An All-Digital Edge Racing True Random Number Generator Robust Against PVT Variations

Kaiyuan Yang, *Student Member, IEEE*, David Blaauw, *Fellow, IEEE*, and Dennis Sylvester, *Fellow, IEEE*

***Abstract*—This paper presents an all-digital true random number generator (TRNG) harvesting entropy from the collapse of two edges injected into one even-stage ring, fabricated in 40 and 180 nm CMOS technologies. A configurable ring and tuning loop provides robustness across a wide range of temperature (−40 °C to 120 °C), voltage (0.6 to 0.9 V), process variation, and external attack. The dynamic tuning loop automatically configures the ring to meet a sufficient collapse time, thereby maximizing entropy. Measured random bits pass all NIST randomness tests across all measured operating conditions and power supply attacks. In 40 nm, the TRNG occupies only 836 μm^2 and consumes 23 pJ/bit at nominal 0.9 V and 11 pJ/bit at 0.6 V.**

***Index Terms*—Cryptography, frequency collapse, model, noise, oscillator, PVT variation, security, true random number generator (TRNG).**

I. Introduction

SECURITY becomes one of the major concerns with the explosion of connected devices and the advent of cloud computing and Internet of Things. High entropy random number is an essential component for information security, which forms the foundation for many cryptographic algorithms used to build cryptosystems. Some common applications are private key for encryption and cryptographic nonces for authentication.

For higher security level in most cryptosystems, true random number generator (TRNG) harvesting entropy from physical sources are preferred over pseudo-random number generator (PRNG) that has a fixed pattern. On-chip TRNG is important for system miniaturization and device noise provides a good entropy source for circuit designers. There have been a variety of designs extracting random number from device noise in literature. The conventional method is to amplify noise directly with a high-gain and high-bandwidth amplifier followed by quantization [1]–[3]. Resistor thermal noise [1], oxide trap noise [2], and SiN device noise [3] have been employed as entropy sources in this scheme. These designs require careful calibrations of the amplifier and ADC to remove bias in generated random numbers. Extensive use of analog designs also makes them less attractive in terms of system integration and technology portability.

Manuscript received September 10, 2015; revised November 12, 2015; accepted December 08, 2015. Date of publication March 01, 2016; date of current version March 29, 2016. This paper was approved by Guest Editor Masato Motomura. This work was supported by the TSMC University Shuttle Program for chip fabrication and NSF.

The authors are with the University of Michigan, Ann Arbor, MI 48109 USA (e-mail: kaiyuan@umich.edu; blaauw@umich.edu; dmcs@umich.edu).

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/JSSC.2016.2519383

Digital TRNGs offer the advantages of easy integration and lower sensitivity to process, voltage, and temperature variations (PVT variation) over conventional analog designs [4]. For mobile and IoT applications, robustness to environmental variations becomes even more critical. Previous works have demonstrated digital TRNGs based on metastability [4], [5], oscillator jitter [6]–[10], and other device noise (e.g., time to oxide breakdown [11]). Metastability-based methods using cross-coupled inverters provide excellent operating frequency and power efficiency, but often require extensive design efforts and run-time calibration to remove systematic and temporal mismatch in devices which are sensitive to environmental variations [4], [5]. A soft oxide breakdown-based TRNG provides high entropy random bits but suffers from low performance and low power efficiency due to the nature of the entropy source [11]. Ring oscillator (RO) jitter-based TRNGs offer design simplicity and portability. Conventional methods using a slow RO to sample a jittery fast RO provide relatively low entropy and low performance due to limited jitter in a single digital RO [6]. This design is also vulnerable to power supply attacks as described in [12]. Efforts to increase entropy of RO-based TRNG include combining outputs of several parallel ROs [7], chaotic ROs with multiple feedback paths (FIRO and GARO) [13], and including a dynamic duty cycle tuning loop to remove bias in outputs [10]. Recent RO-based TRNGs employ new random bit extraction schemes like measuring time for a third-harmonic RO to collapse to fundamental frequency [8] and beat frequency between two ROs running at close frequencies [9]. These new schemes provide better randomness and performance thanks to the new jitter amplification approaches, but robustness was not verified across PVT conditions and could pose difficulties to their applications.

To alleviate the issues of PVT variations, this work presents an all-digital edge racing TRNG based on the collapse time of two racing edges in an even-stage RO with automatic tuning loop, demonstrating extensive robustness against PVT variations and intentional power supply attacks [14]. The usage of oscillation collapse time in an even-stage RO provides three benefits. 1) Easy detection of collapse event: no phase detector is needed and thus there are less nonidealities. 2) Average collapse time is naturally an indicator of the operation condition of the TRNG, on which the automatic tuning loop is based. 3) Tuning does not introduce bias into output bits and a relatively wide target range is acceptable; this eliminates the need for high-resolution tuning and minimizing design complexity and cost. The TRNG has been fabricated in 40 nm CMOS demonstrating 2 Mb/s and 23 pJ/b at nominal 0.9 V while passing all 15 NIST randomness tests across wide operating

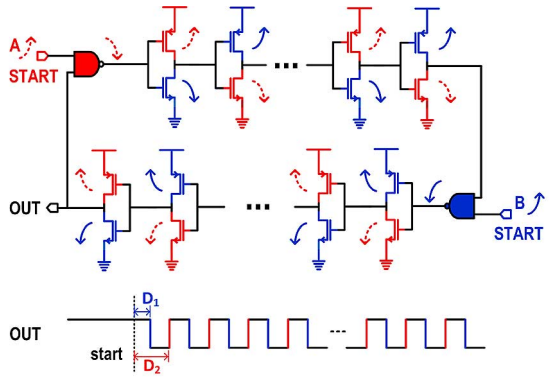


Fig. 1. Concept of TRNG based on frequency collapse of edge racing RO.

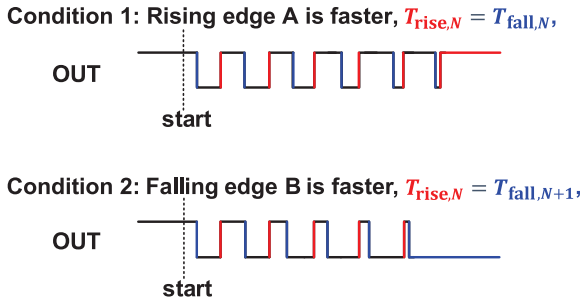


Fig. 2. Even-stage RO collapse conditions and waveforms.

conditions ($-40\,^{\circ}$C to 120 $^{\circ}$C and 0.6 to 0.9 V). A second prototype in 180 nm demonstrates its portability to an older technology commonly used for ultra-low-power applications such as sensor nodes.

