An Asynchronous and Low-Power True Random Number Generator Using STT-MTJ

Ben Perach and Shahar Kvatinsky

Abstract— **The emerging spin-transfer torque magnetic tunnel junction (STT-MTJ) technology exhibits interesting stochastic behavior combined with small area and low operation energy. It is, therefore, a promising technology for security applications, specifically the generation of random numbers. In this paper, STT-MTJ is used to construct an asynchronous true random number generator (TRNG) with low power and a high entropy rate. The asynchronous design enables the decoupling of the random number generation from the system clock, allowing it to be embedded in low-power devices. The proposed TRNG is evaluated by a numerical simulation, using the Landau–Lifshitz–Gilbert (LLG) equation as the model of the STT-MTJ devices. Design considerations, attack analysis, and process variation are discussed and evaluated. We show that our design is robust to process variation, thus achieving an entropy generating rate between 99.7 and 127.8 Mb/s with 6–7.7 pJ per bit for 90% of the instances.**

Index Terms— **Hardware security, magnetic tunnel junction (MTJ), memristors, random number generation, true random number generator (TRNG).**

I. Introduction

SECURITY is a major concern in modern digital systems. One of the main tools used in security is cryptography, which is used to encode information that only authorized entities can access. However, security applications need to be implemented with caution. Changing the cryptographic algorithm or its assumptions, even to a limited extent, can compromise the entire system. One such crucial part of cryptographic algorithms is the generation of the cryptographic keys [1]–[5].

The key of a cryptographic algorithm is the secret of the encryption scheme. The algorithm itself is assumed to be publicly known, and the key is the only missing information needed to reveal the encrypted data [2]. Hence, an adversary will try to obtain the key. Since the key is of finite size, the number of possible values for the key is finite as well, and if this number is too small, an adversary can try them all. In addition, if the adversary has partial knowledge of the key, such as some mathematical conditions between the key bits, this information can be used to reduce the number of options [3]–[5]. Hence, it is desirable to generate a random key with a uniform distribution on all of its possibilities, and so an adversary will have to try all of the options without a defined order. Processes that can generate a random number as the key are called random number generators (RNGs). Note that the design of the RNG itself, as a part of the encryption scheme, is also assumed to be publicly known.

One type of RNG is a TRNG, a true RNG [6], [7]. A TRNG is based on a physically random process (e.g., thermal noise), and the TRNG extracts that randomness to a usable form, such as digital numbers. The TRNG approach for generating random numbers is attractive since the generated number cannot be inferred from the state of the system but can only be predicted from the distribution of the physical random process [1], [2].

Current TNRGs use CMOS logic, such as ring oscillators (ROs) [8] or metastable latches [9], as their source for randomness. However, emerging technologies [10]–[14] offer new and interesting alternatives due to their smaller area and lower power consumption when compared with the transistors. A small-area and low-power TRNG, based on a random process in emerging technologies, will reduce power consumption or enable secure communication for small- or low-power electronic devices (e.g., Internet-of-Things devices and mobile devices). One such technology is the spin-transfer torque magnetic tunnel junction (STT-MTJ) [10], [13], [14]. As an emerging memory technology, STT-MTJ (or STT-MRAM) has relatively low operating energy and small area, and its switching time stochasticity has been thoroughly studied [15].

While the previously proposed STT-MTJ-based TRNGs [16]–[20] require a strict time measurement to achieve high randomness, we propose an asynchronous TRNG. The proposed design relies on discharging a capacitor simultaneously through several STT-MTJ devices. The process ends when the capacitor is sufficiently discharged; the generated random number is extracted from the final state of the STT-MTJ devices. Since the capacitor is discharged asynchronously, the random number generation is independent of a clock signal. The only time measurement required in the TRNG design is the duration to discharge the capacitor, which can be approximated by defining a minimum time. Waiting for more than the minimum time does not influence the randomness of the output. Therefore, this design can be

Manuscript received January 10, 2019; revised April 16, 2019 and June 8, 2019; accepted June 24, 2019. Date of publication August 7, 2019; date of current version October 23, 2019. This work was supported in part by the European Research Council (ERC) through the European Union’s Horizon 2020 Research and Innovation Programme under Grant 757259, in part by the Technion Hiroshi Fujiwara Cyber Security Research Center, and in part by the Israel Cyber Bureau. *(Corresponding author: Ben Perach.)*

The authors are with the Andrew and Erna Finci Viterbi Faculty of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: benperach@campus.technion.ac.il; shahar@ee.technion.ac.il).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2019.2927816

embedded in low-frequency (i.e., low-power) devices without loss of randomness. Furthermore, the use of several STT-MTJ devices increases the extracted randomness in each operation and improves the robustness to process variation.

To evaluate the proposed design, numerical simulations of the stochastic physical model of the STT-MTJ were performed. The randomness of the proposed design and the effect of internal and external influences were measured, including the robustness of the design to process variation. Possible attack venues are discussed and mitigation options are presented. To the best of our knowledge, this is the first study of an STT-MTJ-based TRNG that includes the analysis of attacks, essential for every security-related work.[1]

II. Background

A. True Random Number Generators

RNGs are divided into two main groups, pseudo-RNGs (PRNGs) and TRNGs. PRNGs are deterministic algorithms that only appear to generate a random sequence of numbers. A cryptographic key generated by a PRNG might compromise the encryption since the PRNG outputs are inherently connected [3]–[5], although some PRNGs are considered to be sufficiently secured for cryptographic use [22], [23]. TRNGs are designed to extract a random behavior of some physically random process [6]–[9], resulting in true randomness that can be explained according to some physical laws. The output of the TRNG can only be predicted according to the physical process probability distribution, even if all the information about the system (register values, voltage levels, and so on) is known prior to the TRNG operation.

CMOS TRNGs often use ROs [8], [24] or metastable latches [9] to generate random numbers. An RO is a chain with an odd number of NOT gates, where the output of the last NOT gate is the input of the first NOT gate, resulting in a ring of gates. Since the number of NOT gates is odd, all the outputs of the NOT gates oscillate between logic high and logic low. However, due to noise in the transistors, the rise and fall times of the gates are randomly changed, resulting in frequency variation of the oscillation. RO-based TRNGs use this frequency variation as the source of randomness: for example, they might compare several ROs [24] or measure the time until the occurrence of an RO-related event [8]. TRNGs based on metastable latches force a latch into an unstable equilibrium state and then release it. The stable state that the latch will end in depends on random noise, and therefore, a random number is generated.

The quality of the randomness of the TRNG can be measured by its output properties, which are ideally independent and uniformly distributed. In practice, this need not be the case for cryptographic use since the dependences and nonuniform distribution can be compensated for by postprocessing the output. Such postprocessing methods are referred to as randomness extractors [25], [26]. However, the closer to independent and uniformly distributed the TRNG output is, the simpler the extractor can be. Even if the output is uniform under regular operation, randomness extractors are still commonly used to compensate for real-world effects (e.g., process variations, wear-out, and interference) that might reduce the randomness of the TRNG output [7], [26].

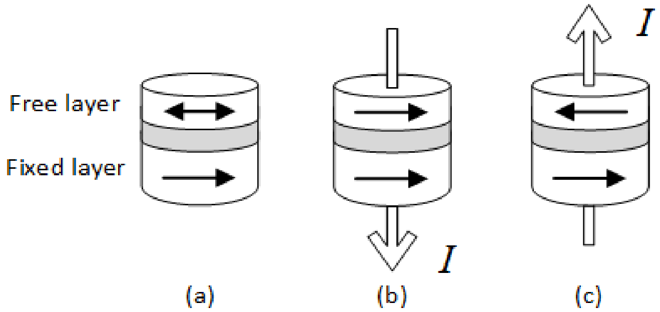


Fig. 1. In-plane STT-MTJ device and its operation. (a) STT-MTJ device structure with the free layer (top), the tunnel barrier layer (middle), and the fixed layer (bottom). (b) P and (c) AP states and the associated currents to switch toward the state.

Other important properties of the TRNG include robustness to process and environmental variations and a high generating rate. An adversary might change environmental parameters (e.g., electromagnetic field and temperature) to interfere with the operation of the TRNG and reduce its randomness. To guarantee that the TRNG is secure, it is critical to identify the underlying random physical process and the factors affecting it and test the TRNG under these factors. In addition, to ensure correct operation, statistical tests, referred to as online tests, are often performed on the TRNG output during run time, which means that they need to be lightly implemented, making them less thorough than the statistical tests performed at design time.

New emerging technologies, such as the STT-MTJ [10], [13], [14], show small area and low operation energy compared with the transistors. In addition, they exhibit ample stochasticity in their operation [15], making them interesting candidates as the randomness source for new TRNG designs.

