Random Bitstream Generation Using Voltage-Controlled Magnetic Anisotropy and Spin Orbit Torque Magnetic Tunnel Junctions

SAMUEL LIU\textsuperscript{1} (Graduate Student Member, IEEE), JAESUK KWON\textsuperscript{1} (Member, IEEE), PAUL W. BESSLER\textsuperscript{1} (Student Member, IEEE), SUMA G. CARDWELL\textsuperscript{2} (Member, IEEE), CATHERINE SCHUMAN\textsuperscript{3} (Senior Member, IEEE), J. DARBY SMITH\textsuperscript{2}, JAMES B. AIMONE\textsuperscript{2} (Senior Member, IEEE), SHASHANK MISRA\textsuperscript{2}, and JEAN ANNE C. INCORVIA\textsuperscript{1} (Senior Member, IEEE)

\textsuperscript{1}Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA
\textsuperscript{2}Sandia National Laboratories, Albuquerque, NM 87123 USA
\textsuperscript{3}Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996 USA

CORRESPONDING AUTHOR: J. A. C. INCORVIA (incorvia@austin.utexas.edu)

This work was supported by the DOE Office of Science (Advanced Scientific Computing Research/Basic Energy Sciences) Microelectronics Co-Design Project COINFLIPS (CO-designed Improved Neural Foundations Leveraging Inherent Physics Stochasticity). The work of Samuel Liu was supported by the National Science Foundation Graduate Research Fellowship Program under Grant 202131125. The work of Paul W. Bessler was supported by the National Science Foundation "Research Experiences for Undergraduates" in accordance with the NSF Program Solicitation NSF 19-582 under Grant 2006753.

This article has supplementary downloadable material available at https://doi.org/10.1109/JXCDC.2022.3231550, provided by the authors.

\section*{ABSTRACT}
Probabilistic computing using random number generators (RNGs) can leverage the inherent stochasticity of nanodevices for system-level benefits. Device candidates for this application need to produce highly random “coinflips” while also having tunable biasing of the coin. The magnetic tunnel junction (MTJ) has been studied as an RNG due to its thermally-driven magnetization dynamics, often using spin transfer torque (STT) current amplitude to control the random switching of the MTJ free layer (FL) magnetization, here called the stochastic write method. There are additional knobs to control the MTJ-RNG, including voltage-controlled magnetic anisotropy (VCMA) and spin orbit torque (SOT), and there is a need to systematically study and compare these methods. We build an analytical model of the MTJ to characterize using VCMA and SOT to generate random bit streams. The results show that both methods produce high-quality, uniformly distributed bitstreams. Biasing the bitstreams using either STT current or an applied magnetic field shows a sigmoidal distribution versus bias amplitude for both VCMA and SOT, compared to less sigmoidal for stochastic write. The energy consumption per sample is calculated to be 0.1 pJ (SOT), 1 pJ (stochastic write), and 20 pJ (VCMA), revealing the potential energy benefit of using SOT and showing using VCMA may require higher damping materials. The generated bitstreams are then applied to two tasks: generating an arbitrary probability distribution and using the MTJ-RNGs as stochastic neurons to perform simulated annealing, where both VCMA and SOT methods show the ability to effectively minimize the system energy with a small delay and low energy. These results show the flexibility of the MTJ as a true RNG and elucidate design parameters for optimizing the device operation for applications.

\section*{INDEX TERMS}
Bitstream, magnetic tunnel junction (MTJ), probabilistic-bit, spin orbit torque (SOT), voltage-controlled magnetic anisotropy (VCMA).

\section*{I. INTRODUCTION}
The magnetic tunnel junction (MTJ) has been widely used in non-volatile random access memory, logic devices, and unconventional computing because of its benefits, e.g., nanosecond speed operation, scalability, thermal stability, high endurance, and compatibility with CMOS [1], [2], [3], [4], [5]. More recently, the MTJ is being studied as a probabilistic-bit circuit element due to its thermally-driven magnetization dynamics [6], [7], [8], [9], [10]. Probabilistic computing leverages the inherent stochasticity of
nanodevices, especially to tackle applications in which incorporating stochasticity can provide system-level energy efficiency [11], [12], for example when the energy cost of high precision is not needed, or because the application itself is probabilistic and could be more efficiently carried out with stochastic hardware. Probabilistic algorithms require large numbers of random bitstreams, where a stochastic device (e.g., switching between two states) is sampled over time. While a logic device may be evaluated by its switching speed and energy, a probabilistic device is evaluated by the quality of the random number generated (i.e., is it a true random number?), the quantity of random numbers that can be generated per unit time, the type of random bitstream that is generated (e.g., is it a distribution of random numbers uniform, Gaussian, etc.), and the controllability of the random number distribution (i.e., can the probability of measuring 0 or 1 be controlled).