This paper is organized as follows. The concept of using frequency collapse in an even-stage RO as entropy source for true random number generation is described in Section II. A mathematical model of the entropy source is also provided in Section II. Detailed implementation of the TRNG prototype and automatic tuning loop against PVT variations are described in Section III. Measurement results of both 40 and 180 nm test chips are provided in Section IV. Finally, this paper is concluded in Section V.

II. Frequency Collapse-Based TRNG

A. Analytical Model of Frequency Collapse in an Even-Stage RO

The main concept of the all-digital PVT-tolerant edge racing TRNG is using the frequency collapse time in an even-stage RO (Fig. 1) as entropy source. Two edges (A, B) are injected through NAND gates into opposite nodes of an even-stage RO simultaneously. Because of even number of stages, "A" is always rising at OUT port while "B" is always falling. For CMOS inverters, rising delay and falling delay are separated and can be changed by process variations. As shown by the arrows in Fig. 1, the two injected edges travel entirely different paths through the ring. Taking device mismatch and random noise into consideration, the time for two edges to travel around the ring are separate accumulations of ideal delay, delay mismatch, and noise. The time points of Nth rising and falling edges at OUT port of an RO with S stages can be expressed as

$$T_{\text{fall},N} = D_1 + \sum_{n=1}^{N}\sum_{i=1}^{S}\left(\text{Ideal\_Delay}_{i,B} + \Delta\text{Delay}_{i,B} + \text{Jitter}_{n,i,B}\right) \quad (1)$$

$$T_{\text{rise},N} = D_2 + \sum_{n=1}^{N}\sum_{i=1}^{S}\left(\text{Ideal\_Delay}_{i,A} + \Delta\text{Delay}_{i,A} + \text{Jitter}_{n,i,A}\right) \quad (2)$$

where D_1 and D_2 are the time for edges "B" and "A" to reach OUT after start (Fig. 1); $\text{Ideal\_Delay}_{i,B}$ and $\text{Ideal\_Delay}_{i,\text{A}}$ are the ideal delay of stage i for edges "B" and "A" not considering process variation and noise; $\Delta\text{Delay}_{i,A}$ and $\Delta\text{Delay}_{i,B}$ are the delay differences in addition to ideal delay due to process variation at stage i for the corresponding edge; $\text{Jitter}_{n,\,i,\,A}$ and $\text{Jitter}_{n,\,i,\,B}$ represent the random delay caused by device noise at stage i during nth iteration of the corresponding edge. Jitter is usually modeled as a random variable following normal distribution $\mathcal{N}\left(0,\,\sigma^2\right)$. As modeled in [15], the variance in inverter delay due to white noise can be expressed as

$$\sigma^2 = \frac{4kT\gamma_N t_{dN}}{I_N\left(V_{DD} - V_t\right)} + \frac{kTC}{{I_N}^2} \quad (3)$$

where t_{dN} is the window that noise is integrated during output transition, I_N is the charging/discharging current, V_t is the threshold voltage, γ_N is a technology-dependent noise coefficient, C is the loading capacitor of the inverter, and k is the Boltzmann constant. The last two terms in (1) and (2) represent nonidealities of the RO and cause one edge to travel faster than the other, thus overtaking the other and collapsing the oscillation. There are two possible collapse conditions depending on the relative amount of delay added to the two edges by process variation (Fig. 2), which can be written as

$$T_{\text{rise},N} = T_{\text{fall},N},\ \text{if } A \text{ is faster than } B \quad (4)$$

$$T_{\text{rise},N} = T_{\text{fall},N+1},\ \text{if } B \text{ is faster than } A. \quad (5)$$

Substitute (1) and (2) into (4) and (5), and considering the fact that $\sum_{i=1}^{S}\text{Ideal\_Delay}_{i,A}$ and $\sum_{i=1}^{S}\text{Ideal\_Delay}_{i,B}$ are identical, the collapse conditions can be expressed as

$$D_2 - D_1 = N \times \sum_{i=1}^{S}\left(\Delta\text{Delay}_{i,B} - \Delta\text{Delay}_{i,A}\right) + \sum_{n=1}^{N}\sum_{i=1}^{S}\left(\text{Jitter}_{n,i,B} - \text{Jitter}_{n,i,A}\right) \quad (6)$$

$$D_2 - D_1 - \sum_{i=1}^{S}\left(\Delta\text{Delay}_{i,B}\right) - \sum_{i=1}^{S}\left(\text{Jitter}_{N+1,i,B}\right) = N \times \sum_{i=1}^{S}\left(\Delta\text{Delay}_{i,B} - \Delta\text{Delay}_{i,A}\right) + \sum_{n=1}^{N}\sum_{i=1}^{S}\left(\text{Jitter}_{n,i,B} - \text{Jitter}_{n,i,A}\right). \quad (7)$$

In (6) and (7), left sides are constant for a given run, the first term on the right side is linear with number of cycles while the second term is accumulation of a normally distributed random variable. The right side can be viewed as a Gaussian random walk with drift in probability theory and, therefore, the model of the collapse event becomes the first-hitting-time model. As a result, the first-hitting-time (collapse time in our case) follows inverse Gaussian distribution [16]. Mean and variance of the

time can, therefore, be derived from the model. Results of both conditions can be expressed in a unified form using absolute values, as shown below

$$\text{mean} = \frac{|D_2 - D_1|}{\left|\sum_{i=1}^{S} \left(\Delta\text{Delay}_{i,B} - \Delta\text{Delay}_{i,A}\right)\right|} \tag{8}$$

$$\text{variance} = \frac{\text{mean}^3}{\lambda}, \lambda = \frac{(D_2 - D_1)^2}{2S\sigma^2} \tag{9}$$

where λ is the shape parameter of the distribution and larger λ indicates less skewness. According to this model, the number of cycles until collapse depends on both systematic delay mismatch and random jitter.

It should be noted that the model above does not consider supply noise, which could be very difficult to precisely model as described in [15]. Here, we consider only low-frequency supply noise from power source or other circuits on chip. Following the analysis in [15], supply noise adds correlated delay variations to all inverters in RO and can be viewed as an additional correlated variation to the $\Delta\text{Delay}_{i,A}$ and $\Delta\text{Delay}_{i,B}$ terms in (1) and (2). However, since the supply variation is common for all stages, variation in the delay difference between the two edges is not significant, which results in small fluctuations in the mean value of collapse time in (8). Despite the modulating of average cycles to collapse by supply noise, the cycles to collapse of a given run still follows inverse Gaussian distribution caused by thermal noise.

B. Systematic Mismatch Versus Random Jitter

As indicated by (8) and (9), the distribution of collapse time depends on the relative magnitude of systematic mismatch and random jitter. If systematic mismatch is small, noise will have a more significant impact, resulting in a longer collapse time with wider distribution. In this case, random bits can be obtained from the RO by recording the number of cycles to collapse. On the other hand, under large systematic mismatch, the RO will collapse in a few cycles with negligible variation. Such systematic behavior is unique for each die and, therefore, can be used to produce a chip ID or PUF [17] but is not useful for extracting random bits.

