From charge to spin and spin to charge: Stochastic magnets for probabilistic switching

Kerem Y. Camsari, Punyashloka Debashis, Vaibhav Ostwal, Ahmed Zeeshan Pervaiz, Tingting Shen, Zhihong Chen, Supriyo Datta and Joerg Appenzeller

***Abstract*—As the rapid pace of Moore's Law has been slowing down, there has been intense activity to "re-invent the transistor". An emerging paradigm is to complement the existing complementary-metal-oxide semiconductor (CMOS) technology with new functionalities, rather than finding a drop-in replacement for it. In this paper, we discuss such a complementary approach that we call Probabilistic Spin Logic (PSL) based on the concept of a probabilistic or p-bit. p-bits fluctuate between 0 and 1 and can be imagined in between deterministic bits that are either 0 or 1 and quantum bits that are a superposition of 0 and 1. Interconnected circuits built out of p-bits (p-circuits) can be broadly useful for Machine Learning and Quantum Computing in the solution of problems that conventional CMOS may not be particularly suited for. While such p-bits can be implemented using standard CMOS technology, we will show that the inherent physics of nanomagnets can naturally provide an energy efficient and scalable p-bit implementation through the use of low-barrier Magnetic Tunnel Junctions (MTJs). In this paper, we provide a general description of p- bits and p-circuits and discuss their applications. We review experimental progress towards constructing p-bits and p-circuits exploiting the inherent stochasticity of nanomagnets, from a physics/device/circuits perspective. In particular, we identify building blocks for "write" and "read" operations that can be used in different combinations to construct functional p-bits and p-circuits. Finally, we discuss prospects and challenges of PSL as an emerging, unconventional computing paradigm for a beyond CMOS era.**

I. Introduction and Broad Overview

WHILE magnetic materials quickly became relevant as storage media in the field of high performance computing [2], semiconducting components have dominated the memory and logic device application space since decades. Scalability and compatibility for integrated chip manufacturability favored dynamic random access memories (DRAMs) over magnetic core memories back in the early 1970s. However, the search for a spintronics-based non-volatile memory continued because of its intrinsic potential advantage for low-power applications. The rise of the Magnetic Tunneling Junction (MTJ) [3], [4] opened the door for such a non-volatile memory and MTJs are today the basic building blocks of magnetic random access memories (MRAMs). MRAM has attracted considerable interest since around 2011 as a non-volatile memory element when it became apparent that Spin-Transfer-Torque (STT) could be employed as a scalable writing mechanism [5]. The latest 256Mb MRAM product from Everspin released in 2016 relies on the STT mechanism in combination with materials which exhibit perpendicular magnetic anisotropy (PMA) that allow for higher integration densities than in-plane magnets. Most recently, Spin-Orbit-Torque (SOT) has been added into the discussion as a means to further improve power consumption by separating the Read and Write path [6] and to create what is called SOT-MRAM.

All authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA.

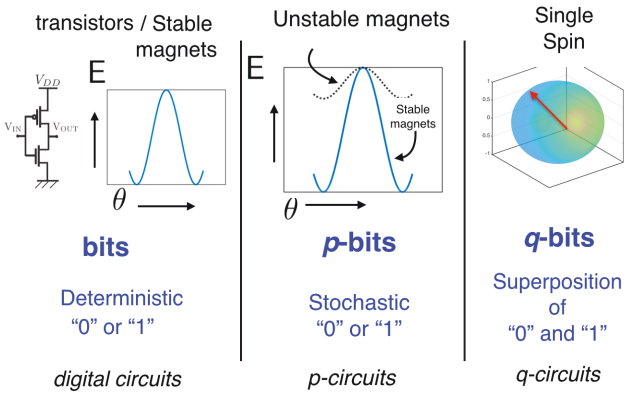

Fig. 1. **Conceptual view of bits, p-bits, q-bits:** Probabilistic or p-bits are conceptually in between deterministic bits and quantum bits or q-bits [1]. Deterministic bits are represented by high (1) / low (0) voltages in transistors, or by stable magnets with two energy minima, while quantum bits are typically represented by the spin of an electron that is a superposition of 0 and 1. In this paper, we show how probabilistic bits can be naturally represented by low barrier nanomagnets whose energy barriers are of the order of the thermal noise.

The above described components: MTJs, PMA materials, and the mechanism of SOT are the central ingredients for the work that will be discussed in this article. The main difference compared to the above however lies in the application space that will be discussed here, namely spintronics-based *computation* rather than *memory*. Memory requires bits represented by *stable magnets* that can store information for long periods of time. We will focus on probabilistic computation requiring probabilistic or p-bits represented by *unstable magnets* (FIG. 1).

About five years ago, as an alternative to the CMOS transistor the so-called Charge-coupled Spin Logic (CSL) [7] was proposed, which became the central effort within the Nanoelectronics Research Initiative (NRI) center Institute for Nanoelectronics Discovery and Exploration (INDEX) [8]. Different from the earlier concept of employing spin currents for everything (All Spin Logic: ASL [9]), CSL was developed based on the idea that interconnects between individual devices could continue using charge currents as in conventional architectures. The active device element provides the charge-to-spin conversion at the write side and the spin-to-charge

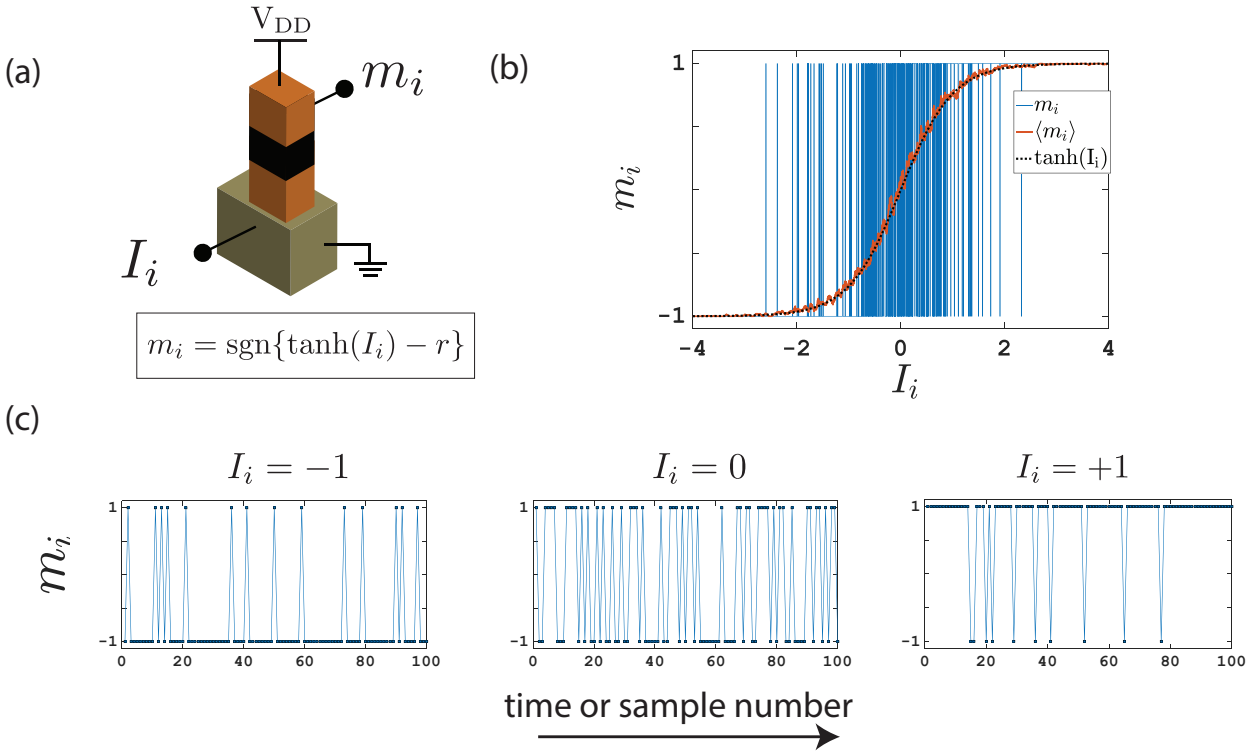

Fig. 2. **p-bit characteristics:** Ideal p-bit device with a dimensionless input I_i and a bipolar output m_i that is always ± 1. r is a random number uniformly distributed between -1 and +1. (b) p-bit output (Solid blue line) as a function of I_i. The average follows a $\tanh(I_i)$ behavior. (c) p-bit outputs at different I_i are shown. In the case of a time-fluctuating p-bit, the x-axis could be time, in the case of an ensemble averaged p-bit the x-axis could be sample number. Reprinted with permission from [1]. Copyright (2017) by the American Physical Society.

conversion at the read side and processes information through the summation of spins in a majority logic approach. Non-volatility is ensured through the use of stable nanomagnets on the read and write side. Combining logic and memory in this way, however, comes at a price of rather high energy consumption as discussed in various benchmarking efforts [10]. The energetic barrier E_B that preserves one or the other magnetic state in the nano-magnet in a non-volatile fashion is also responsible for the higher energy consumption, and reducing E_B can enable more efficient devices that can be a replacement for high performance CMOS circuits [11]. In this article, we will consider an extreme reduction of E_B and make it comparable to the thermal energy kT. Under these conditions noise in the environment can impact the state of the magnet and a new type of computing unit, which we will call a probabilistic or p-bit, becomes possible.

Whether or not emerging spintronics devices can offer a direct replacement of CMOS, for example using low voltage operation [12], is an open question actively investigated at this time [13], [14]. An alternative viewpoint that has been emerging in the last few years is that future electronic devices will involve novel components that work in conjunction with CMOS devices, *augmenting* their functionalities for specialized applications [15]. Indeed, there has been a lot of work in the context of special solvers for computationally hard problems in recent years, leveraging quantum effects, optical devices, digital logic, magnetic and other types of materials [16]–[25]. The objective of this paper is to introduce and report the experimental progress towards such an emerging computing paradigm called Probabilistic Spin Logic (PSL).

