A 4Gbps 0.57pJ/bit Process-Voltage-Temperature Variation Tolerant All-Digital True Random Number Generator in 45nm CMOS

Suresh Srinivasan*, Sanu Mathew, Vasantha Erraguntla*, Ram Krishnamurthy
*Circuits Research Lab, Bangalore Design Lab, Intel Corporation, Bangalore, India
Circuits Research Lab, Hillsboro, OR 97124, USA
suresh.srinivasan@intel.com, sanu.k.mathew@intel.com, ram.krishnamurthy@intel.com

Abstract

This paper describes an all-digital on-die true random number generator implemented in 45nm CMOS technology, with random bit throughput of 4Gbps and total energy consumption of 0.57pJ/bit. A 2-step tuning mechanism enables robust operation in the presence of up to 20% fabrication-time process variation as well as immunity to run-time voltage and temperature fluctuation. The 100% use of digital components enables a compact layout occupying 1024μm^2 with high entropy/bit of 0.94, and scalable operation down to 0.5V, while passing all NIST RNG tests.

1. Introduction

High-entropy randomness is the foundation of data encryption, secure web communications and complex statistical analyses. Random number generators are therefore key blocks of processor platforms, responsible for generating secure keys in cryptography algorithms, session-IDs in secure internet protocols, mobile-internet-device IDs, Monte-Carlo simulations, and various operating system protocols. The quality of random numbers has a significant impact on the vulnerability of security algorithms and the accuracy of statistical simulations. Additionally, performance of secure mail/web servers and kernel scheduling routines is often limited by system stalls caused due to an insufficient pool of random numbers. Therefore there is a need for a high bit-rate high-entropy true random number generator that can be fabricated in a digital CMOS process with robust operation in the presence of process, voltage and temperature (PVT) variations. An all-digital design is motivated by the need for an on-die scalable energy-efficient implementation that can be easily ported to sub-32nm process technologies.

Software generated random numbers are pseudo-random in nature due to their dependence on events such as user-interface interrupts, page-faults, incoming TCP/IP requests and kernel system calls. Since these events are linked to user activity, they can be manipulated by malicious attackers, resulting in security loopholes [1]. True Random Number Generators (TRNGs) on the other hand, extract entropy from natural phenomena, such as device/thermal/flicker noise, or radioactive decay and are therefore immune to side-channel attacks. Among these natural sources, thermal noise is the most widely-used entropy source in TRNG designs. Thermal noise voltage is generated due to small fluctuations in channel current caused by the thermal vibrations of charge carriers in a conducting channel. This noise can be harvested by (i) amplifying the differential voltage across a pair of resistors using several stages of high-gain differential amplifiers and A/D converters [2][3][4], (ii) extracting edge-jitter from parallel chains of long ring-oscillators [5][6] (iii) sampling resolution state/time of meta-stable cross-coupled inverters [7][8] (iv) engineering SiN devices to increase the magnitude of thermal noise [9].

The resistor-amplifier-ADC based designs use analog circuit techniques that do not scale well to future digital process technologies and also require a very stable operating environment, free from deterministic sources of noise such as power-supply fluctuations, neighboring conductor coupling and temperature variations. The entropy of these designs may be disrupted due to random/systematic process variation or device drift due to aging. Mechanisms to counter device variation exist for ring-oscillator and meta-stability based schemes. These designs examine the generated bits to detect deterministic biases and introduce a counter-bias to offset the inherent mismatch. The counter-bias may be in the form of an analog bias voltage generated using a conditionally clocked capacitive charge-pump [7][8] or an external voltage source[5]. Once again, these bias generators are slow, do not scale easily to future process technologies and result in large area and power consumption. Robust operation of these designs in an environment of dynamic voltage scaling (1.1V down to 0.5V) is extremely challenging. As a result, hardware TRNGs are typically fabricated in older process technologies and housed in the chipset, where the communication between the off-chip RNG and processor becomes vulnerable to side-channel attacks and bus snooping.

In this work we propose the first all-digital on-die TRNG generating high-entropy random numbers with a throughput of 4Gbps in 45nm CMOS technology. The proposed TRNG design uses only digital components and is tolerant to both static and dynamic (aging based) process variations and also resilient to dynamic changes in voltage and temperature.