Typically, systematic mismatch dominates in an even-stage RO and makes entropy extraction difficult. Hence, RO-based TRNGs have previously employed an odd-stage number RO where mismatch naturally cancels out [8]. However, we will show in Section III-A that, in fact, the process variation in an even-stage RO can be used as a natural source of tunability to enable a highly adaptive TRNG design that is robust to a wide range of environmental and other factors using an automatic tuning loop.

The relationship between process variation and tunability can be explained by our proposed model as well. $\Delta\text{Delay}_{i,A}$ and $\Delta\text{Delay}_{i,B}$ are the delay differences of each stage due to process variations, which follow independent normal distribution ($\mathcal{N}\left(0, \sigma^2_{\text{variation}}\right)$). Therefore, the denominator term $\sum_{i=1}^{S}\left(\Delta\text{Delay}_{i,B} - \Delta\text{Delay}_{i,A}\right)$ in (8) also follows normal distribution ($\mathcal{N}\left(0, \text{S} \times \sigma^2_{\text{variation}}\right)$). The process variation results

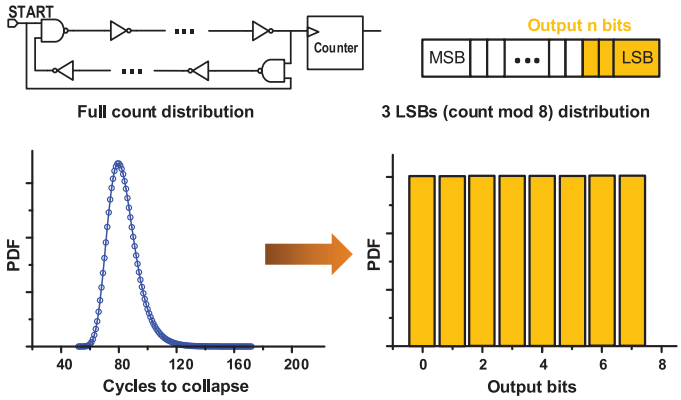


Fig. 3. Random bit generation from collapse time.

in a wide distribution of the mean value in (8), which forms the foundation for our tuning method based on device mismatch. Distribution of the inverse of a normally distributed random variable ($y = \frac{1}{x}, x \sim \mathcal{N}(0, {\sigma_0}^2)$ is

$$\text{pdf}\,(y) = \frac{e^{-\left(\frac{2{\sigma_0}^2}{y^2}\right)}}{\sqrt{2\pi}\sigma_0 \cdot y^2}. \tag{10}$$

Combining (8) and (10), distribution of the average collapse time in (8) across process variations can be calculated as

$$\text{pdf}\,(x_{\text{mean}}) = \frac{e^{-\left(\frac{2|D_2 - D_1|S{\sigma_{\text{variation}}}^2}{{x_{\text{mean}}}^2}\right)}}{\sqrt{2\pi S}\sigma_{\text{variation}} \cdot {x_{\text{mean}}}^2}, x_{\text{mean}} > 0. \tag{11}$$

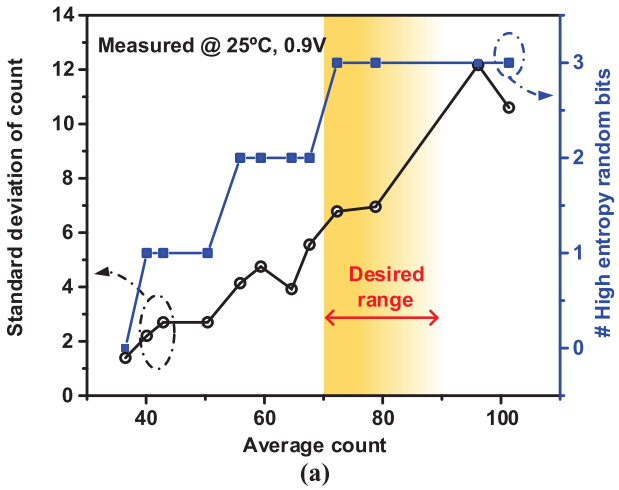
This distribution is highly skewed with long tails and matches the measured distributions in Section IV-B. The peak of the distribution occurs at

$$x_{\text{peak}} = \frac{|D_2 - D_1|}{\sqrt{2\pi S}\sigma_{\text{variation}}} \approx \frac{S \cdot \text{Ideal\_Delay}}{2 \cdot \sigma_{\text{variation}} \cdot \sqrt{2\pi S}} = \frac{S}{2 \cdot \text{Ratio} \cdot \sqrt{2\pi S}} \propto \sqrt{S} \tag{12}$$

where Ratio is the spread (σ/μ) of single stage delay in the oscillator due to local process variations. As discussed previously, we need small systematic variation for random number generation which corresponds to larger average collapse time. A design implication from (12) is that having more stages in RO increases our chance to get an RO with small enough systematic variation for TRNG.

C. Extracting Random Bits From Collapse Time

The analytical model above characterizes the behavior of oscillation collapse in an even-stage RO and indicates that we can get larger variations in collapse time with less systematic variation in RO. To use the frequency collapse concept as entropy source for TRNG, the last step is extracting uniformly distributed random bits from collapse time following inverse Gaussian distribution. Our simple but efficient method shown in Fig. 3 is to take the LSBs of the collapse count as outputs, which is based on dividing the distribution into small enough bins, so



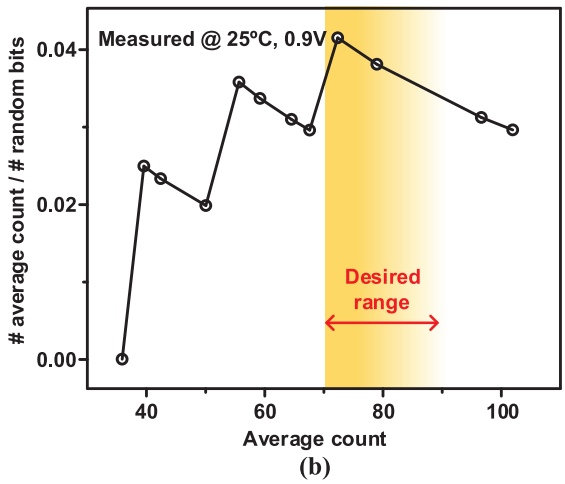


Fig. 4. (a) Standard deviation of cycles to collapse and the number of high entropy random LSBs versus average collapse cycles. (b) Number of random bits divided by average count as an approximation of throughput to illustrate the desired range.

that neighboring bins have negligible differences in probability. Similar strategies have already been applied in previous TRNG designs [8], [9], [11]. The number of LSBs that can be used as random bits depends on the variance in the collapse time distribution, as shown in the measurement results in Fig. 4. Small collapse time with small variations does not provide enough entropy and does not have enough margins, while large collapse time results in low overall random bit throughput (defined as output frequency times number of useful bits) because the number of useful bits does not increase as fast as collapse time. Fig. 4(b) shows the number of random bits divided by average count as an approximation of throughput to show the desired range. Therefore, we target a middle range of collapse time as a result of tradeoffs.