II. P-BITS AND P-CIRCUITS

Probabilistic Spin Logic (PSL) is based on the concept of a probabilistic or p-bit whose output fluctuates between 0 and 1, unlike a deterministic bit that is 0 or 1 at any given time. Indeed, the p-bit can be imagined as an intermediate step between bits and quantum bits (FIG. 1). Quantum or q-bits, which lie at the heart of quantum computing, are delicate superpositions of 0 and 1 that need to be realized at cryogenic temperatures. In contrast, the p-bit is a robust, classical object that can be naturally represented by existing devices at room temperature. As such, an intriguing question to ask is could a subset of the applications targeted by q-bits also be part of a p-bit paradigm? In a pioneering article that stimulated the field of quantum computing [26], Feynman made the observation that efficiently simulating quantum mechanics can be done by a quantum mechanical computer. What is relatively less appreciated is that in the same paper, Feynman sets the stage by making a similar point about probabilistic computing that in order to efficiently simulate a probabilistic many-body problem, using a probabilistic computer is natural. The relevance of the latter statement lies in the fact that many computationally difficult problems are *inherently* probabilistic. Consequently, using the inherent stochasticity of low barrier nanomagnets could allow a compact, energy-efficient hardware

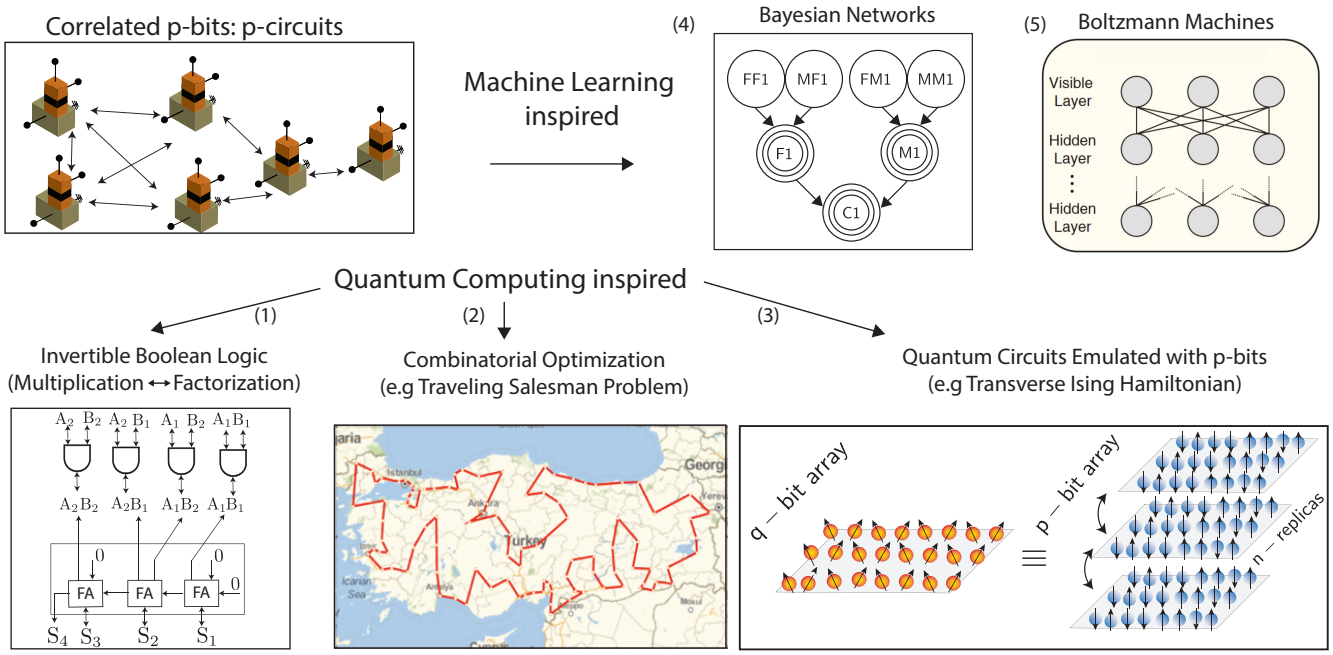

Fig. 3. **Applications of p-bits and p-circuits:** There are two broad applications of correlated p-bits (p-circuits): Applications that are inspired from machine learning and quantum computing as described in the text.

implementation of p-bits to solve these types of problems and is thus highly desirable.

We must of course clearly point out that it is widely believed that any problem that can be efficiently solved by a probabilistic computer can also be efficiently solved by a classical computer equipped with a pseudorandom number generator. In complexity theory, this belief is formally expressed as P=BPP where P and BPP are the classes of problems that can be solved in polynomial time by a classical and probabilistic computer, respectively [**?**]. Further, it is also widely believed that the BPP $\neq$ BQP where BQP is the class of problems that can be solved in polynomial time by a quantum computer, implying quantum computers are more powerful than both classical and probabilistic computers [**?**]. Nevertheless, even though all examples presented in this paper can be simulated in software by classical computers using Eq. 1 and Eq. 2 that we discuss next, the compact implementation of the p-bit can lead to large practical improvements in area and speed in an autonomously operating p-computer [109]. Moreover, some quantum algorithms such as Quantum Annealing can be emulated by a network of p-bits as long as they belong to a special class of systems that are sign-problem free, or stoquastic [57].

An intuitive application of the inherent stochasticity of nanomagnets is true Random Number Generation (RNG). The feasibility of leveraging the stochastic dynamics of stable and unstable nanomagnets for different applications has been thoroughly investigated, both theoretically and experimentally [27]–[35]. Indeed, the p-bit becomes a random number generator when the input is zero. What distinguishes the p-bit, however, is the *tunability* of randomness as a function of an input parameter (FIG. 2). The tunability feature allows p-bits to communicate with one another, distinguishing them from random number generators, to solve computational problems of interest. Computations of such problems have two main ingredients. The first is the tunable randomness functionality as illustrated in FIG. 2. This is sometimes called the "activation function" or the "neuron" in the field of Machine Learning. The second ingredient is the interconnection between p-bits which is sometimes called the "weight matrix" or the "synapse". In the next section, we express these two functionalities mathematically.

A. *Ideal p-bit Behavior*

The ideal PSL equations are described by a neuron-like and a synapse-like equation:

$$m_i = \text{sgn}\{\tanh(I_i) - r\} \quad (1)$$

$$I_i = \sum J_{ij} m_j + h_i \quad (2)$$

where I_i is the dimensionless current that tunes the probability of $+1/-1$ at the output, and r is a random number uniformly distributed between -1 and $+1$. This behavior is illustrated in FIG. 2. Eq. 2 defines the interconnection between p-bits through a weight matrix $[J]$, and h_i is the bias term per p-bit. Expressed in this form, Eq. 1 and Eq. 2 can be viewed as a general stochastic neural network. Indeed, Eq. 1 is sometimes called binary stochastic neuron (BSN) [36].

Eq. 1 and Eq. 2 are essentially the same equations that were proposed by Hinton and colleagues in the context of Boltzmann Machines [**?**], [36]. Boltzmann Machines, sometimes called stochastic Hopfield Networks [37], are typically associated with a powerful learning algorithm useful for unsupervised learning tasks implemented in software. As such,

all examples presented in this paper can be simulated by software models using Eq. 1 and Eq. 2. In the present context, probabilistic or p-bits in are always implied to be hardware building blocks. Hardware p-bits need to be active, *three-terminal* devices that can drive other units according to an interconnection matrix and they constitute a building block for a probabilistic computer that can find use in many areas of interest in Machine Learning and Quantum Computing as we explain in the next subsection (FIG. 3). Indeed, as we will show, probabilistic circuits built out of p-bits do not have to be symmetrical (as all Boltzmann Networks are), in certain cases these directed (Bayesian) networks could be useful for different applications.

B. Broad applications of PSL

Applications with different problems can be encoded to the weights, $[J]$ matrix, so that a dynamical evolution of p-bits finds solutions for applications of interest. Finding a suitable weight matrix for a given problem is not always trivial, though in some cases, for example implementing simple Boolean functions that are described by truth tables, one-shot learning methods exist [1], [38]. For more complicated tasks such as image recognition the proper weights have to be "learned", and this is the main concern of machine learning algorithms. In some cases, the interconnection matrix $[J]$ can heuristically be written down with relative ease, but reaching a solution of interest or more precisely the "ground state" by dynamically evolving the system can become very difficult, requiring special techniques such as annealing and quenching to obtain near ground state solutions [39].

p-circuits can broadly find applications in two disjoint but very active fields of research, namely, Machine Learning and Quantum Computing (FIG. 3). In the context of Machine Learning, hardware p-circuits can enable highly compact and efficient implementations of restricted and unrestricted Boltzmann Machines (BM) as inference networks [40], [41]. While "learning", Boltzmann Machines iterate weights by making use of equilibrium correlations between the "hidden" and "visible" layer of the neural network [36]. It is well-known that obtaining exact equilibrium correlations at each iteration can take a long time in *software* [42]. Thus, constructing *hardware* circuits that naturally evolve toward equilibrium distributions can significantly speed up learning for Boltzmann Machines [43].

Another application in the context of Machine Learning is Bayesian networks, which are directed networks unlike BM's and restricted BMs (RBMs). These networks are typically used to understand causal relationships between probabilistic variables in the real world including applications in forecasting, medical diagnosis and others [45]. Being able to implement such networks electrically through p-bits could accelerate the evaluation of deep correlations as well as being able to adjust variables based on new information [46].

Another application space for p-circuits is concerned with problems that are typically discussed in the context of Adiabatic Quantum Computing (AQC) [47]. This is partly because the algorithms developed for AQC can directly be used by probabilistic circuits operating classically. For example, in the Transverse Ising Hamiltonian a transverse magnetic field is adiabatically turned off to identify the ground state of an Ising Hamiltonian, a process that is called Quantum Annealing (QA) [48]. The resultant Ising Hamiltonian (without the transverse magnetic field) can directly be implemented as a p-circuit to find the ground state of a system that is mapped to, for example, the Traveling Salesman Problem (TSP) [49]. Such a hardware p-circuit can be used to find a feasible solution of a given TSP "instance" by changing the effective "temperature" of the system electrically. The process of slowly changing temperature to guide a system to its ground state is called Classical Annealing (CA). Just like quantum annealers, classical annealers that continuously change the weight matrix can be implemented in hardware exploiting the inherent stochasticity of magnets to solve such Combinatorial Optimization Problems, as shown in panel (2) in FIG. 3 [49]. In addition to changing the effective temperature slowly, other methods, for example rapidly "cooling" (quenching) the effective temperature of a fixed weight network could also be useful [50]. Note that in each case, having electrical control of the temperature parameter allows a wide range of possibilities for different annealing schemes. Indeed, building hardware "Ising Computers" to find the ground state of a problem of interest with or without electrical annealing has received significant attention in recent years [17], [51]–[55].