1063-9667/09 $25.00 © 2009 IEEE
DOI 10.1109/VLSI.Design.2009.69

2. Meta-stability Based TRNG

A cross-coupled inverter (Fig. 1) can be driven towards a meta-stable state by forcing input/output nodes *'a'* and *'b'* to identical values. This is achieved by using a pair of pre-charge devices to initialize both nodes to '1' when the *CLK*=0. At the rising edge of

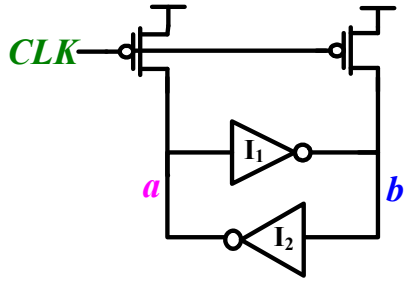

Figure 1: Bi-stable element: Entropy source

CLK, the bi-stable element enters a meta-stable state (Fig 2), where nodes 'a' and 'b' both settle to a value V_{meta}, which represents the intersection point of the VTCs of inverters I_1 and I_2. Resolution to either stable states of (*a=0,b=1*) or (*a=1,b=0*) depends on the magnitude of differential noise at *'a'* and *'b'* during the meta-stable period. A random noise source such as thermal noise at nodes *'a'* and *'b'* will result in a random resolution state during each evaluation phase of the clock. Thus a random bit is generated every cycle. This behavior of the bi-stable element forms the basis

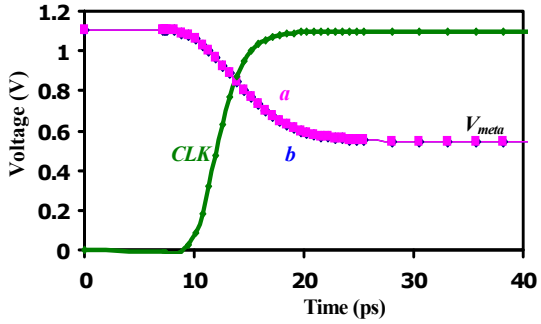

Figure 2: Bi-stable element in meta-stable state

of the proposed digital TRNG. The meta-stable state of the system can be disturbed by two events: (i) Noise on nodes *'a'* and '*b'* (ii) Device mismatch in the cross coupled inverters I_1 and I_2. Figure 3 shows the behavior of the bi-stable element in the presence of noise. Thermal noise magnitudes up to 3mV on nodes *'a'* and *'b'* quickly push the system out of meta-stability. As a result the element settles into the stable state of (*a=1, b=0*). A device mismatch in the cross-coupled inverters can introduce an intrinsic bias that will always push the

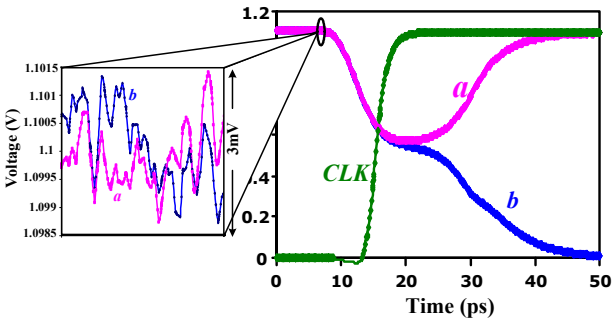

Figure 3: Bistable element behavior with thermal noise

bi-stable element in the direction of this bias, as it comes out of meta-stability. 1% mismatch in the PMOS device (Fig. 4) introduces a static bias that pushes inverter I_1 to resolve towards 1. As a result, even in the absence of thermal noise, the system will always resolve to the state (*a=0, b=1*), disrupting the random behavior of the system. Thermal noise magnitude will now have to be large enough to overcome the intrinsic bias caused due to device mismatch. This imbalance can significantly impact entropy of the generated bits. It should be noted that a systematic drift in PMOS/NMOS strengths may change the absolute values of inverter P/N skew, and impact the meta-stable point (V_{meta}). However, as long the relative P/N skews of I_1 and I_2 match each other, the behavior of the bi-stable element remains unchanged.

