations in switching process. HRS and LRS vary cycle to cycle [11] and the resistance after switching depends on the generation and recombination of oxygen vacancies. This is stochastic process induced by the electric field and the temperature of the oxide under the applied switching voltage [11], [15]. In the proposed TRNG, there are three major sources of entropy, namely, circuit and device noise of CMOS RO, RTN of RRAM, and resistance switching probability of RRAM. Model is based on TiN/Ti/HfOx/TiN RRAM device having a physical oxide thickness t_{ox} of 5 nm. We have used RRAM model based on the experimental data and model fitting from [11] and [22]. Methodology of the proposed TRNG design remain the same given inherent switching variations and RTN noise of RRAM irrespective of the device type and respective models used in the simulation framework. Establishing a consistent model for RRAM-based design is different research problems, which is out of the scope of this paper.

1) Cycle-to-Cycle Variation: We have used the parameters and the equations to model cycle-to-cycle switching variations in RRAM from [11] which are calibrated with experimental data. Current compliance of 100 μA is used for modeling the SET resistance. RESET process is performed by negative ramp voltage and the differential barrier length with the voltage modeled. RESET is a thermally activated process. The temperature increases with the electric power and overcomes the activation energy to switch the state of the device. Switching of the device at an applied RESET voltage is probabilistic activity [15], [24]. To model the variation in the resistance of the RRAM due to defects in the oxide material, we assume Gaussian distribution in the SET resistance of RRAM with the variance of 0.08 [11]. The RESET resistance is calculated by assuming Gaussian distribution of the proportionality coefficient Cxv with variance of 0.034. Cxv models the stochastic variation in the CF rupturing process due to recombination of oxygen vacancies with the ions [10], [11] from cycle to cycle. Due to exponential dependence of RESET resistance on the barrier length, HRS exhibits lognormal distribution characteristics.

2) Random Telegraph Noise: Conduction in the RRAM is explained by TAT of electrons in CF. Due to random distribution in the TAT supporting defects, the current through the RRAM shows stochastic variations with time. The phenomenon responsible for RTN is explained by Balatti *et al.* [17], Puglisi *et al.* [18], and Tseng *et al.* [19] as charging and discharging of the traps at or close to the surface of the CF [Fig. 2(a)]. In Fig. 2(a), red dots show the deactivated traps due to the trapping of electrons. When the trapped electrons are emitted/released, the traps are activated for electron conduction to increase the dc current through RRAM temporarily till the trap is deactivated. Also, the frequency of trap charging increases with the bias voltage (voltage across RRAM) and temperature due to local Joule heating of CF. The trapping/emission time of the defects near the CF junction can be modeled as lognormal distribution [17], [19]. RTN results in current fluctuations through the RRAM with time. However, the relation of the trapping/emission times with the

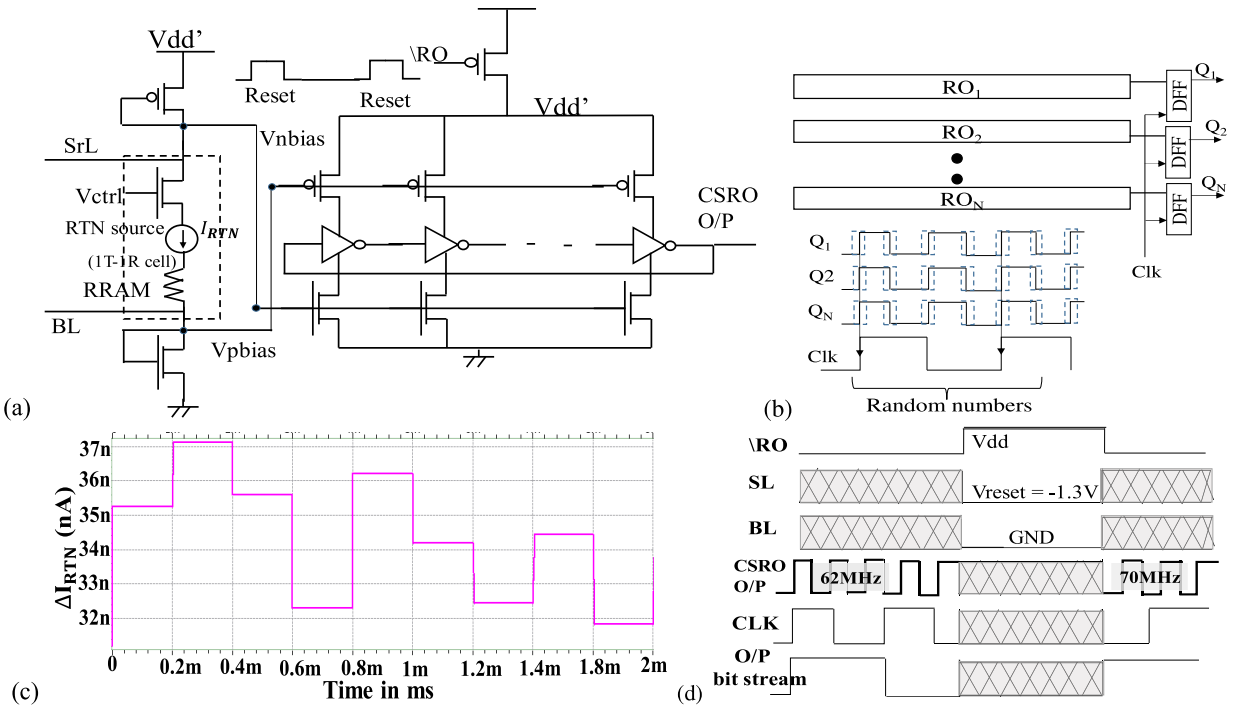

Fig. 3. (a) Proposed TRNG circuit based on CSRO. (b) Illustration of N-bit TRNG. (c) ΔI through RRAM due to RTN with a frequency of 5 kHz. (d) Reconfiguration of TRNG.

