Low-Energy Truly Random Number Generation with Superparamagnetic Tunnel Junctions for Unconventional Computing

D. Vodenicarevic,[1] N. Locatelli,[1] A. Mizrahi,[1,2] J. S. Friedman,[3] A. F. Vincent,[1] M. Romera,[2] A. Fukushima,[4] K. Yakushiji,[4] H. Kubota,[4] S. Yuasa,[4] S. Tiwari,[5] J. Grollier,[2] and D. Querlioz[1,*]

[1]*Centre for Nanoscience and Nanotechnology, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France*

[2]*Unité Mixte de Physique CNRS, Thales, Université Paris-Sud, Université Paris-Saclay, 91767 Palaiseau, France*

[3]*University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas 75080, USA*

[4]*AIST Tsukuba, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8561, Japan*

[5]*School of ECE, Cornell University, Ithaca, New York 14850, USA*

(Received 16 June 2017; revised manuscript received 12 September 2017; published 22 November 2017)

Low-energy random number generation is critical for many emerging computing schemes proposed to complement or replace von Neumann architectures. However, current random number generators are always associated with an energy cost that is prohibitive for these computing schemes. We introduce random number bit generation based on specific nanodevices: superparamagnetic tunnel junctions. We experimentally demonstrate high-quality random bit generation that represents an orders-of-magnitude improvement in energy efficiency over current solutions. We show that the random generation speed improves with nanodevice scaling, and we investigate the impact of temperature, magnetic field, and cross talk. Finally, we show how alternative computing schemes can be implemented using superparamagentic tunnel junctions as random number generators. These results open the way for fabricating efficient hardware computing devices leveraging stochasticity, and they highlight an alternative use for emerging nanodevices.

DOI: 10.1103/PhysRevApplied.8.054045

I. INTRODUCTION

With conventional transistor technology reaching its scalability limits [1], significant effort is involved in the investigation of alternative computing schemes for microelectronics. Many of these emerging ideas, such as stochastic computing [2–6] and certain brain-inspired (or neuromorphic) schemes [7–9], require a large quantity of random numbers. However, the circuit area and the energy required to generate these random numbers are major limitations of such computing schemes. For example, in the popular neuromorphic TrueNorth system [7], one third of the neuron area is dedicated to performing random number generation. Indeed, 10^6 random bits are required at each integration step of the system. More concerning, in stochastic computing architectures, random number generation is typically the dominant source of energy consumption, as the logic performed using the random bits is generally quite simple and efficient by principle. Many practical stochastic computing schemes therefore try to limit the reliance on expensive independent random bits using various techniques, including the sharing or reuse of random bits [10–12]. However, such tricks limit the capabilities of stochastic computing to small tasks, as they introduce correlations between signals.

Most of the aforementioned unconventional computing circuits use pseudorandom number generators. But these pseudorandom number generators either lead to low-quality random numbers or are highly energy and area consuming. A preferable solution would be to rely on "true" random number generators that generate random bits based on physical phenomena that are intrinsically random. However, such truly random number generators are also difficult to realize with minimal energy consumption. This difficulty is due to the fact that most true random number generators function by triggering events whose outcome is intrinsically random. Triggering these events comes with a non-negligible energy cost. The most energy-efficient example uses a bistable CMOS circuit forced into in a metastable state which then randomly falls into one of the two stable states, generating one random bit [13]. It consumes 3 pJ/bit and a circuit area of 4000 μm^2.

In order to reduce this large area footprint, recent proposals suggest leveraging the inherent stochastic programing properties that arise in many of the bistable nanodevices developed for memory applications [14]. This approach was investigated with oxide-based resistive memory devices [15–18], phase-change memory devices

*damien.querlioz@u-psud.fr

[19,20], and magnetic memory devices [21–23], as well as with straintronic memory devices [24]. However, these approaches are based on repeated, energy-intensive programing operations, and they still require high energy for random bit generation. For instance, it requires dozens of pJ/bit to induce a stochastic switch of magnetization in magnetic tunnel junctions with two stable states, as proposed in the "spin-dice" concept, due to the high energy barrier between the magnetic states. Optimized schemes have been proposed [25–27], predicting a further reduction in the energy cost per bit, but they are still bounded by the need for a costly perturbation operation. While proposing a high-quality random number with high throughput, such strategies are no fit for emerging neuro-inspired computing applications like stochastic computing architectures.

A more natural approach would be to extract random numbers directly from thermal noise, as doing so provides randomness at no energy cost. Unfortunately, this approach requires large circuits to amplify thermal noise into a large signal of random bits, and it has not yet been shown to be more energy efficient than the first approach. The lowest-energy solution today is to use jitter as a way to efficiently amplify the noise present in CMOS ring oscillators. The most energy-efficient implementation [28] requires 23 pJ/bit and 375 μm^2.

In this work, we propose the use of nanomagnetic devices that intrinsically amplify thermal noise without external energy supply: superparamagnetic tunnel junctions. These bistable magnetic tunnel junctions are reminiscent of the ones used for magnetic random-access memories (MRAMs) [29]. However, contrary to MRAM cells, the energy barrier between the two magnetic states is very low, and thermal fluctuations induce repeated and stochastic magnetization switching between the two states at room temperature. Therefore, no write operations are required and a low-energy readout of the device state naturally produces random bits. We show that these devices permit the generation of high-quality random numbers at 20 fJ/bit using less than 2 μm^2, which is orders of magnitude more efficient in terms of energy and area than current solutions.

We first show experimentally that superparamagnetic tunnel junctions allow the generation of high-quality random bits with minimal readout circuitry, and that their behavior can be predicted by existing physical models. We then use the model to investigate the influence of device scaling and environmental factors on random bit quality and speed. Circuit simulation enables an estimation of the energy efficiency of random bit generation. Finally, we demonstrate the potential of these devices for unconventional computing through the example task of email message classification using random bits extracted from the experimental data, and we show that they are particularly well adapted to computing schemes trading off speed for ultra-low-energy consumption.