In a matched bi-stable element, a perturbation in the supply voltage presents itself as common-mode noise at nodes *'a'* and *'b'*. Since the meta-stable behavior of the

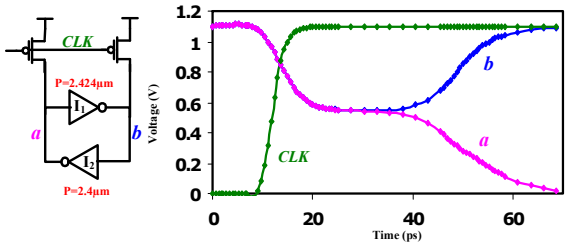

Figure 4: Bi-stable element behavior with 1% device mismatch, no thermal noise

cross-coupled system is affected only by differential noise, supply bounce should not affect the working of the system. Figure 5 shows the impact of 10% power supply noise injected into a matched system, while nodes 'a' and 'b' are at their meta-stable values. Supply noise affects the absolute value of V_{meta}, but does not tip the element out of meta-stability. However, in the presence of mismatch between the inverter devices, common-mode supply noise at V_{cc} will propagate differential components to nodes *'a'* and *'b'*, resulting in the disruption of the system from the meta-stable point. This undesired behavior can be avoided by

minimizing the mismatch between the cross-coupled inverters of the bi-stable element.

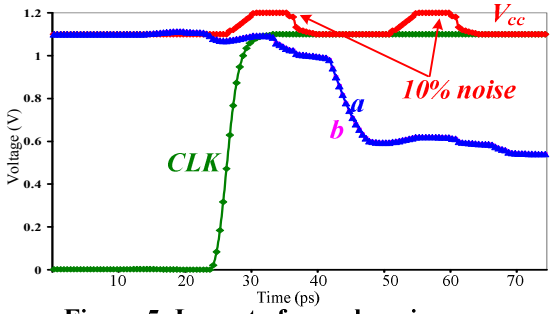

Figure 5: Impact of supply noise on a balanced meta-stable element

Thus we see that in order to accurately harness the entropy of thermal noise using a bi-stable element, we have to provide a mechanism for countering static and dynamic mismatches in the cross-coupled inverters. This would ensure that the resolution state of the system is solely governed by differential thermal noise at the two nodes, resulting in a random series of bits at nodes *'a'* and *'b'*.

3. PVT Tolerant All-Digital TRNG

A bi-stable cross-coupled inverter structure (Fig. 6) is used as the entropy harvester in the proposed all-digital TRNG. During pre-charge phase, a pair of pre-charge PMOS devices forces the system into a meta-stable state. A common NMOS transistor is used as a footer device to eliminate the short-circuit current that may exist due to the conducting NMOS devices during pre-charge. Successful generation of high-entropy random bits from such a system would require tuning mechanisms to counter mismatches between the inverters I_1 and I_2. Two techniques are introduced to handle mismatches that can disrupt the symmetry in the bi-stable element:

A. Coarse-grained Tuning using Programmable Inverters:

Inverters I_1 and I_2 are converted to programmable inverters with 8 digital configuration bits, *pconf[3:0]* and *nconf[3:0]* that control effective drive strengths of the PMOS and NMOS transistors respectively (Fig. 7). These scannable bits are used to conditionally turn ON parallel PMOS/NMOS legs, thereby controlling the P/N skew of the inverter. An increase in NMOS device

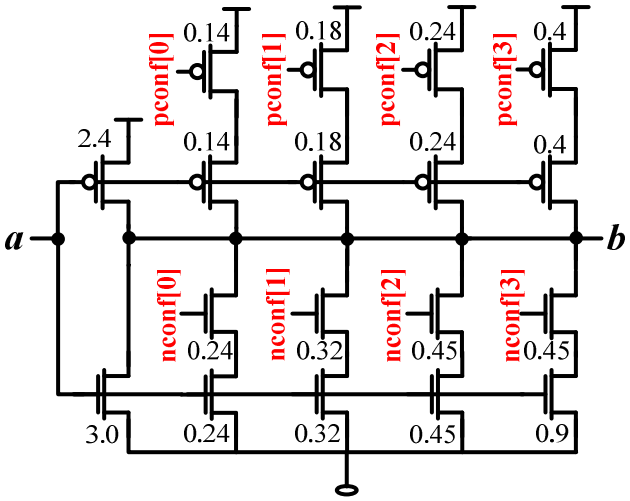

Figure 7: Programmable Inverter with 8 configuration bits

strength in inverter I_1 can be countered by turning ON an appropriate number of parallel NMOS legs in inverter I_2. The legs are weighted to tune out device mismatches in a range of 0-19% with a granularity of ~1% (Table I). This provides a coarse-grained knob for tuning out imbalances in the bi-stable element caused due to device width variation. The configurable inverters can also be programmed to tune out drive strength imbalances caused due to V_t mismatches. V_t variations result in a shift in inverter P/N skew from the