III. All-Digital Implementation of TRNG

The all-digital TRNG comprises only three parts: 1) even-stage RO; 2) control logic with a counter; and 3) automatic tuning loop to counter process, voltage, and temperature (PVT) variations. Such a simple design minimizes design efforts and offers good technology portability.

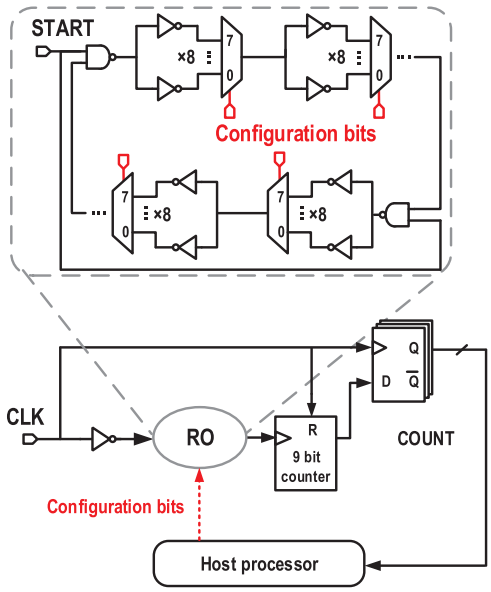


Fig. 5. TRNG block diagram and tunable RO with eight inverters per stage.

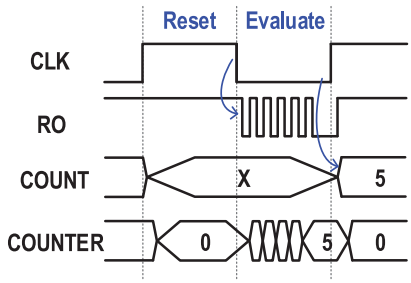


Fig. 6. Operating waveform of TRNG.

A. Tunable Even-Stage Oscillator Based on Device Mismatch

To make the RO working in the desired collapse condition discussed in the previous section, the proposed approach in Fig. 5 replaces each inverter stage in the ring with a set of identically designed inverters and a multiplexer to select one of the inverters for the RO path as specified by configuration bits. Due to process variations, each inverter has slightly different delays in fabricated chips. Through different combinations of inverters, the delay differences between the two edges can be adjusted to meet collapse time requirement. Such a mismatch-based tuning method introduces minimum overhead and provides fine enough tunability to satisfy the requirement. During startup, a simple control program described in Fig. 7 running on the host processor tunes the RO as follows: an LFSR generates a random configuration trial and collects 500–5000 collapse times to calculate their mean and max values. If the mean collapse time is too low, systematic variations are not properly canceled out and a new configuration trial is attempted. Max count value is used to tune the system clock frequency, so that most runs collapse within a given period. Mean and max collapse times that are out of range can also indicate intentional external attack as shown in the measurements in the next section. Once the mean collapse time is in the correct range, the RO is properly tuned and random numbers are generated while the host processor continues to monitor collapse times to adapt to any environmental changes.

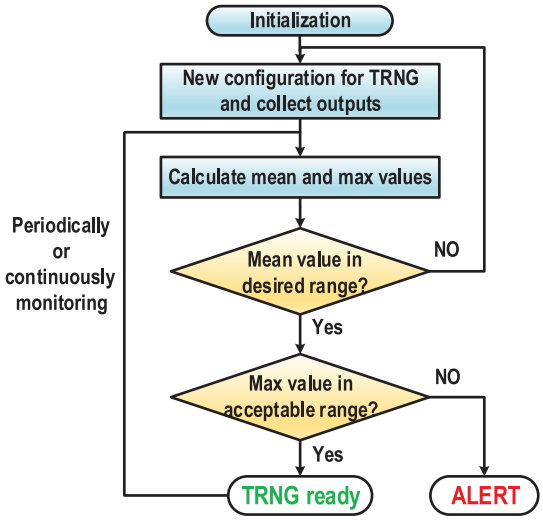


Fig. 7. Automatic tuning FSM of TRNG.

B. TRNG System Implementation

As shown in Fig. 5, the TRNG is implemented using a 32-stage RO with 8 selectable inverters per stage, providing a total of $8^{32} \approx 7.8 \times 10^{28}$ possible RO configurations. The selection of this configuration for prototype is a tradeoff between three major considerations: 1) enough tuning space is desired for robust design; 2) shorter rings takes less area and has higher throughput because the RO runs faster; and 3) according to the model in (12), peak of the average collapse time distribution increases with more stages in the ring, which increases the possibility to find configurations with larger collapse count.

The RO is reset during positive phase of the clock (CLK) and started by the falling edge of CLK as shown in Fig. 6. A 9b counter records the cycles to collapse (COUNT), which is sampled at the rising edge of CLK. Edge collapse automatically stops the counter, eliminating the need for extra phase/frequency detectors (PFD) and other peripheral circuits as in [8] and [9], saving power, area, and potential nonidealities caused by PFD. The frequency of CLK should be chosen to allow most runs collapse in one CLK period. If the RO does not collapse, COUNT will be a fixed maximum value and causes bias in generated random bits. To eliminate these negative impacts, programmable SR latches are included in counter, which sets an invalid flag once the counter value reaches the programmed value. Since TRNGs are typically colocated with an SoC processor, the tuning algorithm can run on the host-processor (off-chip in our tests) although its simplicity makes it suitable for hardware implementation.

IV. Measurement Results

The all-digital TRNG is implemented in 40 nm CMOS GP technology with a nominal voltage of 0.9 V. Measurement results in this section except Section IV-E are all based on the 40 nm prototype. For many mobile and IoT applications, however, older technologies are used because of lower power and cost. To show the portability of the design and the functionality of the TRNG in an older technology with less process variation and noise, a second prototype is fabricated in 180 nm CMOS technology. Measurement results of the 180 nm chip are provided in Section IV-E. Fig. 8 shows the die micrographs of the 40 nm chip with a core area of 836 μm^2 and 180 nm chip with a core area of 7250 μm^2.

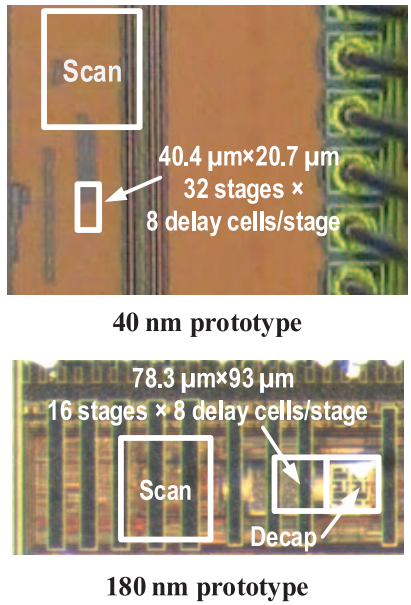


Fig. 8. Die micrographs of 40 and 180 nm TRNG prototypes.