The MTJ has been used as a stochastic-bit circuit element using two main methods: stochastic read and stochastic write. The MTJ is promising for this application because both of these methods offer controllability, where the probability of generating a 0 or 1 can be biased using an applied current or magnetic field [13]. In the stochastic read method, the MTJ is designed with its free layer (FL) near the superparamagnetic limit, such that it thermally fluctuates in time [14], [15], [16]. While bitstream generation can be in the sub-microsecond regime [6], the bitstream generation rate is sensitive to temperature, hindering its use. Alternatively, the MTJ can be designed with a thermally-stable FL, and the amplitude of the switching current used to switch the MTJ via spin transfer torque (STT) will determine the switching probability. This stochastic write method is more favorable for fast bitstream generation speed, scaling to small sizes, and robustness against temperature, but could potentially quickly degrade the tunnel barrier due to high current requirements for probabilistic switching in the ballistic regime.

Here, we design an MTJ device that utilizes voltage-controlled magnetic anisotropy (VCMA) and spin-orbit torque (SOT) [17], [18], [19] to generate random bitstreams. While VCMA and SOT have been studied for this application, their quality, quantity, type, and controllability have not been systematically compared and understood [20], [21], [22]. Since micromagnetic solvers are too resource-intensive to generate large numbers of random bitstreams, we build an analytical model of the VCMA/SOT-MTJ that captures the necessary physics, including device geometry, magnetic material parameters, thermal energy, VCMA, STT, and SOT. We show that probabilistic MTJs based on both VCMA and SOT can produce high-quality, tunable bitstreams that pass the National Institute of Standards and Technology (NIST) uniformity and runs tests for randomness. We compare the energy dissipation and thermal stability of both mechanisms to a stochastic write MTJ. Lastly, we present additional controllability of the VCMA/SOT-MTJs not present in the stochastic read/write MTJs. We show that this control can represent weighted probability problems and can implement simulated annealing in Boltzmann machines applied to the maximum-satisfiability (MAX-SAT) problem.

II. DEVICE OPERATION AND THEORETICAL DETAILS

Fig. 1(a) depicts the VCMA/SOT-MTJ device. The MTJ has perpendicular magnetic anisotropy (PMA) and is in a bottom ferromagnetic (FM) FL, top pinned layer (PL) structure [23], [24], [25], [26], [27]. The various terminals provide knobs to control the FL switching probability, and in turn, the resistance of the MTJ, $R_{MTJ}$, between its two possible resistance states $R_{MTJ} = R_P$ (FM layers parallel P) and $R_{MTJ} = R_{AP}$ (FM layers antiparallel AP). Terminals T2–T4 (grounded, GND) are used to apply the bias voltage $V_b$ for VCMA. Terminals T1 and T2 are used for STT, and terminals T1–T3 for SOT [28], [29], [30], [31], [32], [33], [34].

Fig. 1(b) and (c) depict how the VCMA-MTJ generates random bitstreams by switching into P or AP states after bias voltage is applied. The PMA FL starts out with a double-well energy potential (dotted blue line): its magnetization can be P (0°) or AP (180°) relative to the PL. In Fig. 1(b), turning the voltage on from T2–T4 reduces the PMA and sends the FL magnetization in-plane (90°) through precessional motion and damping, depicted by the red lines. The right side of Fig. 1(b) shows when the voltage is turned back off, PMA is restored, and the FL transitions into one of two potential P or AP states, providing the probabilistic bit 1 or 0.
more strongly than VCMA. This also indicates that in this material system, VCMA is slower to switch than SOT. This is because VCMA relies on a combination of the change in anisotropy as well as damping to switch; more ideal materials for this device type would have a higher damping constant $\alpha$ and VCMA coefficient $K_x$, $\text{vcma}$.