Intriguingly, it has recently been shown that p-bits can approximate the thermodynamic properties of a class of quantum circuits, that includes the class of Transverse Ising Hamiltonians that quantum annealers are based on [56], by generating a number of replicas of the quantum system [57] (displayed in panel (3) in FIG. 3). This implies that even Quantum Annealing, including the transverse magnetic field can be approximated using classical, room temperature p-bits. In software implementations this approach is sometimes called Simulated Quantum Annealing or SQA. There is emerging evidence on relative benefits of SQA compared to CA [**?**], attributed to the SQA's ability of mimicking "tunneling" out of local minima which is not possible in CA though it has also been noted that there might be observable scaling differences between QA and SQA [**?**]. It also seems possible that other quantum systems (such as ferromagnetic Heisenberg models) can be emulated by p-bits in this manner [57], suggesting the possibility using highly scaled room temperature p-circuits replicating thousands of q-bits to approximate a subset of quantum systems that are sign-problem free.

Another interesting possibility for Boolean logic arises when the network is undirected: Unlike traditional Boolean gates that are directed from input to output, there is no distinction between an output and input terminal in a reciprocal network. This allows constructing Boolean gates that not only "find" outputs corresponding to a set of inputs but also a set of inputs that are consistent with a given output! The interesting feature of invertibility, also exhibited by another unconventional computing framework called "memcomputing" [60], [61], does not have an analog in traditional computing and could lead to hardware solutions for many *inverse* problems, where computations are difficult in the reverse direction compared to the forward direction [1] (panel (1) in FIG. 3).

Factorization is one such example, where multiplying two primes is easy while finding the prime factors of a semiprime is believed to be difficult. Factorizers out of invertible p-circuits have been implemented using CMOS emulators [62] and scalable CMOS chips [63]. In this paper, we will discuss another method of factoring by casting p-circuits as general purpose optimizers in Section V-A.

III. Nanomagnets as natural random number generators

In this section, we will discuss how the probabilistic physics of nanomagnets allows for a straightforward implementation for random number generation (RNG) which will be used to implement the p-bit behavior described by Eq. 1. Arguably, the most natural method of obtaining randomness is a single domain magnet with a low energy barrier that spontaneously fluctuates between two binary states (up or down) in the presence of thermal noise. Such nanomagnets fluctuate in time and their magnetization can be read out as a telegraphic resistance signal as shown in FIG. 4a. Another way of obtaining probabilistic behavior is to use a collection of magnets and exploit their average stochastic switching behavior in an ensemble of magnets. There are different methods to implement stochastic devices in this manner as we will discuss in Section III-B below. Interestingly, even high barrier magnets allow for tunable stochasticity as will be discussed in Section III-C.

A. From magnets to devices: Read and write operations

Even though a fluctuating nanomagnet forms the basis of a p-bit by exploiting the inherent noise in the environment, the p-bit, as a device, needs to have electrical inputs and outputs. Therefore the output of a magnetization needs to be transduced into a voltage or current, which we call the "read" operation. Similarly, the p-bit, as a device, needs to allow manipulation of its state by an electrical input, which we call the "write" operation. As an example, the emerging magnetic memory technology, the spin-transfer-torque magneto random access memory (STT-MRAM) uses a Magnetic Tunnel Junction (MTJ) as its basic building block. An MTJ employs two magnets, a reference (fixed) layer whose magnetization is fixed to a particular orientation and a free layer whose magnetization can switch between two stable states. STT-MRAM employs the spin-transfer-torque to write information to the free layer nanomagnet, by depositing a spin-polarized current that is obtained by running an unpolarized charge current through the fixed layer. This spin-polarized current causes a reversal of magnetization of the free layer, flipping the state of the free layer. Reading the state information is achieved by using the same MTJ since the resistance of the MTJ depends on the relative magnetization orientations of the fixed and free layers. An anti-parallel orientation results in a higher resistance $R_{\rm AP}$ compared to a parallel orientation, $R_{\rm P}$. The difference of the two is characterized by the Tunneling Magnetoresistance ratio $\mathrm{TMR} = (R_{\rm AP} - R_{\rm P})/R_{\rm P}$. For present day MTJs, TMR can be around 100-600% [6], [65]. This change in the resistance can be sensed by running a small electrical current through the MTJ.

In this paper, we will also employ the Anomalous Hall Effect (AHE) as another read mechanism, particularly for magnets with perpendicular anisotropy (PMA). AHE allows the reading of a perpendicular magnet in a Hall bar structure, where a small longitudinal read current gives rise to a transverse voltage proportional to the z-directed magnetization in the absence of an external magnetic field. AHE does not require an additional fixed magnetic layer or a well-engineered tunnel junction, making it a convenient read out mechanism. However the magnetoresistance ratios obtained from AHE, i.e. the relative difference in the read out signals produced by different magnetization states are typically much smaller than those obtained from MTJs [66]. AHE, among other mechanisms, is a building block for reading out magnetic information and the devices we demonstrate here using AHE are proof-of-concept devices that open pathways for ultimate MTJ implementations, or other efficient read mechanisms. In addition, in Section IV-B1, we describe a novel reading mechanism that does not involve magnetization information directly but produces a read out voltage based on the magnetic "easy-axis".

The specifics of the read and write mechanisms depend on the particular p-bit implementation and as we will see in Section IV-B2, sometimes the magnetization state is not modulated at all, and the electrical read / write can be achieved by conventional charge-based devices. Next, we first describe how nanomagnets can be used as natural building blocks for RNGs and then describe how these building blocks can be used to engineer electrical p-bit devices with distinct read and write mechanisms.

B. Low barrier nanomagnets as natural RNGs

A single domain nanomagnet is typically characterized by an energy barrier E_B that separates two states with opposite orientations of the magnetic moment. Once the magnet is placed in one of these states, it stays in this state for a given amount of time that is called the retention time, τ. In a simple picture, the retention time has an exponential dependence on the energy barrier [67]:

$$\tau = \tau_0 \exp(E_B/kT) \tag{3}$$

where kT is the thermal energy and τ_0 is a material dependent constant that is called the "attempt time". The retention time τ can be engineered to ensure non-volatility for a desired period of time, typically 10 years or greater in memory applications [65].

In a monodomain magnet, the energy barrier is a result of the total moment that makes up the nanomagnet and an effective anisotropy field which is a preference of the magnetic moment to point in a certain axis that is called the "easy-axis". The anisotropy field is associated with either the shape or the crystal structure of the nanomagnet and E_B is thus given by the following expression:

$$E_B = \frac{1}{2} M_s \Omega H_K \tag{4}$$

where M_s is the saturation magnetization per volume, and Ω denotes volume. $M_s\Omega$ represents the total magnetic moment

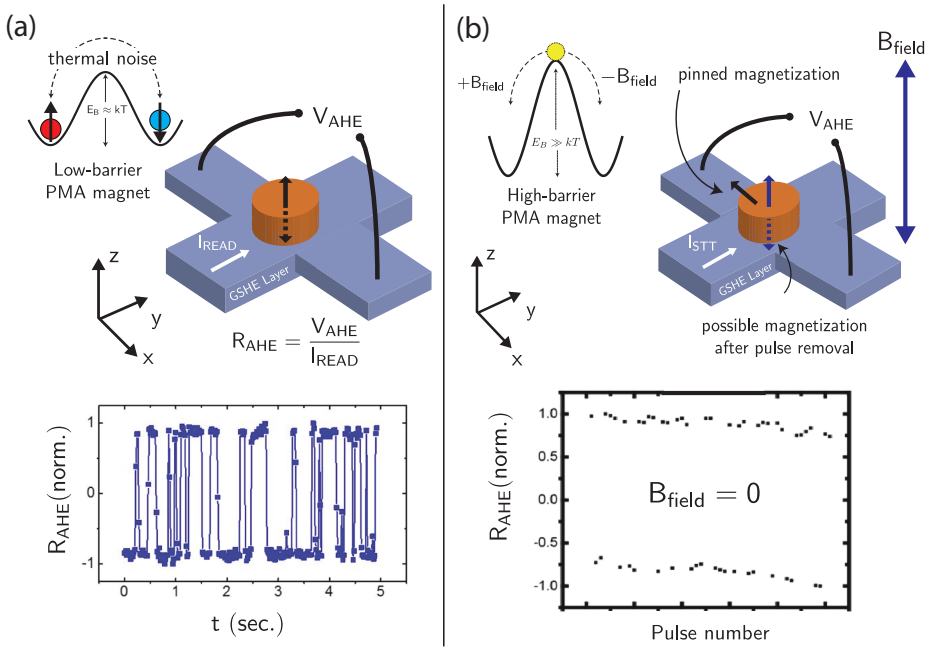

Fig. 4. **Exploiting the inherent stochasticity of nanomagnets**: Two different modes of stochasticity of nanomagnets can be used to generate randomness. (a) The case of a low barrier nanomagnet that is fluctuating in the presence of thermal noise, without any external stimulus. The Hall bar allows the time-varying tracking of a low barrier PMA. (b) A high barrier nanomagnet that is brought into a metastable state and is forced to relax to one of its energy minima. This process is random in the absence of an external input, but could be tuned by a symmetry breaking magnetic field (B_{field}). Reprinted with permission from [64]. Copyright by IEEE (2018).

in the magnetic body, and H_K represents the anisotropy field of the nanomagnet. Using lithographical techniques and different material combinations, both of these factors can be engineered independently to obtain a desired E_B for a particular application [64].