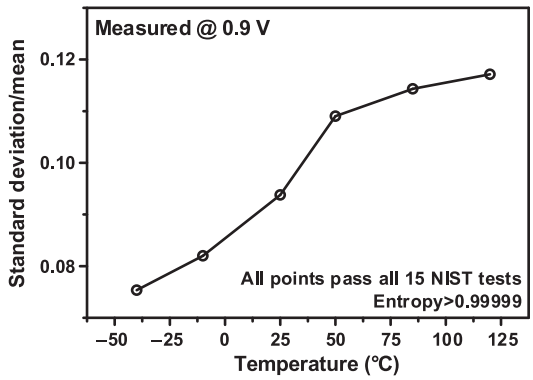


Fig. 9. Measured spread (standard deviation/mean) of cycles to collapse across temperatures.

A. Randomness and Performance of the TRNG

The randomness of the test chip is evaluated by NIST Pub 800-22 RNG testing suite (15 tests) with recommended settings and 100 Mb raw data for each run [18]. The TRNG is robust and passes all NIST tests across all combinations of voltage (0.6 to 0.9 V in 100 mV steps) and temperature ($-40\,^{\circ}$C to 120 $^{\circ}$C in 30 $^{\circ}$C steps) with a required mean-count range of 70 to 90 cycles. As shown in Fig. 9, at lower temperature, the spread (σ/μ) of collapse count is lower due to less thermal noise, but the target mean count range ensures sufficient quality of three LSBs even at $-40\,^{\circ}$C, while enabling successful RO tuning within an acceptable number of configuration trials. Table I shows NIST test suite results of five chips at one of the worst case conditions (0.6 V, $-40\,^{\circ}$C). Throughput is decided by the number of extracted bits per run and system clock frequency. The two factors are contradicting to each other.

TABLE I
MEASURED NIST TEST SUITE RESULTS OF FIVE CHIPS AT WORST CASE CONDITION (−40 °C, 0.6 V)

NIST Pub 800-22, randomness test	Chip #1		Chip #2		Chip #3		Chip #4		Chip #5	
	21 trials		102 trials		34 trials		217 trials		63 trials	
	Pvalue	Pass	Pvalue	Pass	Pvalue	Pass	Pvalue	Pass	Pvalue	Pass
Frequency	0.69	0.987	0.28	0.990	0.13	0.993	0.97	0.993	0.13	0.993
Block frequency	0.12	0.983	0.03	0.983	0.43	0.993	0.45	0.987	0.20	0.987
Cumulativ sum (1)	0.57	0.983	0.09	0.990	0.55	0.987	0.02	0.993	0.25	0.993
Cumulativ sum (2)	0.69	0.993	0.69	0.990	0.28	0.990	0.36	0.990	0.60	0.990
Runs	0.16	0.990	0.90	0.987	0.63	0.997	0.40	0.977	0.28	0.983
Longest runs	0.70	0.990	0.98	0.997	0.86	0.990	0.44	0.993	0.93	0.990
Matrix rank	0.02	0.993	0.12	0.990	0.22	0.983	0.11	0.987	0.12	0.990
FFT	0.39	0.983	0.24	0.980	0.20	0.990	0.52	0.993	0.44	0.993
Serial (1)	0.88	0.983	0.72	0.987	0.84	0.993	0.03	0.990	0.84	0.993
Serial (2)	0.93	0.983	0.21	0.990	0.29	0.987	0.05	0.983	0.57	0.983
Linear complexity	0.93	0.987	0.17	0.990	0.60	0.990	0.40	0.983	0.03	0.977
Non overlapping template	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
Overlapping template	0.19	1.000	0.63	0.980	0.65	0.980	0.57	0.970	0.83	0.980
Universal	0.21	0.990	0.38	0.970	0.26	0.980	0.70	0.960	0.49	0.990
Random excursions	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
Random excursions variant	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
Approximate entropy	0.03	0.960	0.35	0.970	0.26	0.980	0.43	0.990	0.55	1.000

*"PASS" means all sub tests pass minimum requirement.
**Minimum p-value χ^2 is 0.0001. Minimum pass rate is 0.97 for first 10 tests (using 300 × 40 K bits) and 96/100 for the other 5 tests (using 100 × 1 M bits).

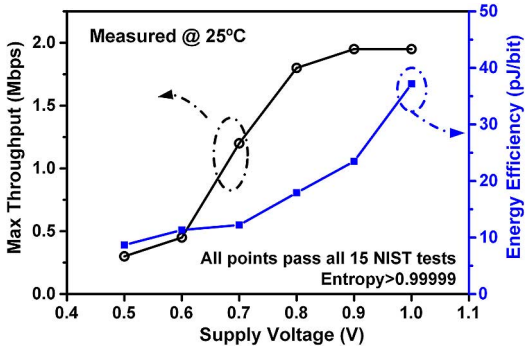


Fig. 10. Measured impacts of supply voltage on throughput of TRNG.

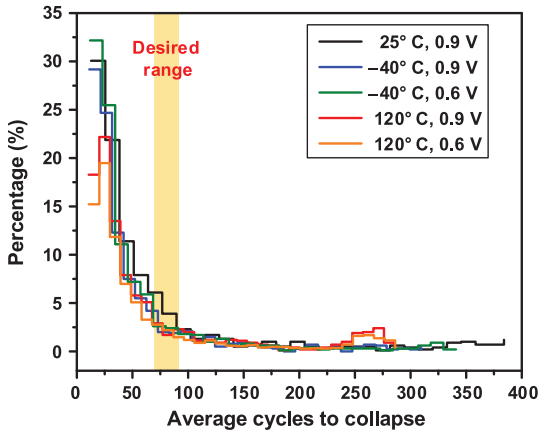


Fig. 11. Measured distributions of average cycles to collapse across random configurations.

Number of high quality bits is decided by the target collapse condition; larger collapse counts provide more random bits but require slower system clock. Measurement shows that finding a region to extract 3 bits is optimal. Once the target region and the number of extracted bits are decided, required system clock frequency and overall throughput can be determined based on RO frequency. The dependence of throughput on environmental conditions is same as that of a single RO. Fig. 10 shows throughput of the TRNG ranges from 300 kbps to 2 Mbps and energy efficiency from 8.7 to 37.2 pJ/b across 0.5 to 1 V at 25 °C. A summary of measurement results and comparisons to prior works are given in Table III.