In Fig. 2(c), the operation of a stochastic write MTJ is also shown. A 1 ns, $-0.25$ mA current pulse is used to stochastically write the device. The pulse duration is chosen to ensure that switching occurs in the ballistic regime; the switching probability is relatively independent of temperature. Due to the difference in operation, a 10 ns, opposite polarity current pulse of 0.10 mA is used to reset the device, with resting periods inserted to match the period of the SOT and VCMA MTJs.

### IV. WEIGHTED BITSTREAM CREATION WITH STT-SOT AND STT-VCMA MTJS

Fig. 3 compares the control of weighted bitstreams generated with SOT, VCMA, and stochastic write methods. For SOT and VCMA in Fig. 3(a) and (b), respectively, the STT current density $J_{\text{STT}}$ and magnetic field $H$ describe a constant bias applied to the device during operation. The pulse and relax time durations are chosen to be the minimum possible while still producing a high-quality distribution. The pulse (relax) times chosen for the SOT and VCMA MTJs, respectively, are 15 (20 ns) and 30 (20 ns). For the stochastic write MTJ distribution in Fig. 3(c), the $x$-axis represent the current density or magnetic field applied during the 1 ns write pulse. A 10 ns pulse of $5 \times 10^{10}$ A/m$^2$ is used to reset the device. A 10 ns is chosen to reduce the reset current density in the interest of preserving the tunnel junction. We analyze the average weight $P$ over many cycles using

$$P = \frac{1}{q} \sum_{i=1}^{q} Z(m_z)$$  \hspace{1cm} (1)

where $m_z$ is the unit magnetization vector in the $z$-direction, $Z(m_z)$ is the output bitstream function that is 0 when $m_z$ is negative and 1 when $m_z$ is positive, and $q$ is the number of samples; here $q = 2000$ samples, repeated for each point while varying current density and magnetic field separately. The first standard deviation is also shown in the colored cloud around each series, which is almost unobservable. This indicates that for the number of samples chosen, a high quality and accurate stochastic weight can be obtained.

Comparing the distributions between the three device types, the current density required to weigh the devices is similar in magnitude for the VCMA and SOT-based devices. Within the material system simulated, the stochastic write MTJ requires currents that are around 500 times larger to stochastically switch the device within 1 ns; this would require larger transistors that can drive larger currents. Comparing the distribution shapes, the SOT and VCMA-based MTJ distributions are approximately sigmoidal, in contrast to that of the stochastic write MTJ. The VCMA/SOT-based

### TABLE 1. Physical parameters used in the model.

| Symbol          | Magnetic constant | Values          |
|-----------------|-------------------|-----------------|
| $\alpha$        | Gilbert damping   | 0.03            |
| $M_s$           | Saturation magnitization | $1.2 \times 10^6$ A/m |
| $H$             | Anisotropy effective field | $1.8 \times 10^4$ Oe |
| $A_e$           | Exchange stiffness | $4 \times 10^4$ erg/cm |
| $P$             | Spin polarization of tunnel current | 0.6 |
| $T$             | Temperature       | 300 K           |
| $k_B$           | Boltzmann constant | $1.38 \times 10^{-23}$ J/K |
| $\mu_0$        | Magnetic permeability | $4\pi \times 10^{-7}$ H/m |
| $e$             | Electron charge   | $1.6 \times 10^{-19}$ C |
| $l_{MO}$        | MgO thickness     | $1.5 \times 10^{-9}$ m |
| $b_{FL}$        | Free layer thickness | $1.1 \times 10^{-9}$ m |
| $a$             | MTJ diameter      | $50 \times 10^{-9}$ m |
| $TMR$           | TMR ration at 0 bias-volt | 150% |
| $\phi$         | MgO potential barrier | $-1.2$ eV |

III. STOCHASTIC MTJ BEHAVIOR

Fig. 2 shows the bitstream generation using SOT, VCMA, and stochastic write. For SOT operation in Fig. 2(a), 0.15 mA was selected as the pulse amplitude since it corresponds to approximately 100 mV applied through the bottom HM, typical of low voltage operation of SOT devices. For the VCMA-operated MTJ, 1.5 V applied across the tunnel barrier was chosen to maximize the damping speed while remaining under breakdown voltage. For both, a 30 ns pulse is applied followed by a 30 ns relaxation period. The blue and green series represent two different runs with the same input, showing stochastic bitstream generation represented by $\hat{z}$ magnetization as a function of time. Comparing the output magnetization during pulses, it is evident that the SOT pins the FL magnetization in the in-plane orientation ($m_z = 0$) using thermal energy at room temperature. Detailed simulation methods can be found in Supplementary Information.