B. STT-MTJ Devices

An STT-MTJ is a device composed of two ferromagnetic layers with a tunnel barrier layer between them [15], [27]. One ferromagnetic layer, the fixed layer, has a fixed magnetization direction. The other ferromagnetic layer, the free layer, can switch its magnetization direction. In this paper, in-plane MTJs are used, where the magnetization direction of the ferromagnetic layers is in the plane of the layers. Fig. 1(a) shows the MTJ structure. The direction of the free-layer magnetization can be changed by a current through the device, and it has two stable states, parallel [P; see Fig. 1(b)] or anti-parallel [AP; see Fig. 1(c)] to the direction of the fixed layer. The direction of the current determines the change in the magnetization direction. Other states (i.e., other directions of the free-layer magnetization) are unstable.

The STT mechanism enables the switching of the orientation of the free-layer magnetization. The electrons passing through a ferromagnetic layer tend to align their magnetic moment in the direction of the magnetization of the layer. Thus, electrons that pass through the fixed layer first are

[1] Attack analysis was previously conducted only on other STT-MTJ-based security devices, such as physically unclonable functions [21].

aligned with its magnetization direction. When these electrons reach the free layer, its magnetization direction shifts toward the P state due to magnetic moment conservation [see Fig. 1(b)]. In the other current direction, electrons are reflected with magnetic moment direction opposite to that of the fixed layer and change the free layer to the AP state [see Fig. 1(c)]. However, a damping process pulls the free-layer magnetization to the closest stable state, requiring a sufficiently strong current for adequate time to enable a switch between the stable states.

The switching process between the P and AP states is random [15] due to the thermal fluctuations in the ferromagnetic layers. Although the current through the MTJ pushes the magnetization of the free layer to a certain stable state (through unstable intermediate states), thermal fluctuations will make the path to that state random, resulting in a random switching time. Even if no current is applied, the state of the STT-MTJ fluctuates constantly since the thermal fluctuations occur regardless of the existence of the current.

The state of the MTJ also determines its resistance, where the P state resistance is marked as R_{ON}, the AP state is marked as R_{OFF}, and $R_{\text{ON}} < R_{\text{OFF}}$. The resistance of the MTJ, when it is in a state other than P or AP, is between R_{ON} and R_{OFF}, and its exact value depends on the state [28]. To determine the state of the MTJ, a low voltage can be applied across it (sufficiently low not to incur a switch), the current can be measured, the resistance of the MTJ can be extracted (by Ohm's law), and the state of the MTJ can be inferred.

To model the operation of the entire STT-MTJ, the magnetization of the free layer is usually approximated to a single domain. The phenomenological Landau–Lifshitz–Gilbert (LLG) equation [29], with the addition of a stochastic term for the thermal fluctuations [30] and Slonczewski's STT term [31], can accurately describe the dynamics of the magnetization of the free layer. For current pulses with low or high current magnitudes, approximations and models for the distribution of the switching time exist [32]–[34]. For current pulses with intermediate current magnitudes, approximations for the switching time distribution and other models are also available [27], [35]. However, there is no model for the switching distribution in the intermediate current region for non pulse waveforms. In the last case, the LLG equation has to be solved numerically.

C. Previously Proposed STT-MTJ-Based TRNGs

Emerging technologies, such as memristors, have been proposed for TRNGs that operate by applying a current pulse through the devices to randomly switch them with approximately 50% probability [36]–[38]. The generated random number in these TRNGs is the state of the memristors at the end of the operation. Similarly, TRNGs based on STT-MTJ [16]–[20] have been proposed. In these designs, the current pulse is controlled by a feedback circuit in order to be robust to process variation and environmental changes. However, only Qu *et al.* [19] have analyzed the effects of process variation and proposed to use several MTJ devices in parallel to mitigate those effects.

Qu *et al.* [39] proposed a differential approach, where two STT-MTJs are connected in series and in reverse orientation. A current pulse is driven through both of the MTJs simultaneously and a dedicated mechanism terminates the pulse when one of the MTJs is switched. The output bit is determined according to the end state of both MTJs. This design was shown to be robust to process variation, operating voltage, and temperature due to the symmetry of the MTJs.

All of the aforementioned TRNGs use controlled current pulses to switch the MTJs. To measure the duration of the current pulses, a clock with a period smaller than or equal to the pulse duration is needed. In addition, the quality of the randomness will be influenced by the ability of the clock signal to accurately measure the pulse duration, further binding and complicating the system. Therefore, these TRNGs are incompatible with systems that have insufficient clock frequency or accuracy, such as low-power systems with a low-frequency clock.

Lee *et al.* [40] proposed to reduce the energy barrier between the P and AP states of the MTJ to enable faster switching with reduced energy. This was achieved by using specially designed MTJ devices. Their proposed TRNG design uses multiple MTJs (to reduce the effect of process variation) with low energy barriers, which were set to an unstable state and released, letting that the MTJs settle randomly to one of the stable states. Vodenicarevic *et al.* [41] proposed a TRNG based on a similar approach, where the STT-MTJ's energy barrier between the P and AP states is sufficiently low to enable spontaneous switching in a reasonable time, without the aid of external stimuli and solely by thermal fluctuations. These two approaches require the use of specially designed MTJs due to the reduced energy barrier [40], and despite the advantages of reduced latency and energy, this approach makes the design more vulnerable to external magnetic fields and, thus, to attacks.

Ghosh [21] comprehensively analyzed spintronics in security applications and showed that the MTJ-based security circuits are sensitive to the effects of an external magnetic field. Hence, it is essential to consider this effect when designing an MTJ-based TRNG to determine the security of such a design and the effects of nearby circuits. However, none of the aforementioned STT-MTJ-based TRNGs that rely on a current pulse [16]–[20], [39] was designed to account for the effects of an external magnetic field and the associated attacks. Lee *et al.* [40] and Vodenicarevic *et al.* [41] only briefly consider the effects of an external magnetic field, and their design relies on a special MTJ device that is more vulnerable to attacks. In this paper, we choose to use the standard MTJ devices that are easier to fabricate in a standard process and have better robustness against attacks. We provide a comprehensive evaluation of the effects of an external magnetic field on the proposed TRNG design.

III. Proposed TRNG Structure and Operation

The proposed TRNG generates N-bit numbers and is composed of a capacitor, N STT-MTJ devices, N sense amplifiers, and transistors that serve as switches, as shown in Fig. 2.

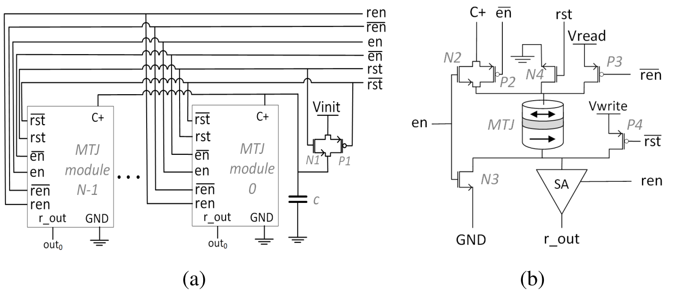


Fig. 2. Proposed TRNG consists of (a) N parallel-connected MTJ-modules and a capacitor. (b) Schematic of the MTJ module.

The TRNG operation consists of three steps, each taking a fixed amount of time. The first step, the Reset step, charges the capacitor C to the V_{init} voltage (using transistors $N1$ and $P1$) and applies a current through the MTJ devices (using transistors $N4$ and $P4$), switching them all to the AP state. The second step, the Enable step, connects C in parallel to all the MTJ devices (using transistors $N2, N3$, and $P2$). This discharges C through the MTJ devices, enabling them to switch to the P state with some probability. During the Enable step, the resistance of an MTJ drops if it is switched, making the capacitor discharge faster. This lowers but does not eliminate the switching probability of the other MTJs. The third step, the Read step, applies a small current through the MTJ devices (using transistor $P3$), and the sense amplifiers determine the state of each. The AP/P states are interpreted as "0"/"1," respectively. Overall, the TRNG outputs an N-bit word.

The proposed TRNG relies on the stochastic switching time of the MTJ as its randomness source. Unlike the previously proposed TRNGs, the randomness extraction operation in the Enable step is asynchronous and does not depend on a strict time measurement. The capacitor is sufficiently discharged during the Enable step to ensure a low probability for further switching until the end of the Read step. Hence, the randomness of the output does not change if the duration of the Enable step is longer than a certain lower bound. Thus, an accurate measurement of the Enable step duration is not required. Note that although the Enable step is done asynchronously, the TRNG still uses a clock signal since a time measurement is still needed to transition between the operation steps.

