A 65-nm CMOS 3.2-to-86 Mb/s 2.58 pJ/bit Highly Digital True-Random-Number Generator With Integrated De-Correlation and Bias Correction

Venkata Rajesh Pamula, *Member, IEEE*, Xun Sun, Sung Min Kim, Fahim ur Rahman, Baosen Zhang, and Visvesh S. Sathe, *Member, IEEE*

***Abstract*—This letter presents a highly digital, technology scalable, and energy-efficient cryptographic-quality true random number generator (TRNG). The proposed architecture presents a balanced approach to TRNG design, relying on a simpler, noncryptographic quality physical random number generator (*phy*RNG) combined with energy-efficient integrated post-processing to de-correlate and de-bias the *phy*RNG bitstream. Operating at a supply voltage (V_{dd}) of 0.53 V, a 65-nm CMOS prototype of the TRNG achieves a peak energy-efficiency of 2.58 pJ/bit. TRNG bitstreams pass all NIST randomness benchmarks over a V_{dd} range of 0.5–1.05 V across −20 °C–100 °C, demonstrating its efficacy and robust operation over a wide V_{dd} and temperature range.**

***Index Terms*—Cryptographic-quality, information security, iterative von Neumann (IVN), Markov chain (MC), metastability, true random number generators (TRNGs).**

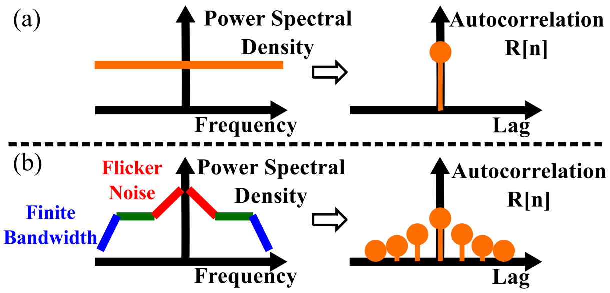


Fig. 1. Spectral properties of device noise in (a) ideal RNG and (b) real RNG.

I. Introduction

EVER-INCREASING volumes of private data exchanged between connected devices drives the need for increasingly efficient, secure communication. *Cryptographic-quality* true random number generators (TRNGs) are essential security primitives to this end. While many secure systems rely on psuedo random number generators, they suffer from vulnerability to malicious attacks due to their predictability [1]. Hardware TRNGs, on the other hand, harness ideally white device noise as an entropy source, either in the form of phase or voltage noise, to generate random bit patterns to provide immunity to security attacks. Several TRNG implementations that exploit device noise in the form of either phase noise [2], [3] or voltage noise [1], [4]) have been reported in the literature. Existing implementations exploit device noise to directly produce true random numbers (TRNs) with little or no integrated post-processing. These approaches require extremely detailed design and runtime control to adapt to environmental variation, and are either limited in bitrate or energy-efficiency.

This letter describes a highly digital, process scalable TRNG that represents a balanced approach to TRNG design. Device noise is extracted using a physical random number generator (*phy*RNG) with substantially relaxed randomness quality requirements, reducing circuit complexity while enhancing robustness, and energy-efficiency. In conjunction, we employ integrated post-processing to de-correlate and de-bias the *phy*RNG bits [5]. Bias reduction is carried out using a combination of coarse offset cancelation, performed in the *phy*RNG, and subsequent bias-elimination using iterative von Neumann (IVN) correction [6]. We propose a Markov-chain-based whitening approach to de-correlate *phy*RNG bits, essential for TRN generation. The proposed architecture was implemented in a 65-nm CMOS test chip and verified to pass all NIST (800-22 and 800-90B) randomness benchmarks.

II. Device Noise Enabled RNGs

An ideal TRNG circuit harnesses thermal noise present in the device, either in the form of jitter or voltage noise to generate the random numbers. Fig. 1(a) shows the spectral properties of device noise in an ideal RNG. An ideal noise source is spectrally white, exhibiting zero autocorrelation for nonzero lags. Ideal TRNG circuits that harness this noise exhibit no bias, resulting an equal probability of generating a bit outcome of 1 or 0 [$P(X = 1) = P(X = 0) = 0.5$]. Real hardware TRNGs, however, exhibit a departure from ideality in two key aspects [Fig. 1(b)]: 1) flicker noise in MOS devices, and finite circuit bandwidth *colors* the noise spectrum at lower and higher frequencies, respectively, leading to correlation in the generated bitstream and 2) process variations introduce bias (particularly in comparator-based techniques), resulting in $P(X = 1) \neq P(X = 0)$. Cryptographic-quality TRNs require very low bias: experiments indicate that even a modest bias of 0.5% degrades entropy sufficiently to readily fail NIST benchmark tests.