1) Reducing net moment: One straightforward approach to obtain a low E_B is to scale the product $M_s\Omega$, which can be achieved by using a material with low saturation magnetization and/or fabricating a nanomagnet with a small volume by lithographic means and proper processing. Magnets of this type are sometimes called to be in the superparamagnetic regime [68].

2) Reducing anisotropy: Another approach to obtain a small energy barrier is by scaling the anisotropy field H_K. This can be achieved by using a material with negligible crystalline anisotropy and a geometry with negligible shape anisotropy. An example of such a structure would be a circular disk from a CoFeB alloy, that ensures that there is no preferred easy axis. Such circular magnets were first discussed in the context of nanomagnetic logic. In these magnets, the demagnetization field keeps the magnet in the plane of the film. However, due to the lack of a magnetic easy-axis, the magnetization randomly rotates within the plane with fluctuations that can occur on a sub-nanosecond time scale due to the large demagnetization field. It is important to note that to obtain this kind of circular magnets, the nanomagnet needs to behave as a monodomain body, avoiding the formation of a vortex structure [69]. This can be achieved by keeping the ratio of diameter to thickness of the nanomagnet low enough such that the dipolar self-energy is less than the exchange energy [69], [70].

3) Combination of reduced anisotropy and moment: In practice, a combination of these two approaches can be used to achieve the desired low barrier nanomagnet characteristics. For example, a perpendicular magnetic anisotropy (PMA) system presents another way to reduce H_K to obtain a low barrier nanomagnet. In the case of the so-called "interfacial" PMA magnets, the effective uniaxial anisotropy is given by:

$$H_K^{\text{eff}} = K_s/M_s(1/t_{\text{PMA}} - M_s/2\mu_0) \tag{5}$$

where the first term arises due to an interface anisotropy (K_s $[J/m^2]$) and the second term arises from the demagnetization field [71]. The competition between these two opposing fields can be optimized to obtain a very small H_K^{eff} by tuning the thickness of the PMA layer (t_{PMA}). As an example of this approach, we show a Ta/CoFeB/MgO PMA with a thickness of t_{PMA}=1.3 nm. FIG. 4a.

C. High barrier nanomagnets as natural RNGs

The stochastic switching behavior of a collection of high barrier nanomagnets can be exploited to generate randomness in different ways [27], [28], [72]–[77]. One specific approach to obtain tunable stochasticity is to initialize the magnetization m of a *single* nanomagnet to be at the top of the energy "hill" by an external force such as spin-orbit-torque or voltage control, and then repeat measurements on this single device to extract statistics from a *collection* of measurements.

In the case of a PMA magnet, this corresponds to forcing m to lie in the plane, the hard-axis for the PMA magnet ($\pm x$ direction in FIG. 4b). Once the external force is removed, the magnetization relaxes back to one of the two valleys of energy corresponding to the easy-axes of the magnet.

This experiment has also been realized in a Ta/CoFeB/MgO system but the magnetic stack is appropriately designed to have a smaller thickness ($t_{\rm PMA}$=1.0 nm) to ensure strong perpendicular magnetic anisotropy as can be seen in Eq. 5. The writing mechanism is through the so-called "spin-orbit-torque" that arises due to the charge current flowing in the ($\pm y$) direction in heavy metals such as Pt, Ta, W [79], [80]. It is well-known that the giant spin Hall effect (GSHE) of the Ta/CoFeB structure can deterministically switch the PMA magnetization in the presence of a symmetry breaking magnetic field that is typically applied in the direction of a charge current or artificially generated by an additional magnetic layer (see for example, [80]–[83]). In the absence of any such field, and if the current density through the Ta layer is large enough, the SOT exerted on the magnet pulls its magnetization entirely in plane ($\pm x$ direction), collinear with the spin current polarization direction. Once the current pulse is removed, the settled magnetization state is read by means of the anomalous Hall effect (AHE) using the Ta Hall bar underneath the magnetic island [84]. As shown in FIG.4b, at zero external field, when hard axis initialization is performed repeatedly by applying current pulses to the Ta underlayer, the final magnetization state is found to be approximately equally probable in the "up" or the "down" direction. If this repeated pulsing is performed in the presence of an external magnetic field, the magnetization relaxes to one direction more frequently than the other, depending on the field strength, leading to a sigmoidal response of the ensemble average. It is important to note that variations or misalignments in the direction of the magnetic fields could lead to systematic biases in the up/down probability obtained in this manner. These variations would generally cause shifts in the sigmoid making it non-symmetric with respect to the tuning parameter. Such shifts in the sigmoid can be corrected in the synaptic equation through the electrically applied individual bias terms, h_i.

In the next section, we describe a fully electrical device that can create such a symmetry breaking magnetic field.

IV. Implementing p-bits in hardware

So far we have discussed how nanomagnets can be used as natural random number generators. In this section, we discuss device designs with electrical input and output terminals that use such nanomagnets, where different designs will be equipped with different "read" and "write" mechanisms. We will classify hardware implementations of p-bits into two categories: current-controlled or voltage-controlled. Current-controlled p-bits are low input impedance devices, e.g. devices that include for the manipulation (write) of the nanomagnet, a GSHE layer that is used to generate spin-orbit-torque as discussed in the previous section. Voltage-controlled p-bits are high input impedance devices whose "write" operation involves electrical field control, by means of – for example – metal-oxide-semiconductor (MOS) transistors or the magnetoelectric (ME) effect. In traditional electronics, voltage controlled devices, for example MOS transistors compared to bipolar junction transistors, have been more successful since they allow reducing power consumption dramatically. This observation likely applies in the context of hardware p-bits as well, but it is worth taking a look at the circuit aspects of p-bits rather than just this device argument. In fact, when a number of current controlled devices are interconnected through a "synapse" circuit (Eq. 2), current controlled devices can exploit a natural weighted current summation from large resistors to a small sink resistor. This could be a critical advantage for current controlled devices since they would not need additional circuitry, such as transimpedance amplifiers to convert currents to voltages for performing the synaptic operation [85]. It is thus worthwhile to consider both current-controlled and voltage-controlled devices for the time being and next, we will present various p-bit implementations where the input and output terminals of Eq. 1 are represented by the terminal currents and voltages of the device.

A. Current controlled p-bits

1) Oersted ring based p-bit: The structure shown in FIG. 4b can be turned into an all-electrical p-bit device by an elegant design that provides the symmetry breaking magnetic field electrically. This is achieved by constructing a metallic ring around the magnet which generates a $\pm z$ directed magnetic field when a current flows through it, as shown in FIG. 5 [86]. The metallic ring is placed on top of the metallic GSHE layer and electrically isolated from it by means of an SiO_2 layer that is not shown in the figure. FIG. 5b displays the average magnetization as a function of the current that controls the magnetic field through the Oersted ring. The continuous magnetization plot is obtained by averaging the discrete magnetization at different trials, i.e. after repeating the same experiment described in Section III-C multiple times at certain discrete currents through the ring. FIG. 5c-d show the response of a single device as a function of different trials. The response of the device can be mapped to Eq. 1 using appropriate device variables, and this mapping can then be used to build hardware p-circuits, interconnected to implement neural network like functionalities, as we will discuss later in Section V-B.

2) GSHE-based p-bit: A more natural implementation of a p-bit is by the use of low barrier nanomagnets that have their magnetization fluctuate in the presence of thermal noise, as discussed in Section III-B. To realize an all-electrical tunable p-bit, these magnetic fluctuations need to be converted into currents or voltages, in the "read" part of the p-bit. A common method of converting magnetic information to voltages is by the use of Magnetic Tunnel Junctions (MTJ's) as mentioned in Section III-A. FIG. 6 shows a conceptual device that makes use of an MTJ as the "read" part of the p-bit with a free layer that can be manipulated by a current through the GSHE layer underneath, as part of the "write" operation [1]. An important characteristic of this design is that it resembles the basic unit cell of the SOT-MRAM technology but replaces the free layer of the MTJ with a low barrier magnet. At zero input current, the MTJ resistance would ideally be fluctuating with approximately 50/50 probability between a high resistive state, that is defined by an antiparallel arrangement of the magnetic moments of the low barrier magnet and the fixed magnet and a low resistive state, where the two magnetization directions are parallel. Note that in practice the situation of

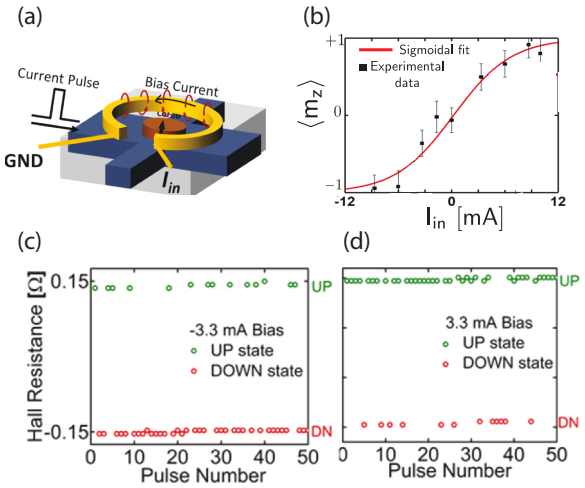

Fig. 5. **An all-electrical p-bit using high barrier nanomagnets** (a) A p-bit with an SOT clock and a symmetry breaking Oersted field generated by the circular ring. (b) Average magnetization at different input currents on the Oersted ring. (c) and (d) shows the magnetization state at different trials at -3.3 mA and +3.3 mA. Figure adapted from [86].