IV. Evaluation

A. Measure for Randomness

Determining whether a sequence of numbers is random is considered difficult [42]. To try and overcome this problem, the standard statistical test suites, such as the NIST SP 800-22 [42], are usually used to inspect for random properties. However, the proposed TRNG is evaluated using a simulation, which generates the TRNG outputs (N-bit words) in an independent and identically distributed (i.i.d.) manner. Each output is generated independently by numerically solving the stochastic differential equation system of the TRNG, which uses a different sequence of computer-generated random thermal noise for each TRNG output word. For more information about and justifications for the simulation (see Section IV-B). Hence, the simulation is designed, such that there are no dependences between bits from different TRNG outputs, but only the dependences between bits in the same output word, reducing the dependence checks to within an output word.

To measure the dependences between bits in the same output word, we use two measures: the Shannon entropy and the min-entropy of the output words. Entropy quantifies the amount of surprise in the outcome of the experiment. The higher the entropy, the more surprise in the experiment, i.e., the experiment is more random. For an i.i.d. source with values from a finite set $\mathcal{X}$ with probability distribution function $p : \mathcal{X} \rightarrow [0, 1]$, the Shannon entropy per word is $-\sum_{x\in\mathcal{X}} p(x) \log_2 p(x)$ (with the definition that $0 \cdot \log_2 0 = 0$) and the min-entropy per word is $\min_{x\in\mathcal{X}}(-\log_2 p(x))$. Both entropies are measured in bits. If the number of elements in the finite set $\mathcal{X}$ is m, then both the Shannon and min entropies get a value in the range $[0, \log_2 m]$, where 0 entropy is achieved on the deterministic distribution and the $\log_2 m$ entropy is achieved on the uniform distribution. Taking $\mathcal{X}$ to be the set of N-bit words results in $\log_2 m = N$. Hence, the maximum entropy for an i.i.d. N-bit TRNG is N.

When there are dependences between bits in the TRNG output, some output words will be more likely than others and the distribution of a single TRNG output word will deviate from the uniform distribution, resulting in lower entropy than the maximum. Stronger dependence increases the deviation from the uniform distribution and lowers the entropy. Hence, entropy is a measure for the dependences in a single TRNG output word, and an entropy close to maximum means low dependences.

The min-entropy is a lower bound to the Shannon entropy (with equality achieved on the uniform and deterministic distributions), and it is the lowest amount of randomness a single sample of a random variable can give. Randomness extractors are sometimes designed to extract an output for every input, so the correct measure here is the min-entropy of their source [25]. The Shannon entropy is the expected randomness from a random variable. For an i.i.d. source, the Shannon entropy plays an important role in bounding the number of uniformly distributed bits that can be extracted from n samples [43], [44]. Hence, the Shannon entropy gives us a notion of how many samples are required to extract a certain degree of randomness, while the min-entropy gives us the worst case randomness of a single sample.

B. Simulation Methodology

We evaluated our TRNG with Monte Carlo simulations for the Enable step for different topologies, each with a different number of MTJ devices (different N values). The simulation numerically solves the differential equation system of the MTJs (stochastic LLG equations) and the capacitor. The LLG equations are used since a non pulse current waveform is passed through the MTJ devices. The transistors $N2, N3$, and $P2$ were modeled by a constant resistance. The equations were solved using a standard midpoint scheme [45], assuming that no external magnetic field (unless otherwise

TABLE I

SIMULATED CIRCUIT PARAMETERS. $R_{N2,N3,P2}$ IS THE MODELED TOTAL EFFECTIVE RESISTANCE OF TRANSISTORS $N2$, $N3$, AND $P2$

Feature	Value	Feature	Value
NFET operation gate voltage	$1.5V$	R_{on}	1000Ω
PFET operation gate voltage	$0V$	R_{off}	2500Ω
$R_{N2,N3,P2}$, 2-bit TRNG	4450Ω	V_{init}	$0.8V$
$R_{N2,N3,P2}$, 4-bit TRNG	3440Ω	T_{enable}	$10ns$
$R_{N2,N3,P2}$, 6-bit TRNG	2640Ω	C	$10pF$
$R_{N2,N3,P2}$, 8-bit TRNG	1960Ω	Temp.	$300K$

stated) and the stochastic term were interpreted in the sense of Stratonovich. We could not obtain an analytic expression for the TRNG output distribution.

Each iteration of the Monte Carlo simulation produces the TRNG output binary word. For each measurement of entropy, 2000 iterations were conducted. The probability of each TRNG output was evaluated as its frequency of appearance. However, when the parameters of the simulated MTJs were identical (i.e., with no device-to-device variations), the probability was evaluated as the frequency of the corresponding Hamming weight divided by the number of outputs with the same Hamming weight,[2] thus increasing the accuracy. Note that each iteration of the Monte Carlo simulation produces a single N-bit output of the TRNG, meaning that the simulation produces the probability of a single output of the TRNG.

When this simulation model is used to extract the entropy of the TRNG, the TRNG is assumed to be an i.i.d. source since the model does not include the correlation between consecutive runs of the TRNG. This assumption is justified since the MTJs are always in the AP state prior to the Enable step, regardless of the output of the last run. Furthermore, thermal fluctuations occur constantly; hence, the exact position of the magnetization (around the AP state) at the beginning of the Enable step is itself random. This results in a fresh start in every run.

Nevertheless, consecutive runs will be correlated in a real-world TRNG. Some correlation will be caused by changes in the operation parameters but not by true causality between samples. If, for example, during the sampling of the TRNG, a nearby circuit will periodically work with cycle time T, the magnetic field on the MTJs will change with the same period. As a result, samples with time T between them will show correlation just because they have the same distribution, not because they have true causality. Similar effects can come in the form of temperature, voltage, and so on.

The STT-MTJ devices are modeled as device C from [27], a standard in-plane STT-MTJ that can be used for memory design and has the lowest switching current in [27]. The circuit parameters are listed in Table I. Different values for the modeled effective resistance of transistors $N2, N3$, and $P2$ were simulated for each topology and were chosen to maximize the entropy; the results are shown in Table I. To find the size of the transistors and verify the accuracy of the constant resistance model, we performed circuit simulations (with resistors instead of MTJs) in Cadence Virtuoso using a 28-nm GlobalFoundries process. The Virtuoso simulations showed that the transistors transition quickly at the beginning of the Enable step. Since this transition time is faster than the switching time of the MTJs, it can be ignored. During the rest of the Enable step, the total resistance of the transistors is approximately constant, verifying that our constant resistance model of the transistors is acceptable for the evaluation of our design.

Since the resistance of the interconnect is a few ohms per μm, while the resistances of the transistors and MTJs are in the order of k**Ω**, we neglect the wire resistance. The difference in resistance due to the difference in wire lengths between the MTJ modules is also in the order of a few ohms and is well within the process variation considerations in Section IV-D3. Hence, it is neglected as well. In addition, we measured the parasitic capacitance and leakage currents in our design. The simulation shows that the parasitic values are several orders of magnitude lower than the non parasitic values, and hence, they are ignored.

The number of MTJ devices in the design, i.e., N, was restricted to 2, 4, 6, and 8 because the larger N is, the shorter the discharge time of the capacitor. Hence, a stronger current should flow through the MTJ devices to maintain the same switching probability, requiring a lower transistor resistance. The lower resistance will shorten the discharge time further, but the resulting switching probability of the MTJs will increase. However, a lower transistor resistance means a larger transistor size. A larger N value can improve the performance of the TRNG (see Section IV-D). However, the simulation was restricted to $N \leq 8$ since the transistor sizes required for the $N = 8$ topology are considered large (width of a few hundred nano-meters). A larger N value can be achieved by considering different system parameters (e.g., higher V_{init} and larger capacitor).

An example of a single iteration of the simulation with $N = 4$ is shown in Fig. 3 ($N = 4$ was chosen for clarity). The currents through the MTJs during the simulated Enable step are shown in parallel to the magnetization of the free layer (in a single dimension, the dimension of the fixed-layer magnetization). When an MTJ switches to the P state, its resistance drops and the current through it increases. When the current reaches the P state, the current through the MTJ and the magnetization of the free layer remain steady since the current through the MTJ continues to direct the state to the P state.

We compare our results to the CMOS-based TRNGs. Since the previously proposed STT-MTJ-based TRNGs require a high-frequency clock or a modified STT-MTJ device and do not include process variation or external magnetic field influence in their evaluation, we have not compared our results to those designs.

C. NIST Statistical Test Suite

The National Institute of Standards and Technology (NIST) SP 800-22 rev.1a [42] test suite is commonly used to evaluate the RNGs. The suite is composed of several statistical tests, each operating on a string of bits. The test indicates whether

[2] From symmetry, outputs with the same Hamming weight have the same probability.