current fluctuation is still unclear and is determined to be randomly distributed. Variation of RTN current directly related to the fluctuations of current through RRAM [22]. Essentially, RTN is a multilevel low-frequency noise in the RRAM of kilohertz range. RTN can be characterized by the factorial hidden Markov model [25] by superposing the multiple two-level RTNs. However, this does not provide a deterministic circuit model that could be adapted for circuit analysis. RTN being a truly random process in RRAM, leading to read current fluctuations, exhibits no deterministic behavior which could be modeled without direct access to RRAM. Modeling RTN as normal distribution component in RRAM current is the simplistic model for circuit analysis. In this paper, multilevel RTN in HRS state (RESET) is modeled as variable current source [I_{RTN} through RRAM in Fig. 3(a)] with 20%–30% variation in the steady current in HRS state by fitting in normal distribution curve shown in Fig. 2(b) [18], [19], [22]. This RTN model follows the RTN current measurements in [22]. The frequency of current fluctuation is affected by the temperature which is due to longer trapping and emission periods of electron at lower temperature compared to those at higher temperatures [19]. Therefore, available entropy due to RTN is temperature dependent.

III. Proposed TRNG

In this section, we describe the proposed TRNG and perform qualitative and quantitative analysis.

A. Details of the Proposed TRNG

The proposed TRNG based on CSRO is shown in Fig. 3(a). The delay of the inverters in CSRO can be controlled to adjust the frequency of oscillations. The principle of delay control is based on current starving of the inverters by controlling the gate voltage of the additional control transistors [16] stacked in nMOS and pMOS networks, respectively [Fig. 3(a)]. Gate voltages of these two series transistors are derived from a bias circuit. In the proposed TRNG, we embed RRAM in the bias circuit to control the gate bias voltages randomly as dictated by the RTN and cycle-to-cycle switching variations of RRAM.

The bias circuit is shown in the inset of Fig. 3(a). It consists of RRAM and access transistor (1T-1R structure) for programing the RRAM as required. nMOS is sized to carry the compliance current of RRAM. V_{ctrl} of the access transistor can be connected to constant voltage greater than threshold voltage of nMOS device during normal TRNG operation. Frequency of programing depends on the quality of bit stream generated with time and RRAM switching speed. In 22-nm technology, the width of the diode connected transistors ($W_p = 400$ nm and $W_n = 200$ nm) in the bias network is chosen to keep the voltage across the RRAM below 300 mV under the highest HRS of the RRAM (3 MΩ). The operating voltage of the TRNG is 1 V. We assume worst case conditions for voltage drop estimation across RRAM for process variation tolerance. The bias voltages $V_{n\text{bias}}$ vary in proportion to current through the RRAM. $V_{p\text{bias}}$ varies in negative correlation with RRAM current and $V_{n\text{bias}}$, i.e., $V_{p\text{bias}} = V_{\text{dd}} - (V_{\text{RRAM}} + V_{\text{ds}}(V_{\text{ctrl}}) + V_{n\text{bias}})$. $V_{n\text{bias}}$ and $V_{p\text{bias}}$ voltage nodes in the bias network are such that the voltage at $V_{n\text{bias}}$ varies around V_{dd} and $V_{p\text{bias}}$ varies around ground voltage. This ensures that nMOS and pMOS in the inverters chain are operated around respective threshold voltages. Variation of CSRO frequency with the current through RRAM is explained as follows. When the RTN

current increases, the current through the bias network, V_{ds}, of diode-connected nMOS increases proportionally by increasing $V_{p\text{bias}}$ node voltage. At the same time, $V_{n\text{bias}}$ decreases proportionally. Because of increasing the pMOS gate voltage and decreasing the nMOS gate voltage in the inverter stack, delay of the inverters increases. Consequently, the frequency of the oscillator decreases. Thus, current variations in RRAM due to RTN induce respective differential change in $V_{p\text{bias}}$ and $V_{n\text{bias}}$. Differential change in bias voltages in turn changes the delay of the inverter chain and, thus, the frequency of the CSRO. It should be noted that the direction of inverter delay differential depends on the net effect of strength of pMOS and nMOS delay control transistors and bias voltages. In this paper, we have used 2:1 ratio for pMOS to nMOS sizing. The speed of inverters varies in the direction of $V_{p\text{bias}}$. These variations are stochastic in nature, and thus, data sampled by the sampling clock is random due to stochastic variations in operation of CSRO. The output of multiple ROs is provided to D-flip flops which sample the outputs using a sampling clock, as shown in Fig. 3(b).

B. Sampling Frequency

Sampling frequency determines the rate of generation of random numbers. Minimum sampling frequency is dictated by the frequency of oscillations generated by CSRO. Sampling frequency must be selected at least half of that of CSRO oscillations to avoid the duplication of the bits in the random bit stream. Theoretically, sampling frequency up to several megahertz can be selected for CSRO oscillations greater than 10 MHz. We have selected 6 MHz of sampling frequency for the CSRO oscillations in ∼60–70-MHz range. Sampling frequency can be selected during the time of design by estimating the frequency of CSRO from the initial delay of the inverters. Sampling frequency can also be selected dynamically to improve the statistical properties of the bit stream [26]. A technique based on built-in self-test (BIST) is proposed to measure the statistical properties of RO-based TRNGs in [26]. However, it requires on-chip clock generator with dynamically adjustable frequency and BIST with logic for testing statistical properties which adds to the design complexity and additional cost. Frequency of random bit stream is theoretically limited by the number of inverter stages in the CSRO and frequency of the various sources of entropy in the TRNG. Circuit and device noise depends on bias voltage, temperature, junction capacitance of MOS devices and scales proportionally with the number of stages in a single-ended CSRO [27]. To achieve synergistic effect of circuit noise and RTN and high-speed generation of random numbers, we limit the sampling frequency in the range of megahertz.

C. Configurability

Frequency of CSRO can be dynamically configured by altering the parameters in the bias circuit, which varies the current starved by the delay inverters. For this purpose, we embed 1T-1R cell in the bias circuit. Due to exponential dependency of HRS current on the barrier length, a small change in the barrier length manifests as significant change in the resistance unlike in LRS where current is linearly dependent on the barrier length. HRS exhibits the higher cycle-to-cycle variability and RTN compared to LRS [18], [19]. Therefore, in the design, we RESET the RRAM for reconfiguration. It should be noted that the cycle-to-cycle switching parameter variations and RTN are uncorrelated but concurrent in nature [18].