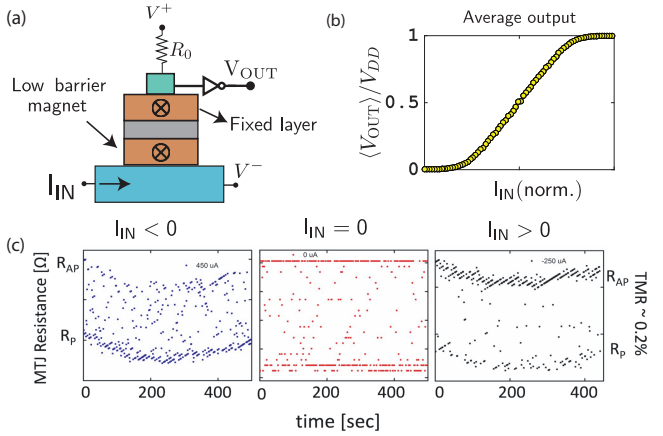

Fig. 6. **An all-electrical p-bit using low barrier nanomagnets** (a) A proposed device using a low barrier nanomagnet as the free layer of an MTJ. The fluctuating read voltage is amplified by an inverter to produce a sigmoidal response at the output as a function of an input current at the SOT layer. (b) The *simulated* time-averaged response of the device. (c) The *experimental* response of the structure (without the inverter) at different input currents. (See [87] for details).

50/50 fluctuations might be achieved at a finite write current on the MTJ due to dipolar pinning fields. This may cause a constant offset to the sigmoid along the x-axis in FIG. 6b, which can be corrected at the synaptic function stage. The fluctuations, i.e. the probability of finding the system in its parallel or anti-parallel state, are then controlled by the spin-orbit-torque generated by a current flowing in the heavy metal layer, tuning the magnetization of the free layer. A small read voltage across a resistive voltage divider converts the resistance fluctuations into a voltage, which is then amplified by an inverter to obtain the electrically controllable tunable randomness which is a key distinguishing feature of the p-bit, as we discussed in Section II (FIG. 6b).

Note that in this case, the low barrier nanomagnet needs to be an in-plane magnet because of the spin-orbit-torque physics. While devices consisting of stable MTJs and GSHE layers have been experimentally reported previously, the spin-orbit-control of an MTJ with a stochastic free layer has not been experimentally demonstrated before, to the best of our knowledge. FIG. 6c shows initial experimental results on such a device, where the resistance fluctuations of the MTJ are plotted as a function of time at different input currents. As expected, a negative input current $I_{\rm IN}$ results in a higher probability to find the system in its low resistive state, while a positive input current $I_{\rm IN}$ results in a higher probability to find the system in its high resistive state, proving experimentally the tunability of the random fluctuations of the nanomagnet. It is worth noting that the above experimental device is by no means optimized or designed to operate at the expected sub-nanosecond switching speeds discussed before. In fact, the effective energy barrier E_B in this prototype device was rather large, between 15-20 kT, allowing the monitoring of the magnetic fluctuations.

Moreover, the device in FIG. 6 exhibited a low tunnel magnetoresistance (TMR), as evident from the small resistance difference between parallel and the anti-parallel magnetic state, which is simply a result of the non-ideal sample fabrication approach used for this prototype demonstration. While there is ample space for improved device designs, the above initial experimental demonstration clearly shows a path forward towards a p-bit implementation that exhibits all the desired features as discussed in the early sections of this article, i.e. in particular the electrically tunable randomness that is at the heart of this device idea. The principal argument to build such a device is the compact implementation of the complex tunable randomness functionality in an all electrical device. Designing the same functionality in standard CMOS would require a large area footprint and external random number generators. The experimental details of this particular structure can be found in [87].

B. *Voltage controlled p-bits*

1) Magnetoelectric p-bit: Voltage control of magnetism has attracted enormous research interest in recent years since it promises low energy dissipation in magnetic logic and memory devices (see, for example [88]–[90]). One way of controlling magnetic properties by electric fields is the use of "multi-ferroic" heterostructures that couple different phenomena. For example, a piezoelectric (PE) material that converts voltages to strain and a magnetostrictive nanomagnet that couples strain to a change in its magnetic properties can be used to control magnetic properties with voltages. Such a structure that combines a CoFeB magnet with a piezoelectric material $Pb(Mg_{1/3}Nb_{2/3})_{0.7}Ti_{0.3}O_3$ (PMN-PT) is shown in FIG. 7a. To maximize the effect, it is desirable to choose a magnet with a large magnetostriction coefficient and a piezeoelectric with a large PE coeffcient. The (011) cut PMN-PT is such an anisotropic material with large piezoelectric coefficients with different signs for d_{31} and d_{32} [91]. Therefore, when an out-of-plane electric field is applied in the $\hat{z}$ direction, this produces an in-plane strain with different signs in the $\hat{x}$ and $\hat{y}$ direction. If the strain on the PMN-PT transfers to

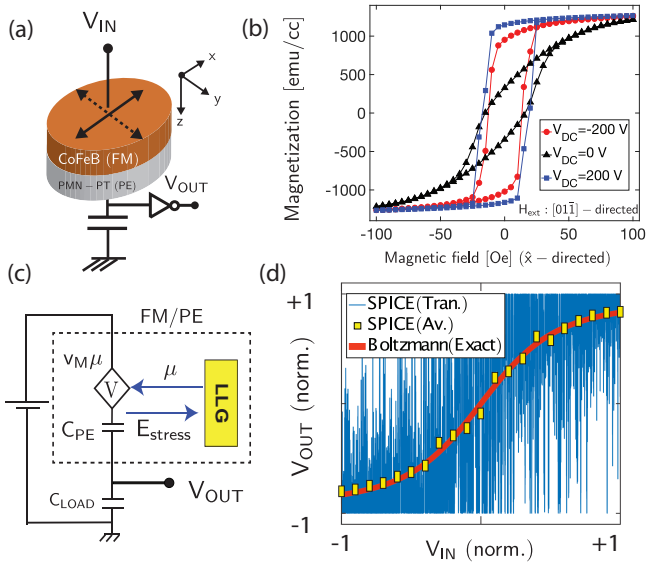

Fig. 7. **Towards a voltage controlled p-bit:** (a) A proposed schematic of a piezoelectric / magneotstrictive p-bit. The fluctuations of a low barrier nanomagnet are turned into voltage fluctuations by an inverse magnetoelectric effect. (b) Experimental measurement of magnetization vs. an applied field under different voltages indicate how the magnetic anisotropy of the magnet is affected by the applied voltage, that causes a strain in the PE, that in turn modulates the magnetic anisotropy of the magnet through magnetostriction [94]. (c) An equivalent circuit for evaluating the magnetoelectric (ME) read and write effects, self-consistently. The charge on the PE capacitor causes a stress-induced change in the magnetic energy (E_{stress}) that enters the magnetization dynamics (LLG equation), while the change in magnetization (μ) causes a change in the capacitor voltage through a dependent source ($v_M\mu$). (d) Simulated response of the device shown in (a) in SPICE. The details of this simulation can be found in [95]. Reprinted Fig. 7 (c) and (d) with permission from [95]. Copyright (2018) by the American Physical Society.

the CoFeB nanomagnet without any loss, applying a vertical electric field stretches or compresses the magnetic film. This strain modulates the easy-axis anisotropy of the magnet. In an exaggerated scenario, the magnet can be imagined to be stretched to become an ellipse with a major (easy) axis in the direction of the solid black arrow and for the other voltage polarity the magnet becomes an ellipse with a major axis in the direction of the dashed black arrow (FIG. 7a). Indeed, the change of magnetic properties of an unpatterned CoFeB film on PMN-PT under various applied electric field conditions is clearly visible in experimental data shown in FIG. 7b. A positive voltage enhances the easy-axis anisotropy of the magnetic field in the $[01\bar{1}]$ direction, making the M-H curve in FIG. 7b more "square-like". Note that due to the ferroelectric nature of the PMN-PT film, the piezoelectric coefficients can also change under large electric fields, which is why the response to voltage is symmetric for $\pm$ 200 V. An important aspect of the magnetoelectric and the inverse magnetoelectric effect is that excluding multi-terminal geometries and complex pulsing schemes [89], [92], the magnetic anisotropy can only be rotated by 90 degrees, due to the time-reversal symmetric nature of the stress term in the magnetic energy. For this reason, deterministic 180 degree switching often employs additional mechanisms, such as spin-transfer-torque to perform deterministic 180 degree switching [93].

In a set of recent experiments [94], [96], it was shown that the *reciprocal* effect of the magnetoelectric effect can be used as a means to "read" the magnetic state. The stress induced modulation of the magnetic anisotropy enters the magnetic energy through a term that is proportional to the magnetic easy axis $(m_x^2 - m_y^2)$ as well as the charge induced in the PE material. For example, for a magnet with a uniaxial anisotropy the magnetic energy can be written as:

$$E = H_K(1 - m_x^2) + Qv_M(\mu) \tag{6}$$

where μ is the pseudo-magnetization $m_x^2 - m_y^2$, H_K is the anisotropy energy, $\pm x$ is the easy-axis, Q is the charge that is developed on the PE material due to an externally applied electric field and v_M is the magnetoelectric coefficient that is a combination of the piezoelectric and magnetostrictive parameters. The magnetic field modulation due to an applied voltage that enters the magnetization dynamics is obtained from $-\nabla_m E/M_s\Omega$ where ∇_m is a gradient with respect to the magnetization. It is clear from Eq. 6 that in addition to a change in magnetic anisotropy due to an applied charge Q, a change in μ reciprocally causes a change in Q, since $\partial E/\partial\mu = Qv_M$ and this change can be detected electrically, as shown in [94], [96]. This reciprocal mechanism provides an opportunity of reading out the easy-axis information of the magnet without the use of secondary layers or magnetic tunnel junctions. As such, the inverse magnetoelectric could be an additional building block that can be used in combination with other "write" mechanisms to build functional devices.

FIG. 7c illustrates the self-consistent circuit model that captures the direct and the reciprocal effect, and FIG. 7d displays the simulated response of a circular low barrier nanomagnet without a preferred easy-axis. In the presence of noise, the magnetization is changing stochastically in the plane of the magnet, and these fluctuations are converted into a fluctuating voltage at the output, that can be amplified by an inverter. Applying an input voltage can pin the pseudo-magnetization to one axis or the other, which provides the required tunability. This example illustrates how the physics of low barrier nanomagnets can be creatively combined with the physics of the magnetoelectric effect that can become a compact, low-power alternative p-bit using the direct and inverse magnetoelectric effects without the use of tunnel junctions.