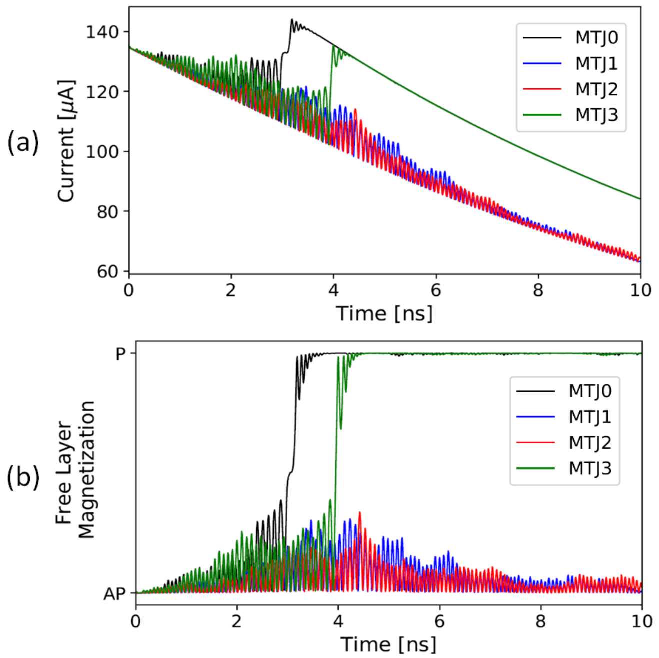


Fig. 3. Example for a single iteration of the simulation for $N = 4$. (a) Current through the MTJs during the Enable step. (b) Magnetization direction of the free layer in the dimension of the fixed-layer magnetization during the Enable step. In this example, MTJ0 and MTJ3 are flipped to the P state, and the TRNG output word will be 1001.

TABLE II

Results of the NIST Test Suite [42] for the Proposed TRNG for Different Numbers of Bits (N). P-Value Threshold Is 0.0001 and the Success Rate Threshold Is 1004/1024 $\approx$ 0. 980 for Testing 1024 Sequences. The Proposed TRNG Passed All the Tests

Test Name	$N = 2$		$N = 4$		$N = 6$		$N = 8$	
	P value	Succ. Rate	P value	Succ. Rate	P value	Succ. Rate	P value	Succ. Rate
Frequency	.28	.993	.0079	.992	.69	.992	.54	.985
Block Frequency	.73	.995	.43	.990	.53	.985	.20	.988
Runs	.038	.990	.22	.988	.068	.983	.25	.992
Longest Run	.87	.990	.34	.993	.073	.990	.28	.990
Serial (1)	.17	.988	.029	.982	.081	.989	.14	.985
Serial (2)	.52	.994	.17	.987	.44	.989	.58	.987
Approximate Entropy*	.28	.982	.00078	.984	.75	.984	.065	.986
Cusum (1)	.026	.990	.0011	.990	.016	.988	.00067	.983
Cusum (2)	.16	.992	.0082	.992	.0014	.989	.055	.985

*This is not the same entropy as computed in the rest of this paper; see [42] for details.

this string of bits is likely to come from a uniform i.i.d. source. To conduct the test suite, a number of bit sequences are retrieved from the generator to be tested (sequences do not share bits between them). Each test is run on all sequences, resulting in a P-value for each sequence on each test. The P-value quantifies the distance between the test result for that sequence and the expected result for a uniform i.i.d. sequence. A sequence is said to pass a test if the P-value for that sequence and test is above a threshold. Then, two scores are given for each test, a success rate and a P-value. The success rate is the proportion of sequences that pass the test, while the P-value is a number between 0 and 1. This number quantifies the distance of the sequences' P-value distribution from the expected P-value distribution of a uniform i.i.d. source for that test. Note that even a perfect source will not produce a perfect success rate and P-value (i.e., 1) for a test due to statistical deviations. A generator is said to pass a test if both the test's success rate and the P-value are above a threshold.

The proposed TRNG does not have a uniform distribution since the output words with different Hamming weight have different probabilities, while the output words with the same Hamming weight have the same probability (due to the symmetry). Since the design was calibrated for maximum entropy, the words with the minimum (all zeros) and maximum (all ones) Hamming weights have the lowest probabilities, while the words with the half zeros and half ones have the highest probability. This results in a bias toward certain bit patterns when a sequence of output words is considered as a sequence of bits. Hence, the NIST test suite is expected to fail on the proposed TRNG without postprocessing. Failing the NIST test suite does not mean that the TRNG is not random; it means only that the TRNG distribution does not appear to be uniformly distributed. Furthermore, non uniformity does not mean that the TRNG cannot be used for cryptography. It means only that postprocessing might be a prerequisite for this use. Postprocessing might be used regardless to compensate for process variation and interference effects [7], [26].

To show that the proposed TRNG is sufficiently random, the TRNG output words are postprocessing by a simple and reversible function before evaluation by the test suite. The reversibility of the processing indicates that the new bit sequence has the same information as the original sequence. To define the postprocessing, the TRNG output word sequence is denoted by $\{w_1, w_2, w_3, \ldots\}$, and the processed word sequence is denoted by $\{z_1, z_2, z_3, \ldots\}$, each of which are N-bit words. Then, the postprocessing is defined as $z_1 = w_1$ and $z_i = w_i \oplus z_{i-1}$ for $i > 1$ ($\oplus$ is the bit-wise XOR operation). The bit sequence is taken as the bits of $\{z_i\}_{i \geq 1}$, starting from z_1, in a big-endian fashion. This operation is reversible (since $w_1 = z_1$ and $w_i = z_i \oplus z_{i-1}$ for $i > 1$) and simple to implement (requiring an N-bit register and N XOR gates). Note that this postprocessing is not a qualified cryptographic hash function or randomness extractor. It is used solely to demonstrate that the TRNG has sufficient randomness to pass the test in the NIST test suite.

For each topology of the TRNG, 1024 sequences of 1024 bits each (a total of 2^{20} bits) were generated for the NIST test suite. The results of the P-value and success rates for the tests with their thresholds are listed in Table II. To perform all of the tests in the suite, a substantially larger number of bits is needed, which will require an unfeasible amount of time to produce with the TRNG simulation. Therefore, the tests listed in Table II are those that can be run with the produced number of bits. All topologies passed all executed tests.

D. Entropy Per Output

We evaluated the entropy of the TRNG for different design, environmental, and process parameters.

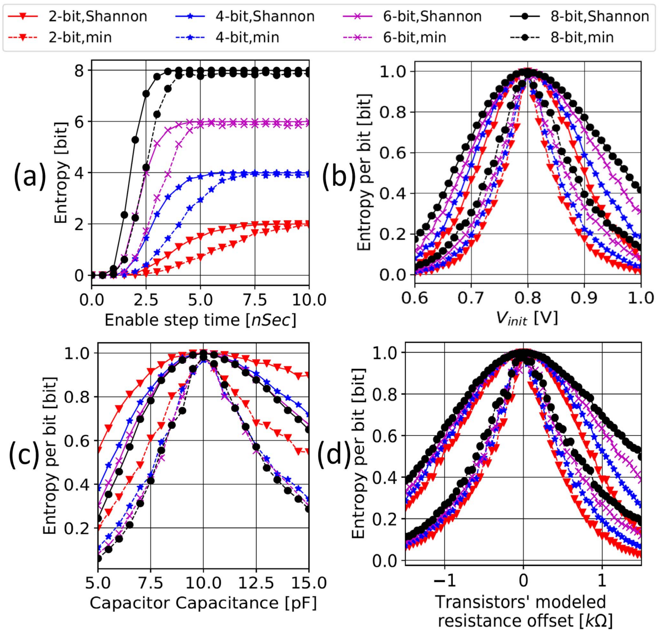


Fig. 4. Entropy of the TRNG for different (a) enable step time duration, (b) V_{init}, (c) C, and (d) offset in the total effective resistance of transistors $N2$, $N3$, and $P2$.

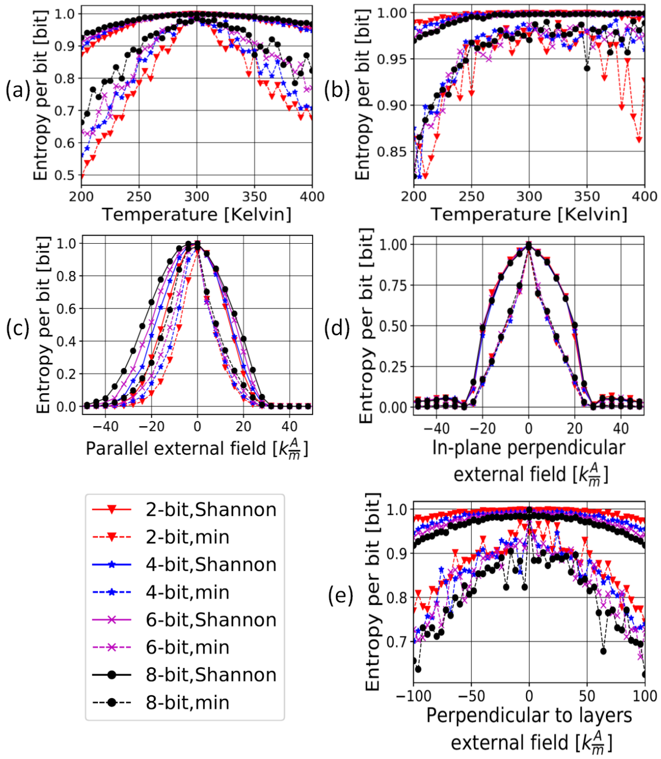


Fig. 5. Effect of temperature (a) noncharging MTJ resistance and (b) including a change in MTJ resistance, where R_{ON} is constant and TMR $= ((R_{\text{OFF}} - R_{\text{ON}})/R_{\text{ON}})$ changes at a linear rate of $-0.4(\%/^\circ K)$. Effect of a constant external magnetic field (fixed direction and varying magnitude) (c) parallel to the fixed-layer magnetization, (d) in-plane and perpendicular to the fixed-layer magnetization, and (e) perpendicular to the layer's plane.