For programing the RRAM, we halt the operation of CSRO by power gating pMOS transistor connected to power supply. Power gating transistor is driven by a pulse with pulse of width equal to the write time. The nMOS transistor controlled by V_{ctrl} in the regular operation is connected to V_{dd} or a constant voltage. RRAM is RESET by applying V_{reset} ramp voltage of -1.3 V across the electrodes from SL and BL signals. RRAM demonstrates switching time of ∼10 ns which adds penalty of one cycle with CSRO frequency of up to 100 MHz in the worst possible scenario of write conditions. The primary advantage of the proposed TRNG over other RRAM-based TRNGs is high-speed generation [9], [10] of random bit stream and the frequency of TRNG is independent of the write time of RRAM [10]. By choosing a reset voltage at probabilistic switching voltage and using probability switching model of RRAM, the entropy can be further improved. The reconfiguration feature can also be exploited to recover from adversary attacks by generating new random numbers. However, this requires additional circuitry to detect the adversary attacks and activate the write operation of RRAM. The TRNG is reconfigured in regular intervals after generating a set of random numbers (10^6) under default conditions without any assistance to detect adversary attacks for simplicity of the solution. In a secure environment, TRNG can be operated without RESET operation for generating millions of random numbers and reconfigured to generate a new set of random numbers.

By applying the probabilistic switching voltage instead of RESET voltage -1.3 V the RRAM undergoes probabilistic switching. The RRAM remains RESET or changes to SET state by the applied switching voltage [23], [24]. This kind of switching could improve the randomness in the oscillator frequency further after configuration. This is out of the scope of this paper and will be explored in our future work.

The frequency of programing pulse is at least few 1000 times slower than frequency of CSRO oscillations. Typically, TRNG is configured after generating a few sets of random numbers. Within a single configuration cycle, circuit noise and RTN acts as a source of randomness to generate jitter in oscillations. Between different configuration cycles RRAM switching parameters' variation and respective RTN synergistically contribute to entropy in the TRNG system. Table I presents the comparative analysis of the proposed TRNG with other spintronics and RRAM TRNGs. Power consumed in the proposed TRNG is to operate a CSRO and RRAM switching for configuration which is comparable to the power consumed by RRAM memory in an Internet of Things (IoT) device. Area of the proposed TRNG is larger compared to other RRAM switching-based TRNGs which consume power in analog-to-digital converters and comparators requiring more power and design time. Therefore, the proposed TRNG is suitable for security primitives in embedded systems and IoT for secure design.

TABLE I
COMPARATIVE ANALYSIS OF TRNG

	Methodology	Source of entropy	Speed	Advantages	Drawbacks
Spin dice [4] Perturb and tracking [5]	Reset and probabilistic switching voltage for programming. And [5] eliminates reset every cycle conditionally from the previous o/p sample. **3 phases**: reset, perturb and read	Probabilistic switching of Magnetic Tunnel Junction (MTJ)	MHz	Ultra-low voltage switching operation of MTJ.	Speed is limited by reset [4], probabilistic switching time; Delay, area and power overhead of tracking system [5].
Balatti et. al. [7]	Stochastic set process by a random set pulse with median of set voltage distribution	Stochastic switching process	kHz-MHz	Broader Resistance distribution compared to MTJ.	Accurate switching voltage control, and Slow switching limits the speed.
Balatti et. al. [8]	Probabilistic switching of a pair of RRAM (series/parallel). **3 phases**: set, reset and read.	RRAM probabilistic switching	~0.16kHz	No biasing of random bit.	Slow switching speed. Requires analog parts: Comparator
Yang et. al. [17]	RRAM cell in bias circuit of sampling frequency oscillator. Biasing circuit uses a current source of >50uA. Requires ~200mV of bias voltage differential.	RTN of RRAM	MHz-GHz	Sample frequency generator based TRNG.	FoM of RRAM requirements are in contrary for storage and TRNG applications. Interface noise and routing congestions would mitigate the effect of RRAM RTN.
Yang et. al. [38]	Counting the number of cycles to reset RRAM indicates variation in write speed	RESET speed variation	kHz-MHz	Write speed variation is reliable source of entropy.	Complex design to count the number of clock cycles for write with feedback. Binary decision based on analog quantity leads to repetition of bits in the random stream.
Proposed TRNG	RRAM current makes the current through the bias network of CSRO which in turn generates phase noise and jitter in the oscillations.	RTN, electronic noise in RO and cycle-cycle reset switching variations	MHz-GHz	High speed not limited by RRAM switching speed. Simple CSRO based design. No analog parts	Requires power gating; and larger area compared other RRAM based TRNGs.