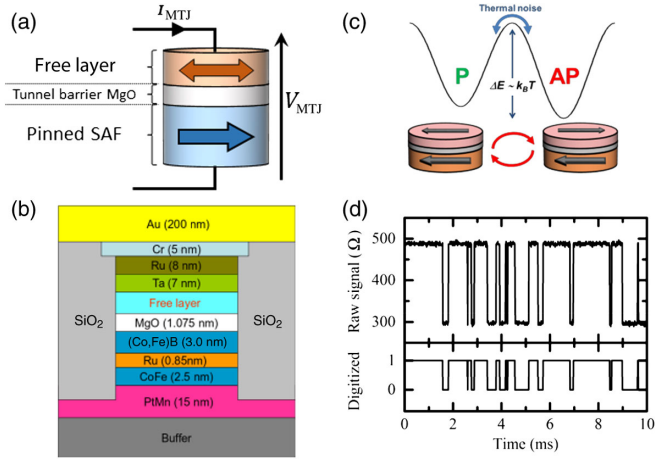


FIG. 1. Structure and behavior of superparamagnetic tunnel junctions. (a) Basic structure of the measured superparamagnetic tunnel junctions and readout setup. (b) Detailed stack of the junctions. (c) Representation of the two stable magnetic states, and the associated energy barrier. (d) Experimental resistance trace and thresholding operation.

II. EXPLOITING THE STOCHASTIC BEHAVIOR OF SUPERPARAMAGNETIC TUNNEL JUNCTIONS

Superparamagnetic tunnel junctions are bistable spintronic nanodevices composed of a high-stability pinned nanomagnet and a low-stability "free" nanomagnet, separated by a tunnel oxide layer [Fig. 1(a)]. Their structure is highly similar to the magnetic tunnel junctions used as the basic cells of MRAMs. The devices we measure are fabricated by sputtering, with a standard magnetic tunnel junction process, with the CMOS-compatible stack detailed in Fig. 1(b). *E*-beam lithography patterning is then performed to produce 50×150 nm^2 elliptic pillars.

The free magnet has two stable states, parallel (P) and antiparallel (AP) relative to the pinned layer [Fig. 1(c)]. Through the tunnel magnetoresistance effect [30], the electrical resistance of the junction in the AP state R_{AP} is higher than the resistance in the P state R_P. This effect is traditionally measured through the tunnel magnetoresistance (TMR) coefficient defined by $R_{AP}/R_P = 1 + \mathrm{TMR}$.

The lateral dimensions of the device are chosen so that the effective energy barrier between the two stable states is not very high compared to $k_B T$. Unlike the case of MRAMs, for which the magnetization direction of the free magnet is highly stable and can be switched only by proper external action, the magnetization direction of the superparamagnetic free magnet spontaneously switches between its two stable states due to low stability relative to thermal fluctuations [Fig. 1(c)] [31,32]. Here, no bias or perturbation scheme is required to provoke these random fluctuations—only temperature.

Resistance-versus-time measurements are done on junctions by applying a small, 10-μA constant current through

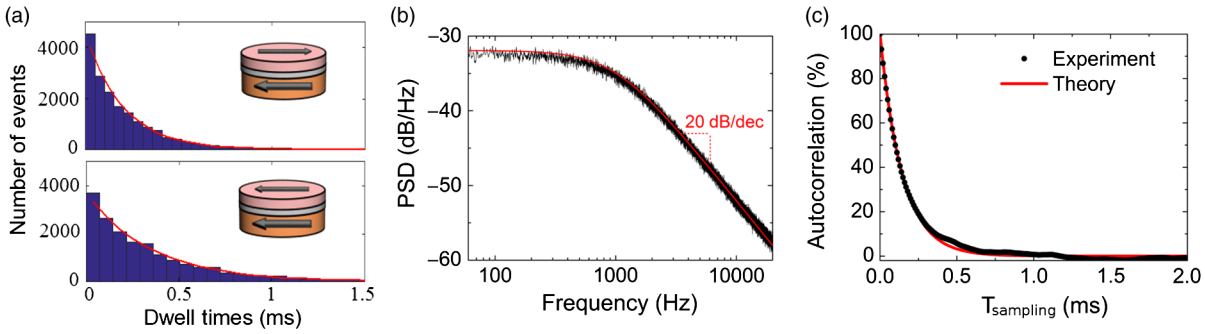


FIG. 2. Statistics of the experimental superparamagnetic tunnel junction signal. (a) Experimental histograms of the dwell times in (top panel) antiparallel [(AP) high-resistance] and (bottom panel) parallel [(P) low-resistance] states, for a superparamagnetic magnetic tunnel junction measured over 10 s. (b) Experimental power spectrum density (PSD) of the resistance signal. (c) Autocorrelation of the experimental resistance signal as a function of the signal sampling period.

the junction. Such a small current amplitude is chosen to have a negligible influence on the magnetic behavior of the device [33] and to maximize its lifetime while providing a clear signal. Figure 1(d) shows a sample from the time evolution of the electrical resistance of a junction measured at room temperature, as well as a binarized version, obtained by thresholding. We see that the resistance follows two-state fluctuations analogous to a random telegraph signal. The mean frequency of fluctuations is strongly related to the shape and material properties of the junction [34].

Figure 2(a) shows the histograms of the dwell times in the 1 (AP) and 0 (P) states, obtained through measurement of a superparamagnetic tunnel junction over a 10-s period. We see that these histograms can be fitted by an exponential law, which is characteristic of a Poisson process. Figure 2(b) presents the power spectrum density of the same signal, superimposed with the expected power spectrum density of a random telegraph signal based on a Poisson process. Excellent agreement between the measured results and the hypothesis of a Poisson process is seen.