1) Design Parameters: Include the Enable step duration time, V_{init}, C, and the effective modeled resistance of $N2$, $N3$, and $P2$. The simulation results of the entropy for different design parameters are shown in Fig. 4. The design parameters trade off between the different performance measures of the TRNG (entropy, operation time, area, and power) while using the same MTJ devices. An important observation is that our design can, in the ideal case, reach nearly the maximum possible entropy (1-bit entropy per MTJ device).

Fig. 4(a) shows the effect of the Enable step duration, as mentioned in Section III. When the duration is sufficiently long, the randomness of the TRNG is maximal. This allows the TRNG to be used in low-frequency devices, where the time measurement has a low resolution.

It is evident that a small change in V_{init}, from its designed value of 0.8 V, can change the entropy. The value of the initial capacitor voltage affects the duration of the Reset step and the entropy throughput of the TRNG (see Section IV-E). However, reasonable variations in the capacitance of the capacitor (less than 0.5 pF, or approximately 5%) have a little effect on the entropy since the capacitance is relatively large.

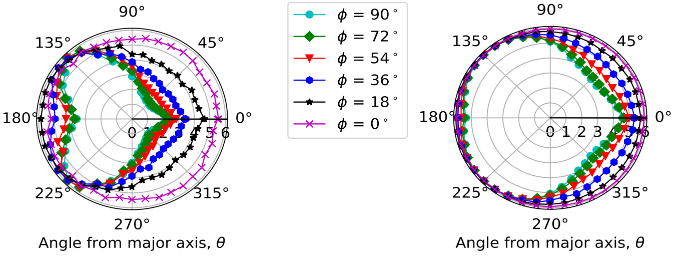
In addition, we can conclude that modeling the open transistors by constant resistance is tolerable. Deviations in the range of 250 Ω from the designed value do not reduce the entropy much. From the Cadence Virtuoso simulation (see Section IV-B), we verified that the actual effective resistance of $N2$, $N3$, and $P2$ fluctuates well within the range of ±250 Ω around the designed value.

2) Environmental Parameters: These parameters are external to the TRNG and can be altered by an adversary (see Section V). The effect of temperature and the external magnetic field on the MTJ devices is considered here, and their influence on the entropy is shown in Fig. 5.

In the LLG equation, the temperature affects only the thermal fluctuations of the MTJ. However, the resistance of the MTJ is temperature-dependent [46], [47]. While the resistance of the MTJ in the P state is roughly constant, the resistance of the AP state changes more considerably with the temperature [46]. The TMR $= ((R_{\text{OFF}} - R_{\text{ON}})/R_{\text{ON}})$ changes in an approximately linear manner around $300^\circ K$, with a rate of $-0.2(\%/^\circ K)$–$-0.4(\%/^\circ K)$ according to [46]. To investigate the temperature-dependent behavior of the proposed TRNG design, we simulated different temperatures while maintaining R_{ON} constant, and changing R_{OFF} to produce a TMR change with a linear rate of $0(\%/^\circ K)$ (i.e., constant TMR) [see Fig. 5(a)] and $-0.4(\%/^\circ K)$ [see Fig. 5(b)].

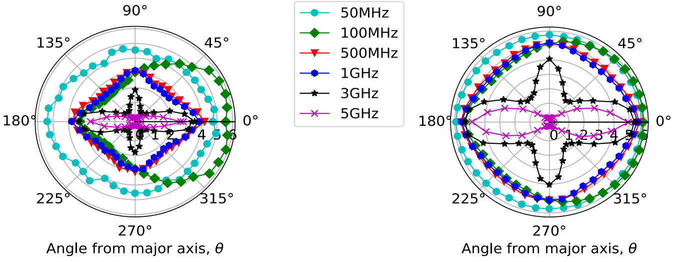
If the TMR rate is $0(\%/^\circ K)$, only the thermal fluctuations are affected by the temperature change. In this case, a degradation in the TRNG entropy is evident when the temperature deviates from $300^\circ K$. The switching probability of the MTJs changes with the temperature (as can be seen by the min-entropy in Fig. 5(a) and as reported in other works [47]–[49]), but the Shannon entropy of the TRNG output word remains high in the examined temperature range.

If the TMR rate is $-0.4(\%/^\circ K)$, the entropy of the TRNG is better than in the $0(\%/^\circ K)$ case. When the temperature is lower (higher) than $300^\circ K$, the resistance of R_{OFF} increases (decreases), resulting in a lower (higher) initial current through the MTJs but with a longer (shorter) discharge time for the



(a) (b)

Fig. 6. Effect of a constant external field in different angles and a fixed magnitude of 10 k(A/m) on the entropy of a 6-bit TRNG. ϕ is the angle of the field from the axis perpendicular to the MTJ plane, and θ is the in-plane angle from the magnetization of the fixed layer. (a) Min-entropy. (b) Shannon entropy.



(a) (b)

Fig. 7. Effect of an alternating external field on multiple in-plane directions and a fixed magnitude of 10 k(A/m) on a 6-bit TRNG. (a) Min-entropy. (b) Shannon entropy.

capacitor. The longer (shorter) discharge time results in a longer (shorter) time for a non negligible switching probability current, which might negate the reduced (increased) switching probability produced by the thermal fluctuations. These two effects, the thermal fluctuations and the current waveform through the MTJ, interact in a non trivial way with the entropy due to their different non linear characteristics.

This analysis shows that the Shannon entropy of the proposed TRNG design behaves well under temperature changes.

When considering the external magnetic field, a sufficiently high field can reduce the entropy to 0, as shown in Fig. 5(b) and (c). Every device using MTJs will be susceptible to a strong enough magnetic field, which can be exploited by an adversary (see Section V for further discussion). When the external magnetic field is perpendicular to the fixed-layer magnetization (the fixed-layer magnetization of all MTJs is parallel), from the symmetry of the MTJ, the entropy is expected to be symmetrical around zero magnetic field, as shown in Fig. 5(c) and (d). In addition, since an in-plane MTJ is used, the external field perpendicular to the MTJ layers has little influence on the entropy. Therefore, when designing the TRNG circuit, we would like to position nearby wires in the same plane as the TRNG and use vias as short as possible and positioned as far from the TRNG as possible.

To further consider an external magnetic field, we simulated the effect for other directions of the field relative to the fixed-layer magnetization of the MTJs. Fig. 6 shows the effect of the direction of an external field with a magnitude of 10 k(A/m) on a 6-bit TRNG. The worst effect is for approximately $\theta = 45^\circ \backslash 315^\circ$, rather than in the direction of the fixed-layer magnetization ($\theta = 0^\circ \backslash 180^\circ$) or perpendicular to it ($\theta = 90^\circ \backslash 270^\circ$). Fig. 7 shows how an alternating external field applied from multiple directions, with a fixed magnitude of 10 k(A/m) and different frequencies, affects the 6-bit TRNG. Since a field perpendicular to the MTJ affects the entropy much less than other directions, the alternating field simulation was limited only to the in-plane directions. Note that the Enable step duration is 10 ns (see Table I), which corresponds to a 100-MHz frequency.

An interesting result is that the performance under an alternating external field in the range of 50 MHz–1 GHz is actually better than the performance under a constant external field. This implies that nearby circuits at frequencies up to 1 GHz will induce minimal performance loss. Another observation is that an alternating external field above a certain frequency is very effective. Countermeasures against external fields are discussed in Section V.

3) Process Variation: Process variation in the TRNG will result in different switching probabilities for the different MTJs, biasing some to switch with higher probabilities and some with lower probabilities, resulting in bias for certain output words. However, the i.i.d. property of the TRNG is not affected by process variation (see Section IV-B): it affects only the distribution of an output word but not the dependences between different output words. Hence, the methodology presented in Section IV-A is used in this section as well.