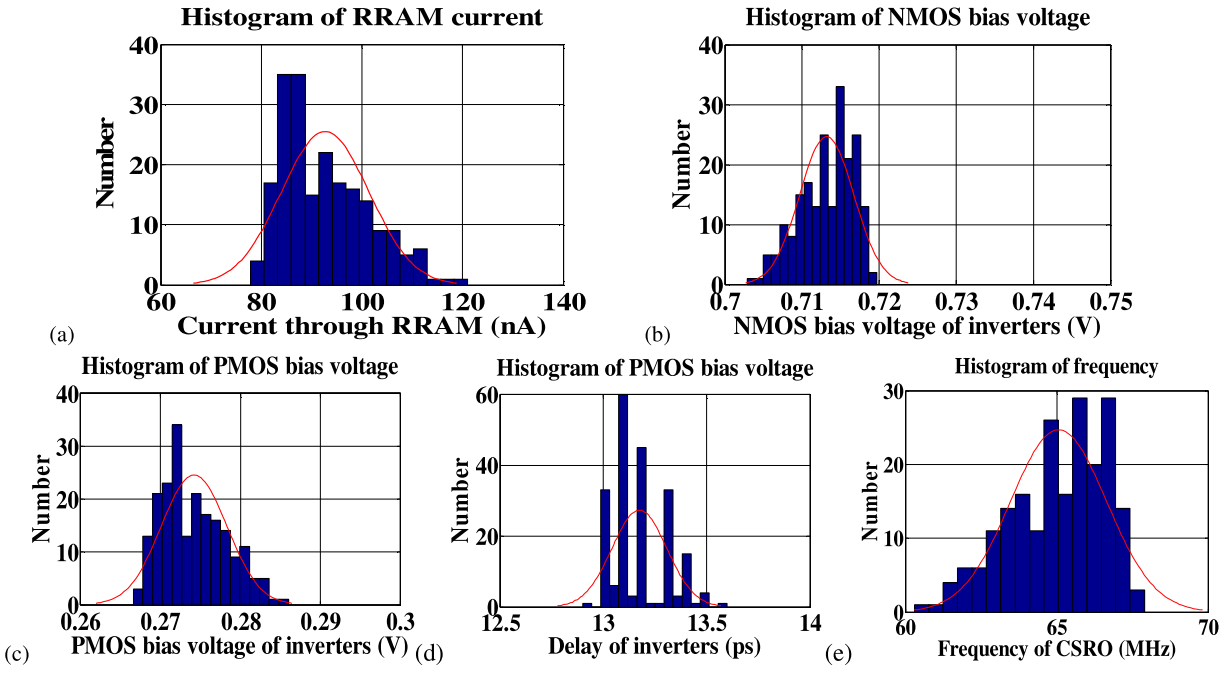

Fig. 4. Histogram over 200 configuration cycles. (a) RRAM current. (b) nMOS bias voltage. (c) pMOS bias voltage. (d) Delay of inverters. (e) Frequency of CSRO.

IV. SIMULATION RESULTS

We present the simulation results of the proposed TRNG using 22-nm PTM models of MOS transistors and Verilog-A model of RRAM (Section II). Fig. 4 shows the histogram of RRAM current, bias voltages and delay of inverters, and frequency of a CSRO at 200 different RESET cycles for reconfiguring the resistance of RRAM. The current through RRAM in different RESET cycles varies from ~83 to 116 nA

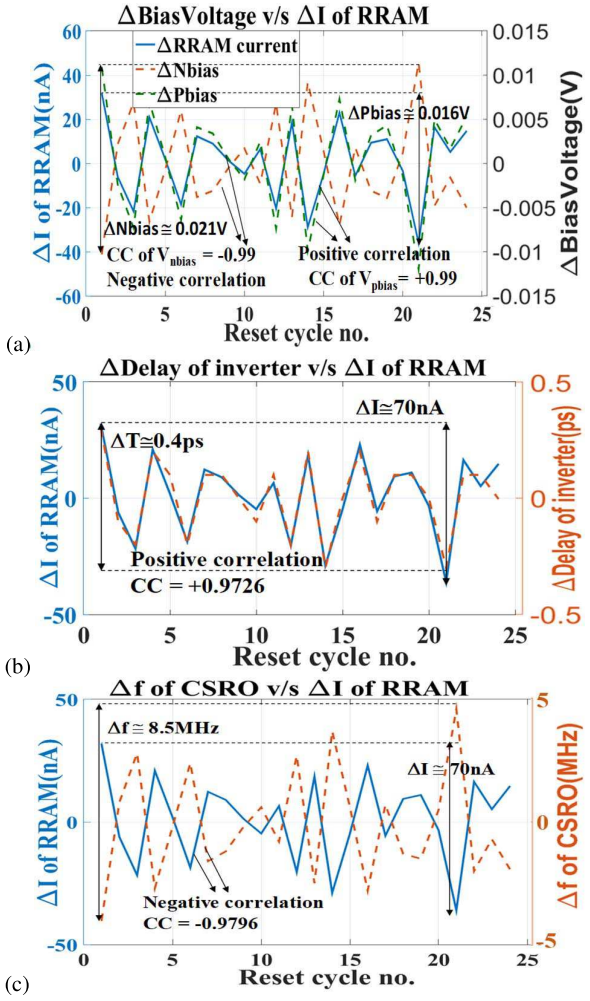

Fig. 5. Correlation plot of current through RRAM over 25 RESET configuration cycles with (a) bias voltages, (b) delay of inverter, and (c) frequency of oscillations.

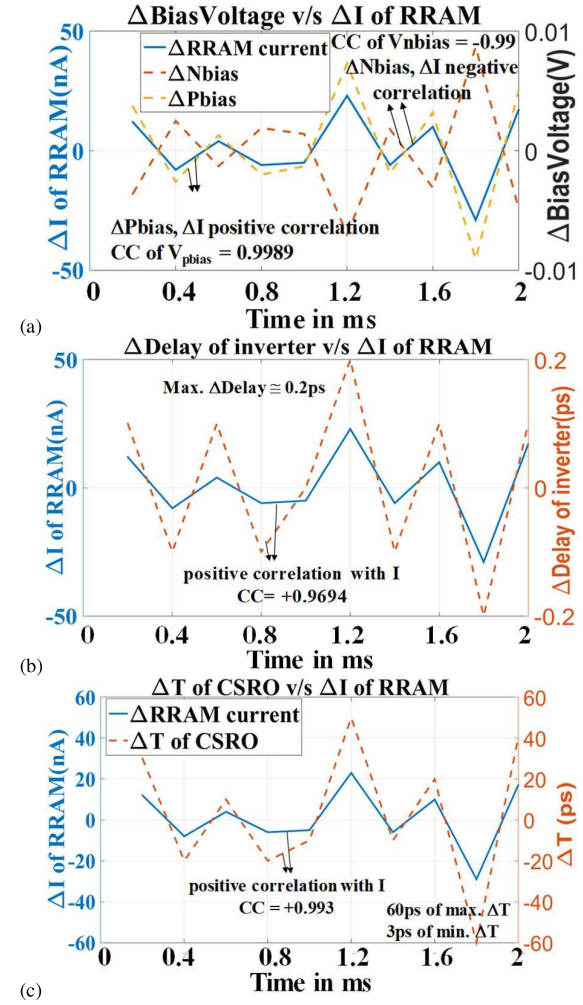

Fig. 6. Differential current through RRAM resulting differential change in (a) bias voltages, (b) delay of inverters, and (c) jitter (ΔT) in oscillations due to RTN at 30 °C.