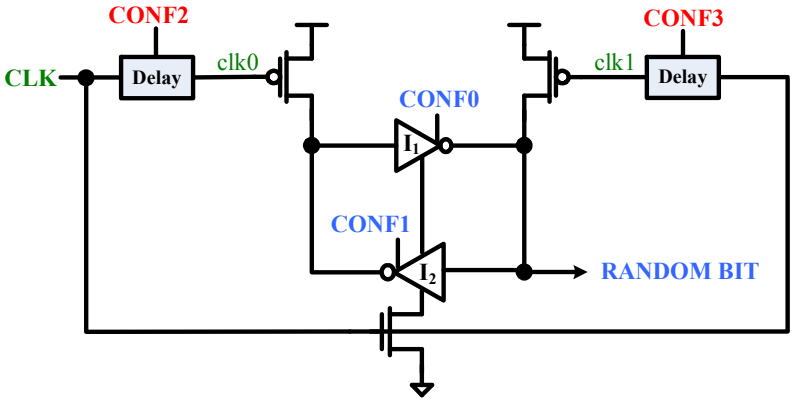

Figure 6: All-digital TRNG Organization

nominal designed value. This shift can be countered using either NMOS or PMOS configuration bits. Choosing the appropriate value of *nconf[3:0]* enables tuning out up to 158mV V_t mismatch (Table II). *Pconf[3:0]* can also be used to compensate for V_t

Table I: Tuning out device size mismatch

conf[3:0]	% NMOS size mismatch	% PMOS size mismatch
0000	0	19
0001	1	18
0010	3	17
0011	4	16
0100	5	14
0101	6	13
0110	7	12
0111	8	11
1000	10	8
1001	11	7
1010	12	6
1011	13	5
1100	14	3
1101	15	2
1110	16	1
1111	17	0

mismatches up to 24mV at a finer granularity than *nconf[3:0]*. Thus with the right combination of *pconf[3:0]* and *nconf[3:0]* a large range of process variation induced imbalances can be countered.

B. Fine-grained Tuning using delayed clocks:

A finer granularity of tuning can be achieved by separating out the pre-charge clocks and introducing a tunable skew between them using programmable delay generators (Fig. 8). By controlling the pre-charge release times for nodes 'a' and 'b', relative delay is introduced between the start of evaluation phase at each node, thereby injecting a directional bias on one side of the bi-stable element. This bias may be used to counter inherent voltage/temperature biases present in the system. Configuration bits *conf[3:0]* in the programmable clock delay element (Fig. 9) are scanned in through the scan-chain to control the PMOS drive strengths of the clock inverter, providing a tunable rising clock edge. This tuning mechanism provides finer granularity of control, enabling mismatch compensation as small as 0.05% in NMOS devices and 0.35% in PMOS devices. It will be demonstrated in section 4 that this level of granularity is sufficient to effectively tune out all residual mismatches, enhancing the sensitivity of the system to differential thermal noise.

Table II: Tuning out V_t mismatch using nconf

nconf	NMOS Vt mismatch (mV)	PMOS Vt mismatch (mV)
0001	1.5	7.8
0010	2.7	13.8
0011	3.1	23.8
0100	4.0	31.8
0101	6.0	41.8
0110	8.0	47.8
0111	10.0	57.8
1000	11.0	81.8
1001	12.0	91.8
1010	14.0	97.8
1011	14.5	107.8
1100	16.0	117.8
1101	18.0	137.8
1110	20.0	147.8
1111	21.0	157.8