Random bits can be extracted by sampling the voltage across the device at a constant frequency. The voltage is initially sampled at 100 kHz, and bitstreams with slower sampling rates are obtained by subsampling the initial bitstream. To evaluate the quality of the obtained random bits, the device is measured for over 2.5 days, producing 21.2 gigabits. No external magnetic field is applied during the measurement.

III. OPTIMIZING THE QUALITY OF RANDOM BITS

The sampling frequency needs to be chosen carefully relative to the mean switching frequency of the junction, defined as $F_{\mathrm{MTJ}} = 1/(\tau_1 + \tau_0)$, where τ_1 and τ_0 are the mean dwell times in states 1 and 0, respectively. F_{MTJ} is measured to be 1.66 kHz ($\tau_1 + \tau_0 \approx 604\ \mu$s). Figure 2(c) presents the correlation of consecutive bits extracted at different sampling rates. This result is superimposed on the one theoretically expected from a Poisson process. At high sampling frequency, subsequent bits are naturally autocorrelated (at $F_{\mathrm{sampling}} = 100$ kHz, correlation reaches 92.8%), and they can therefore not be used for applications. This correlation decreases exponentially with the sampling period, which can therefore be chosen based on the correlation requirements on the random numbers.

As observed in Fig. 2(a), the AP and P states possess an asymmetric stability: the device spends more time, on average, in the P state than in the AP state, which corresponds to a mean state (mean of the binarized signal) of 60.5%. This asymmetry can be connected to the stray field induced by the pinned magnetic layer structure, which is present in all magnetic tunnel junctions [35]. This biasing field offsets the junction mean state from the ideal 50% value required for most applications, and it is subject to device-to-device variations.

In order to eliminate this bias and any residual bit correlation, a "whitening" of the random bits is therefore required. To achieve this operation, we make use of a standard technique: combining several bitstreams into a single one using XOR gates. It can be shown (see Fig. S10 of the Supplemental Material for the mathematical derivation [36]) that the autocorrelation after XOR whitening is the product of the individual autocorrelations of the combined signals. The autocorrelation therefore decreases exponentially with the number of combined magnetic-tunnel-junction (MTJ) bitstreams, and it is always lower than the autocorrelation of any of the combined signals. In the same way, the mean state of the whitened bitstream gets exponentially closer to 50% with the number of XOR-combined bitstreams and always stays closer to perfect balance than any of the bitstreams in combination. As a reference, a more advanced but heavy stateful whitening technique (referred to here as the Blum technique [37]) is also applied to the raw measurements.

As an illustration, we consider bits extracted at a frequency of 5 kHz. The bitstream is then divided into

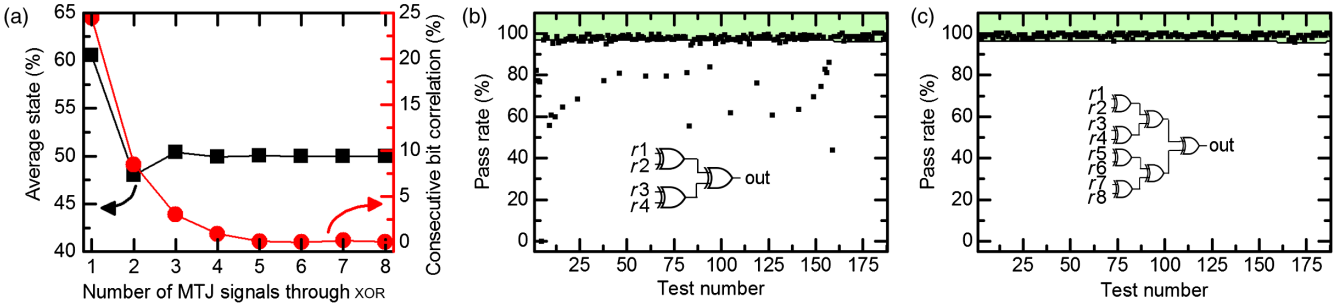


FIG. 3. Whitened experimental random bitstream quality assessment. (a) Mean-state and consecutive bit autocorrelation as functions of the number of independent superparamagnetic tunnel junction signals combined by XOR. NIST STS randomness quality test results on experimental data whitened by (b) XOR4 and (c) XOR8 at an $F_{\text{sampling}} = 5$ kHz sampling frequency. When all test results are in the green area, the bitstream is consistent with cryptographic quality.

chunks of equal length which are used as independent signals and XOR combined bit by bit for the XOR whitening process. We plot in Fig. 3(a) the consecutive bit correlation and the mean state of the whitened bitstream as functions of the number of signals combined by XOR. The correlation and the mean-value bias decrease with the number of XOR-combined signals. With 4 bitstreams (XOR4), the resulting consecutive bit correlation drops under 1% and the mean value reaches 49.9%. For 8 bitstreams (XOR8), the autocorrelation is below 0.06% and the mean state reaches 50%, with a standard deviation of 0.5%. These results suggest that XOR whitening can correct correlation and mean-value issues.

However, in order to fully evaluate the quality of a whitened bitstream, the signal autocorrelation and the mean state are not sufficient metrics. We therefore use the standardized National Institute of Standards and Technology Statistical Test Suite (NIST STS) [38], which evaluates the quality of the random bitstream against 188 tests. The NIST STS computes the statistics of bitstreams, such as mean value, autocorrelation, standard deviation, estimated entropy, and pattern occurrence frequencies, and determines whether they are consistent with perfect randomness. The NIST STS tests also look for the presence of repeated structures, linear dependencies, and other behaviors unexpected in a perfectly random bitstream.

To perform the NIST STS tests, the bitstream to be tested, measured over 2.5 days, is divided into 10^6-bit sequences. Each chunk is then tested independently, and the pass rate (the percentage of 10^6-bit sequences passing the test) is computed for each of the 188 tests. Figures 3(b) and 3(c) show the results for XOR4- and XOR8-whitened bitstreams, respectively. For a bitstream to be consistent with cryptographic quality, the pass rates of all tests should lie in the green region [38], corresponding to the expected minimal pass rate provided by the NIST STS, depending on the number of tested chunks. We can see that bits extracted by XOR8 whitening pass this requirement (this is also the case with the Blum technique), while, with XOR4 whitening, only a fraction of the tests is consistent with the cryptographic quality of the random bits [39].