in random steps which induce respective differential changes in the delay of inverters. nMOS and pMOS bias voltages [Fig. 4(b) and (c)] undergo differential change in each of the configuration. The delay of the inverter varies in tandem with the current through the RRAM/bias network [Fig. 4(c) and (d)] exhibiting positive correlation. Frequency of CSRO changes in the range of ∼62–67 MHz in unpredictable random steps [Fig. 4(c)]. Also, the frequency of CSRO varies in negative correlation (± and vice versa) with respect to current through RRAM. Positive and negative correlation can be observed from the similar/complement distribution pattern and peaks in the histogram. It can also be observed that the pMOS and nMOS bias voltage differentials vary in negative correlation (± and vice versa) with each other. pMOS bias voltage varies proportional to current through the RRAM [Fig. 5(a)]. The bias voltages demonstrate a differential change of few millivolts (21 and 16 mV), which induce a proportional change in the delay of the inverters and frequency of CSRO. Fig. 5(b) illustrates the cycle-to-cycle differential change in the delay of inverters, and it varies in the direction of $V_{p\text{bias}}$. Fig. 5(c) shows the frequency of CSRO varying with current through the RRAM. As the current through RRAM decreases frequency of CSRO increases and vice versa.

RTN in the RRAM induces jitter in the oscillations which leads to randomness in the bit stream sampled from the CSRO oscillations. To illustrate the effect of RTN on jitter in CSRO oscillations, we have plotted the differential current flow through the RRAM and bias voltages with time [Fig. 6(a)] and the time period of the oscillations with time. Also, differential change in delay of the inverter is plotted due to the respective change in the current through RRAM [Fig. 6(b)]. The delay of inverter changes in the range of ±10–±200 fs exhibiting a maximum differential change of 200 fs. Jitter in the range as low as 3 ps to as high as 60 ps is observed [Fig. 6(c)]. This additional jitter due to RTN in the RRAM acts as a source of randomness to produce the random bit stream when CSRO oscillation is sampled by a clock of stable frequency. Correlation coefficient (CC) in Figs. 5 and 6 shows ∼100% (CC ≈ ±1) negative/positive correlation of the parameters with RRAM current.

We present NIST test results of a 10-bit TRNG with 100 000 random data samples (1 000 000 bits in a stream) to validate the randomness of the data generated. From Fig. 7, it can be noted that the p-value in the NIST tests is greater than 0.01, which indicates sufficient randomness in the generated bit stream.

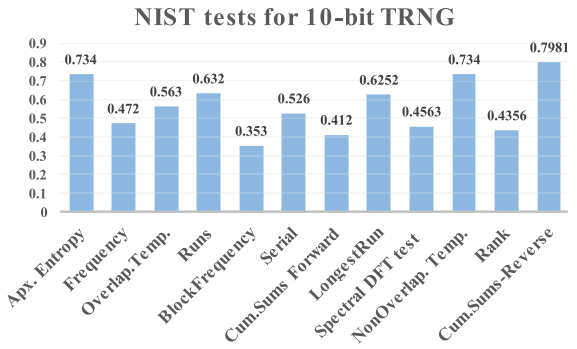

Fig. 7. NIST test results on bit stream from 10-bit TRNG.

V. Adversary Attacks on TRNGs

In this section, we discuss the adversary attacks on the RO-based TRNGs and RRAM-based TRNGs. We also discuss the robustness of the proposed design against these attacks.

A. Background on Attacks and Prevention

Several adversary attacks such as frequency injection attack, attack over the network [28], electromagnetic waves emission based [29], [30], and fault attacks [21], [29]–[32] have been investigated in the literature. Researchers have also proposed techniques such as error correction, induction of nonlinearity in the response [30], attacks detection from frequency, bit stream monitoring, and recovery using *RC* filters closer to the power supply [31], [32] to safeguard against these adversary attacks. Attacks such as frequency injection attacks from the power rails could be avoided by keeping the power rails not accessible to the adversary externally. This can be achieved by deriving the voltage from a dedicated on-chip power supply [33]. In this paper, we focus our discussion on potential attacks on RRAM-based TRNGs.

B. Vulnerabilities: Temperature Sensitivity of RTN

RRAM-based TRNGs are vulnerable to the adversary attacks due to the sensitivity of RRAM characteristics to temperature and voltage. RTN of RRAM is associated with the charge and discharge time of traps in the CF. The frequency of charge and discharge is dependent on the Joule heating of the CF and the ambient temperature. Fig. 8 illustrates the simulation results for the effect of RTN at 5 °C. At cooler temperature, the charge and discharge time of the electrons in the traps are longer which reduces the rate of change of RRAM current [19]. Hence, change in the RRAM current varies at the rate of few hertz (1–4 times/s). Very few traps available are responsible for RTN decreasing the variation range of RRAM current in few nanoampere (ΔI of RRAM is 0.5 nA–1 nA), as shown in Fig. 7. Differential change in the bias voltages, delay, and frequency is reduced by ∼25×, ∼10×, and ∼40×, respectively, compared to the variations at 30 °C. There are flat regions in the frequency plot where the differential change in the frequency is almost zero due to degradation in the entropy available at cooler temperature. Attack model is discussed in Section V-C.