Table 1 lists the parameters and values used in the simulation to represent a standard CoFeB PMA MTJ. The diameter of the circular MTJ is 50 nm, tunnel magnetoresistance (TMR) is chosen to be a reasonable value for PMA MTJs, TMR = 150%, and the FL thickness is chosen to be 1.1 nm. The STT critical switching current density ($J_s = I_c/A$, $I_c$ is critical switching current and $A$ is the MTJ area) for the chosen parameters evaluates in the range of $1 \times 10^{11}$ – $4 \times 10^{12}$ A/m$^2$, reasonable for CoFeB [35], [36], [37], [38], [39].
MTJs are centered with a weight $P = 0.5$ at zero bias. In an experimental system, this may not be observed due to the presence of stray fields, but this can potentially be alleviated using stack engineering. This centering of $P = 0.5$ at zero bias can be advantageous when producing a uniformly distributed true random number, shown in Section V.

V. RANDOM BITSTREAM QUALITY

An immediately attractive application for devices that can stochastically output random bits is as a building block for true random number generators (TRNGs). TRNGs have been used in a wide range of applications from neuromorphic computing to hardware security. Therefore, it is important for these random bitstream generators to output high-quality random numbers.

Fig. 4(a)–(c) shows a kernel density estimation of a distribution of random numbers generated to perform the NIST uniformity test for the quality of random numbers [40]. For each of the device types, they are weighted to $P = 0.5$ and sampled eight times continuously to produce an 8-bit number, repeated to produce 10 000 samples. This is also done for $P = 0.6$ to compare with the result from a purposely biased generator. From the plots, all three types of devices can produce any number between 0 and 255 at an approximately equal probability for $P = 0.5$. The quality of the distribution is evaluated by first computing a normalized $\chi^2$

$$\chi^2 = \frac{N}{\sum_i (O_i - E)^2}{E}$$

where $N$ is the number of possible output numbers (number of bins), $O_i$ is the observed frequency within a given bin, and $E = N/q$ is the expected frequency, where $q$ is the total number of samples. The $\chi^2$ value is a normalized sum of squared deviations from an expected distribution. If the value is small enough, then the observed quantity deviates very little from the expected distribution. For VCMA, SOT, and stochastic write MTJs, the computed $\chi^2$ values are respectively 272.02, 268.11, and 246.25 for a system with 255 degrees of freedom. This corresponds to $p$-values of 0.2215, 0.2742, and 0.6415 for a $\chi^2$ goodness-of-fit test. These $p$-values are much larger than 0.01, indicating that all three devices can produce bitstreams that do not significantly deviate from the uniform distribution.

This is further corroborated in the run tests shown in Fig. 4(d)–(f). This test checks for the length of repeated instances of 1’s within a set of continuous samples, chosen to be 10 000. The blue triangles depict the proportion of instances as a function of run length for devices biased to $P = 0.5$. In all cases, the distribution matches the description $P_1 = (1/2)^x$, where $x$ is the length of the run of 1’s and $P_1$ is the probability that a run of 1’s of length $x$ is observed. Distributions are also presented for devices biased to $P = 0.7$, showing deviation away from the expected distribution. The tests show that all three devices produce bitstreams that are identically distributed, even when weighted.