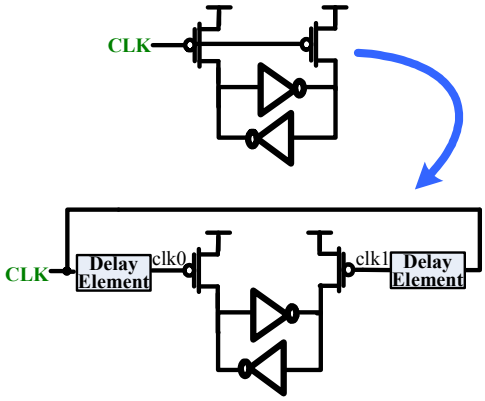

Figure 8: Separate pre-charge clocks for finer control

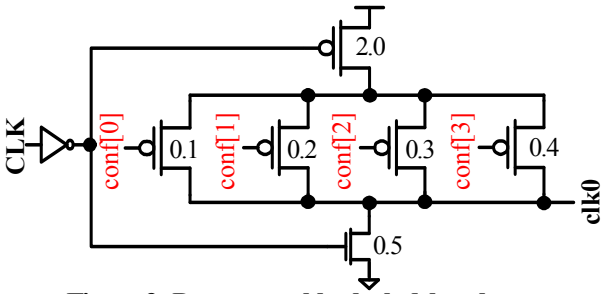

Figure 9: Programmable clock delay element

4. 45nm CMOS Implementation Results

In a 1.1V, 45nm CMOS process [10] the TRNG operates at 4GHz and generates 1 random bit each cycle, resulting in a throughput of 4Gbps, and total power consumption of 2.26mW, with leakage component of 0.3mW. RMS thermal noise at the inverter nodes is ~1.5mV. Six RNG tests (Table III) from the NIST suite [11] were applied to a bit-stream of 500 consecutive bits generated by the TRNG. Each test generates a *P-value*, an indicator of bit-stream entropy. A threshold of *P-value>0.01* is the essential pass criteria for a test. A frequency test *P-value*=1 indicates a bit-stream with equal number of 1's and 0's.

Table III: NIST Test Suite

Frequency test	Measures ratio of 1's vs. 0's in bitstream
Block frequency test	Frequency test at a block level
FFT	Examines periodicity of sequence
Long runs	Flags long runs of 0's or 1's
Cumulative sums	Computes sub-sequence partial sums
ACF Plot	Bitstream autocorrelation for lags 0-200

Table IV shows the *p-values* for the test suite in a perfectly balanced TRNG (0% mismatch) at 1.1V, 110°C. These results indicate that the magnitude and entropy of raw thermal noise at the inverter nodes, with the device sizes shown in Fig. 7, is high enough to generate a random bitstream that easily passes all NIST tests. As explained earlier, the presence of mismatch can reduce the sensitivity of the design to this thermal noise. A 2-step tuning mechanism is used to eliminate the effect of any process-variation-induced imbalance in the system. In the first step, the coarse-grained control is adjusted by scanning in 16 configuration bits (CONF0 and CONF1 in Fig. 6). Figure 10 shows the effect of coarse-grained tuning on the frequency test results in the presence of NMOS device mismatch ranging from 0% to 17%. Effective mismatches in V_t, transistor width/length, via/line resistance are simulated by varying the NMOS device size on inverter I_1 (or I_2) from its nominal value of 3µm to a max value of 3.51µm at a granularity of 0.05%. The programmable inverter I_2 (or I_1) is then configured to counter this mismatch. Coarse-grained tuning successfully counterbalances all mismatches that fall in the neighborhood of settings in Table I. This condition is represented by the peaks in *p-value* in Figure 10. It may also be observed that though the granularity of adjustment in this step is ~1%, we obtain passing *p-values* for mismatches that fall outside this window. This is due to the fact that the magnitude of thermal noise is large enough to independently overcome up to 0.4% imbalance in device sizes. Mismatches greater than 0.4% result in failing *p-values*. These mismatches require the fine-grain control provided by the clock delay generator. In the 2nd step of tuning, the relative skews of pre-charge clocks *clk0* and *clk1* (Fig. 8) is adjusted. At the end of this step, devices can be balanced to within 0.05% of perfect balance, well within the range of thermal noise sensitivity. Mismatches that caused the frequency test to fail in the first step of tuning are handled by scanning in the appropriate clock configuration bits (CONF2 and CONF3 in Fig. 6). At the end of the second tuning step, mismatches have been minimized to the extent that random thermal noise injects sufficient entropy into the

