A low-power and high-speed True Random Number Generator using generated RTN

James Brown, Rui Gao, Zhigang Ji*, Jiezhi Chen*(2), Jixuan Wu(2), Jianfu Zhang, Bo Zhou, Qi Shi, Jacob Crowford, and Weidong Zhang
Faculty of Engineering and Technology, Liverpool John Moores University, Liverpool, UK, email: z.ji@ljmu.ac.uk
(2) School of Information Science and Engineering, Shandong University, Jinan, P. R. China, email: chen.jiezhi@sdu.edu.cn

Abstract: A novel True Random Number Generator (TRNG), using random telegraph noise (RTN) as the entropy source, is proposed to address speed, design area, power and cost simultaneously. For the first time, the proposed design breaks the inherent speed limitation and generates true random numbers up to 3Mbps with ultra-low power. This is over 10 times faster than the state-of-the-art RTN-TRNG **[6]**. Moreover, the new design does not require selection of devices and thus avoids the use of large transistor array and laborious post-selection process. This reduces the circuit area and the cost. The proposed TRNG has been successfully validated on three different processes and they all passed the National Institute of Standards and Technology (NIST) tests, making it a suitable candidate for future cryptographically secured applications in the internet of things (IoT).

Introduction: True random number generators (TRNGs) harvest physical randomness as entropy sources and are heavily used in cryptography and security **[1]**. However, their power consumption and design complexity are often high **[1-5]**. The recently-proposed TRNG using Random Telegraph Noise (RTN) **[6-8]** has been considered as an ideal solution for future IoT applications (**Fig.1**), due to its simplicity, low power and robustness against temperature and supply voltage variations. Their practical use, however, is hindered by two major deficiencies: 1) *Device Selectivity*: RTN-TRNGs is based on one nano-scaled transistor exhibiting clear RTN signal by one pre-existing trap. However, the percentage of such devices in one wafer is very low (**Fig.2a&b**) **[9]**. A large transistor array is usually needed in the design, out of which one transistor will be selected manually by tuning the circuit after fabrication. This leads to larger design area and higher cost **[10-11]**. 2) *Low speed*: The speed of the RTN-TRNG has strong correlation with τ, the sum of the times to capture and emission (τc, τe) **[8]**. The opposite voltage dependence of τc and τe imposes an inherent limit for speed. *In this work, a novel RTN-TRNG is proposed to tackle the above drawbacks without extra design penalty. The design is successfully applied on three different processes (**Table II**) and maximum NIST-validated bit rate of 3Mbps has achieved.*

RTN generation and characterization: In addition to pre-existing traps, some traps generated by electrical stress can also induce RTN **[12-13]**. This provides a pathway to 'insert' RTN into any nano-scaled device. One nFET from process A1 is used as an example in the following. After hot carrier stress for 50s, a new trap can be generated, which still exists after accelerated recovery (**Fig.3a&b**). This generated trap shows clear RTN (**Fig.3c-e**). It is found in our process, A1, that although RTN only exists in 17% of the fresh devices, clear RTN can be observed in more than 80% of the devices after hot carrier stress (**Fig.4**). Therefore, the bulky transistor array and the post-selection **[6]** can be avoided when the TRNG entropy is taken from these generated traps. Similar to the pre-existing ones, these generated traps also show strong voltage dependence (**Fig.5a**) and are highly stable (**Fig.5b**). Their profile extracted from the RTN measurements **[14]** suggest that they could be away from the Si/dielectric interface, which further supports their nature of generation **[15]** (**Fig.6a&b**).

RTN acceleration with AC operation: All the existing RTN-TRNGs operate under DC condition and are slow. The voltage tuning is usually applied to optimize its bit rate. Due to the opposite voltage dependence, no matter what gate voltage is applied, either τc or τe will increase, hindering further improvement (**Fig.7a**) **[6-8]**. To tackle this dilemma, we propose to operate RTN under AC condition: The current is sensed under VgL for half cycle after VgH is applied for the other half (**Fig.8**). This allows τc and τe being controlled by VgH and VgL independently (**Fig.7a**). The reduction of τc and τe is only limited by measurement accuracy (for the lowest-allowable VgL) and device reliability (for the highest-allowable VgH). Compared with DC operation, AC operation can easily accelerate RTN by hundreds of times (**Fig.7a**). A clear difference can be observed for the same nFET when operating under DC (**Fig.9a-c**) and AC (**Fig.9d-f**) respectively. It has been reported that τc and τe can also be reduced when the applied frequency increases **[16]**. Such frequency dependence can also be observed in our measurements (**Fig.9g-i**). This can further reduce τc and τe, leading to a faster TRNG. The amplitude of RTN is also found to be large under AC condition (**Fig.9a-i**), because the sensing voltage, VgL, is already in the subthreshold region where strong percolation is expected **[17]**. What is worth noting is that such sub-threshold sensing scheme also naturally reduces power consumption.