2) Embedded MRAM-based p-bit: Another highly compact voltage controlled p-bit design is shown in FIG. 8a. The voltage control of this design stems from using an n-type metal oxide semiconductor (NMOS) transistor in conjunction with an unstable MTJ [97]. As apparent from FIG. 8, this design again allows to create a tunable random number generator similar to the cases discussed in FIG. 2, FIG. 5b, FIG. 6b, and FIG. 7d. However, different from all the other p-bit designs discussed above, here the NMOS transistor defines the device operation to a large extent. The MTJ is simply used to create a fluctuating resistor. The device operates by making the NMOS resistance much larger or much smaller than the fluctuating MTJ resistance to provide tunability.

FIG. 8a shows the 1T/1MTJ circuit diagram of this p-bit, where the MTJ resistance changes stochastically in the presence of thermal noise, and the NMOS resistance is controlled

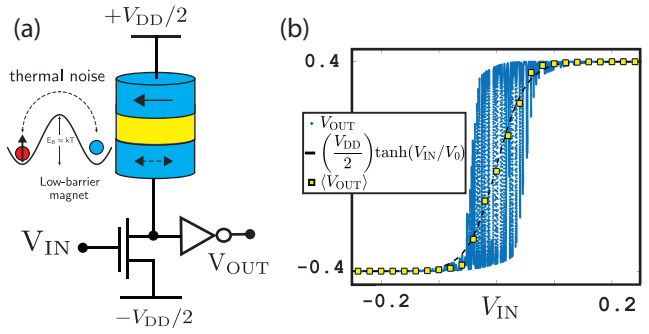

Fig. 8. **Embedded MRAM based p-bit:** (a) A compact implementation of a low barrier nanomagnet (LBM)-based p-bit, where the basic structure of the p-bit slightly modifies the 1T/1MTJ cell of the emerging STT-MRAM technology by making the free layer of the MTJ an LBM. (b) Shows the instantaneous and the long time average of the output at different input voltages simulated in SPICE, by combining transistor models with magnetization dynamics. Reprinted with permission from [97]. Copyright by IEEE (2017).

by the input voltage. In fact, as we will show in FIG. 9b, this circuit design can be emulated by discrete electronic components. FIG. 8b shows the input-output characteristics of this p-bit when the input voltage is swept slowly from $-\mathrm{V_{DD}}/2$ to $+\mathrm{V_{DD}}/2$. The time-averaged output voltage approximately shows the desired tunable randomness, although the response is not rail-to-rail for all input voltages in the stochastic window. Due to the highly noise tolerant nature of probabilistic circuits, this does not seem to cause any significant problem as shown by us through detailed p-circuit simulations for different applications [1], [62].

C. Comparisons of different p-bit realizations

In this section, we have shown and demonstrated different conceptions of a three-terminal, *tunable* RNG. We have considered current-controlled and voltage-controlled designs. In the current-controlled designs both high barrier and low barrier nanomagnets were used to obtain the p-bit functionality. In the case of high barrier magnets, an attractive feature is the use of technologically mature, stable nanomagnets whose variations are relatively low and under control. Also, the hard axis initialization is robust against device to device variations in the critical current density [64], [86]. The physics of such stable nanomagnets also have been comprehensively understood and analyzed since almost all magnetic applications use non-volatile stable magnets. Even though the fluctuations of low barrier nanomagnets have received some attention [98], the theory is by no means at the level of stable magnets. For example, Brown in his seminal paper that describes thermal fluctuations dismisses the low barrier limit as "the case of least interest" [99]. Recent interest in low barrier nanomagnets has led to theoretical developments (See for example, [100]–[103]). One surprising result has been that circular in-plane magnets can fluctuate with sub-ns rates due to a precessional-like fluctuation mechanism [101] and they can be very hard to pin with spin-currents, making them robust against read-disturb issues [102]. Some of these read-disturb issues in the context of a GSHE-driven, low barrier nanomagnet based p-bit [1] are discussed in [104]. These combined features make circular in-plane magnets ideal candidates for the 1T/1MTJ design shown in Section. IV-B2, though it has been pointed out that they may be prone to an increased level of variations [**?**], [103]. Numerical results suggest that when energy barriers of such low barrier nanomagnets are comparable or less than about a few kT, fluctuation rates only weakly depend on the size of the energy barrier [101]. Among alternatives, the 1T/1MTJ design seems to be the most technologically mature, given that embedded MRAM is now close to commercialization by several foundries at this time [105]–[107].

Another contrast between high and low barrier magnets is related to biasing and clocking circuitry. The use of high barrier nanomagnets requires each p-bit to be updated by additional control circuitry that makes the operation of a p-circuit non-autonomous or sequenced. On the other hand, the use of low-barrier nanomagnets removes the need for such sequencing and control circuitry because each p-bit can flip *autonomously*. For symmetric networks, the update order does not matter as long as the synaptic roundtrip time is faster than the fastest p-bit in the system, ensuring the network converges to the expected Boltzmann distribution for a given weight matrix [108]. A sequencerless, autonomous operation is an attractive feature of using low-barrier nanomagnets [109].

Finally, even though the purely voltage-controlled p-bit design discussed in FIG. 7 has not yet been experimentally demonstrated, it is worth analyzing the prospects of similar capacitively driven designs that exploit the voltage control of magnetism (VCM). This can lead to extremely energy efficient building blocks and is an active area in the context of emerging beyond-CMOS devices [12]–[14]. An example of a probabilistic bit that makes use of such a voltage-controlled mechanism has been proposed recently in [44].

V. Implementing p-circuits in hardware

In previous sections, we outlined different hardware implementations of individual p-bits making use of the stochastic behavior of nanomagnets. While a single p-bit can be used as a true tunable random number generator for different applications, interconnected p-bits can implement a much bigger set of problems as we discussed in Section II-B. In this section, we will discuss experiments involving more than one p-bit that are interconnected through a synaptic network and show how applications inspired by Quantum Computing and Machine Learning can be targeted with p-circuits.

Before we show specific examples, a few general comments regarding the *dynamical* behavior of p-circuits are in order. The behavior of Eq. 1-2 is such that any p-circuit will evolve towards the low-energy states of a given system specified by the eigenstates of the $[J]$ matrix. For symmetrically connected networks, Eq. 2 can be derived from a classical cost function, E, such that $I_i = -\partial E/\partial m_i$. For linear synapses that are described by Eq. 2, the cost function becomes a general, Ising-like Hamiltonian $E(\mathbf{m}) = -\sum_{i,j} J_{ij} m_i m_j$. This means that when p-circuits evolve in time, individual p-bits are ultimately distributed according to a Boltzmann distribution described by $\rho(\mathbf{m}) \propto \exp[-\beta E(\mathbf{m})]$ where β is an effective inverse-temperature. The number of steps that is required for the network to reach Boltzmann distribution,

namely the "mixing time", depends on the network topology and the effective temperature, with no generic answers [110], [111]. What a hardware accelerator can do is to significantly speed up the time for a step. For the hardware p-circuit to follow the Boltzmann distribution, however, Eq. 2 needs to be computed faster than the fastest p-bit in the system [108]. This is a reasonable restriction since typical delays due to resistive/capacitive crossbars or CMOS circuits are of the order of picoseconds while magnetic fluctuations are of the order of nanoseconds or at best, hundreds of picoseconds. Note that as long as the synapse is fast, the order of updates in a symmetric network does not matter (provided that updates are not simultaneous as in "Gibbs sampling" [1], [112]), and this allows an autonomous, clockless operation of p-circuits, greatly simplifying their design. This feature also makes them robust to variations because p-bits can have varying fluctuation rates, as long as they are not too fast compared to synapse delays.

A. *Factorization with p-circuits*

FIG. 9a displays a generic structure of a neural network where 4 asynchronously running p-bits are interconnected by a weight logic unit that performs the synaptic function for a given problem. In this example, a p-bit design that emulates the 1T/1MTJ design discussed in Section IV-B2 is implemented. The design uses a stochastic MTJ emulator with a 2 by 1 analog multiplexer (MUX) to choose a high (R_{AP}) or low (R_P) resistance connected to a supply voltage. The "select" signal of the multiplexer is generated by a microcontroller that produces a telegraphic bitstream with equal up and down probability to choose R_{AP} or R_P. In addition, the select signal is also programmed to simulate MTJ "pinning", that is, when a large current flows through the circuit when the transistor is on, the microcontroller chooses a select signal to favor one of the resistive paths, rather than producing a bitstream with equal up and down probability. The circuit shown in FIG. 9b represents a low-level, compact emulation of the 1T/1MTJ p-bit and is used to build an asynchronous 4 p-bit network that is constructed to perform integer factorization. A separate Arduino microcontroller is used to read the instantaneous p-bit voltages and to produce a voltage that is routed back to the p-bits as their respective inputs. Since this voltage (according to Eq. 2) is generally analog, a Digital to Analog Converter (DAC) is used in conjunction with the Arduino.