State-of-the-art hardware TRNG implementations have largely focused on eliminating this bias in the *phy*RNG to improve its entropy ($H_{phy\mathrm{RNG}}$), resulting in a complexity-energy dissipation-bitrate trade-off [1]–[3]. Yet, barring [4], correlation in the bitstream remains a critical but un-addressed challenge.

The proposed TRNG addresses both bias and correlation based on two key observations.

1) Suppressing *phy*RNG bias to levels below thermal noise as required by cryptographic-quality TRNGs is exacting in terms of complexity and power dissipation.

Manuscript received December 2, 2018; revised January 14, 2019; accepted January 28, 2019. Date of publication January 31, 2019; date of current version May 17, 2019. This paper was approved by Associate Editor Alvin Leng Sun Loke. This work was supported in part by NSF under Grant CNS 171449, and in part by Semiconductor Research Corporation-T3S under Contract 2017-TS-2774. *(Corresponding author: Venkata Rajesh Pamula.)*

The authors are with the Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: pamula@uw.edu).

Digital Object Identifier 10.1109/LSSC.2019.2896777

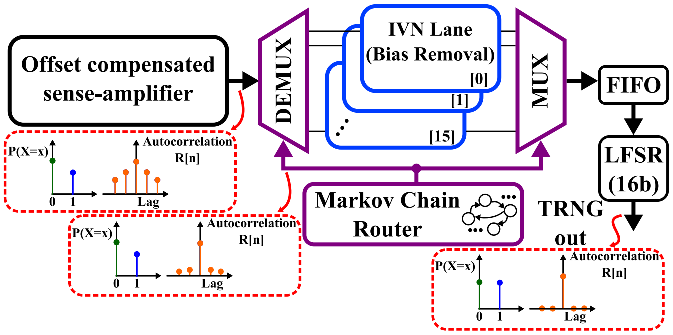


Fig. 2. Architecture of the proposed TRNG.

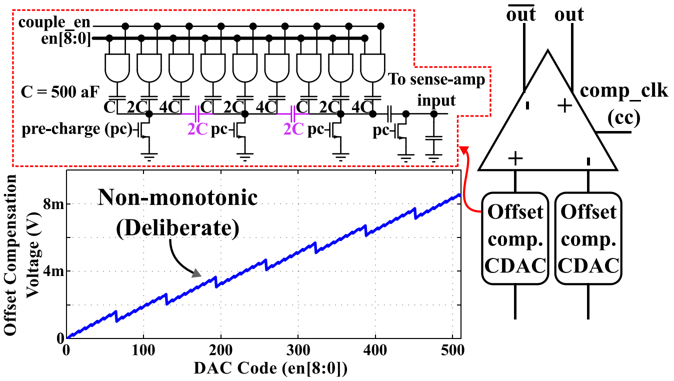


Fig. 3. Schematics of the sense-amp-based *phy*RNG with offset cancelation capacitive DAC and its transfer characteristics.

2) TRNG circuit approaches that directly harness CMOS device noise in the *phy*RNG without further post-processing will be colored by flicker noise and finite-bandwidth effects, leading to autocorrelation in the output bitstream.

We, therefore, propose an approach that engineers an effective balance between *phy*RNG quality (in terms of bias and autocorrelation), and the quantity of post-processing required in order to produce a robust TRNG with maximum efficiency at a target bitrate.

III. Proposed TRNG Architecture

Fig. 2 shows the high level architecture of the proposed TRNG which embeds a strong-ARM latch-based sense-amplifier as the *phy*RNG. The output bits of *phy*RNG are de-correlated through a Markov chain (MC) followed by de-biasing through IVN corrector. The resulting bitstream is finally passed through a 16-bit LFSR to remove any residual correlation. It must be noted that both the bias and the autocorrelation of the bistream reduces as it moves through the datapath.