RTN-TRNG design and validation: Since RTN is produced at VgL under AC condition, two nFETs are used (**Fig. 10**). By applying 180°-shifted gate biases, two RTNs are generated every half cycle. After amplification and digitization, they can be combined together through a transmission gate. For a given trap, it is known that the sum of the times to capture and emission ($\tau c+\tau e$), when averaged, is a constant against time (Eqns in **Fig.11**). Therefore, by toggling only at rising edge of the RTN, the new trace can be obtained with 1s and 0s of equal probability, making it truly random without complicated post-processing **[8]**. After sampling at a given clock frequency, the random number stream is generated with high entropy (**Fig.12**). It passes all the NIST tests at the maximum speed of 2Mbps (**Fig.13a&b**), over 10 times faster than state-of-the-art one **[6, 18]**.

Yield and applicability to process: The yield is estimated by evaluating 14 TRNGs. Over 90% pass at 1Mbps and almost 50% even reach 3Mbps making it readily usable in practice (**Fig.14**). The same design is also applied on another two processes, A2 and A3 (**Table II**). They also passed the NIST tests (**Table III**).

Conclusions: A novel RTN-TRNG design is demonstrated. By injecting randomness into the transistors through electrical stress and operating under AC domain, the new design provides a solution to address speed, design area, power consumption, reliability and cost simultaneously. The proposed TRNG has passed the NIST tests, making it readily applicable for cryptographically secured applications.

Acknowledgement: This work is supported by Engineering and Physical Science Research Council of UK (EP/L010607/1) and China key Research and Development Program (2016YFA0201802). The authors would like to thank imec for providing samples for this research.

References: **[1]** M. Bucci *et al*, p. 403, IEEE Trans. Comput 2003. **[2]** C. Tokunaga *et al*, JSSC 2008. **[3]** S. Srinivasan *et al*, p. 203, VLSI-C 2010. **[4]** N. Liu *et al*, p. 203, VLSI-C 2011. **[5]** K. Yang *et al*, p. 280, ISSCC 2014. **[6]** R. Brederlow *et al*, p. 79, ISSCC 2006. **[7]** T. Figliolia *et al*, p.17, ISCAS 2016. **[8]** A. Mohanty *et al*, p.2248, TVLSI 2017. **[9]** C. Chen *et al*, p. 190, IRPS 2011. **[10]** M. D. Giles *et al*, p. 501, VLSI 2015. **[11]** N. Tega *et al*, p.630, IRPS 2011. **[12]** M. Toledano-Luque *et al*, p.978, VLSI 2012. **[13]** R. Gao *et al*, p.778, IEDM 2016. **[14]** T. Nagumo *et al*, p.628, IEDM 2010. **[15]** C. Lu *et al*, p. 936, TED 2014. **[16]** R. Wang *et al*, p. 978, VLSI-T 2012. **[17]** L. Gerrer, *et al*, p. 226, ESSDERC 2015. **[18]** NIST, Pub 800-22, 2001

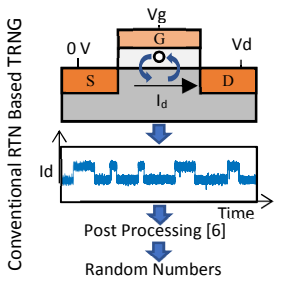


Fig. 1 Conventional method for true random number generation using RTN in nano-scaled transistors.

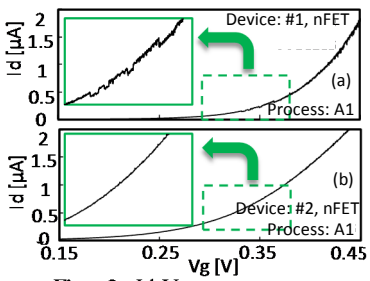


Fig. 2 Id-Vg sweep on two devices **(a)** with and **(b)** without RTN. Process A1 is used.