VI. TEMPERATURE AND ENERGY COMPARISONS

Because all three device types rely on thermal fluctuations as the source of stochasticity, the effect of temperature changes on biasing of the device is a concern. Fig. 5(a) shows the weight stability of all three types of devices biased to levels of 0.3, 0.5, and 0.7, with 500 samples per point. For all three devices, there is essentially no temperature dependence for a biasing of 0.5 even up to 500 K, though it is important to note that the deviation from 0.5 for the VCMA-MTJ could indicate that 500 samples are not enough to generate a well-calibrated weight. The lack of temperature dependence for a bias of 0.5 is due to the physical switching mechanism; a higher temperature does not change the average magnetization when the device is sampled and forced to the in-plane state. As a result, whether a 0 or a 1 is sampled remains a 50% probability. At biases of 0.3 and 0.7, while the apparent weight of the stochastic MTJ remains close to the original at 300 K, there is a clear trend toward 0.5 for the VCMA and SOT-based MTJs. This is because as the thermal fluctuations get larger, a larger current is required to bias the device past the range of

FIGURE 2. Stochastic switching dynamics of RNGs. For (a) SOT, (b) VCMA, and (c) stochastic write MTJ device types, the normalized out-of-plane magnetization vector $m_z$ as a function of time is shown for 20 samples over 2000 ns. Two independent runs are shown in green and blue to show stochasticity. The sampling control methods as a function of time are shown in the plots above.
FIGURE 3. Probabilistic weight control of stochastic devices. The probability of sampling a 1 is shown as a function of STT current $J_{\text{STT}}$ and magnetic field $H_z$ for (a) SOT, (b) VCMA, and (c) stochastic write MTJ device types, where each point is generated by sampling a given device 2000 times. The weight in response to STT current (magnetic field) is shown in blue (orange). The cloud around each distribution shows the first standard deviation (almost unobservable).

For the material system simulated, the energy dissipation is calculated as follows:

$$E = \int_{t_1}^{t_2} \left( \frac{V_b^2}{R_{\text{MTJ}}} + I_{\text{STT}}^2 R_{\text{MTJ}} + I_{\text{SOT}}^2 R_{\text{HM}} \right) \ast dt$$

(3)

where $R_{\text{MTJ}}$ and $R_{\text{HM}}$ are the parallel resistance of the MTJ and resistance of the HM layer, respectively, $I_{\text{SOT}}$ is the SOT current applied through the HM, $I_{\text{STT}}$ is the current applied through the junction, and $t_1$ and $t_2$ are the beginning and end times of one sample. Fig. 5(b) shows the energy dissipation necessary to produce one sample for the three device types as a function of the weight. Due to the contrasting physical methods of operation, the energy dissipation between the three device types is on different orders of magnitude. For the stochastic write MTJ, there is an increase in energy dissipation with respect to weight due to the larger amplitude current necessary to set a larger weight. This is slightly present for the SOT and VCMA methods, but to a much lesser extent because only a small STT current bias is necessary to weigh the devices. The VCMA-MTJ dissipates on the order of 20 pJ/sample, comparatively higher than that of the stochastic write MTJ at around 1 pJ/sample and the SOT-MTJ at around 0.1 pJ/sample. This is mostly because of the relatively slow damping of the magnetization toward the in-plane superposition, requiring a relatively high-voltage pulse of $V_b = 1.5$ V applied for 30 ns, compared to the short write (1 ns) and reset (10 ns) pulses of the stochastic write MTJ and low-current, 15 ns pulse of the SOT-MTJ. As a result, for future development of VCMA-MTJs in this application, materials with high damping $\alpha$ or VCMA coefficient $K_{s,\text{vcma}}$ can reduce the energy cost and increase sampling speed.

VII. APPLICATION (GENERATING SAMPLES FROM A DISTRIBUTION)

While probabilistic bit streams can be used for a variety of applications, one widely applicable function is to use random number generators (RNGs) to represent a given distribution, which all three device types described can feasibly accomplish. Suppose the distribution we would like to sample from is that of a four-sided die that rolls 0 with probability 1/2, and rolls 1, 2, and 3 with probability 1/6 each. These outcomes can be identified with the outcomes of two coinflips that can be heads (H) or tails (T): 0 corresponds to TT, 1 to TH, 2 to HT, and 3 to HH. If $p$ is the probability of H for coin 1 and $q$ is the probability of H for coin 2, then to represent the desired probability distribution, the following needs to hold:

$$pq = \frac{1}{2}$$

$$p(1 - q) = \frac{1}{6}$$

$$q(1 - p) = \frac{1}{6}$$

$$(1 - p)(1 - q) = \frac{1}{6}.$$  (4)
This is an over-constrained set of equations, rendering no solution for $p$ and $q$, i.e., we do not yet know how to bias the probability of H:T of our device from 50:50 to something else to achieve this probability distribution.