A. Sense-Amplifier-Based phyRNG

Fig. 3 shows the schematic of the *phy*RNG used in the TRNG test chip. A strong-ARM latch-based sense-amplifier is offset compensated through adaptive mixed-signal feedback using a 9-bit capacitive DAC (CDAC) (Fig. 3). The CDAC is capable of compensating offsets up to 8.5 and 17 mV in high resolution and low resolution modes, respectively, with a step size less than 500 μV. The CDAC is deliberately designed for a nonmonotonic transfer characteristic to avoid "missing codes," ensuring that the offset cancelation resolution remains well within the noise voltage of the *phy*RNG. Notably, *phy*RNG choice is not constrained in the proposed architecture—Our specific implementation was influenced by the availability of necessary building-blocks. Although the IVN corrector in the TRNG is capable of de-biasing the input bitstream, reducing *phy*RNG bias enables higher post bias-removal bitrate and reduced energy dissipation per random bit (E_b).

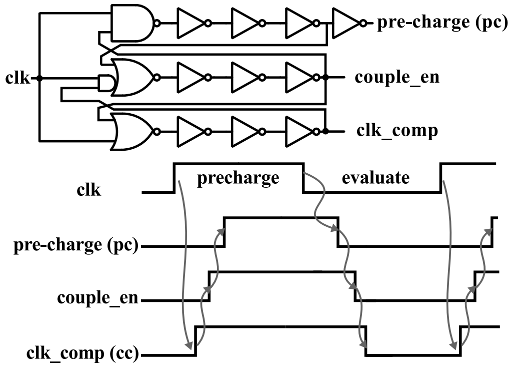


Fig. 4. Three-way interlocked pulse generator and timing diagram.

The sense-amplifier employs pMOS devices at the input pair to reduce flicker noise, minimizing *phy*RNG output autocorrelation. To reduce energy dissipation, the unit capacitors (C) in the CDAC are realized using custom fringe metal capacitors of 500 aF each. The reduced capacitance also reduces time constants associated with sense-amplifier precharge, allowing precharge to occur over more time-constants, even at high operating frequencies. A three-way interlocked pulse generator (Fig. 4) is proposed to robustly generate the necessary control signals for the *phy*RNG with the desired relative timing relationships.

B. Energy-Efficient Post-Processing

The output bits from the *phy*RNG are routed to an energy-efficient post-processor which de-correlates and removes the residual bias in the incoming bitstream. Correlation removal, necessary for TRN generation, must be performed before IVN-based de-biasing. Bias-elimination using IVN requires that incoming bits be independent and identically distributed (i.i.d). Correlation is suppressed by routing the *phy*RNG bits through an MC-based router, which sends each incoming *phy*RNG bit to one of up to 16 bias removal channels based on the Markov state. Autocorrelation can be modeled by an MC (Fig. 5). A 2^n state MC models the conditional probability of generating a 0 or 1 based on the history of previous n output bits. The Markov router exploits this observation to (architecturally) assign a unique de-biasing module to each Markov state – bits generated by the *phy*RNG in a given Markov state all go to the same IVN module, so that each IVN receives bits with identical *phy*RNG history and therefore identical statistics, conforming to i.i.d requirements. We implemented a 16 state MC is implemented to remove autocorrelation up to a lag of 4.

Each IVN module processes *phy*RNG bits to produce de-biased output bits. Provided with i.i.d bits at entropy $H_{phy\text{RNG}}$ and bitrate b, IVN generates output bits with $h = 1$ at a bitrate of $kH_{phy\text{RNG}}b$. The bitrate loss resulting from a finite-size IVN implementation is modeled by k ($0 \leq k < 1$). Standard von Neumann correction generates un-biased outputs from biased inputs by examining bit-pairs and producing an output bit of 0 and 1 corresponding to input bit-pairs of 01 and 10, respectively. IVN extends this idea by using a tree-like structure to analyze longer input patterns [5], [6]. The IVN tree must be optimally pruned to maximize k for a given number of processing nodes (N) allowed area and dynamic energy dissipation (E_{dyn}) considerations. The addition of each node to an optimally constructed