B. Random Search Performance of the Tuning Loop

Since the tuning loop is based on random search, the number of trials to achieve desired configuration is random and affects the setup time of the TRNG. For each RO configuration, an average collapse time can be measured. Distributions of this value across 5000 configurations and 5 environmental

TABLE II
HIT RATE OF RANDOM SEARCH UNDER DIFFERENT ENVIRONMENTAL CONDITIONS

Temp (°C)	VDD (V)	Possibility in 70–90 range
25	0.9	8%
−40	0.9	6%
−40	0.6	5%
120	0.9	5%
120	0.6	6%

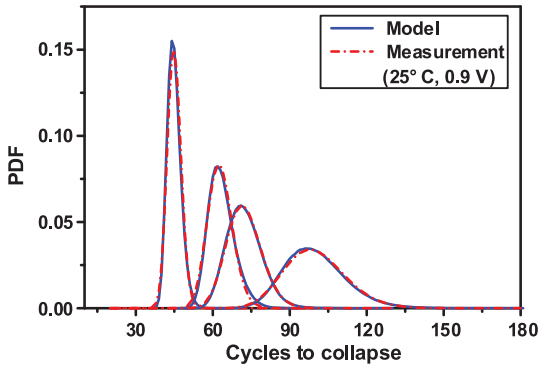


Fig. 12. Measured distribution of cycles to collapse versus analytical model derived from measured mean and variances.

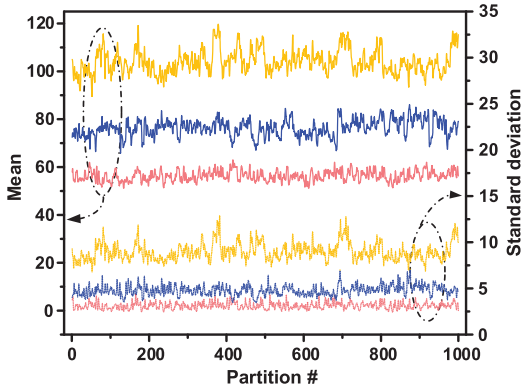


Fig. 13. Fluctuation of collapse time mean and standard deviation values of three runs with different collapse times.

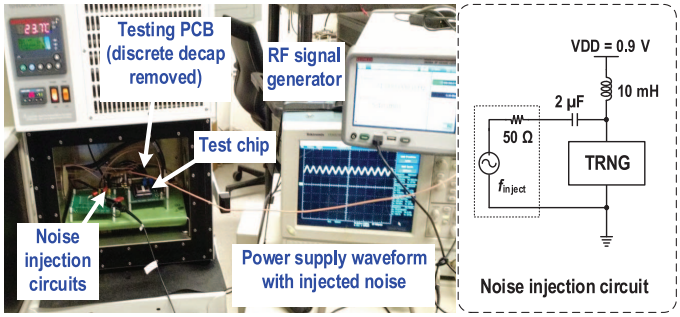


Fig. 14. Supply injection attack testing setup and schematic of noise injection circuits.

conditions are shown in Fig. 11. It can be seen that distributions are very similar across environmental variations. Small differences in distributions come from different variance of inverter delay under different conditions. Inverters have larger delay variations at lower VDD and lower temperature, according to simulation results. From the measured distribution, the possibility that a random configuration falls in desired range is calculated and shown in Table II. Assume the trials are random and independent, the probability that first success occurs at nth trial is

$$P(n) = (1-p)^{n-1}p \tag{13}$$

where p is the probability of success for one trial. In our case, worst case p can be approximated as 5% from Table II. The

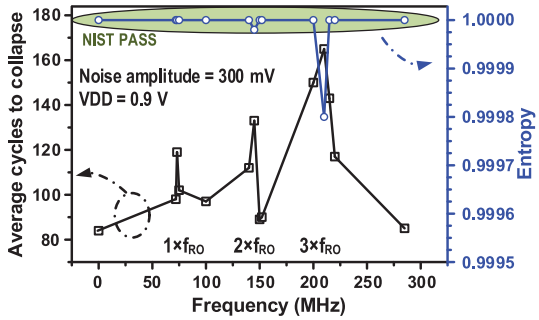


Fig. 15. Measured impacts of supply noise frequency on randomness of the TRNG.

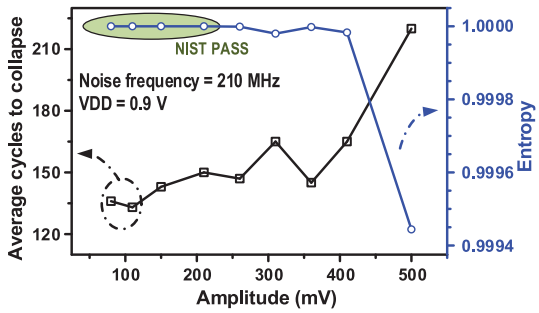


Fig. 16. Measured impacts of supply noise amplitude on randomness of the TRNG.

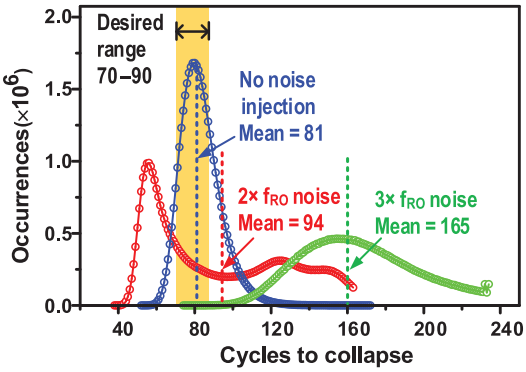


Fig. 17. Cycles to collapse distribution before and after supply noise injection of 300 mV.

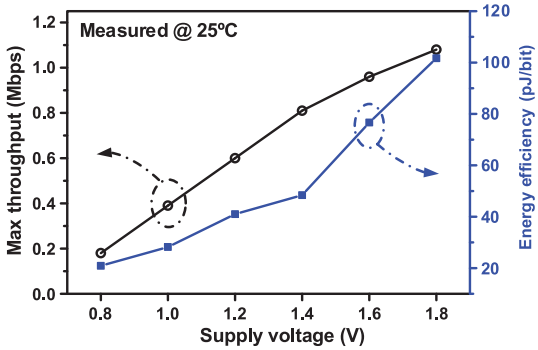


Fig. 18. Measured throughputs and energy efficiencies across supply voltages for 180 nm chip.