To obtain the desired distribution, we employ a hidden process that induces dependence between observable coins. That is, we assume there is some hidden process that alters the coin’s bias, such that for a fraction $r$ of the time coin P has a probability of heads $p_1$, and coin Q has a probability of heads $q_1$. For the remainder of the time the coins are sampled, the coins have probabilities of heads $p_2$ and $q_2$, respectively.

We term this method the hidden dependence method. This yields the following system of equations that can now be solved:

$$
 rp_1 q_1 + (1 - r) p_2 q_2 = \frac{1}{2}, \\
 rp_1 (1 - q_1) + (1 - r) p_2 (1 - q_2) = \frac{1}{6}, \\
 r (1 - p_1) q_1 + (1 - r) (1 - p_2) q_2 = \frac{1}{6}, \\
 r (1 - p_1) (1 - q_1) + (1 - r) (1 - p_2) (1 - q_2) = \frac{1}{6}.
$$

The Library of Evolutionary Algorithms in Python (LEAP) is used to solve the equations and determine $r$, $p_1$, $q_1$, $p_2$, $q_2$ to within a desired tolerance: in this example, it comes out to requiring $r = 0.64925303$, $p_1 = 0.8326947$, $q_1 = 0.84802849$, $p_2 = 0.3542616$, and $q_2 = 0.32530298$. The precision of the quantities depends on the precision needed for the probability distribution. To map out the desired probability distribution, we sample two devices for each device type to represent coin P and coin Q, every 500 times.

The value $r$ is the proportion of time two coins are selected with the probability of heads $p_1$ and $q_1$; for the remainder of the time, coins are used with a probability of heads $p_2$ and $q_2$. When sampling, we randomly select between these tunings such that the proportion $r$ is achieved on average. Each time, when $0 \leq r < 0.64925303$ we bias coin P to $p_1 = 0.8326947$ and then flip it to see if we achieve H or T; similarly, we bias coin Q to $q_1 = 0.84802849$ and then flip it to see if we achieve H or T. When $0.64925303 \leq r \leq 1$, we bias coin P to $p_2 = 0.3542616$ and then flip it to see if we achieve H or T; and, we bias coin Q to $q_2 = 0.32530298$ and then flip it to see if we achieve H or T. Fig. 6(a) shows the resulting probability density of TT, TH, HT, and HH over the 500 coinflips for each of the three devices, showing we approximately achieve the desired probability distribution. We see that there is some deviation from the exact probability distribution. This is in part because the devices are weighted using lookup tables constructed from Fig. 3 with approximately 2% noise, similar to how
The ability to control the VCMA and SOT MTJ device distributions allows their use as neurons in Boltzmann machines for simulated annealing strategies to solve non-deterministic polynomial-time hard (NP-hard) optimization problems. An optimization problem can be translated into a weight matrix, where the solutions to the problem can be solved by minimizing the energy of the system. This is accomplished by using tunable random bit generators as neurons. The outputs of the random bits are applied in a matrix-vector multiplication (MVM) with the set weights, which are then recursively input back into the neurons. The system then eventually settles into preferred states that depict the solution. Since an optimization problem can have many solutions, stochasticity mediated by “temperature” is utilized to check many possible answers by sampling repeatedly. This “temperature” is gradually adjusted in the simulated annealing process to obtain an optimized result. Therefore, devices that can produce controllable distributions that emulate temperature effects are good device candidates.

The magnitude of the sampling pulse can be used to adjust the effective temperature of the system, shown in Fig. 6(b) for the SOT-MTJ. A higher amplitude SOT current pulse ($4 \times 10^{11}$ A/m$^2$) increases the stochasticity across the applied STT current range, thus resulting in a higher “temperature” while a lower amplitude pulse ($2 \times 10^{11}$ A/m$^2$) results in a lower “temperature.” Analogously, the VCMA-MTJ displays the same behavior with changes in voltage amplitude.