Variations in the MTJs and in the transistors operating in the Enable step have been considered. Section IV-D1 shows that reasonable variation in capacitor C does not change the entropy; hence, C is not considered for process variation. For the transistors, we modeled variations in their fixed resistance. For the MTJ, we modeled variations in the physical size of the devices: major and minor axis length (the MTJ shape is an ellipse cylinder [27]), the thickness of the free layer, and the thickness of the oxide layer (tunnel barrier layer). We generated 1000 different instances for each TRNG topology and their entropy was evaluated. The parameters were drawn independently of a Gaussian distribution with mean as the designed value (listed in Table I and in [27]) and a standard deviation of 5% [50].

The geometry of the MTJ affects its demagnetization factors [27] and resistance. The method presented in [51] was used to compute the new demagnetization factors. The MTJ resistance is proportional to the exponent of the oxide layer thickness (t_{ox}) and inversely proportional to its area (A) [50], i.e., $R_{\text{ON}}, R_{\text{OFF}} \propto ((e^{\rho \cdot t_{\text{ox}}})/A)$ (ρ is a constant that depends on the device technology). The thickness of the oxide layer appears in the simulation only as a part of the MTJ resistance. Since the oxide layer thickness and the coefficient ρ are unavailable for the simulated device, the process variation of the oxide layer was evaluated as an additional variation in the MTJ resistance by a Gaussian distribution with a standard deviation of 5%. The simulation results are listed in Table III. The results show that the entropy per bit increases with the number of MTJs, resulting in a twofold increase in the TRNG total entropy.

Even though most TRNG instances will have sufficiently high entropy under process variation, some TRNG instances

TABLE III
ENTROPY RESULTS WITH PROCESS VARIATION SHOWING THE AVERAGE, STANDARD DEVIATION, MEDIAN, AND THE 10TH PERCENTILE

	Shannon Entropy per Bit				Min-Entropy per Bit			
N	Avg.	sd	Med.	P_{10}	Avg.	sd	Med.	P_{10}
2	0.74	0.19	0.76	0.46	0.46	0.21	0.47	0.17
4	0.79	0.12	0.80	0.64	0.51	0.14	0.51	0.33
6	0.82	0.08	0.83	0.72	0.54	0.10	0.55	0.41
8*	0.86	0.06	0.86	0.78	0.58	0.08	0.59	0.47

*For the 8-bit TRNG, we simulated 6000 iterations.

might still produce low entropy and be unusable. To protect the TRNG from such an event, several TRNG instances should be fabricated together. Fortunately, the largest area is consumed by the capacitor C and the sense amplifier (see Section IV-F), which can be shared among the different fabricated MTJ modules (the implications of this suggestion are not discussed in this paper). Thus, the process variation robustness can be increased with a low area overhead.

E. Entropy Generating Rate

In many systems, a large number of entropy bits are required. In this case, the entropy generating rate (entropy bits per second) is the desired performance metric. Nevertheless, the TRNG is not required for the entire operation of the system since there is no need for random words most of the time. The generating rate gives a notion of the delay time between starting the TRNG and having the desired number of entropy bits, enabling the system to react faster. Since many TRNG generated numbers are involved, the generating rate refers to the Shannon entropy. The entropy generating rate is measured in units of entropy bits per second, which is the amount of entropy produced by the TRNG in a second.

To determine the entropy generation rate, all three operation steps should be considered. The Reset step duration is dominated by the capacitor charging time. If we model the passgate $P1-N1$ connecting the capacitor to V_{init} (see Fig. 2) as a resistor of 1.5 KΩ, the capacitor charging time from 0 to 0.79 V (98.8%) is 66 ns. The duration of the Enable step is 10 ns (see Table I). In the Read step, the states of the MTJs are read using sense amplifiers. If we take the read latency of 2.8 ns reported by [52], an output is produced every 78.8 ns.

For an 8-bit TRNG with process variation, 90% of instances have an entropy generation rate between 79.2 and 101.5 Mb/s. This rate can be improved by terminating the Enable step earlier [as shown in Fig. 5(a)] and by reducing the charging time (since the capacitor is not fully discharged immediately after an operation). For an 8-bit TRNG, the entropy generation rate can be improved to 99.7–127.8 Mb/s for 90% of the instances. However, the actual generation rate of our proposed design depends on the system clock used to measure the duration of the steps; the computed times represent the best case.

F. Area and Energy

To estimate the area of the TRNG, we evaluate the area for each component in the proposed circuit. The area of a single STT-MTJ device is 0.003 μm^2 [27]. To find the capacitor C area, we modeled it as a MOS capacitor with GlobalFoundries 28-nm technology and obtained an area of 400 μm^2 using Cadence Virtuoso. We evaluated the sense amplifier area from [52] as the summed area of its transistors. All transistors were evaluated with a width of 500 nm and minimum length (transistor sizes are not specified in [52]). This size upper bounds the area of the transistors shown in Fig. 2. The resulting area of an MTJ module is 0.395 μm^2. The sense amplifier in [52] uses an additional capacitor, but its area is relatively small since the read duration is 2.8 ns. Hence, each MTJ module area was approximated as 0.6 μm^2. Table IV lists the TRNG area for different N values.

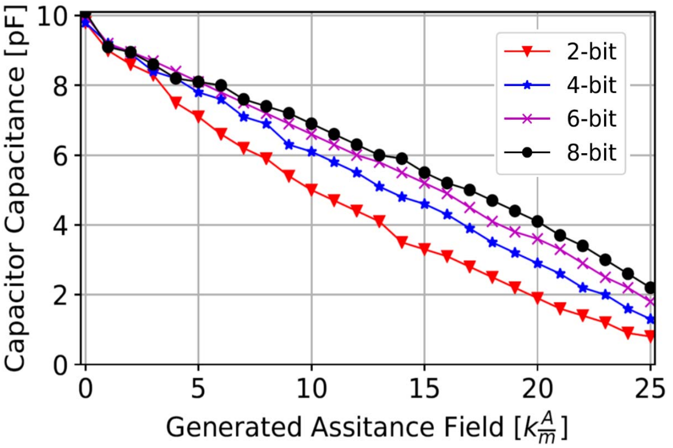


Fig. 8. Capacitor capacitance as a function of the generated assistance magnetic field while preserving the entropy. The field is in the direction of the fixed-layer magnetization.

To reduce the area, the number of sense amplifiers can be lowered by reading the MTJs sequentially or by sharing sense amplifiers with a nearby MTJ memory array. However, the major area contributor is the capacitor C. Using a capacitor other than a MOS capacitor to reduce the area is left as future work. Alternatively, it is possible to lower the capacitance (and hence the size of the capacitor) while maintaining the same switching distribution, but this requires reducing the number of MTJs and/or using larger transistors (reducing their resistance) to preserve the same current through the MTJs. Another option is to use MTJ devices with a higher switching probability for the same current (as done in [40] and [41]), which will require a smaller capacitor but will increase the sensitivity of the design to external magnetic fields. A different option is to use the generated magnetic field by a dedicated wire, similar to the solution suggested in [53] for the STT-MRAM memory. The magnetic field will raise the switching probability; hence, a smaller current will be needed to produce the same entropy, also resulting in a smaller capacitor. Fig. 8 shows the relationship between the generated magnetic field (in the direction of the fixed-layer magnetization) and the capacitor required to achieve high entropy per bit (close to 1). In addition, the generated magnetic field might be dynamically adjusted to compensate for the effects of process variation and parameter drift of the circuit. However, such a solution might incur an area overhead and design complications. Precise analysis of such a solution is left for future research.

The energy of the TRNG in the Reset and Read steps is for capacitor charging, switching from P to AP of the MTJs (at the Reset step), and for the read operation. The Enable step only uses the energy stored in the capacitor. The energy required to charge the capacitor is the energy that the capacitor holds, 3.2 pJ, plus another 3.2 pJ consumed on the passgate connected to it (transistors $N1$ and $P1$ in Fig. 2), for a passgate effective resistance of 1.5 KΩ and a Reset step time of 66 ns. The MTJs have a write energy of 4.5 pJ and a read energy of 0.7 pJ [52]. Table IV lists the energy per bit and power for different N values.

TABLE IV

COMPARISON OF ENTROPY GENERATION RATE, AREA, ENERGY, AND POWER OF THE PROPOSED TRNG* WITH CMOS TRNGS

	Entropy Generation Rate $[Mb/s]$	Area $[\mu m^2]$	Energy per bit $[\frac{pJ}{entropy\text{-}bit}]$	Power $[mW]$
2-bit TRNG	15.4-33.4	401.2	6.16-13.4	0.2
4-bit TRNG	40.9-63.8	402.4	5.9-9.2	0.38
6-bit TRNG	66.7-92.7	403.6	5.8-8.0	0.54
8-bit TRNG	94.0-120.6	404.8	5.7-7.3	0.7
Yang *et al.* [8]	23.16	375	23	0.54
Srinivasan *et al.* [9]	2400	4004	2.9	7

*The numbers for the proposed TRNG are presented for 90% of the instances and for the highest generating rate.