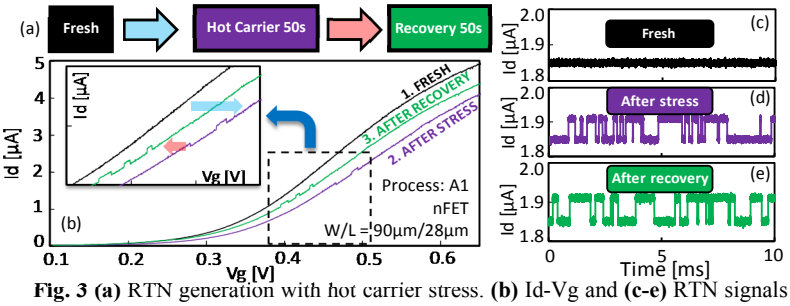


Fig. 3 (a) RTN generation with hot carrier stress. **(b)** Id-Vg and **(c-e)** RTN signals under constant Id=1.8μA after each step. Stress: Vg=Vd=1.9V,50s,125°C. Recovery: Vg=Vd=-1.5V/0V,50s,125°C. All measurements are at 27°C.

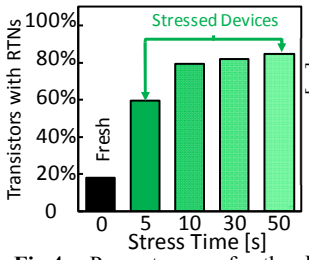


Fig.4 Percentage of the devices with observable RTN after different time of stress.

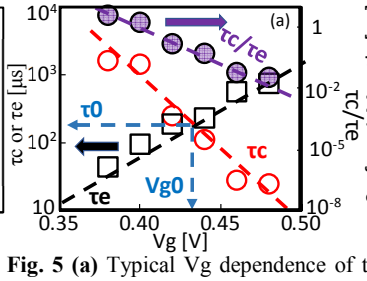
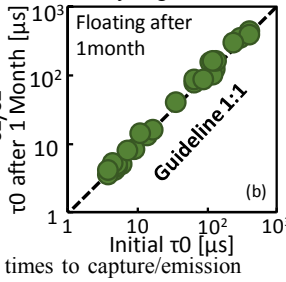
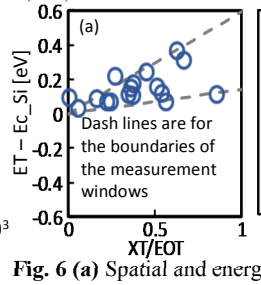


Fig. 5 (a) Typical Vg dependence of times to capture/emission (τc/τe) of the generated RTN. **(b)** Comparison of τ0 between fresh and floating over 1month after stress.





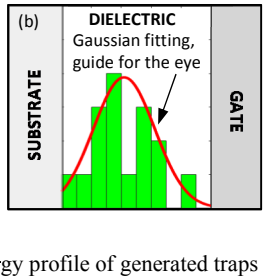
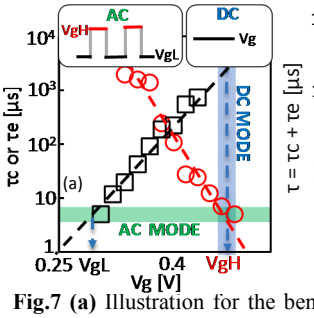


Fig. 6 (a) Spatial and energy profile of generated traps using method in ref. 14. **(b)** The histogram of trap numbers against the spatial location.



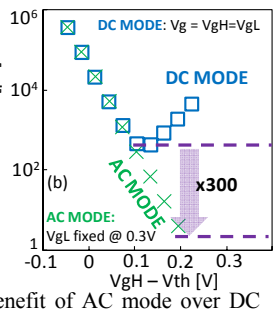


Fig.7 (a) Illustration for the benefit of AC mode over DC mode to accelerate RTN. **(b)** The comparison of the time constant, τ = τc+τe, between DC and AC operation modes.

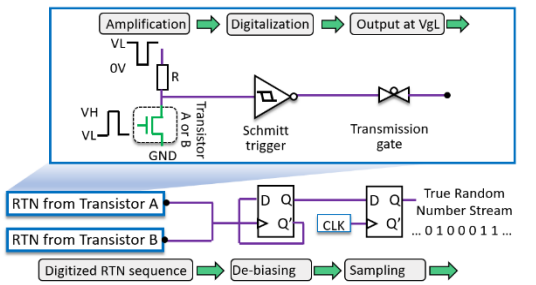


Fig.10 The circuit schematic for the proposed RTN-based TRNG. Two transistors are used to generate RTN alternatively within one voltage period.

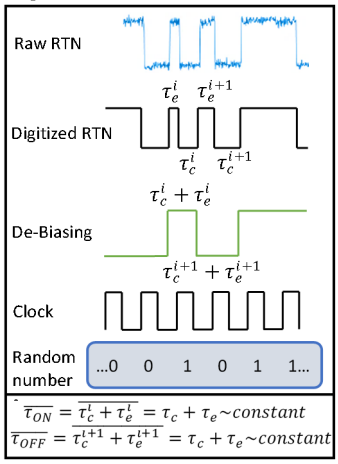


Fig.11 Procedure to convert RTN signal to random bit stream. By using the intrinsic property of RTN (Equations), true randomness is obtained without post-processing.