Table IV: Test Results with 0% mismatch, 1.1V, 110°C

Frequency test	*Pvalue* = 1.00	**Pass**
Block frequency test	*Pvalue* = 0.608	**Pass**
FFT	*Pvalue* = 0.746	**Pass**
Long runs	*Pvalue* = 0.676	**Pass**
Cumulative sums	*Pvalue* = 0.629/0.38	**Pass**

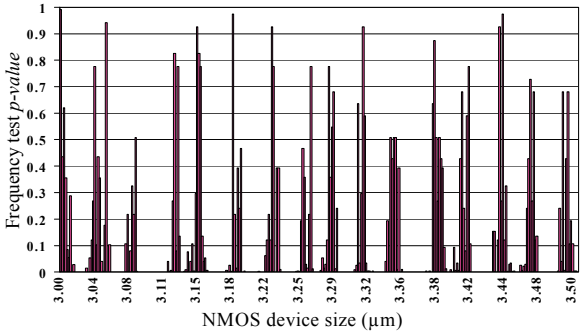

Fig 10: Coarse-grained tuning to counter device mismatch

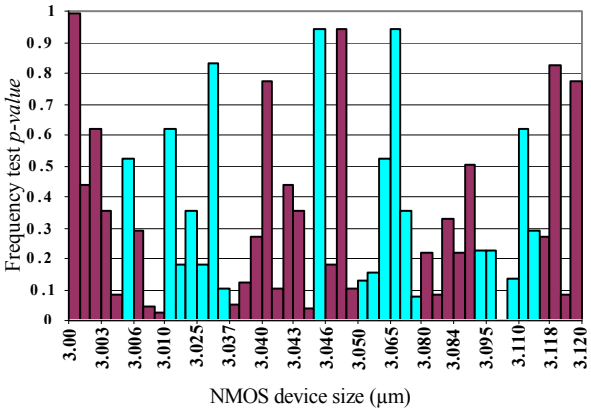

Fig 11: Fine-grained tuning using clock delay generator

Table V: NIST Test Results with 20% mismatch

Operating Conditions	**Vcc=1.1V, T=110°C**	
Configuration	**Nconf0=0000, Nconf1=1111 Pconf0=0000, Pconf1=0000 Clkconf0=0010, Clkconf1=1100**	
Frequency test	*Pvalue* = 0.33	**Pass**
Block Frequency test	*Pvalue* = 0.38	**Pass**
FFT test	*Pvalue* = 0.88	**Pass**
Long Runs test	*Pvalue* = 0.96	**Pass**
Cumulative Sums test	*Pvalue* = 0.48/ 0.30	**Pass**

bit-stream (Fig. 11). Similar results are obtained for all NIST tests on a run of 500 consecutive bits obtained from the TRNG with a worst-case mismatch of 20% on the NMOS device in inverter I_1. (Table V). The 2-step

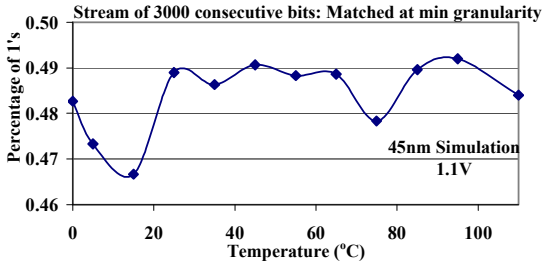

Fig. 12: Effect of temperature on randomness

tuning process minimizes effective device mismatch to a small enough magnitude such that any remnant imbalance is overcome by thermal noise. In this condition, the system is also immune to common mode noise events like power supply noise or temperature variations (Fig. 12). The all-digital nature of the TRNG results in good V_{cc} scaling behavior. Sensitivity of the RNG to thermal noise increases at low voltage,

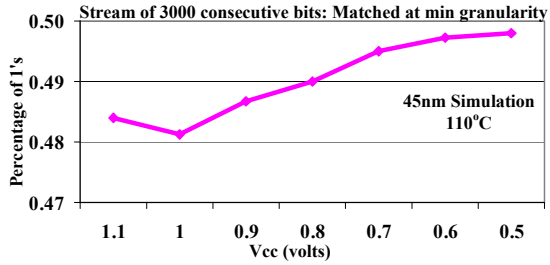

Fig. 13: Impact of V_{cc} on randomness

resulting in probabilities approaching ideal values as supply approaches 0.5V (Fig. 13). NIST autocorrelation tests at different voltages and temperatures with 20% mismatch show zero correlation with 95% confidence for up to 200 lags (Fig. 14). This confirms the

