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

V. Rajesh Pamula, Xun Sun, Sung Kim, Fahim ur Rahman, Baosen Zhang, and Visvesh S. Sathe

University of Washington, Seattle, WA, USA; E-mail: pamula@uw.edu

Abstract

We present a robust, all-digital True Random Number Generator (TRNG) architecture that efficiently combines low-quality physical random number generators (PRNGs) with integrated de-correlation and de-biasing. A 65-nm CMOS TRNG test chip demonstrates NIST test-suite compliance across 0.5–1.0 V supply voltage (V_{dd}) and -20–100 °C, even with significant PRNG entropy (H) degradation. The measured 2.58 pJ/bit is the lowest among all-digital NIST-compliant TRNGs. In terms of energy, area and performance metrics, this digital implementation is especially suited for advanced CMOS process nodes.

Introduction

The growing volume of private data exchanged between networked devices and heightened need for privacy and security are driving the demand for efficient, scalable and high-quality security primitives such as TRNGs. Hardware TRNGs achieve randomness by extracting device noise in the form of either phase noise [1], [2] or noise during metastability resolution [3]. Traditional techniques have largely focused on eliminating variation-induced bias by aggressively enhancing PRNG quality, which incurs either increased complexity and energy dissipation [1]–[3], or reduced bit rate [1], [2]. Correlation poses another significant un-addressed challenge. Ideal TRNGs extract white noise such as un-filtered device thermal noise. However, limitations in the form of MOS $1/f$ noise and finite circuit bandwidth inherently result in shaped (non-white) noise, leading to randomness degradation in the form of output bit correlation [1]–[4] (Fig. 1a). This paper presents a process-scalable TRNG architecture that employs more energy-efficient but less ideal PRNGs, and relies on integrated entropy extraction algorithms to de-correlate (whiten) and de-bias PRNG bits (Fig. 1b). Additional contributions of this work include a Markov Chain (MC)-based whitening scheme, and an efficient VLSI architecture for whitening and bias removal.

Architecture and Implementation

Fig. 2 outlines the proposed approach. Integrated post-processing relaxes the requirement for the PRNG output bias β, defined for output X as P(X=1) = 0.5+β. Values of β in excess of 0.1 can be tolerated, in contrast to typical values of $<10^{-3}$ in prior work. An efficient StrongARM latch employing modest offset cancellation produces raw PRNG bits of sufficient quality. A von Neumann corrector [4], [5] can only eliminate bias when its input bits are independent (uncorrelated) with consistent signal statistics, requiring whitening (de-correlation) before bias removal. This work proposes using a Markov Chain [5]—which tracks the probability of producing a ‘1’ or ‘0’ output in any state as a function of prior outcomes (Fig. 3)— to remove finite-lag autocorrelation. A 4-bit MC router selects one of 2^4 Iterative von-Neumann corrector (IVN) modules (“lanes”), based on the current MC state, to receive PRNG bits. Bits routed to any lane therefore have identical four-bit histories and similar statistics. This approach removes the influence of the four prior bit outcomes, sufficiently suppressing correlation for NIST compliance.

Probabilistically generated IVN output bits are buffered using a three-issue FIFO. A linear feedback shift register (LFSR) further suppresses residual correlation. The IVN, consisting of a tree of *T-* and *P-modules*, iteratively operates on input bit pairs (Fig. 4) to produce bias-corrected outputs at a rate higher than traditional von Neumann correction. Given a PRNG producing bits with entropy H_{PRNG} at bit rate B, the IVN state machine generates outputs with $H = 1$ at a reduced rate of $k{\cdot}B{\cdot}H_{PRNG}$ (Fig. 2). For $0.25 \leq k < 1$, k models rate loss from a finite sub-tree IVN implementation and grows with tree size. The choice of IVN sub-tree depth is determined by the best trade-off between performance, area, and energy efficiency. Fig. 5 illustrates the simulated effect of MC-based whitening and the H vs. bit rate trade-off associated with IVN.