Table I presents more-comprehensive results: the proportion of tests with passing rates that are consistent with cryptographic quality are given for XOR-whitened bitstreams at different sampling frequencies and numbers of XOR-combined bitstreams. The results confirm that the quality of the whitened bitstream increases for lower sampling frequencies (less correlation) and higher numbers of XOR-combined bitstreams (less correlation and lower bias). Higher numbers of XOR-combined bitstreams therefore allow for a further increase of the sampling rate while still passing all the NIST statistical tests, at the expense of more circuit area and energy consumption. XOR8 at $F_{\text{sampling}}/F_{\text{MTJ}} = 3.0$ appears to be an optimal choice, with 100% of the tests being consistent with cryptographic quality and the highest sampling frequency. A more-comprehensive analysis of the impact of the number of XOR-ed bitstreams is presented in Fig. S1 of the Supplemental Material [36].

Consistent results (presented in Fig. S2 of the Supplemental Material [36]) are observed on a second sample, measured over 1.5 days, producing 8.96 gigabits.

TABLE I. NIST Statistical Test Suite results for the whitened experimental random bitstream. Percentage of NIST STS tests satisfying cryptographic quality requirements for different numbers of combined bitstreams, and different sampling frequencies.

F_{sampling}	$F_{\text{sampling}}/F_{\text{MTJ}}$	Raw	XOR2	XOR4	XOR8
100 kHz	60.4	0	10.1	10.1	10.1
20 kHz	12.1	0.5	0.5	10.6	12.2
9.1 kHz	5.5	1.1	10.6	10.6	88.3
5.9 kHz	3.6	1.1	1.1	16.5	100
5 kHz	3.0	1.1	1.1	72.9	100
1.9 kHz	1.1	1.1	14.4	97.9	100
0.9 kHz	0.54	1.1	14.4	98.4	100
0.7 kHz	0.42	1.1	16.0	97.9	100
0.5 kHz	0.30	1.1	16.0	98.4	100

IV. SCALING CAPABILITIES OF THE RANDOM NUMBER GENERATORS IN TERMS OF SPEED AND ENERGY CONSUMPTION

A further study of the potential of superparamagnetic tunnel junctions for random number generation requires a realistic model of the device. In the literature, at low electric current, magnetic-tunnel-junction switching is usually described by an Arrhenius-Néel two-state analysis, modeling a thermally activated magnetic switching [40]. The mean switching rates in each state are then described by

$$r_{0\to1} = 1/\tau_0 = f_0 \exp\left(-\frac{\Delta E_{0\to1}}{k_B T}\right)$$
$$r_{1\to0} = 1/\tau_1 = f_0 \exp\left(-\frac{\Delta E_{1\to0}}{k_B T}\right), \quad (1)$$

where $f_0 = 1$ GHz is the magnetic attempt frequency, and $\Delta E_{0\to1}$ and $\Delta E_{1\to0}$ are the energy barriers associated with each transition [see Fig. 1(c)]. Our experimental results suggesting that superparamagnetic tunnel junction switching is a Poisson process are consistent with this model.

The superparamagnetic tunnel junctions that we characterized experimentally in this study are slow devices. They can be used to generate random bits at kilohertz frequencies, sufficient for real-time brain-inspired systems like those found in Ref. [7], but not for high-performance applications. In our (50×150)-nm superparamagnetic tunnel junctions, we identify that the switching occurs through nucleation and propagation of a magnetic domain, probably seeded by fluctuations in a subset of grains within it [31] (see Fig. S3 of the Supplemental Material [36]). By contrast, recent experiments on perpendicular-magnetic-anisotropy (PMA) magnetic tunnel junctions have shown that aggressively scaled devices (having diameters smaller than 35 nm) switch at the scale of the whole volume [34]. Therefore, in the context of random number generators, extreme scaling of the nanodevices appears as providential, as smaller volumes and areas are directly linked to a lower magnetization stability of the free magnet [41], increasing random-bit-generation speed exponentially. This beneficial impact of scaling effects is in sharp contrast to MRAMs, where conservation of stability with extreme scaling presents an important challenge [42].

From the study described in the previous section, we observe that a 25% correlation between consecutive bits can be efficiently whitened out by XOR8 and allow generated random numbers to pass all of the NIST STS tests. This consideration, together with the model, allows us to evaluate quantitatively the speed of scaled random bit generators based on superparamagnetic tunnel junctions by evaluating the maximum sampling frequency to keep the correlation $\rho^c_{X,X+1} \lesssim 25\%$ (see Fig. S4 of the Supplemental Material for details [36]):

$$F^{\max}_{\text{sampling}} \approx 3F_{\text{MTJ}} = \frac{3}{2} f_0 \exp\left(-\frac{\Delta E}{k_B T}\right), \quad (2)$$

where ΔE is the energy barrier separating the two states. $\Delta E = K_{\text{eff}}(D)\pi(D^2/4)t$ is derived as a function of the device diameter D, where $t = 1.6$ nm is the free-magnet thickness and the effective anisotropy $K_{\text{eff}}(D)$ is derived while considering interfacial anisotropy and bulk anisotropies, using experimental values from Ref. [34]. Figure 4(a), based on this derivation, shows that random bits could be generated at up to tens of megahertz for energy barriers below $5k_BT$, corresponding to a diameter of 8 nm.