G. Comparison to Other TRNGs

Table IV compares our proposed design with two different state-of-the-art CMOS-based TRNGs. The proposed TRNG has a high entropy generation rate and low energy per bit compared with the CMOS TRNGs with a similar area and power.

The proposed TRNG is not compared with the previously proposed STT-MTJ-based TRNGs since they used specially designed STT-MTJ devices or did not consider process variation effects or external magnetic field effects (as discussed in Section II-C).

Yang *et al.*'s [8] TRNG is based on ROs with a 28-nm CMOS process node. The design of the TRNG includes several ROs, each controlling a 14-bit counter to measure an event (private for each RO). The resulting counter is random due to the noise in the RO and serves as the output. Only a subset of the counter's bits is used to produce a uniform distribution. This TRNG has a similar area and power as our proposed TRNG. The proposed 8-bit TRNG has 4× to 5× higher entropy rate and 3× to 4× better energy per bit than the TRNG of Yang *et al.* [8].

Srinivasan *et al.*'s [9] TRNG is based on a metastable latch with a 45-nm CMOS process node. It has 10× larger area and consumes 10× more power. The metastable latch TRNG is $20\times to 24\times$ faster and consumes $2\times to 2.6\times$ less energy per bit than our 8-bit proposed design. Nevertheless, the design proposed by Srinivasan *et al.* [9] generates a single random bit every clock cycle, requiring the use of a 2.4-GHz clock to achieve the high entropy rate. The use of a high rate clock and the relatively large area make this design impractical for low-power and low-frequency devices.

V. DEALING WITH AN ADVERSARY

Consider an attack model where the attacker can change the environmental conditions of the TRNG. For example, the attacker can place a fixed magnet in proximity to the TRNG to control the external magnetic field, or use an antenna, or remotely control a circuit close to the TRNG (such as a processor). However, we assume that the adversary does not have physically invasive access.

The temperature has little effect on the TRNG (see Section IV-D2) and, therefore, is not an interesting attack venue. On the other hand, the external magnetic field on the TRNG can decrease the entropy substantially. Passive shielding can mitigate the effect of an external field. Prior work [54]–[56] on passive shielding demonstrated this for MTJ-based memories.

A different approach to interference is detection. This can be done using online tests that check for a certain amount of randomness. Once the randomness is below a specific threshold, an error should be sent to the operating entity, informing it of nearby interference or an ongoing attack. This solution will not prevent the attack, but it might convert it to a denial-of-service attack.

VI. CONCLUSION

In this paper, we presented an asynchronous TRNG that utilizes the random switching time of STT-MTJ devices. The TRNG was comprehensively evaluated in simulations using the physical equations describing the STT-MTJs. The evaluation showed that by increasing the number of STT-MTJs in the design, the TRNG can have greater entropy per output and better resilience to process variation. Furthermore, the design achieves better throughput than current CMOS TRNGs, with lower energy per bit and similar die area and power dissipation. However, the MTJ devices are susceptible to attacks controlling the external magnetic field, requiring the use of additional countermeasures.

REFERENCES

[1] D. Eastlake, J. Schiller, and S. Crocker. (Jun. 2005). *RFC4086: Randomness Requirements for Security*. Accessed: Jan. 2018. [Online]. Available: https://tools.ietf.org/html/rfc4086

[2] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*. Boca Raton, FL, USA: CRC Press, 1996.

[3] I. Goldberg and D. Wagner. (Jan. 1996). *Randomness and the Netscape Browser*. Accessed: Jan. 2018. [Online]. Available: https://people.eecs.berkeley.edu/~daw/papers/ddj-netscape.html

[4] Z. Gutterman, B. Pinkas, and T. Reinman. (Jul. 2006). *Open to Attack: Vulnerabilities of the Linux Random Number Generator, Black Hat*. [Online]. Available: https://www.blackhat.com/presentations/bh-usa-06/BH-US-06-Gutterman.pdf

[5] J. Kelsey, B. Schneier, D. Wagner, and C. Hall, "Cryptanalytic attacks on pseudorandom number generators," in *Proc. 5th Int. Workshop Fast Softw. Encryption*, 1998, pp. 168–188.

[6] C. K. Koc, *Cryptographic Engineering*. Springer, 2008.

[7] S. Bhunia and M. Tehranipoor, "Hardware security primitives," in *Hardware Security*, S. Bhunia and M. Tehranipoor, Eds. San Mateo, CA, USA: Morgan Kaufmann, 2019, pp. 311–345. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128124772000174

[8] K. Yang, D. Fick, M. B. Henry, Y. Lee, D. Blaauw, and D. Sylvester, “A 23 Mb/s 23 pJ/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] S. Srinivasan *et al.*, “2.4 GHz 7 mW all-digital PVT-variation tolerant true random number generator in 45 nm CMOS,” in *Proc. Symp. VLSI Circuits*, Jun. 2010, pp. 203–204.

[10] H. Akinaga and H. Shima, “Resistive random access memory (ReRAM) based on metal oxides,” *Proc. IEEE*, vol. 98, no. 12, pp. 2237–2251, Dec. 2010.

[11] H.-S. P. Wong *et al.*, “Phase change memory,” *Proc. IEEE*, vol. 98, no. 12, pp. 2201–2227, Dec. 2010.

[12] H.-S. P. Wong *et al.*, “Metal-oxide RRAM,” *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, Jun. 2012.

[13] M. Wang *et al.*, “Current-induced magnetization switching in atom-thick tungsten engineered perpendicular magnetic tunnel junctions with large tunnel magnetoresistance,” *Nature Commun.*, vol. 9, Feb. 2018, Art. no. 671.

[14] G. Hu *et al.*, “STT-MRAM with double magnetic tunnel junctions,” in *IEDM Tech. Dig.*, Dec. 2015, pp. 26.3.1–26.3.4.

[15] T. Devolder *et al.*, “Single-shot time-resolved measurements of nanosecond-scale spin-transfer induced switching: Stochastic versus deterministic aspects,” *Phys. Rev. Lett.*, vol. 100, Feb. 2008, Art. no. 057206.

[16] E. I. Vatajelu, G. Di Natale, and P. Prinetto, “STT-MTJ-based TRNG with on-the-fly temperature/current variation compensation,” in *Proc. IEEE 22nd Int. Symp. On-Line Test. Robust Syst. Design (IOLTS)*, Jul. 2016, pp. 179–184.

[17] A. Fukushima *et al.*, “Spin dice: A scalable truly random number generator based on spintronics,” *Appl. Phys. Express*, vol. 7, no. 8, 2014, Art. no. 083001.

[18] S. Oosawa, T. Konishi, N. Onizawa, and T. Hanyu, “Design of an STT-MTJ based true random number generator using digitally controlled probability-locked loop,” in *Proc. NEWCAS*, Jun. 2015, pp. 1–4.

[19] Y. Qu, J. Han, B. F. Cockburn, W. Pedrycz, Y. Zhang, and W. Zhao, “A true random number generator based on parallel STT-MTJs,” in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, Mar. 2017, pp. 606–609.

[20] Y. Wang, H. Cai, L. A. B. Naviner, J.-O. Klein, J. Yang, and W. Zhao, “A novel circuit design of true random number generator using magnetic tunnel junction,” in *Proc. IEEE/ACM Int. Symp. Nanosc. Archit. (NANOARCH)*, Jul. 2016, pp. 123–128.

[21] S. Ghosh, “Spintronics and security: Prospects, vulnerabilities, attack models, and preventions,” *Proc. IEEE*, vol. 104, no. 10, pp. 1864–1893, Oct. 2016.

[22] E. Barker and J. Kelsey, “Recommendation for random number generation using deterministic random bit generator,” NIST, Gaithersburg, MA, USA, Tech. Rep. SP 800-90A, Revision 1, Jun. 2015.

[23] W. Kan, “Analysis of underlying assumptions in NIST DRBGs,” *IACR Cryptol. ePrint Arch.*, vol. 2007, p. 345, Sep. 2007.

[24] Y. Lao, Q. Tang, C. H. Kim, and K. K. Parhi, “Beat frequency detector–based high-speed true random number generators: Statistical modeling and analysis,” *J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 1, Apr. 2016, Art. no. 9.

[25] S. P. Vadhan, “Pseudorandomness,” *Found. Trends Theor. Comput. Sci.*, vol. 7, nos. 1–3, pp. 1–336, 2011.

[26] S.-H. Kwok, Y.-L. Ee, G. Chew, K. Zheng, K. Khoo, and C.-H. Tan, “A comparison of post-processing techniques for biased random number generators,” in *Information Security Theory and Practice. Security and Privacy of Mobile Devices in Wireless Communication*, C. A. Ardagna and J. Zhou, Eds. Berlin, Germany: Springer, 2011, pp. 175–190.