TABLE III
TRNG PERFORMANCE SUMMARY AND COMPARISON WITH RECENT PUBLICATIONS

	This work				[9]	[8]	[4]	[11]	[5]	[2]
	0.9 V	0.6 V	1.8 V	0.8 V						
Technology	40 nm		180 nm		65 nm	28 nm	45 nm	65 nm	130 nm	120 nm
Entropy source	Jitter in oscillator				Jitter in oscillator	Jitter in oscillator	Meta-stability	Time to oxide breakdown	Meta-stability	Meta-stability
Bit rate (Mb/s)	2	0.45	1.08	0.18	2	23.16	2400	0.011	0.2	0.2
Tested operating conditions	0.6 to 0.9 V –40 to 120 °C		0.8 to 1.8 V		0.8 to 1.2 V	N/A	0.28 to 1.35 V	N/A	N/A	N/A
NIST Pass	All				All	All	All	All	5	–
Area (μm^2)	836		7250		6000	375	4004	1200	36 300	9000
Power (mW)	0.046	0.005	0.109	0.0037	0.13	0.54	7	2	1	0.05
Efficiency (nJ/bit)	0.023	0.011	0.101	0.021	0.066	0.023	0.0029	181.81	5	0.25
Post processing	No				No	No	No	No	No	Yes
Frequency attack robustness	Yes (up to 500 mV)				N/A	Yes (filter)	N/A	N/A	N/A	N/A

expectation of the distribution in (13) could be calculated as $1/p$ and, therefore, it should take 20 trials on average to find a proper configuration hitting target range. In addition to this statistical approximation, it was found that the worst case throughout the whole testing process was 315 trials to meet targets.

C. Validation of TRNG Analytical Model

Measurement results also confirm the validation of the TRNG model presented in Section II-A. In Fig. 12, the measured distribution of cycles to collapse with different settings are shown in red lines while inverse Gaussian distribution calculated from measured mean and variance values are shown in blue lines. The correspondence of the two lines confirms the distribution of collapse time proposed in our model.

For simplicity and small area, no voltage regulation or dedicated decaps are included on-chip and no external regulator is added to testing board, which emulates a worst case noisy environment in SoC. In practical applications, decaps and, possibly, voltage regulators can be added to suppress supply noise. As discussed in Section II, supply noise affects average collapse time causing fluctuations in the distribution. To better illustrate the effects, three random runs at nominal 0.9 V and 25 °C with different configurations are divided into 1000 partitions (1K runs in each partition) with average values and standard deviations of each partition plotted in Fig. 13. Fluctuation of average collapse time exists in our measurement with a spread (σ/μ) around 5%. In fact, all testing results in Section IV are measured in same noisy environments and proved that the overall distribution can still be approximated by inverse Gaussian distribution and is capable to generate high-quality random bits.

D. Supply Noise Injection Attack to the TRNG

Supply noise injection is an effective attack technique to compromise TRNGs by injection locking [12]. Supply noise at multiples of RO frequency locks the oscillation with smaller jitter and greatly reduces entropy harvested by conventional RO-based TRNG. To test the robustness of the proposed TRNG to injection locking, we implement a noise injection testing setup by coupling a sine wave to the dc supply voltage (Fig. 14).

As shown in Fig. 15, injection locking occurs at harmonics of the ring frequency (f_{RO}) and impacts both the collapse count and the bit entropy. Here, the injection locking not only reduces the jitter of the oscillation, but also "locks" the distance between the two edges in the RO, causing the RO to collapse slower. When directly applying such a supply noise injection attack to a single configuration of the TRNG, it fails to pass NIST tests. Figs. 15 and 16 illustrate the impacts of injected supply noise frequency and amplitude on the average collapse time. Shannon entropy of generated random bits is shown by the right y-axis as a simple indicator of the randomness under different conditions. As can be seen, injection locking happens at multiples of RO frequency, and the effect is most severe at $3 \times f_{RO}$ and begins to degrade randomness after supply noise peak-to-peak amplitude is larger than 250 mV. However, since the injection locking shifts the mean collapse count outside the specified range and changes the distribution of cycles to collapse (Fig. 17), the control loop can automatically detect and reject the harvested bits. It then selects new configurations that provide slightly different oscillation frequencies, restoring the desired average count value and randomness. Hence, while operating the control loop, all NIST tests are passed with injection peak-to-peak amplitudes up to 500 mV at the worst case injection frequency of $3 \times f_{RO}$. If more sophisticated attacks are used, it is possible that the tuning loop cannot restore normal operation of the TRNG but the attack can still be detected to minimize damage to the secure system.

E. Measurement Results of 180 nm TRNG Prototype

180 nm devices have much smaller conducting current and, therefore, have less delay variation due to noise, which poses

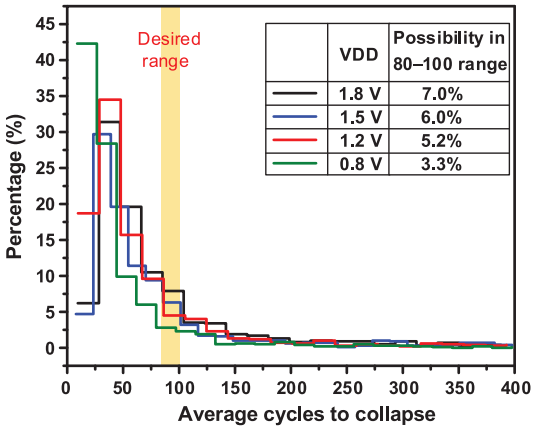


Fig. 19. Measured distributions of average cycles to collapse across random configurations for 180 nm chip.

difficulties to TRNG designs. Measurement results show that a configuration with 75 average cycles to collapse has a standard deviation of 4 in 180 nm; whereas, in 40 nm, the standard deviation is 6.8. To overcome the decrease in variance, RO needs to be tuned to a region with 80–100 average cycles to collapse, which ensures the entropy of three LSBs. Fig. 18 shows the measured throughput from 1.08 to 0.18 Mbps and energy efficiencies from 101.7 to 28.9 pJ/b across 1.8 to 0.8 V supply voltages. NIST tests confirm the randomness of the design from 1.8 down to 0.8 V. A summary of the measurement results is included in Table III.

Another concern about the TRNG in 180 nm is, less process variation, which is used to tune collapse condition in the proposed design. Fig. 19 shows the distributions of average cycles to collapse at different supply voltages. It can be seen that lower supply voltage shifts the distribution left and makes it less possible to find a proper configuration. Compared to 40 nm results in Fig. 11, less process variation shifts the distribution right which compensates the increase in target values. The overall hit rate of random search is similar to that of 40 nm design.

V. Conclusion

This work demonstrated an all-digital TRNG harvesting entropy from the frequency collapse event of two edges injected into an even-stage RO. The cycles to collapse serves as a good indicator of the quality of generated random bits. A configurable ring based on device mismatch and an automatic tuning loop based on cycles to collapse provides robustness across a wide range of temperature (-40 to 120 °C), voltage (0.6 to 0.9 V), process variation, and external attack. Due to the simplicity and robustness of the tuning scheme, no delicate tuning circuits or extensive voltage regulation is needed. Therefore, it is simple to port the all-digital design to other technologies. Measurement results prove that the design is portable to 180 nm technology commonly used for ultralow-power sensor applications. Tested chips pass all NIST randomness tests across all measured operating conditions and power supply attacks. The all-digital TRNG occupies only 836 μm^2 in 40 nm technology while consuming 23 pJ/bit at nominal 0.9 V and 11 pJ/bit at 0.6 V.