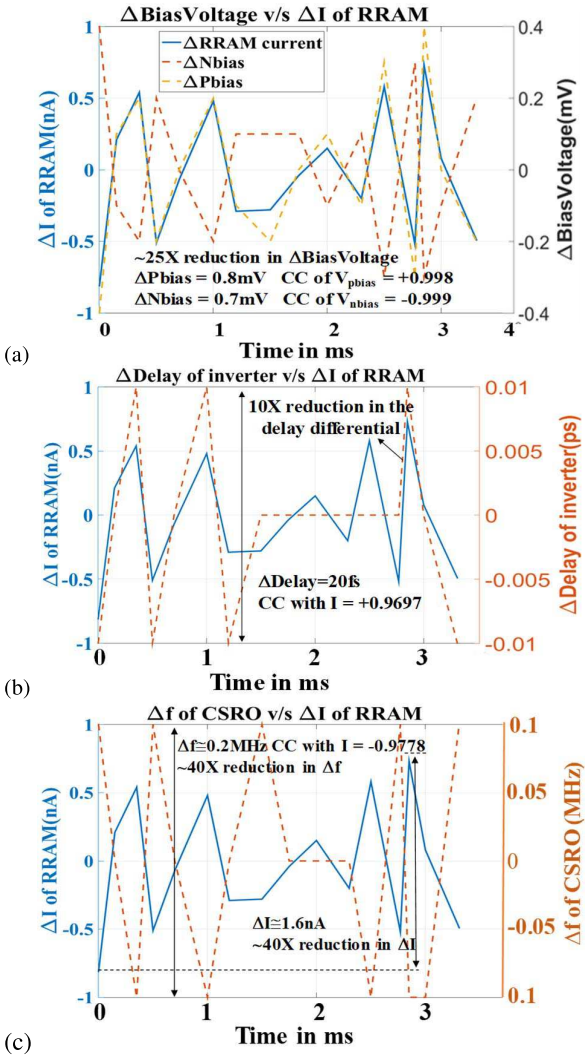

Fig. 8. Differential current through RRAM resulting differential change in (a) bias voltages, (b) delay of inverters, and (c) frequency (Δf) in oscillations due to RTN at 5 °C.

C. Attack Model: Cooling and Model Building

Entropy decreases (due to RTN) considerably at low temperature [19] which affects the quality of random numbers generated. This makes the underlying cryptographic system vulnerable to adversary attacks [2], [3], [34]. Adversary can cool the chip by nitrous oxide and control the temperature of the chip which would eventually affect the entropy of the RRAM-based TRNGs. Although the vulnerability to machine learning (ML)-based model building attacks on TRNGs is still unproven, TRNGs could be vulnerable to model-based attacks similar to physically unclonable functions [35].

D. Countermeasures: Temperature Sensing, Configurability, and Correction

By using an on-chip temperature sensor to sense the ambient temperature and configuring the TRNG at an adaptive frequency depending on the temperature could safeguard against temperature-based attacks. Diode temperature sensor proposed in [36] can be employed for on chip temperature

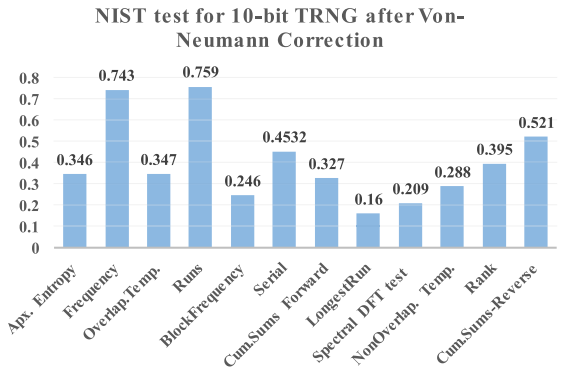

Fig. 9. NIST tests on 10-bit TRNG within RESET cycle when RTN noise is set to zero with Von Neumann correction.

sensing. It should also be noted that RRAM demonstrates lower durability of few million cycles compared to other nonvolatile memories [9] which affects the productive lifetime of TRNG in the security chip under insecure operating environments. Research efforts have been dedicated toward making RRAM a commercial memory device high endurance of $\sim 10^{10}$ cycles and 10 years have been achieved with HfO2/Ti cap bipolar RRAM in HRS and LRS switching [39]. With $\sim 10^{10}$ cycles of configuration, TRNG can be used to generate $\geq 10^{10}$ sets of random number streams. Assuming configuration of 1000 times a day in the lifetime of TRNG offers a lifetime of $\sim 10^{10}/(365 \times 1000) = 27\,397.3$ years. In case of few millions of cycles of endurance $\sim 10^{6}/(365 \times 1000) = 2.739$ years of operation is offered by the proposed TRNG design.

Reconfiguration at regular intervals after generating a few random bit streams makes it almost impossible for the adversary to predict the new CSRO frequency. This can be used to safeguard against the model building attacks. Furthermore, RRAM switching speed adds to speed overhead in such scenario. In the applications where the speed of TRNG renders it useless, the Von Neuman correction technique [37] is employed within RESET cycle to compensate for the reduction in the entropy with configuration frequency of kilohertz. We applied the Von Neumann correction on the bit stream generated after setting the effect of RTN to zero (no RTN current source) in current fluctuations of RRAM on 10-bit TRNG (Fig. 9).

VI. Conclusion

We proposed a high-speed (kilohertz–megahertz), reconfigurable CSRO-based TRNG for on-chip applications. It exploits the RTN low frequency noise in RRAM and cycle-to-cycle switching parameter variations as the source of entropy. We propose a technique to reconfigure the system to recover against adversary attacks. Configurability makes the model building and ML attacks harder. The 10-bit random data stream is validated successfully for sufficient randomness using NIST test suite. The speed of the designed TRNG is 6 MHz and energy/bit is 22.8 fJ.

Acknowledgment

The authors would like to thank Dr. F. M. Puglisi for his help on thorough understanding of the RRAM model.

References

[1] M. Stipčević and Ç. K. Koç, "True random number generators," in *Open Problems in Mathematics and Computational Science*. New York, NY, USA: Springer-Verlag, 2014, pp. 275–315.

[2] A. Arslan, S. Kardas, S. Aldirmaz, and S. Ertürk, "Are RNGs achilles' heel of RFID security and privacy protocols?" Int. Assoc. Cryptol. Res., Tech. Rep. 2016/1130, 2016.

[3] Y. Ma, J. Lin, and J. Jing, "On the entropy of oscillator-based true random number generators," in *Proc. Cryptogr. Track RSA Conf.*, 2017, pp. 165–180.