Factoring integers can be achieved by different weights using p-circuits. For example, in the past we have shown that factorizers can be obtained by constructing binary multipliers and then use the invertibility feature of p-circuits to factor numbers [1], [108]. Here, we demonstrate another method to factor integers by casting factoring as an optimization problem [113]–[115]. The idea is to construct an energy function that optimizes a function like $E = (XY - Z)^2$ where X and Y are the prime factors of a semiprime Z. Once the energy function is identified, it can be used to obtain input currents to p-bits by computing the change in total energy as a function of magnetization $(-\partial E/\partial m_i)$ for each p-bit. Note that this method requires a different synapse function for every given semiprime, while a generic multiplier operating in reverse can

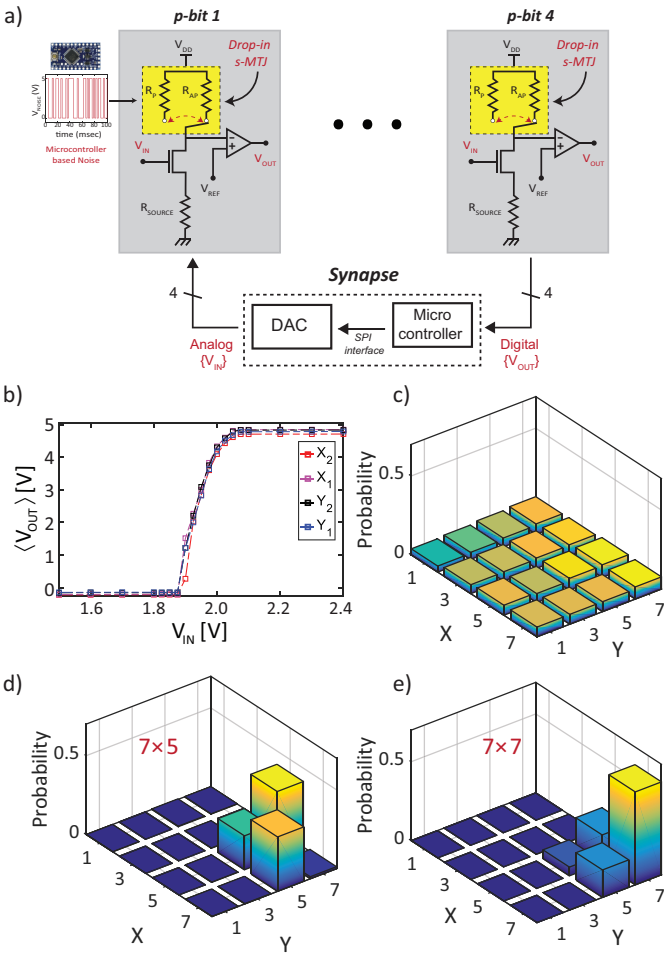

Fig. 9. **p-circuits as general purpose optimizers:** (a) General p-circuit architecture where a network of p-bits are interconnected by a CMOS-based weight logic. and a conceptual circuit level emulation of the 1T/1MTJ p-bit described in FIG. 8a. As a proof-of-concept p-bit, this design is implemented using a 2 by 1 MUX that receives a telegraphic random signal from a microcontroller to randomly select a resistance of R_{AP} and R_P. (b) Experimental average output of emulated p-bit. (c) Experimental results obtained from the autonomous 4 p-bit network for the *uncorrrelated* case where each p-bit is fluctuating randomly. (d-e) Experimental results for the *correlated case* to factor number 35 = 5 × 7 = 7 × 5, and 49 = 7 × 7.

factor any number that it can multiply. If the weight logic is easily programmable as in FIG. 9a, factoring different numbers can be achieved by reprogramming the synapses. In FIG. 9b-c, two examples, i.e. factoring 35 and 49 are shown that were obtained by the MUX-based p-circuit. The MUX based circuit is a stand in for actual MTJs, indeed an 8 bit p- circuit using MTJs has recently been implemented in hardware for the problem of factorization [116]. Factoring by optimizing a cost function serves as one example of a broad class of optimization problems that can be solved by p-circuits in hardware.

An important feature of the cost function implemented here is its non-linearity. Typically, the energy function involves 2-body interactions involving terms such as $E = \sum_{ij} m_i m_j$ that commonly arise in Ising type Hamiltonians [117]. These 2-body interactions produce *linear* synapses since the dimensionless input that is computed from $(-\partial E/\partial m)$ produces linear equations that can be written as matrix products, $\{I\} = [J]\{m\}$, as in Eq. 2. However, in general, the cost function

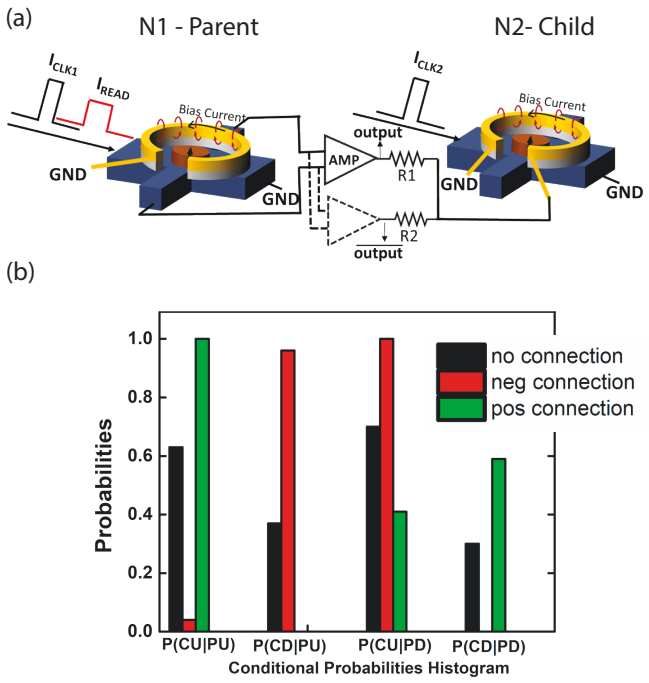

Fig. 10. **Hardware Bayesian circuits with p-bits:** (a) Experimental implementation of a 2 p-bit network to implement a directed p-circuit using the p-bit described in FIG. 5. (b) Experimentally obtained conditional probability histogram of device N2 (Child) when device N1 (Parent) is up or down, for example, CU/PU denotes the probability of Child node being up when the Parent node is set to the up state. Three cases are given corresponding to no connection, positive connection and negative connection as defined in the text (Adapted from [86]).

could include higher order interactions (such as $m_i m_j m_k$) that can lead to non-linear synapses with more functionality [118]. Such non-linear synapses do not seem to be easily implementable by traditional crossbar arrays but could be realized by high level CMOS-based devices.

Solving the factorization problem, unlike many other optimization problems, requires to reach the absolute ground state of the problem rendering even very close approximations useless [114]. In general, however, finding approximate solutions to optimization problems are of great interest and our general objective here is to show how autonomously operating p-circuits can serve as energy-efficient hardware accelerators for this class of problems.

B. Bayesian Network with magnetic p-circuits

As a final experimental example, in FIG. 10 we present an all-electrical implementation of a 2 p-bit Bayesian circuit using the high-barrier clocked p-bit introduced in FIG. 5. Bayesian networks are commonly used to process causal relations between random variables, where a parent node controls a child node, based on a prior probability distribution. A Conditional Probability Table (CPT) describes the causal, probabilistic relationship.

For networks involving multiple layers with many "parent" nodes controlling multiple "child" nodes, the calculation of causal relationships between each node can become computationally intractable using analytical approaches like Bayes theorem [119], [120]. Therefore, building and utilizing hardware Bayesian networks can help accelerate these functions.

As a simple example of how hardware Bayesian networks can be constructed, we use 2 of the p-bits that were introduced in FIG. 5 to demonstrate a conditional probability network as shown in FIG. 10. Spin-orbit torque (SOT) clocks are applied for a sequential hard-axis initialization first starting with device N1 (Parent) similar to the operation of the individual p-bit described in Sections III-C and IV-A1. Next, the same initialization approach is applied to device N2 (Child). As before, magnetization states are read using Anomalous Hall Effect (AHE) with a small read current. The AHE voltages are amplified to $\pm V_{\rm DD}$ using a discrete operational amplifier (OP-AMP), in order to drive the next stage. The synaptic function is implemented by using discrete resistors R1 and R2, such that during the hard-axis initialization of N2, the magnetization state information of N1 is provided as an input to the Oersted ring of N1 using an OP-AMP through resistances R1 and R2. A positive connection corresponds to R1$=$ 810 Ω and R2 is open-circuited (R2 $\rightarrow \infty$) and a negative connection corresponds to the case where R1 is open-circuited and R2 $=$ 810 Ω, while no connection corresponds to both R1 and R2 being open-circuited. FIG. 10 shows correlations between the child and the parent where the probability of child being UP (DN) given parent is UP (DN) is high for a positive connection. For a negative connection, on the other hand, the conditional probability of child being UP (DN) given parent is DN (UP) is high. Note that the CPT is uniquely determined by the choice of resistors and can be adjusted to implement other CPT's [46]. Details on the operation of this particular two p-bit circuit can be found in reference [86].

VI. Prospects and Challenges

We have discussed several alternatives of building hardware p-bits and p-circuits by exploiting the inherently stochastic nature of nanomagnets. It is important to note that fully digital implementations of Eq. 1 and Eq. 2 are possible and p-circuits using 50 [108] to 500 [121] p-bits have been demonstrated using standard CMOS units such as microcontrollers and FPGAs unlike the nanodevice implementations discussed in this paper. A key finding is that the inherent stochasticity of nanomagnet based p-bits can enable highly compact and efficient hardware implementations of p-circuits that would consume a much larger area and power in corresponding CMOS implementations [40]. Random number generation is typically achieved using standard CMOS devices in Linear Feedback Shift Register (LFSR) circuits that consume a large amount of area and provide pseudo random numbers, as opposed to true random numbers that are generated by the stochasticity of nanomagnets. Detailed Energy-Delay evaluations of some of the p-bits discussed here support this view [102].

While p-bits lend themselves naturally to implementations of Binary Stochastic Neurons (BSNs), we have not discussed the synaptic operation in detail, except briefly illustrating CMOS synapses in Section V-A. Using emerging nanodevices replacing CMOS synapses, such as memristors or novel nan-

odevices to accelerate the synaptic function involves a rich literature (see for example, [85], [122]–[124]).

At this time, there are many alternatives to implement synapses, for example resistive networks to implement p-circuits of the type we discussed in FIG. 10 or CMOS synapses of the type discussed in FIG. 9. As mentioned in the beginning of Section V, an asynchronous implementation of p-circuits, without any global clocks and sequencers, requires the synaptic operation to be carried out much faster than the activation function. To address this, in the future, memristive or capacitive networks could be integrated on to the p-bit activation directly [125], somewhat similar to the True North architecture [126], to speed up the synaptic function. In a recent FPGA-based demonstration [109], it has been argued that an autonomously functioning probabilistic computer based on an embedded MRAM design could achieve 10^{15} flips per second that could be useful for solving difficult optimization problems.