The system-level diagram of the Boltzmann machine for simulated annealing along with a presented optimization problem is shown in Fig. 6(c). The case when using a digital systolic array is shown in the figure, where an analog-to-digital converter (ADC) converts the analog output of the neuron before the MVM and a digital-to-analog converter (DAC) converts the digital output into an analog signal for the device input. The chosen problem is a MAX-SAT problem that attempts to find a combination of $X$, $Y$, $Z$ Boolean variables that can satisfy the largest number of clauses. The system consists of six neurons (random-bit MTJs) $V_{BM}$ to represent the three variables and their complements, along with a $6 \times 6.16$-bit floating point weight matrix $W_{BM}$. The clauses are translated into weight matrix values as described by Bojnordi and Ipek [42]. The system energy is calculated using $E_{BM} = V_{BM}^T W_{BM} V_{BM}$ to visualize the annealing process in Fig. 6(d).

As shown in Fig. 6(d), only a small number of temperature steps are necessary to converge to the optimized solution, though the annealing strategy will differ according to the problem. In this case, for both SOT and VCMA, the system temperature is initially set to be high and is decreased to the minimum value in equal increments over five iterations. For SOT, the initial maximum sampling current density is set to be $5 \times 10^{11}$ A/m$^2$ and is decreased to a minimum of $1 \times 10^{11}$ A/m$^2$. For VCMA, the initial maximum sampling voltage is set to 1.5 V and is decreased to a minimum of 1.1 V. Fig. 6(d) shows the result of 100 iterations of each of the annealing process, resulting in 100% accuracy in identifying the minimum system energy solution of $X'$, $Y'$, $Z$ at $E_{BM} = -33$. The darker lines depict multiple overlapping iterations. Each iteration took an average energy dissipation of 18.5 pJ.
FIGURE 6. Stochastic bitstream applications in distribution representation and simulated annealing. (a) Two devices of each device type are used to construct an arbitrary distribution calculated from the Hidden Correlation Bernoulli Coins method, showing a good match with the ideal distribution (black dashed line) for the probability of getting Tails-Tails (TT), Tails-Heads (TH), etc. for the two MTJs. (b) Change in effective temperature by modulating sampling pulse amplitude for a SOT device. Each point is an average of 500 samples. (c) Schematic of a Boltzmann machine for simulated annealing. The clauses selected for the MAX-SAT test are also shown. MTJs are depicted by the circles; arrows show the MTJ output that is digitized for MVM with the weight matrix; and arrows show the recursive input back into the MTJs. Pink lines show the temperature change as shown in (b). (d) Simulated annealing process of Boltzmann machines constructed from both SOT and VCMA devices for 100 cycles each. The correct solution is also shown in the box above.

for the SOT-MTJ and 305.3 pJ for the VCMA-MTJ. The combination of tunability and low energy indicates that both MTJ types can be used effectively as neurons for simulated annealing implementations.

IX. CONCLUSION

In summary, we have presented comprehensive simulations of an MTJ-based device that can utilize either VCMA or SOT along with thermal fluctuations to generate high-quality bitstreams, enabling their use as TRNGs. We show that the output bitstream probability can be controlled using STT current or an external magnetic field and compare the controllability with a stochastic write MTJ. The quality of output bitstreams using all three device types is evaluated on the NIST uniformity and runs tests, with all three devices performing past the threshold necessary for a TRNG. The energy dissipation, delay, and thermal stability are compared between the device types. With the presented model parameters, the VCMA-MTJ has the largest delay and energy dissipation, and we evaluate that material optimizations for higher damping $\alpha$ and higher VCMA coefficient $K_{vcma}$ could bring device performance to parity. We then present two example applications for the devices. Firstly, the devices were shown to be able to generate an arbitrary probability distribution. Secondly, the VCMA and SOT MTJs were shown to have a degree of controllability emulating temperature by modulating pulse duration, allowing use as stochastic neurons to perform simulated annealing in a Boltzmann machine. Both device types were shown to effectively minimize the system energy with a small delay and low energy. This work proposes and evaluates novel device architectures that demonstrate promise in implementing tunable RNGs that can be the building blocks of stochastic and neuromorphic systems.

ACKNOWLEDGMENT

This article describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by the National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under Contract DE-NA0003525. (Samuel Liu and Jaesuk Kwon contributed equally to this work.)