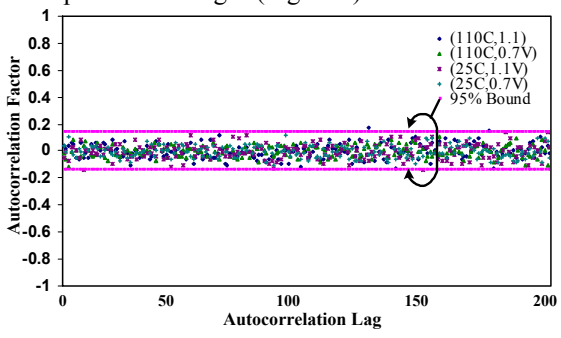

Fig. 14: Effect of Voltage/Temperature on Autocorrelation

robustness of this design to common-mode events like voltage/temperature variations.

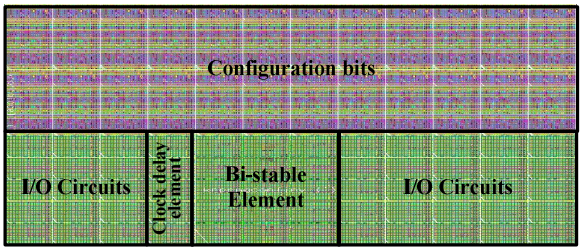

Fig 15: 45nm All-digital TRNG layout

5. Summary and Conclusion

An all-digital TRNG targeted for an on-die encryption engine with high tolerance to PVT variations is implemented in 45nm CMOS technology. A 2-step coarse/fine tuning mechanism provides tolerance to 20% mismatch in device characteristics, enabling 4Gbps throughput, generating an output bit-stream that passes NIST RNG tests with entropy/bit of 0.94. The use of 100% digital components ensures a compact layout measuring 1024μm^2 (Fig. 15) with a low energy consumption of 0.57pJ/bit. The design shows good V_{cc} scaling behavior with entropies approaching ideal values at supply voltages around 0.5V and good robustness to temperature and voltage variations.

6. Acknowledgments

The authors thank C. Dike for discussions and valuable feedback and M. Haycock, G. Taylor, S. Borkar and J. Schutz for encouragement and support.

7. References

[1] Z.Gutterman et al, "Analysis of the Linux RNG", *IEEE Symposium on Security and Privacy*, pp 371-385, May 2006.

[2] R. Brederlow, et al, "A low-power TRNG using random telegraph noise of single oxide-traps", *ISSCC Dig. Tech. Papers*, pp. 536-537,Feb., 2006.

[3] B. Jun and P. Krocher, "The Intel RNG", White Paper, *http://www.cryptography.com/intelRNG.pdf*, 1999

[4] C. Petrie and J. Connelly, "Noise-based IC RNG for applications in cryptography," *IEEE Trans. Circuits & Systems-I*, vol. 47, no. 5, pp.615-621, May 2000

[5] M. Bucci, L. Germani, R. Luzzi et al., "A high-speed oscillator-based truly random number source for cryptographic applications on smart card IC," *IEEE Trans. on Computers*, vol. 52, pp. 403-409, April 2003.

[6] S. Fujita, et al., "Si nanodevices for RNG circuits for cryptographic security", *ISSCC Dig. Tech. Papers*, pp. 294-295, Feb. 2004.

[7] D. Kinniment and E. Chester, "Design of an on-chip random number generator using metastability," *Proc. ESSCIRC.*, pp. 595-598, Sep., 2002.

[8] C. Tokunaga, et al, "TRNG with a metastability-based quality control," *ISSCC Dig. Tech. Papers*, pp. 404-405, Feb. 2007.

[9] Mari Matsumoto et al, "1200μm^2 Physical RNG based on SiN MOSFET for Secure Smart-Card Application", *ISSCC Dig. Tech. Papers*, pp. 414-415, Feb. 2008.

[10] K. Mistry et al., A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging, *Proc. IEEE IEDM*, Dec. 2007, pp. 247-250.

[11] 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*, 2001