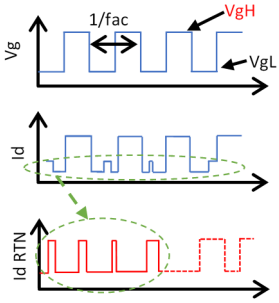


Fig. 8 Illustration of the procedure for measuring RTN at AC operation mode.

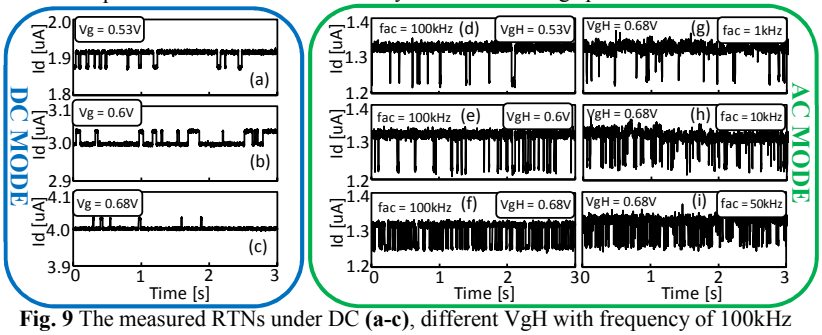
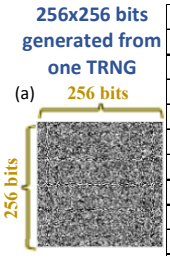
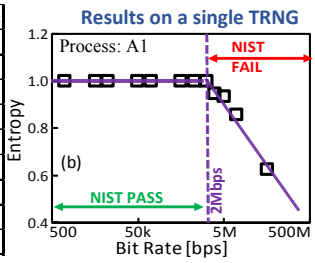


Fig. 9 The measured RTNs under DC **(a-c)**, different VgH with frequency of 100kHz **(d-f)** and different frequencies, with VgH = 0.68V. All the measurements are taken at 27°C. For AC mode, VgL = 0.28V is always used.

	A1	A2	A3
Technology	22nm	28nm	45nm
W/L	90nm/28nm	90nm/70nm	
EOT	1.0nm	1.3nm	1.45nm
Dielectric	HKMG	HKMG	HKMG
Structure	FinFET	Planar	Planar



Test ID	Test Name	p Value	Result
T1	Monobit test	0.5342	PASS
T2	Frequency tests (block)	0.2133	PASS
T3	Cumulative sums Test	0.7399	PASS
T4	Run Test	0.5341	PASS
T5	Longest Run (block)	0.9915	PASS
T6	DFT Test	0.1222	PASS
T7	Non-overlapping test	0.3505	PASS
T8	Entropy Test	0.3505	PASS
T9	Serial Test	0.2133	PASS



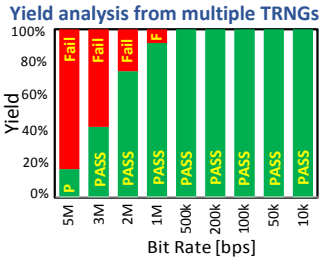


Fig. 12 (a) Bitmap of the random bits generated with 2Mbps bit rate. **Table. I** Results summary from the NIST tests for process A1 where α = 0.01, bit rate = 2MHz, bit sequence length = 40k, and number of bit streams = 100. **(b)** Entropy of the random a bit stream generated from process A1 at different bit rates.

Fig. 13 NIST test yield at different bit rates. 14 TRNGs are randomly chosen for the analysis. Process A1.

Test ID	p value	
	Process A2	Process A3
T1	0.0431 (Pass)	0.7399 (Pass)
T2	0.9114 (Pass)	0.9114 (Pass)
T3	0.7399 (Pass)	0.1223 (Pass)
T4	0.2133 (Pass)	0.9915 (Pass)
T5	0.3505 (Pass)	0.7399 (Pass)
T6	0.7399 (Pass)	0.9114 (Pass)
T7	0.9114 (Pass)	0.5341 (Pass)
T8	0.2133 (Pass)	0.7399 (Pass)
T9	0.5341 (Pass)	0.3505 (Pass)

Table. II (Top) Summary of the three different processes used in this work. **Table. III (Bottom)** Summary of the NIST tests for processes A2 and A3. Where α = 0.01. The bit stream is generated at 2Mbps.