In addition, in a final system, specialized transistor-based electronics needs to be associated with the superparamagnetic tunnel junctions to read their states without interfering with the random-bit-generation quality. Here, we consider a precharge sense amplifier circuit [(PCSA); see Fig. 4(b)], a CMOS circuit originally proposed as a MRAM-read circuit [43]. We simulate this circuit using standard integrated circuit design software (CADENCE tools) and the transistor models of a 28-nm commercial technology. The superparamagnetic tunnel junctions are modeled using a compact (VERILOG-A–based) model implementing the Arrhenius-Néel model. The results of circuit simulation [Fig. 4(c)] show that the read energy is relatively independent from superparamagnetic tunnel junction resistance, and very low

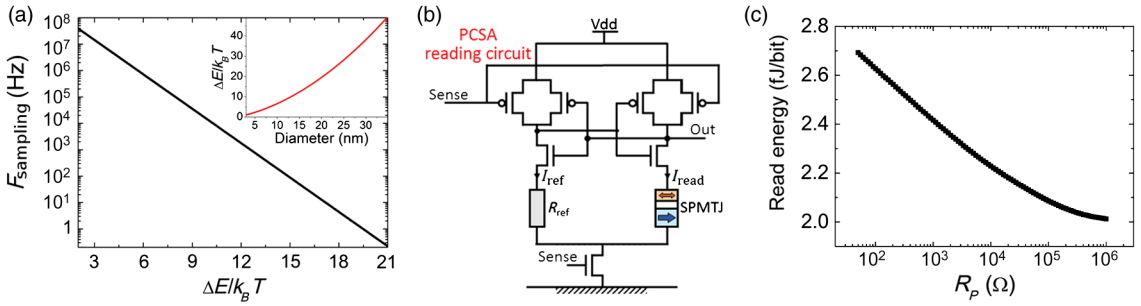


FIG. 4. Sampling rate and readout circuitry. (a) Effect of scaling the energy barrier on the ideal sampling frequency, based on the device model. (Inset) The energy barrier as a function of the junction diameter for PMA MTJs. (b) Precharge-sense-amplifier (PCSA) circuit for reading the state of a superparamagnetic tunnel junction (SPMTJ). (c) PCSA reading energy as a function of the superparamagnetic tunnel junction P state resistance R_P.

(approximately 2 fJ/bit). We also evaluate the read disturb effect of the PCSA. Reading the state of a junction can potentially affect random bit generation through the spin-torque effect. Based on the spin-torque model of Ref. [40], its impact on the mean state is around 10^{-6} for junctions such as the one we characterized experimentally. It would stay below 0.1% for ultrascaled junctions functioning at high frequencies, as shown in Fig. S11 of the Supplemental Material [36]. This small effect would therefore be corrected by whitening.

Evaluating the energy consumption of random bit generation requires taking into account the whitening process. As XOR whitening combines multiple junction states per generated bit, it requires multiple read operations per generated bit. XOR8 reads 8 junctions to generate a bit and requires 20 fJ/bit on average (including the XOR gate operation). In terms of area, in a 28-nm technology, the layout of a full XOR8 random bit generator takes less than 2 μm^2. XOR4 whitening would require 9.8 fJ/bit and a 1-μm^2 area.

These results show the potential of superparamagnetic tunnel junctions for state-of-the-art low-energy random number generation.

V. SENSITIVITY OF THE RANDOM NUMBER GENERATORS TO PERTURBATIONS

Although superparamagnetic tunnel junctions allow random number generation with minimal energy, their sensitivity to external perturbations must be carefully evaluated.

First, as the stochastic switching of superparamagnetic tunnel junctions is thermally activated, temperature directly affects their switching rates. Figure 5(a), based on the model introduced in the previous section, shows the temperature dependence of the maximum sampling frequency for several values of the effective barrier. Higher temperatures produce better random numbers: as temperature increases, the superparamagnetic tunnel junction switching rates increase accordingly, thus allowing faster sampling frequencies. Devices should therefore be sized based on their lowest operation temperature.

Superparamagnetic tunnel junctions are also sensitive to magnetic fields. Figure 5(b) shows the experimental mean state of a superparamagnetic junction as a function of the external magnetic field. Fields of a few oersteds shift the mean state to a level that cannot be corrected by XOR8 whitening. Magnetic shielding is therefore necessary for applications. Such technology (based on Mumetals) has already been developed for MRAM.

Finally, a challenge regarding scalability and integration is that closely packed superparamagnetic tunnel junctions can interact by dipolar interaction, which could lead to correlations in random numbers. In the case of perpendicularly magnetized superparamagnetic tunnel junctions, using the previously introduced model, we determine that the critical center-to-center distance between two superparamagnetic tunnel junctions guaranteeing negligible cross talk [44], corresponding to less than a $\rho_c = 0.1\%$ cross-correlation, is given by (see Fig. S5 of the Supplemental Material for details [36])

$$d_c = \left(\frac{\mu_0 (M_S V)^2}{4\pi k_B T \tanh^{-1}(\rho_c)}\right)^{1/3}. \tag{3}$$

Figure 5(c) shows the evolution of this critical distance at room temperature as the diameter of the junctions is scaled down. It falls below 100 nm for ultimately scaled 10-nm-diameter devices, which constitutes a layout design rule, and which would naturally be respected if the junctions were associated with PCSA circuits.

VI. USING SUPERPARAMAGNETIC TUNNEL JUNCTIONS FOR UNCONVENTIONAL COMPUTING

To illustrate the potential of superparamagnetic tunnel junctions for unconventional computing, we use the experimental whitened random bitstreams as inputs for a modern stochastic circuit [Fig. 6(a) and Ref. [3]] that performs