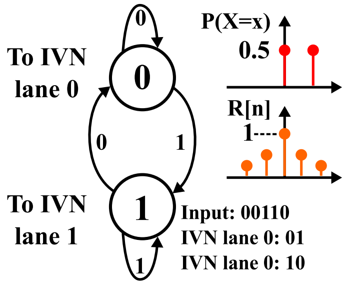
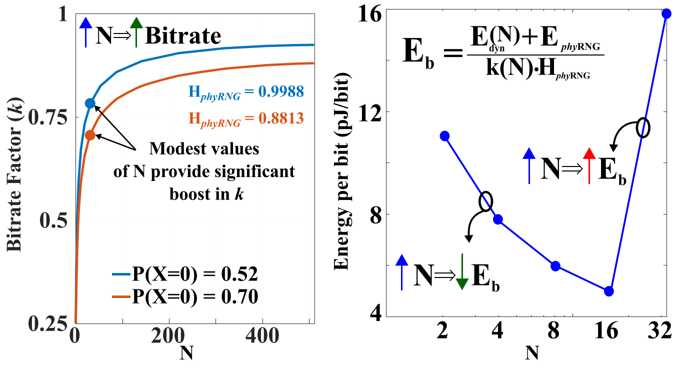


Fig. 5. 2-state MC modeling of an unbiased but correlated process. Incoming *phy*RNG bits are routed to either IVN lane 0 or lane 1 depending on the Markov state.



(a) (b)

Fig. 6. Impact of the number of nodes (N) of *optimally pruned* IVN tree on (a) bitrate factor (normalized bitrate) and (b) energy per bit (E_b).

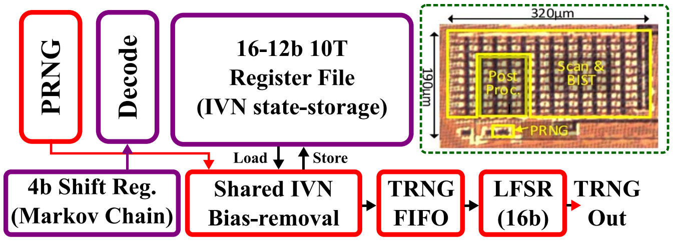


Fig. 7. VLSI mapping of the TRNG architecture and its chip micrograph.

TABLE I
MEASURED (PARTIAL) NIST RANDOMNESS TEST SUITE RESULTS

	Pass Rate for NIST 800-22[1]			
Voltage (V)	1.05		0.53	
Temperature (°C)	-20	100	-20	100
Runs	1.0	0.99	0.98	0.97
Longest Run	0.98	0.96	1.0	1.0
Non-Ovl. Temp.	0.98	0.98	0.98	0.97
Approx. Entropy	0.98	0.98	1.0	0.98
Serial (1)	1.0	1.0	0.99	1.0
[1] All “PASS” (96/100 required to PASS)				

tree has a successively lower probability of generating a random bit, causing the increase in k to saturate Fig. 6(a). Simulations show that a reduction in E_b due to increased k [Fig. 6(b)] at the cost of higher E_{dyn} leads to an optimal choice of N.

The combination of the MC and its interface to the IVN (Fig. 7) is mapped efficiently in VLSI by exploiting the observation that only one of the IVN modules is active in any given cycle. This enables time-multiplexing of the IVN bias-correction logic, storing the state of each IVN module in a register file of 16 × 12b (corresponding to $N = 16$) and reading the register line that constitutes the state of each to the selected IVN module. The register file address is indexed by the Markov state that the system is currently in. Since the Markov state is modeled as a trace of prior bits, the address generation logic reduces to a 4-bit shift register comprising of previously generated *phy*RNG bits. The de-biased bits are then buffered into a multi-issue FIFO and routed through a 16-bit LFSR to remove the residual correlation.