Fig. 6 shows the VLSI implementation of the proposed TRNG. Capacitive coupling-based offset cancellation removes up to 17 mV of offset with 9 bits of resolution. An efficient load-store architecture is employed to take advantage of only one IVN lane receiving and processing a PRNG bit produced in any cycle. A 16×12-bit register file (RF) stores the state of the pruned IVN tree for each of the 16 IVN lanes. In each cycle, an RF entry corresponding to the MC router-chosen IVN lane is read (Load) and processed together with the PRNG bit using shared IVN logic to derive the TRNG bits. An updated IVN state for the selected lane is subsequently written back (Store) into the RF. Since the MC state reflects a 4-bit PRNG history, a shift-register trivially performs the task of an MC router, producing RF addresses.

Silicon Results

At $V_{dd} = 1.0$ V, the 65-nm TRNG test chip (Fig. 10) operates at 200 MHz, corresponding to an output bit rate of 86 Mb/s. The minimum energy point (MEP) is observed at $V_{dd} = 0.53$ V, producing TRNG bits at 5.76 Mb/s and 2.58 pJ/bit (Fig. 7). Across V_{dd} of 0.5–1.0 V, the TRNG produces high-quality random bits with a worst-case measured entropy of 0.999996. This corresponds to an entropy gap (1-H) of $4{\times}10^{-6}$ even with significant PRNG entropy degradation (Fig. 8). The simplified PRNG accounts for less than 4.6% of total system dissipation across the TRNG frequency range (4.4–200 MHz). As the digital post-processor dominates area and energy dissipation, this architecture will greatly benefit from implementation in more advanced process nodes. The effectiveness of the TRNG whitening architecture is demonstrated in Fig. 9. TRNG randomness was benchmarked using the NIST suite. Sixteen randomness criteria were successfully validated under MEP (0.53 V) and high-performance (1 V) configurations, both at -20 and 100 °C (Table 1). Compared to related efforts (Table 2), the proposed TRNG achieves the lowest energy/bit among digital TRNGs with robust operation across V_{dd} and temperature.

References

[1] E. Kim *et al.*, *ISSCC,* 2017, pp. 144–145.
[2] K. Yang *et al.*, *ISSCC,* 2014, pp. 280–281.
[3] S. Mathew *et al.*, *JSSC,* pp. 1695–1704, Jul. 2016.
[4] V. Rozic *et al*, *HOST,* 2016, pp. 37–42.
[5] Y. Perez, Ann. Statistics, vol. 20(1), pp. 590–597, 1992.
[6] L. Alberto, “Probability, statistics, and random processes for electrical engineering (2017)”.
[7] K. Yang *et al., JSSC,* pp. 1022–1031, Apr. 2016.

Acknowledgment

This work was funded by the NSF under grant number CNS 171449.

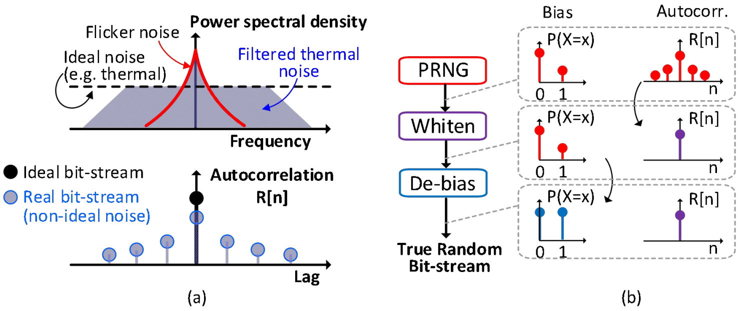


Fig.1. (a) Finite circuit bandwidth and 1/*f* noise leads to correlated bits. (b) Proposed TRNG architecture.