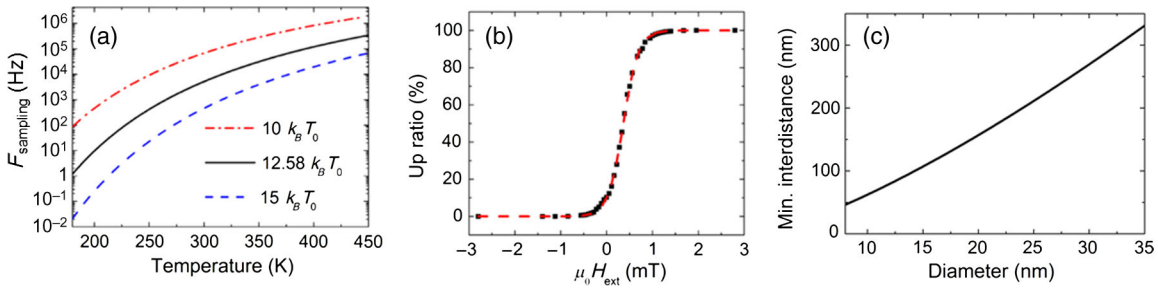
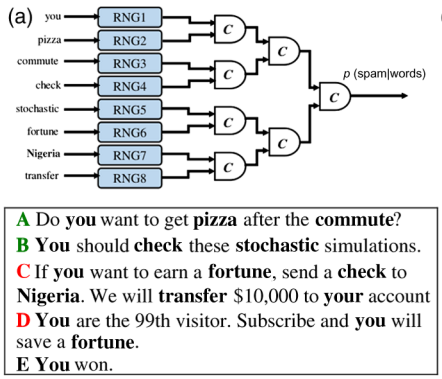
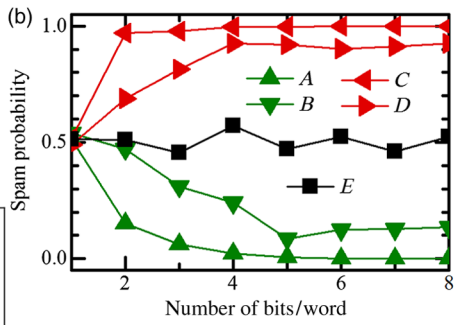


FIG. 5. External perturbations and cross-talk effects. (a) Theoretical curve of the maximum sampling frequency for high-quality random bit generation, as a function of temperature, for different junction stabilities (the black curve corresponds to the junction characterized in Figs. 2 and 3). (b) Black symbols indicate the experimental mean state of the junction (up ratio) as a function of the applied magnetic field (the red dotted line represents theoretical values). (c) Theoretical minimal distance between superparamagnetic tunnel junctions allowed to prevent cross talk, as a function of the superparamagnetic tunnel junction diameter.





(c)

Method	Messages correctly classified
No whitening	60%
XOR4	100%
XOR8	100%
Blum	100%

FIG. 6. Email classification with stochastic computing using whitened experimental random bitstreams. (a) Stochastic email classification circuit, and email messages to classify. One "RNG" block includes several random bit generators in order to provide bits with controllable probability. (Note that boldface in the sample message highlights the words present in the dictionary.) (b) Resulting spam probability as a function of the number of random bits per word using XOR4-whitened experimental 5-kHz data over 2000 iterations. (c) Spam classification success rates for different whitening techniques for 5-kHz sampling, using 8 bits/word and 2000 iterations.

Bayesian inference as a non-Turing machine. As a pedagogical task, we use this circuit to classify email messages as either spam or not spam [sample messages are presented in Fig. 6(a)], as was recently introduced in Ref. [3].

The approach uses a dictionary of known words with their associated occurrence rates in spam and nonspam messages. Each word of the dictionary has an associated probabilistic binary generator whose probability of drawing a 1 is set to different values depending on the presence (or absence) of the word in the presented sentence. As our random bit generators provide bitstreams with mean values of 0.5, multiple random bit generators are needed to create a probabilistic binary generator [see the random-number-generator (RNG) block in Fig. 6(a), which is detailed in Fig. S6 of the Supplemental Material [36]]. The outputs of these generators are then combined using C elements to perform an approximate Bayesian inference [3]. The time average of the output gives the probability that the presented message is spam.

Figure 6(b) gives the spam probability inferred using XOR4-whitened bitstreams and shows that the more random bit generators that are used per word, the more precisely the probabilistic binary generator can be tuned, and the better the prediction is. Also, the longer the output averaging time, the more accurate the answer for the system is. A trade-off for maintaining low energy consumption is found for 8 random bit generators/word and averaging over 2000 samples (see Figs. S7 and S8 of the Supplemental Material [36]).

Because of its reliance on multiple stages of binary bitstream combination, and fine generator probability tuning, this circuit is sensitive to the quality of the underlying random number generator. We test the circuit using raw 5-kHz-sampled experimental bitstreams, as well as its XOR4- and XOR8-whitened versions. When the bits are not whitened, the circuit does not perform satisfactorily [see Fig. 6(c) and Fig. S7 of the Supplemental Material [36]]. Using bits whitened with XOR8, the circuit performs as well as the referenced Blum whitener, successfully classifying all messages. Furthermore, XOR4, which does not pass all NIST STS tests, also provides perfect classification while requiring less energy.

These results highlight the potential of the approach for low-energy applications. Using the results of the previous section, circuit simulation with 8 random bit generators/word and 2000 clock cycles shows that a message can be classified using only nanojoule energy (the exact value depends on the number of words in the dictionary; see Fig. S9 of the Supplemental Material [36]). This simple study shows that superparamagnetic tunnel junctions can be used for efficient random number generation for low-power probabilistic computing.

VII. CONCLUSION

In this work, we show that the natural dynamics of superparamagnetic tunnel junctions produces random telegraph signals that can be read and turned into high-quality random bitstreams with minimal energy and circuit overhead while staying fully compatible with standard CMOS fabrication processes.

The whitening process turning these measurements into usable random bitstreams implies energy and area overhead. However, while the referenced Blum whitening would add important CMOS overhead, XOR adds very little. XOR8 and Blum both provide high random bit quality consistent with cryptographic requirements, but XOR8 fits better with low-energy applications, as it typically requires only 20 fJ/bit and 2 μm^2, orders of magnitudes less than the current state of the art. This efficiency comes at the cost of speed. Scaled superparamagnetic tunnel junctions could generate random bits at speeds of dozens of megahertz, which is slower than higher-energy random bit generators, but sufficient for many unconventional computing schemes in very-low-power consumption contexts such as the

Internet of things. This efficiency also comes at the cost of a certain sensitivity of random bit generation to the environment, making it prone to attacks. Random bit generation based on superparamagnetic tunnel junctions is therefore much better suited for unconventional computing than for cryptographic applications.