TABLE II
MEASURED NIST 800-90B TEST SUITE RESULTS

NIST SP 800-90B Entropy Assessment	Results of 1Mb bitstream (score, dof)
IID Permutation	PASS (NA, NA)
Chi-square Independence	PASS (1892, 2047)
Chi-square Goodness of fit	PASS (5.83, 9)
LRS Test	PASS (NA, NA)
Min. Entropy	0.996
Restart Test	PASS (NA, NA)
dof: degrees of freedom, NA: Not applicable	

IV. MEASUREMENT RESULTS

The TRNG test-chip is fabricated in a 65-nm CMOS low-power process and operates at a nominal voltage and frequency of 1 V and 200 MHz, respectively. Occupying a total area of 0.01 mm^2 excluding scan and BIST (Fig. 7), the TRNG operates over a wide V_{dd} range of 0.5–1.05 V and yields a corresponding throughput of 3.2 Mb/s and 86 Mb/s, respectively. The minimum energy point (MEP) was observed at V_{dd} of 0.53 V, with the TRNG operating at 5.76 Mb/s while dissipating 8.33 μW of power. The corresponding E_b achieved by the TRNG is 2.58 pJ/bit [Fig. 8(a)]. The sense-amplifier-based *phy*RNG consumes only 4.6% of the total system power, most of which is dominated by the digital post-processor, across the operational frequency range of 4.4–200 MHz. Therefore, the proposed TRNG architecture greatly benefits from technology scaling for reducing both area and power consumption.

TRNG robustness to V_{dd} variation is characterized by sweeping V_{dd} from 0.5–1.05 V. As shown in Fig. 8(b), the TRNG generates high quality random bits with a measured entropy (H_{TRNG}) exceeding 0.999996 over the entire voltage range even with significantly degraded *phy*RNG entropy as indicated by the increased *phy*RNG entropy gap (1-H) at lower V_{dd}. The measured autocorrelation factor (ACF) of 100M consecutive bits with lags 1–50 is shown in Fig. 8(c). The MC de-correlator and the LFSR post-processing results in the bitstream ACF quickly descending to the ACF floor (10^{-4}) for lags > 1, confirming the mutual independence of the individual bits in the bitstream. TRNG quality is further accessed through NIST randomness benchmark suites (both 800-22 and 800-90B). One hundred bitstreams each of 1M bits were generated under both MEP (0.53 V) and high throughput (1.05 V) configurations at -20 °C and 100 °C. All bitstreams successfully pass all the NIST benchmarks. Table I shows the pass rate for worst performing five out of a total of 16 randomness tests, while the results of NIST 800-90B are presented in Table II. Complete NIST 800-22 test suite results are presented in [5]. Finally, the performance of the TRNG is summarized and compared against the *state-of-the-art* in Table III. Among digital TRNG implementations, the proposed TRNG achieves the highest E_b while, remaining robust across V_{dd} and temperature variations.

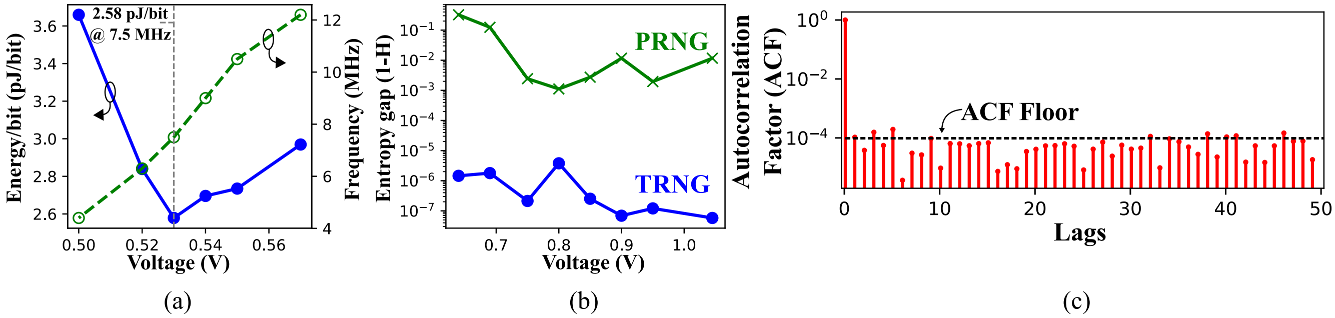


Fig. 8. Measured test-chip performance (a) E_b versus V_{dd} and operating frequency for MEP characterization. (b) *phy*RNG and TRNG entropy gap (1-H) across V_{dd}. *phy*RNG offset cancelation was performed at 1.05 V and disabled for lower V_{dd}. A significant entropy gap exists even at 1.05 V, and is maintained as *phy*RNG entropy degrades with V_{dd} scaling. (c) Autocorrelation of 100M consecutive bits.