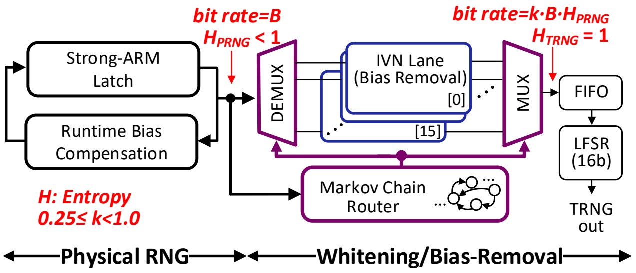


Fig. 2. Proposed Markov-based whitening and IVN-based bias-removal architecture.

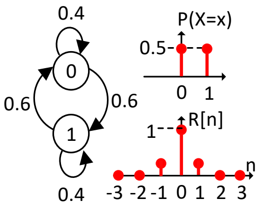


Fig. 3. Example 2-state MC modeling an unbiased but correlated process.

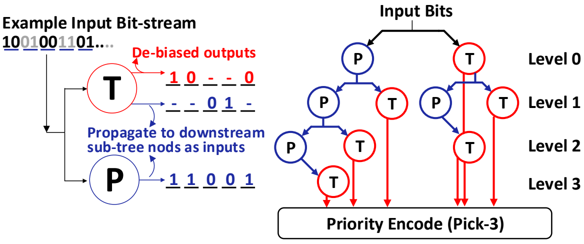


Fig. 4. P- and T-state machines process successive bits. Pruned IVN tree implemented in design.

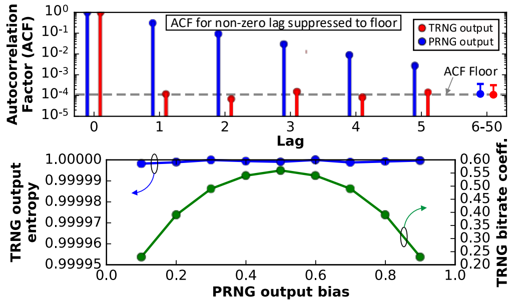
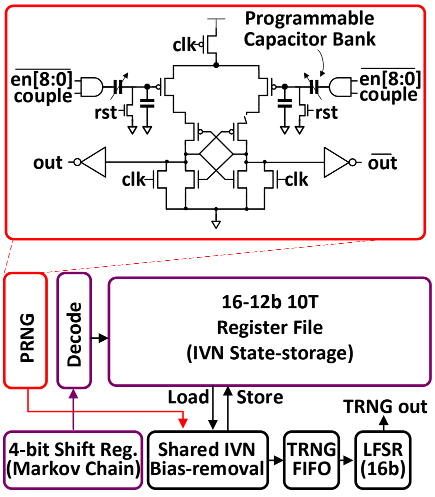


Fig. 5. Markov-based whitening and IVN simulation. IVN removes bias by trading off output bit rate.



Programmable Capacitor Bank
clk
en[8:0] couple
rst
out
clk
PRNG
Decode
16-12b 10T Register File (IVN State-storage)
Load
Store
TRNG out
4-bit Shift Reg. (Markov Chain)
Shared IVN Bias-removal
TRNG FIFO
LFSR (16b)

Fig. 6. Simplified PRNG schematic and efficient load-store VLSI architecture for whitening + bias removal.

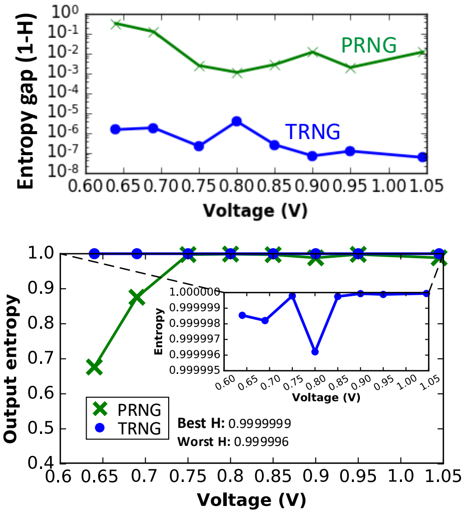


Fig. 8. Measured entropy and entropy gap (lower is better) of PRNG and TRNG bits vs. V_{dd}. TRNG outputs remain robust even as PRNG entropy degrades with V_{dd}.