The evaluation of the probabilistic email classifier circuit also suggests that, in many alternative computing schemes, lower-quality whitening can be used successfully to achieve extreme energy efficiency without degrading performance. At design time, a balance between random number quality, generation speed, and energy consumption can be freely chosen to suit the target application. This flexibility is especially important in the context of modern Bayesian inference systems [45,46], but also for embedded circuits and Internet-of-things applications that are designed to work at low frequencies and low energies.

This study shows, through the example of superparamagnetic tunnel junctions acting as natural noise amplifiers, that emerging nanodevices could be used as highly efficient sources of true randomness for a wide range of applications.

ACKNOWLEDGMENTS

This work is supported by the European Research Council Starting Grant NANOINFER (Grant No. 715872), by the BAMBI EU collaborative FET Project grant (FP7-ICT-2013-C; Project No. 618024), by a public grant overseen by the French National Research Agency (ANR) as part of the Investissements d'Avenir program (Labex NanoSaclay; Grant No. ANR-10-LABX-0035), by the ANR grant CogniSpin (Grant No. ANR-13-JS03-0004), and by the French Ministère de l'écologie, du développement durable et de l'énergie. The authors thank J. Droulez and P. Bessière for the fruitful discussion.

[1] R. Courtland, Transistors could stop shrinking in 2021, IEEE Spectrum **53**, 9 (2016).

[2] A. Alaghi and J. P. Hayes, Survey of stochastic computing, ACM Trans. Embedded Comput. Syst. **12**, 92 (2013).

[3] J. S. Friedman, L. E. Calvet, P. Bessiere, J. Droulez, and D. Querlioz, Bayesian inference with Muller *C*-elements, IEEE Trans. Circuits Syst., I **63**, 895 (2016).

[4] A. Morro, V. Canals, A. Oliver, M. L. Alomar, and J. L. Rossello, Ultra-fast data-mining hardware architecture based on stochastic computing, PLoS One **10**, e0124176 (2015).

[5] T. J. Hamilton, S. Afshar, A. van Schaik, and J. Tapson, Stochastic electronics: A neuro-inspired design paradigm for integrated circuits, Proc. IEEE **102**, 843 (2014).

[6] C. Winstead, V. C. Gaudet, A. Rapley, and C. Schlegel, in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2005), Adelaide, Australia, 2005* (IEEE, New York, 2005), p. 1116.

[7] P. A. Merolla *et al.*, A million spiking-neuron integrated circuit with a scalable communication network and interface, Science **345**, 668 (2014).

[8] W. Maass, Noise as a resource for computation and learning in networks of spiking neurons, Proc. IEEE **102**, 860 (2014).

[9] M. Suri, D. Querlioz, O. Bichler, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, and B. DeSalvo, Bio-inspired stochastic computing using binary CBRAM synapses, IEEE Trans. Electron Devices **60**, 2402 (2013).

[10] B. D. Brown and H. C. Card, Stochastic neural computation. I. Computational elements, IEEE Trans. Comput. **50**, 891 (2001).

[11] A. Morro, V. Canals, A. Oliver, M. L. Alomar, and J. L. Rossello, Ultra-fast data-mining hardware architecture based on stochastic computing, PLoS One **10**, e0124176 (2015).

[12] S. S. Tehrani, W. J. Gross, and S. Mannor, Stochastic decoding of LDPC codes, IEEE Commun. Lett. **10**, 716 (2006).

[13] S. K. Mathew, S. Srinivasan, M. A. Anders, H. Kaul, S. K. Hsu, F. Sheikh, A. Agarwal, S. Satpathy, and R. K. Krishnamurthy, 2.4 Gbps, 7 mW all-digital PVT-variation tolerant true random number generator for 45 nm CMOS high-performance microprocessors, IEEE J. Solid-State Circuits **47**, 2807 (2012).

[14] J. Rajendran, R. Karri, J. B. Wendt, M. Potkonjak, N. McDonald, G. S. Rose, and B. Wysocki, Nano meets security: Exploring nanoelectronic devices for security applications, Proc. IEEE **103**, 829 (2015).

[15] C. Y. Huang, W. C. Shen, Y. H. Tseng, Y. C. King, and C. J. Lin, A contact-resistive random-access-memory-based true random number generator, IEEE Electron Device Lett. **33**, 1108 (2012).

[16] S. Balatti, S. Ambrogio, Z. Wang, and D. Ielmini, True random number generation by variability of resistive switching in oxide-based devices, IEEE J. Emerging Sel. Top. Circuits Syst. **5**, 214 (2015).

[17] Y. Wang, W. Wen, H. Li, and M. Hu, in *Proceedings of the 25th Edition on Great Lakes Symposium on VLSI (GLSVLSI '15), Pittsburgh, 2015* (Association for Computing Machinery, New York, 2015), p. 271.

[18] M. Hu, Y. Wang, W. Wen, Y. Wang, and H. Li, Leveraging stochastic memristor devices in neuromorphic hardware systems, IEEE J. Emerging Sel. Top. Circuits Syst. **6**, 235 (2016).

[19] E. Piccinini, R. Brunetti, and M. Rudan, Self-heating phase-change memory-array demonstrator for true random number generation, IEEE Trans. Electron Devices **PP**, 1 (2017).

[20] X. Fong, M. C. Chen, and K. Roy, in *Proceedings of the 72nd Device Research Conference, Santa Barbara, 2014* (IEEE, New York, 2014), p. 103.