[27] A. F. Vincent, N. Locatelli, J. O. Klein, W. S. Zhao, S. Galdin-Retailleau, and D. Querlioz, “Analytical macrospin modeling of the stochastic switching time of spin-transfer torque devices,” *IEEE Trans. Electron Devices*, vol. 62, no. 1, pp. 164–170, Jan. 2015.

[28] J. C. Slonczewski, “Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier,” *Phys. Rev. B, Condens. Matter*, vol. 39, pp. 6995–7002, Apr. 1989.

[29] T. L. Gilbert, “A phenomenological theory of damping in ferromagnetic materials,” *IEEE Trans. Magn.*, vol. 40, no. 6, pp. 3443–3449, Nov. 2004.

[30] J. L. García-Palacios and F. J. Làzaro, “Langevin-dynamics study of the dynamical properties of small magnetic particles,” *Phys. Rev. B, Condens. Matter*, vol. 58, pp. 14937–14958, Dec. 1998.

[31] J. C. Slonczewski, “Current-driven excitation of magnetic multilayers,” *J. Magn. Magn. Mater.*, vol. 159, nos. 1–2, pp. L1–L7, 1996.

[32] D. M. Apalkov and P. B. Visscher, “Spin-torque switching: Fokker-Planck rate calculation,” *Phys. Rev. B, Condens. Matter*, vol. 72, Nov. 2005, Art. no. 180405.

[33] Z. Li and S. Zhang, “Thermally assisted magnetization reversal in the presence of a spin-transfer torque,” *Phys. Rev. B, Condens. Matter*, vol. 69, no. 13, 2004, Art. no. 134416.

[34] J. Z. Sun, “Spin-current interaction with a monodomain magnetic body: A model study,” *Phys. Rev. B, Condens. Matter*, vol. 62, no. 1, pp. 570–578, 2000.

[35] Y. Zhang *et al.*, “Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions,” *IEEE Trans. Electron Devices*, vol. 59, no. 3, pp. 819–826, Mar. 2012.

[36] H. Jiang *et al.*, “A novel true random number generator based on a stochastic diffusive memristor,” *Nature Commun.*, vol. 8, p. 882, Oct. 2017.

[37] T. Zhang *et al.*, “High-speed true random number generation based on paired memristors for security electronics,” *Nanotechnology*, vol. 28, no. 45, 2017, Art. no. 455202.

[38] Y. Wang, W. Wen, H. Li, and M. Hu, “A novel true random number generator design leveraging emerging memristor technology,” in *Proc. ACM 25th Ed. Great Lakes Symp. VLSI*, 2015, pp. 271–276.

[39] Y. Qu *et al.*, “Variation-resilient true random number generators based on multiple STT-MTJs,” *IEEE Trans. Nanotechnol.*, vol. 17, no. 6, pp. 1270–1281, Nov. 2018.

[40] H. Lee, F. Ebrahimi, P. K. Amiri, and K. L. Wang, “Design of high-throughput and low-power true random number generator utilizing perpendicularly magnetized voltage-controlled magnetic tunnel junction,” *AIP Adv.*, vol. 7, no. 5, 2017, Art. no. 055934.

[41] D. Vodenicarevic *et al.*, “Low-energy truly random number generation with superparamagnetic tunnel junctions for unconventional computing,” *Phys. Rev. Appl.*, vol. 8, Nov. 2017, Art. no. 054045.

[42] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, and E. Barker, “A statistical test suite for random and pseudorandom number generators for cryptographic applications,” NIST, Gaithersburg, MA, USA, Tech. Rep. SP 800-22, Revision 1a, Apr. 2010.

[43] R. Renner and S. Wolf, “Simple and tight bounds for information reconciliation and privacy amplification,” in *Proc. 11th Int. Conf. Theory Appl. Cryptol. Inf. Secur. (ASIACRYPT)*. Berlin, Germany: Springer-Verlag, 2005, pp. 199–216.

[44] T. Holenstein and R. Renner, “On the randomness of independent experiments,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1865–1871, Apr. 2011.

[45] M. D’Aquino, C. Serpico, G. Coppola, I. D. Mayergoyz, and G. Bertotti, “Midpoint numerical technique for stochastic Landau-Lifshitz-Gilbert dynamics,” *J. Appl. Phys.*, vol. 99, no. 8, 2006, Art. no. 08B905.

[46] V. Drewello, J. Schmalhorst, A. Thomas, and G. Reiss, “Evidence for strong magnon contribution to the TMR temperature dependence in MgO based tunnel junctions,” *Phys. Rev. B, Condens. Matter*, vol. 77, no. 1, Jan. 2008, Art. no. 014440.

[47] Y. Wang, H. Cai, L. A. B. Naviner, Y. Zhang, J. O. Klein, and W. S. Zhao, “Compact thermal modeling of spin transfer torque magnetic tunnel junction,” *Microelectron. Rel.*, vol. 55, nos. 9–10, pp. 1649–1653, 2015.

[48] E. I. Vatajelu, G. Di Natale, and P. Prinetto, “Security primitives (PUF and TRNG) with STT-MRAM,” in *Proc. IEEE VTS*, Apr. 2016, pp. 1–4.

[49] M. N. I. Khan, A. S. Iyengar, and S. Ghosh, “Novel magnetic burn-in for retention and magnetic tolerance testing of STTRAM,” *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 26, no. 8, pp. 1508–1517, Aug. 2018.

[50] J. Li, C. Augustine, S. Salahuddin, and K. Roy, “Modeling of failure probability and statistical design of spin-torque transfer magnetic random access memory (STT MRAM) array for yield enhancement,” in *Proc. 45th Annu. Design Autom. Conf. (DAC)*, 2008, pp. 278–283.

[51] D. A. Goode and G. Rowlands, “The demagnetizing energies of a uniformly magnetized cylinder with an elliptic cross-section,” *J. Magn. Magn. Mater.*, vol. 267, no. 3, pp. 373–385, 2003.

[52] Q. Dong *et al.*, “A 1 Mb 28 nm STT-MRAM with 2.8 ns read access time at 1.2 V VDD using single-cap offset-cancelled sense amplifier and *in-situ* self-write-termination,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 480–482.

[53] R. Patel, X. Guo, Q. Guo, E. Ipek, and E. G. Friedman, “Reducing switching latency and energy in STT-MRAM caches with field-assisted writing,” *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 1, pp. 129–138, Jan. 2016.

[54] K. Yamada, “Anisotropic magnetic shielding effectiveness of magnetic shielded package,” *IEEE Trans. Magn.*, vol. 53, no. 11, Nov. 2017, Art. no. 8500104.

[55] W. Wang and Z. Jiang, "Magnetic shielding design for magnetoelectronic devices protection," *IEEE Trans. Magn.*, vol. 44, no. 11, pp. 4175–4178, Nov. 2008.

[56] E. Paperno, H. Koide, and I. Sasada, "Charts for estimating the axial shielding factors for triple-shell open-ended cylindrical shields," *IEEE Trans. Magn.*, vol. 37, no. 4, pp. 2881–2883, Jul. 2001.

Ben Perach received the B.Sc. degree in mathematics from The Hebrew University of Jerusalem, Jerusalem, Israel, in 2010, and the M.Sc. degree in electrical engineering from Tel Aviv University, Tel Aviv, Israel, in 2017. He is currently working toward the Ph.D. degree in electrical engineering at Technion–Israel Institute of Technology, Haifa, Israel.

His current research interests include computer architecture with a focus on the processor design, and also field-programmable gate arrays, security, and data networks.

Shahar Kvatinsky received the B.Sc. degree in computer engineering and applied physics and the M.B.A. degree from The Hebrew University of Jerusalem, Jerusalem, Israel, in 2009 and 2010, respectively, and the Ph.D. degree in electrical engineering from Technion–Israel Institute of Technology, Haifa, Israel, in 2014.

From 2006 to 2009, he was with Intel, Jerusalem, as a Circuit Designer and was a Postdoctoral Research Fellow with Stanford University, Stanford, CA, USA, from 2014 to 2015. He is currently an Assistant Professor with the Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion–Israel Institute of Technology. His current research interests include circuits and architectures with emerging memory technologies and the design of energy-efficient architectures.

Dr. Kvatinsky was a recipient of the 2015 IEEE Guillemin-Cauer Best Paper Award, the 2015 Best Paper of *Computer Architecture Letters*, the Viterbi Fellowship, the Jacobs Fellowship, the ERC Starting Grant, the 2017 Pazy Memorial Award, the 2014 and 2017 Hershel Rich Technion Innovation Awards, the 2013 Sanford Kaplan Prize for Creative Management in High Tech, the 2010 Benin prize, and seven Technion Excellence Teaching Awards. He is an editor of the *Microelectronics Journal*.