TABLE III
TRNG Performance Summary and Comparison With the State-of-the-Art

	This work		ISSCC'17 [3]		ISSCC'14 [2]	JSSC'16 [4]		JSSC'16 [7]		JSSC'17 [8]	VLSI'18 [9]
Technology	65nm		65nm		65nm	14nm		40nm		180nm	14nm
Entropy Source	Metastability		Jitter		Jitter	Metastability		Jitter		Chaotic Map	Metastability
NIST Pass	All		All (3 LSBs)		All	All		All		All	All
Voltage	0.53	1.0	1.08	1.2	0.9	0.4	0.75	0.6	0.9	0.6	0.65
Bitrate (Mbps)	3.2	86.0	8.2	9.9	23.16	8.6	162.5	2.0	0.45	0.27	1480
Efficiency (pJ/bit)	2.58	6.08	35.47	42.17	23	3	9.23	11	23	0.30	2.5
Power (μW)	8.33	523	289	418	159	27	1500	5	46	0.394	3700
Area (mm^2)	0.01		0.00092		0.00037	0.00101		0.00083		0.21	0.0021
V$_{dd}$-robust	Yes		Yes		Yes	Yes		Yes		Yes	Yes
Temp.-robust	Yes		Not reported		Not reported	Not reported		Yes		Not reported	Not reported

V. Conclusion

We present a digital TRNG architecture that relies on a *phy*RNG with relaxed quality and robustness requirements by incorporating integrated post-processing. The efficiency and robustness benefits afforded by such a *phy*RNG, combined with an efficient VLSI mapping of de-correlation and de-biasing algorithms allows for a scalable, efficient TRNG implementation. A test-chip demonstration of the architecture yields a robust, fully NIST compliant, highly digital TRNG that passes all the NIST randomness benchmark tests over a V_{dd} range of 0.5–1.05 V across a temperature range of -20 °C to -100 °C, achieving an ultralow E_b of 2.58 pJ/bit at 0.53 V.

References

[1] 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.

[2] K. Yang *et al.*, "16.3 A 23Mb/s 23pJ/b fully synthesized true-random-number generator in 28nm and 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Techn. Papers (ISSCC)*, 2014, pp. 280–281.

[3] E. Kim, M. Lee, and J.-J. Kim, "8.2 8Mb/s 28Mb/mJ robust true-random-number generator in 65nm CMOS based on differential ring oscillator with feedback resistors," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2017, pp. 144–145.

[4] S. K. Mathew *et al.*, "μRNG: A 300-950 mV, 323 Gbps/W all-digital full-entropy true random number generator in 14 nm FinFET CMOS," *IEEE J. Solid-State Circuits*, vol. 51, no. 7, pp. 1695–1704, Jul. 2016.

[5] V. R. Pamula *et al.*, "An all-digital true-random-number generator with integrated de-correlation and bias correction at 3.2-to-86 Mb/s, 2.58 pJ/bit in 65-nm CMOS," in *Proc. IEEE Symp. VLSI Circuits*, 2018, pp. 1–2.

[6] V. Rožic, B. Yang, W. Dehaene, and I. Verbauwhede, "Iterating Von Neumann's post-processing under hardware constraints," in *Proc. IEEE Int. Symp. Hardw. Orient. Security Trust (HOST)*, 2016, pp. 37–42.

[7] K. Yang, D. Blaauw, and D. Sylvester, "An all-digital edge racing true random number generator robust against PVT variations," *IEEE J. Solid-State Circuits*, vol. 51, no. 4, pp. 1022–1031, Apr. 2016.

[8] M. Kim, U. Ha, Y. Lee, K. Lee, and H.-J. Yoo, "A 82-nW chaotic map true random number generator based on a sub-ranging SAR ADC," *IEEE J. Solid-State Circuits*, vol. 52, no. 7, pp. 1953–1965, Jul. 2017.

[9] S. Satpathy *et al.*, "An all-digital unified static/dynamic entropy generator featuring self-calibrating hierarchical Von Neumann extraction for secure privacy-preserving mutual authentication in IoT mote platforms," in *Proc. IEEE Symp. VLSI Circuits*, 2018, pp. 169–170.