[4] B. Jun and P. Kocher, "The Intel random number generator," Cryptogr. Res., San Francisco, CA, USA, White Paper, 1999.

[5] Y. Lao, Q. Tang, C. H. Kim, and K. K. Parhi, "Beat frequency detector–based high-speed true random number generators: Statistical modeling and analysis," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 1, 2016, Art. no. 9.

[6] A. Fukushima *et al.*, "Spin dice: A scalable truly random number generator based on spintronics," *Appl. Phys. Exp.*, vol. 7, no. 8, p. 083001, 2014.

[7] W. H. Choi *et al.*, "A magnetic tunnel junction based true random number generator with conditional perturb and real-time output probability tracking," in *IEDM Tech. Dig.*, Dec. 2014, pp. 5–12.

[8] Y. Wang, W. Wen, H. Li, and M. Hu, "A novel true random number generator design leveraging emerging memristor technology," in *Proc. 25th Ed. Great Lakes Symp. VLSI*, 2015, pp. 271–276.

[9] S. Balatti, S. Ambrogio, Z. Wang, and D. Ielmini, "True random number generation by variability of resistive switching in oxide-based devices," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 5, no. 2, pp. 214–221, Jun. 2015.

[10] S. Balatti *et al.*, "Physical unbiased generation of random numbers with coupled resistive switching devices," *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 2029–2035, May 2016.

[11] F. M. Puglisi, P. Pavan, A. Padovani, and L. Larcher, "A compact model of hafnium-oxide-based resistive random access memory," in *Proc. Int. Conf. IC Design Technol. (ICICDT)*, pp. 85–88, May 2013.

[12] G. Niu *et al.*, "Material insights of HfO_2-based integrated 1-transistor-1-resistor resistive random access memory devices processed by batch atomic layer deposition," *Sci. Rep.*, vol. 6, Jun. 2016, Art. no. 28155.

[13] S. Yu, "Resistive random access memory (RRAM)," *Synthesis Lect. Emerg. Eng. Technol.*, vol. 2, no. 5, pp. 1–79, 2016.

[14] D. Ielmini and R. Waser, Eds., *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*. Hoboken, NJ, USA: Wiley, 2015.

[15] S. Yu, X. Guan, and H.-S. P. Wong, "On the switching parameter variation of metal oxide RRAM—Part II: Model corroboration and device design strategy," *IEEE Trans. Electron Devices*, vol. 59, no. 4, pp. 1183–1188, Apr. 2012.

[16] G. S. Jovanović and M. K. Stojčev, "Current starved delay element with symmetric load," *Int. J. Electron.*, vol. 93, no. 3, pp. 167–175, 2006.

[17] S. Balatti, S. Ambrogio, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Voltage-dependent random telegraph noise (RTN) in HfO_x resistive RAM," in *Proc. IEEE Int. Rel. Phys. Symp.*, Jun. 2014, pp. MY.4.1–MY.4.6.

[18] F. M. Puglisi, L. Larcher, P. Pavan, A. Padovani, and G. Bersuker, "Instability of HfO_2 RRAM devices: Comparing RTN and cycling variability," in *Proc. IEEE Int. Rel. Phys. Symp.*, Jun. 2014, pp. MY.5.1–MY.5.5.

[19] Y. H. Tseng *et al.*, "Modeling of electron conduction in contact resistive random access memory devices as random telegraph noise," *J. Appl. Phys.*, vol. 111, no. 7, p. 073701, 2012.

[20] J. Yang *et al.*, "A low cost and high reliability true random number generator based on resistive random access memory," in *Proc. 11th Int. Conf. ASIC (ASICON)*, Nov. 2015, pp. 1–4.

[21] A. T. Markettos and S. W. Moore, "The frequency injection attack on ring-oscillator-based true random number generators," in *Cryptographic Hardware and Embedded Systems—CHES 2009*. Berlin, Germany: Springer-Verlag, 2009, pp. 317–331.

[22] D. Veksler *et al.*, "Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics," in *IEDM Tech. Dig.*, Dec. 2012, pp. 9.6.1–9.6.4.