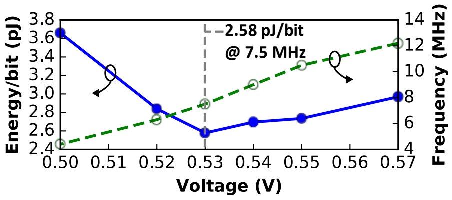


Fig. 7. Energy/bit and operating frequency vs. V_{dd}.

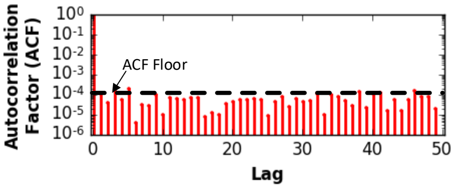


Fig. 9. Measured TRNG autocorrelation on 100M consecutive bits (log scale).

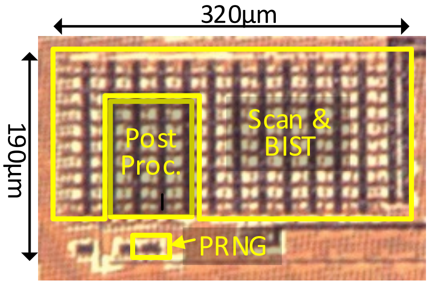


Fig. 10. Die micrograph of test chip.

Table 1. NIST randomness benchmark measurements. Each test uses 100 runs of 1Mb.

	Pass Rate for NIST Pub 800-22 Tests (All "PASS")*			
	Nominal		Min. Energy	
Voltage (V)	1.0		0.53	
Temperature (˚C)	-20	100	-20	100
Frequency	0.98	0.98	1.0	0.99
Block Frequency	1.0	1.0	0.99	1.0
Cumulative Sums	0.98	0.98	1.0	0.99
Runs	1.0	0.99	0.98	0.97
Longest Run	0.98	0.96	1.0	1.0
Rank	1.0	1.0	1.0	1.0
FFT	1.0	0.99	0.99	0.97
Non-Overlap. Template	0.98	0.98	0.98	0.97
Overlap. Template	0.98	1.0	0.99	1.0
Universal	1.0	0.99	0.99	0.99
Approximate Entropy	0.98	0.97	0.98	0.98
Rand. Excursions	1.0	0.98	1.0	0.98
Rand. Excursions Variant	1.0	0.98	0.98	0.98
Serial (1)	1.0	0.99	0.98	0.97
Serial (2)	1.0	0.99	0.99	0.99
Linear Complexity	0.97	1.0	1.0	0.98

*96/100 required to PASS, except Random Excursions (63/67)

Table 2. Comparison with related works. This work achieves the lowest energy/bit among digital TRNGs.

	This work		ISSCC'17 [1]		ISSCC'14 [2]	JSSC'16 [3]		JSSC'16 [7]	
Technology	65nm		65nm		28nm/65nm*	14nm		40/180nm*	
Entropy Source	Metastability		Jitter		Jitter	Metastability		Jitter	
NIST Pass	All		All (3 LSBs)		All	All		All	
V_{dd}	0.53	1.0	1.08	1.2	0.9	0.4	0.75	0.6	0.9
Bit rate (Mbit/s)	3.2	86.0	8.2	9.9	23.16	8.6	162.5	2.0	0.45
Efficiency (pJ/bit)	2.58	6.08	35.47	42.17	23	3	9.23	11	23
Power (µW)	8.33	523	289	418	159	27	1500	5	46
Area (mm^2)	0.01		0.00092		0.00037	0.00101		0.00083	
V_{dd}-robust	Yes		Yes		Yes	Yes		Yes	
Temp-robust	Yes		Not reported		Not reported	Not reported		Yes	