References

[1] C. S. Petrie and J. A. Connelly, "A noise-based IC random number generator for applications in cryptography," *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.*, vol. 47, no. 5, pp. 615–621, May 2000.

[2] R. Brederlow, R. Prakash, C. Paulus, and R. Thewes, "A low-power true random number generator using random telegraph noise of single oxide-traps," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2006, pp. 1666–1675.

[3] M. Matsumoto, S. Yasuda, R. Ohba, K. Ikegami, T. Tanamoto, and S. Fujita, "1200 μm2 physical random-number generators based on SiN MOSFET for secure smart-card application," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2008, pp. 414–624.

[4] S. K. Mathew *et al.*, "2.4 Gbps, 7 mW All-digital PVT-variation tolerant true random number generator for 45 nm CMOS high-performance microprocessors," *IEEE J. Solid-State Circuits*, vol. 47, no. 11, pp. 2807–2821, Nov. 2012.

[5] C. Tokunaga, D. Blaauw, and T. Mudge, "True random number generator with a metastability-based quality control," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 78–85, Jan. 2008.

[6] M. Bucci, L. Germani, R. Luzzi, A. Trifiletti, and M. Varanonuovo, "A high-speed oscillator-based truly random number source for cryptographic applications on a smart card IC," *IEEE Trans. Comput.*, vol. 52, no. 4, pp. 403–409, Apr. 2003.

[7] 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.

[8] K. Yang, D. Fick, M. B. Henry, Y. Lee, D. Blaauw, and D. Sylvester, "A 23Mb/s 23pJ/b fully synthesized true-random-number generator in 28 nm and 65 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 280–281.

[9] Q. Tang, B. Kim, Y. Lao, K. K. Parhi, and C. H. Kim, "True random number generator circuits based on single- and multi-phase beat frequency detection," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC'14)*, 2014, pp. 1–4.

[10] T. Amaki, M. Hashimoto, and T. Onoye, "A process and temperature tolerant oscillator-based true random number generator with dynamic 0/1 bias correction," in *Proc. IEEE Asian Solid-State Circuits Conf. (A-SSCC'13)*, Nov. 2013, pp. 133–136.

[11] N. Liu, N. Pinckney, S. Hanson, D. Sylvester, and D. Blaauw, "A true random number generator using time-dependent dielectric breakdown," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2011, pp. 216–217.

[12] 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*, Berlin, Germany: Springer, 2009, pp. 317–331.

[13] M. Dichtl and J. D. Golić, "High-speed true random number generation with logic gates only," in *Cryptographic Hardware and Embedded Systems*, Berlin, Germany: Springer, 2007, pp. 45–62.

[14] K. Yang, D. Blaauw, and D. Sylvester, "A robust -40 to 120 °C all-digital true random number generator in 40 nm CMOS," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2015, pp. 248–249.

[15] A. A. Abidi, "Phase noise and jitter in CMOS ring oscillators," *IEEE J. Solid-State Circuits*, vol. 41, no. 8, pp. 1803–1816, Aug. 2006.

[16] J. L. Folks and R. S. Chhikara, "The inverse Gaussian distribution and its statistical application—A review," *J. Roy. Stat. Soc. Ser. B Methodol.*, vol. 40, no. 3, pp. 263–289, Jan. 1978.

[17] K. Yang, Q. Dong, D. Blaauw, and D. Sylvester, "A physically unclonable function with BER $< 10^{-8}$ for robust chip authentication using oscillator collapse in 40 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015 pp. 254–255.

[18] National Institute of Standards and Technology, *A Statistical Test Suite for the Validation of Random Number Generators and Pseudo Random Number Generators for Cryptographic Applications*, Pub. 800–22, 2010.

Kaiyuan Yang (S'13) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2012, and the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2014. He is currently working toward the Ph.D. degree at University of Michigan.

His research interests include energy-efficient VLSI design and hardware security.

David Blaauw (M'94–SM'07–F'12) received the B.S. degree in physics and computer science from Duke University, Durham, NC, USA, in 1986, and the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 1991.

He was with Motorola, Inc., Austin, TX, USA, where he was the Manager of the High Performance Design Technology Group. Since August 2001, he has been on the Faculty of the University of Michigan, where he is a Professor. He has authored/coauthored over 450 papers and holds 40 patents. His research interests include VLSI design with particular emphasis on ultralow-power and high-performance design.

Dr. Blaauw was the Technical Program Chair and General Chair for the International Symposium on Low-Power Electronic and Design. He was also the Technical Program Co-Chair of the ACM/IEEE Design Automation Conference and a member of the ISSCC Technical Program Committee.

Dennis Sylvester (S'95–M'00–SM'04–F'11) received the Ph.D. degree in electrical engineering from the University of California, Berkeley, CA, USA, in 1999.

He is a Professor of Electrical Engineering and Computer Science with the University of Michigan, Ann Arbor, MA, USA, and the Director of the Michigan Integrated Circuits Laboratory (MICL), a group of 10 faculty and more than 70 graduate students. He has held research staff positions with the Advanced Technology Group, Synopsys, Mountain View, CA, USA, Hewlett-Packard Laboratories, Palo Alto, CA, USA, and visiting professorships at the National University of Singapore, Singapore, and Nanyang Technological University, Singapore. He is the Co-Founder of Ambiq Micro, Austin, TX, USA, a fabless semiconductor company developing ultralow-power mixed-signal solutions for compact wireless devices. He has authored/coauthored over 375 articles along with one book and several book chapters. He holds 20 U.S. patents. His research interests include the design of millimeter-scale computing systems and energy-efficient near-threshold computing.

Dr. Sylvester serves on the Technical Program Committee of the IEEE International Solid-State Circuits Conference and previously served on the Executive Committee of the ACM/IEEE Design Automation Conference. He also serves as a Consultant and Technical Advisory Board Member for electronic design automation and semiconductor firms in his research areas. He has served as an Associate Editor for the IEEE TRANSACTIONS ON CAD and the IEEE TRANSACTIONS ON VLSI SYSTEMS, and Guest Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. He was the recipient of the NSF CAREER Award, the Beatrice Winner Award at ISSCC, an IBM Faculty Award, an SRC Inventor Recognition Award, and eight Best Paper Awards and Nominations. He is the recipient of the ACM SIGDA Outstanding New Faculty Award and the University of Michigan Henry Russel Award for distinguished scholarship. His dissertation was recognized with the David J. Sakrison Memorial Prize as the most outstanding research in the UC-Berkeley EECS Department