[23] F. M. Puglisi, L. Larcher, G. Bersuker, A. Padovani, and P. Pavan, "An empirical model for RRAM resistance in low- and high-resistance states," *IEEE Electron Device Lett.*, vol. 34, no. 3, pp. 387–389, Mar. 2013.
[24] A. Chen and M.-R. Lin, "Reset switching probability of resistive switching devices," *IEEE Electron Device Lett.*, vol. 32, no. 5, pp. 590–592, May 2011.
[25] F. M. Puglisi and P. Pavan, "Factorial hidden Markov model analysis of random telegraph noise in resistive random access memories," *ECTI Trans. Electr. Eng., Electron., Commun.*, vol. 12, no. 1, pp. 24–29, 2014.
[26] S. U. Hussain, M. Majzoobi, and F. Koushanfar, "A built-in-self-test scheme for online evaluation of physical unclonable functions and true random number generators," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 2, no. 1, pp. 2–16, Jan./Mar. 2016.
[27] A. Hajimiri, S. Limotyrakis, and T. H. Lee, "Phase noise in multi-gigahertz CMOS ring oscillators," in *Proc. IEEE Custom Integr. Circuits Conf.*, May 1998, pp. 49–52.
[28] A. P. Johnson, R. S. Chakraborty, and D. Mukhopadhyay, "A novel attack on a FPGA based true random number generator," in *Proc. Workshop Embedded Syst. Secur. (WESS)*, 2015, Art. no. 6.
[29] P. Bayon, L. Bossuet, A. Aubert, and V. Fischer, "Fault model of electromagnetic attacks targeting ring oscillator-based true random number generators," *J. Cryptograph. Eng.*, vol. 6, no. 1, pp. 61–74, 2016.
[30] P. Bayon *et al.*, "Contactless electromagnetic active attack on ring oscillator based true random number generator," in *Proc. Int. Workshop Constructive Side-Channel Anal. Secure Design*, 2012, pp. 151–166.
[31] B. Sunar, W. J. Martin, and D. R. Stinson, "A provably secure true random number generator with built-in tolerance to active attacks," *IEEE Trans. Comput.*, vol. 56, no. 1, pp. 109–119, Jan. 2007.
[32] E. Bühl and M. Ihle, "A fault attack robust TRNG," in *Proc. IEEE 18th Int. On-Line Test. Symp. (IOLTS)*, Jun. 2012, pp. 114–117.
[33] S. Köse and E. G. Friedman, "Distributed power network co-design with on-chip power supplies and decoupling capacitors," in *Proc. 13th Int. Workshop Syst. Level Interconnect Predict. Workshop*, Jun. 2011, pp. 1–5.
[34] S. Durrant. (1999). *Random Numbers in Data Security Systems*. [Online]. Available: hup:Z/wu-w.inte
[35] U. Rührmair, F. Sehnke, J. Sölter, G. Dror, S. Devadas, and J. Schmidhuber, "Modeling attacks on physical unclonable functions," in *Proc. 17th ACM Conf. Comput. Commun. Secur.*, 2010, pp. 237–249.
[36] M. Mansoor, I. Haneef, S. Akhtar, A. de Luca, and F. Udrea, "Silicon diode temperature sensors—A review of applications," *Sens. Actuators A, Phys.*, vol. 232, pp. 63–74, Aug. 2015.
[37] S.-H. Kwok, Y.-L. Ee, G. Chew, K. Zheng, K. Khoo, and C.-H. Tan, "A comparison of post-processing techniques for biased random number generators," in *Proc. IFIP Int. Workshop Inf. Secur. Theory Practices*, 2011, pp. 175–190.
[38] J. Yang, Y. Lin, Y. Fu, X. Xue, and B. A. Chen, "A small area and low power true random number generator using write speed variation of oxidebased RRAM for IoT security application," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2017, pp. 1–4.
[39] Y. Y. Chen *et al.*, "Endurance/retention trade-off on $HfO_2/Metal$ cap 1T1R bipolar RRAM," *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 1114–1121, Mar. 2013.
[40] Y. Yoshimoto, Y. Katoh, S. Ogasahara, Z. Wei, and K. Kouno, "A ReRAM-based physically unclonable function with bit error rate< 0.5% after 10 years at 125° C for 40 nm embedded application," in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2016, pp. 1–2.

Rekha Govindaraj received the B.E. degree (honors) from Visweswaraya Technological University, Belgaum, India, in 2009 and the M.Tech. degree from IIT Kharagpur, Kharagpur, India, in 2012. She is currently working toward the Ph.D. degree at the LOGICS Lab, University of South Florida, Tampa, FL, USA.

She was a System On Chip Design Engineer with Qualcomm Inc., for two years. She has authored or coauthored several conference papers and a journal paper. She holds a U.S. patent (U.S. patent 9 543 013). Her current research interests include low-power VLSI circuits and system design.

Ms. Rekha is a Student Member of the National Academy of Inventors USF Chapter, and the Florida Gamma Chapter of Tau Beta Pi, which were offered for her excellent scholastic achievements.

Swaroop Ghosh (SM'13) received the B.E. degree (honors) from IIT Roorkee, Roorkee, India, in 2000, the M.S. degree from the University of Cincinnati, Cincinnati, OH, USA, in 2004, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA, in 2008.

From 2008 to 2012, he was a Senior Research and Development Engineer of Advanced Design, Intel Corporation, where he focused on low-power and robust embedded memory design in scaled technologies. From 2012 to 2016, he was a Faculty Member at the University of South Florida, Tampa, FL, USA. Since 2016, he has been an Assistant Professor at Penn State University, State College, PA, USA. His current research interests include low-power circuits, hardware securities, and digital testing for nanometer technologies.

Dr. Ghosh was a recipient of the DARPA Young Faculty Award in 2015, the ACM SIGDA Outstanding New Faculty Award in 2016, the USF Outstanding Research Achievement Award in 2015, and the College of Engineering Outstanding Research Achievement Award in 2015. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I and a Senior Editorial Board Member of the IEEE JOURNAL OF EMERGING TOPICS ON CIRCUITS AND SYSTEMS (JETCAS). He served as the Lead Guest Editor for the IEEE JETCAS. He has also served in the technical program committees of ACM/IEEE conferences such as DAC, ICCAD, CICC, DATE, ISLPED, GLSVLSI, Nanoarch, and ISQED. He has organized ACM/IEEE DAC Ph.D. Forum in 2015 and 2016.

Srinivas Katkoori (S'95–M'97–SM'03) received the Ph.D. degree from the University of Cincinnati, Cincinnati, OH, USA, in 1998.

He is currently an Associate Professor of Computer Science and Engineering at the University of South Florida, Florida, Tampa, FL, USA. He has authored over 100 peer-reviewed journal and conference papers. He holds one U.S. patent (6 963 217). His current research interests include VLSI design, CAD, high-level synthesis, field-programmable gate array-based synthesis, IC reliability, evolutionary algorithms, and hardware securities.

Dr. Katkoori is a Senior Member of the Association for Computing Machinery (ACM). He was a recipient of the 2001 NSF Career Award, the 2002–2003 USF Outstanding Faculty Research Achievement Award, the 2005 Outstanding Engineering Educator Award from the IEEE Florida Council (Region 3), the 2007–2008 USF Undergraduate Teaching Award, and the 2013 USF Jerome Krivanek Distinguished Teacher Award. Besides NSF, his research sponsors include Honeywell, NASA JPL, Department of Defense, Florida DOT, and Florida High Tech Corridor Funding. He serves on technical committees of several VLSI conferences and is a peer reviewer for many VLSI journals. Two papers he has coauthored were nominated for best paper awards at 2003 ASPDAC and 2014 IFIP/IEEE VLSI SOC conferences. He served on ACM SIGDA Board from 2010 to 2013 as a Treasurer and an Associate Editor for the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS from 2006 to 2010, and since 2015, he has been serving as the Vice-Chair of the IFIP Working Group (10.5).