[21] A. Fukushima, T. Seki, K. Yakushiji, H. Kubota, H. Imamura, S. Yuasa, and K. Ando, Spin dice: A scalable truly random number generator based on spintronics, Appl. Phys. Express **7**, 083001 (2014).

[22] W. H. Choi, Y. Lv, J. Kim, A. Deshpande, G. Kang, J.-P. Wang, and C. H. Kim, in *Proceedings of the 2014 IEEE International Electron Devices Meeting, San Francisco, 2014* (IEEE, New York, 2014), p. 12.5.1.

[23] S. Oosawa, T. Konishi, N. Onizawa, and T. Hanyu, in *2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS), Grenoble, 2015* (IEEE, New York, 2015), p. 1.

[24] M. Barangi, J. S. Chang, and P. Mazumder, Straintronics-based true random number generator for high-speed and energy-limited applications, IEEE Trans. Magn. **52**, 1 (2016).

[25] Y. Kim, X. Fong, and K. Roy, Spin-orbit-torque-based spin-dice: A true random-number generator, IEEE Magn. Lett. **6**, 1 (2015).

[26] A. Sengupta, A. Jaiswal, and K. Roy, in *Proceedings of the 74th Annual Device Research Conference (DRC), Newark, DE, 2016* (IEEE, New York, 2016), p. 1.

[27] 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. **7**, 055934 (2017).

[28] K. Yang, D. Fick, M. B. Henry, Y. Lee, D. Blaauw, and D. Sylvester, in *Proceedings of the 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, 2014* (IEEE, New York, 2014), p. 280.

[29] D. Apalkov, A. Khvalkovskiy, S. Watts, V. Nikitin, X. Tang, D. Lottis, K. Moon, X. Luo, E. Chen, A. Ong, A. Driskill-Smith, and M. Krounbi, Spin-transfer torque magnetic random access memory (STT-MRAM), J. Emerging Technol. Comput. Syst. **9**, 13 (2013).

[30] J. Z. Sun and D. C. Ralph, Magnetoresistance and spin-transfer torque in magnetic tunnel junctions, J. Magn. Magn. Mater. **320**, 1227 (2008).

[31] W. Rippard, R. Heindl, M. Pufall, S. Russek, and A. Kos, Thermal relaxation rates of magnetic nanoparticles in the presence of magnetic fields and spin-transfer effects, Phys. Rev. B **84**, 064439 (2011).

[32] A. Mizrahi, N. Locatelli, R. Lebrun, V. Cros, A. Fukushima, H. Kubota, S. Yuasa, D. Querlioz, and J. Grollier, Controlling the phase locking of stochastic magnetic bits for ultra-low power computation, Sci. Rep. **6**, 30535 (2016).

[33] 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 **62**, 164 (2015).

[34] H. Sato, E. C. I. Enobio, M. Yamanouchi, S. Ikeda, S. Fukami, S. Kanai, F. Matsukura, and H. Ohno, Properties of magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure down to junction diameter of 11 nm, Appl. Phys. Lett. **105**, 062403 (2014).

[35] J. Hayakawa, S. Ikeda, Y. M. Lee, R. Sasaki, T. Meguro, F. Matsukura, H. Takahashi, and H. Ohno, Currrent-induced magnetization switching in MgO barrier based magnetic tunnel junctions with CoFeB/Ru/CoFeB synthetic ferrimagnetic free layer, Jpn. J. Appl. Phys. **45**, L1057 (2006).

[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.8.054045 for extra characterization information, measurements on a second device, and whitening-related mathematical demonstrations.

[37] M. Blum, Independent unbiased coin flips from a correlated biased source—A finite state Markov chain, Combinatorica **6**, 97 (1986).

[38] J. Soto, in *Proceedings of the 22nd National Information Systems Security Conference, Arlington, VA, 1999*, Vol. 10 (NIST, Gaithersburg, MD, 1999), p. 12.

[39] The NIST tests also include a uniformity condition on the distribution of P values among tested sequences [38]. This condition was passed for all tests for the sequences processed by Blum and XOR8.

[40] A. Mizrahi, N. Locatelli, R. Matsumoto, A. Fukushima, H. Kubota, S. Yuasa, V. Cros, J. Kim, J. Grollier, and D. Querlioz, in *Proceedings of the IEEE International Magnetics Conference (INTERMAG 2015), Beijing, 2015* (IEEE, New York, 2015), p. 1.

[41] H. Sato, T. Yamamoto, M. Yamanouchi, S. Ikeda, S. Fukami, K. Kinoshita, F. Matsukura, N. Kasai, and H. Ohno, *Proceedings of the 2013 IEEE International Electron Devices Meeting, Washington, DC, 2013* (IEEE, New York, 2013), p. 3.2.1.

[42] T. M. Maffitt, J. K. DeBrosse, J. Gabric, E. T. Gow, M. C. Lamorey, J. S. Parenteau, D. R. Willmott, M. A. Wood, and W. J. Gallagher, Design considerations for MRAM, IBM J. Res. Dev. **50**, 25 (2006).

[43] W. Zhao, C. Chappert, V. Javerliac, and J. P. Noziere, High speed, high stability and low power sensing amplifier for MTJ/CMOS hybrid logic circuits, IEEE Trans. Magn. **45**, 3784 (2009).

[44] A. Neiman, L. Schimansky-Geier, F. Moss, B. Shulgin, and J. J. Collins, Synchronization of noisy systems by stochastic signals, Phys. Rev. E **60**, 284 (1999).

[45] A. Coninx, P. Bessière, E. Mazer, J. Droulez, R. Laurent, M. A. Aslam, and J. Lobo, *Proceedings of the IEEE International Conference on Rebooting Computing (ICRC 2016), San Diego, 2016* (IEEE, New York, 2016), p. 1.

[46] M. Faix, R. Laurent, P. Bessiere, E. Mazer, and J. Droulez, Design of stochastic machines dedicated to approximate Bayesian inferences, IEEE Trans. Emerging Top. Comput. **PP**, 1 (2016).