There are many challenges ahead for p-circuits to become efficient high-level accelerators for a broad class of probabilistic functions. Overcoming variations at the individual device level, realizing better TMR ratios to improve power dissipation and the integration of synapses with local p-bits are tasks ahead, to just name a few. On the algorithmic end, it is imperative to find weight matrices that are sparse and discrete so that the requirements on the synaptic functions are relaxed. An intriguing near-term direction of p-circuits is the use of embedded MRAM technology, perhaps integrated with a fast synaptic function such as a GPU that implements a $>$ 1000 p-bit network. Such a probabilistic circuit could become a highly efficient "p-computer" for applications in the active fields of Quantum Computing and Machine Learning.

Acknowledgement

This work was supported in part by the Center for Probabilistic Spin Logic for Low-Energy Boolean and Non-Boolean Computing (CAPSL), one of the Nanoelectronic Computing Research (nCORE) Centers as task 2759.003 2759.004, and 2759.005, a Semiconductor Research Corporation (SRC) program sponsored by the NSF through CCF 1739635.

References

[1] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, "Stochastic p-bits for invertible logic," *Phys. Rev. X*, vol. 7, p. 031014, Jul 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.7.031014

[2] A. Moser, K. Takano, D. T. Margulies, M. Albrecht, Y. Sonobe, Y. Ikeda, S. Sun, and E. E. Fullerton, "Magnetic recording: advancing into the future," *Journal of Physics D: Applied Physics*, vol. 35, no. 19, p. R157, 2002.

[3] T. Miyazaki and N. Tezuka, "Giant magnetic tunneling effect in fe/al2o3/fe junction," *Journal of Magnetism and Magnetic Materials*, vol. 139, no. 3, pp. L231–L234, 1995.

[4] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions," *Physical review letters*, vol. 74, no. 16, p. 3273, 1995.

[5] M. Gajek, J. Nowak, J. Sun, P. Trouilloud, E. O'sullivan, D. Abraham, M. Gaidis, G. Hu, S. Brown, Y. Zhu *et al.*, "Spin torque switching of 20 nm magnetic tunnel junctions with perpendicular anisotropy," *Applied Physics Letters*, vol. 100, no. 13, p. 132408, 2012.

[6] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. Piramanayagam, "Spintronics based random access memory: a review," *Materials Today*, vol. 20, no. 9, pp. 530–548, 2017. [Online]. Available: https://doi.org/10.1016/j.mattod.2017.07.007

[7] S. Datta, S. Salahuddin, and B. Behin-Aein, "Non-volatile spin switch for boolean and non-boolean logic," *Applied Physics Letters*, vol. 101, no. 25, p. 252411, 2012.

[8] [Online]. Available: https://www.src.org/program/nri/index/

[9] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature nanotechnology*, vol. 5, no. 4, p. 266, 2010.

[10] C. Pan and A. Naeemi, "An expanded benchmarking of beyond-cmos devices based on boolean and neuromorphic representative circuits," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 3, pp. 101–110, 2017.

[11] J. Grollier, D. Querlioz, and M. D. Stiles, "Spintronic nanodevices for bioinspired computing," *Proceedings of the IEEE*, vol. 104, no. 10, pp. 2024–2039, 2016.

[12] S. Manipatruni, D. E. Nikonov, and I. A. Young, "Beyond cmos computing with spin and polarization," *Nature Physics*, vol. 14, no. 4, p. 338, 2018.

[13] S. Manipatruni, D. E. Nikonov, C.-C. Lin, T. A. Gosavi, H. Liu, B. Prasad, Y.-L. Huang, E. Bonturim, R. Ramesh, and I. A. Young, "Scalable energy-efficient magnetoelectric spin–orbit logic," *Nature*, vol. 565, no. 7737, p. 35, 2019.

[14] R. Ramesh, S. Manipatruni, and I. Young, "Electric-field control of magnetism," *MRS Bulletin*, vol. 44, no. 4, pp. 288–294, 2019.

[15] I. A. Young and D. E. Nikonov, "Principles and trends in quantum nano-electronics and nano-magnetics for beyond-cmos computing," in *Solid-State Device Research Conference (ESSDERC), 2017 47th European*. IEEE, 2017, pp. 1–5.

[16] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, "Evidence for quantum annealing with more than one hundred qubits," *Nature Physics*, vol. 10, no. 3, pp. 218–224, Mar. 2014.

[17] M. Yamaoka, C. Yoshimura, M. Hayashi, T. Okuyama, H. Aoki, and H. Mizuno, "A 20k-Spin Ising Chip to Solve Combinatorial Optimization Problems With CMOS Annealing," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 1, pp. 303–309, Jan. 2016.

[18] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, "A fully-programmable 100-spin coherent Ising machine with all-to-all connections," *Science*, p. aah5178, Oct. 2016.

[19] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K.-i. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, "A coherent Ising machine for 2000-node optimization problems," *Science*, vol. 354, no. 6312, pp. 603–606, Nov. 2016.

[20] T. Okuyama, M. Hayashi, and M. Yamaoka, "An Ising Computer Based on Simulated Quantum Annealing by Path Integral Monte Carlo Method," in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, Nov. 2017, pp. 1–6.

[21] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, "Intrinsic optimization using stochastic nanomagnets," *Scientific Reports*, vol. 7, p. 44370, Mar. 2017.

[22] R. Hamerly, T. Inagaki, P. L. McMahon, D. Venturelli, A. Marandi, T. Onodera, E. Ng, C. Langrock, K. Inaba, T. Honjo, K. Enbutsu, T. Umeki, R. Kasahara, S. Utsunomiya, S. Kako, K.-i. Kawarabayashi, R. L. Byer, M. M. Fejer, H. Mabuchi, D. Englund, E. Rieffel, H. Takesue, and Y. Yamamoto, "Experimental investigation of performance differences between coherent Ising machines and a quantum annealer," *Science Advances*, vol. 5, no. 5, p. eaau0823, May 2019.

[23] T. Wang and J. Roychowdhury, "Oscillator-based Ising Machine," *arXiv:1709.08102 [physics]*, Sep. 2017, arXiv: 1709.08102.

[24] ——, "OIM: Oscillator-Based Ising Machines for Solving Combinatorial Optimisation Problems," in *Unconventional Computation and Natural Computation*, ser. Lecture Notes in Computer Science, I. McQuillan and S. Seki, Eds. Springer International Publishing, 2019, pp. 232–256.

[25] A. Raychowdhury, A. Parihar, G. H. Smith, V. Narayanan, G. Csaba, M. Jerry, W. Porod, and S. Datta, "Computing With Networks of Oscillatory Dynamical Systems," *Proceedings of the IEEE*, vol. 107, no. 1, pp. 73–89, Jan. 2019.

[26] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, no. 6-7, pp. 467–488, Jun. 1982.

[27] 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," *Applied Physics Express*, vol. 7, no. 8, p. 083001, 2014.

[28] W. H. Choi, Y. Lv, J. Kim, A. Deshpande, G. Kang, J.-P. Wang, and C. H. Kim, "A magnetic tunnel junction based true random number generator with conditional perturb and real-time output probability tracking," in *Electron Devices Meeting (IEDM), 2014 IEEE International*. IEEE, 2014, pp. 12–5.

[29] Y. Kim, X. Fong, and K. Roy, "Spin-orbit-torque-based spin-dice: A true random-number generator," *IEEE Magnetics Letters*, vol. 6, pp. 1–4, 2015.

[30] 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 Advances*, vol. 7, no. 5, p. 055934, 2017.

[31] B. Parks, M. Bapna, J. Igbokwe, H. Almasi, W. Wang, and S. A. Majetich, "Superparamagnetic perpendicular magnetic tunnel junctions for true random number generators," *AIP Advances*, vol. 8, no. 5, p. 055903, 2018. [Online]. Available: https://doi.org/10.1063/1.5006422

[32] D. Vodenicarevic, N. Locatelli, A. Mizrahi, J. Friedman, A. Vincent, M. Romera, A. Fukushima, K. Yakushiji, H. Kubota, S. Yuasa, S. Tiwari, J. Grollier, and D. Querlioz, "Low-Energy Truly Random Number Generation with Superparamagnetic Tunnel Junctions for Unconventional Computing," *Physical Review Applied*, vol. 8, no. 5, p. 054045, Nov. 2017.

[33] D. Vodenicarevic, N. Locatelli, A. Mizrahi, T. Hirtzlin, J. S. Friedman, J. Grollier, and D. Querlioz, "Circuit-Level Evaluation of the Generation of Truly Random Bits with Superparamagnetic Tunnel Junctions," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–4.

[34] P. Debashis and Z. Chen, "Tunable random number generation using single superparamagnet with perpendicular magnetic anisotropy," in *2018 76th Device Research Conference (DRC)*. IEEE, 2018, pp. 1–2.

[35] A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yuasa, J. Grollier, and D. Querlioz, "Neural-like computing with populations of superparamagnetic basis functions," *Nature communications*, vol. 9, no. 1, p. 1533, 2018. [Online]. Available: http://doi.org/10.1038/s41467-018-03963-w

[36] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A Learning Algorithm for Boltzmann Machines*," *Cognitive Science*, vol. 9, no. 1, pp. 147–169, Jan. 1985.

[37] J. J. Hopfield, "Hopfield network," *Scholarpedia*, vol. 2, no. 5, p. 1977, 2007.

[38] D. J. Amit and D. J. Amit, *Modeling brain function: The world of attractor neural networks*. Cambridge university press, 1992. [Online]. Available: https://doi.org/10.1016/0166-2236(90)90155-4

[39] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[40] R. Zand, K. Y. Camsari, S. D. Pyle, I. Ahmed, C. H. Kim, and R. F. DeMara, "Low-energy deep belief networks using intrinsic sigmoidal spintronic-based probabilistic neurons," in *Proceedings of the 2018 on Great Lakes Symposium on VLSI*. ACM, 2018, pp. 15–20. [Online]. Available: https://doi.org/10.1145/3194554.3194558

[41] R. Zand, K. Y. Camsari, S. Datta, and R. F. DeMara, "Composable probabilistic inference networks using mram-based stochastic neurons," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 15, no. 2, p. 17, 2019.

[42] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted boltzmann machines for collaborative filtering," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 791–798.

[43] R. Faria, J. Kaiser, O. Hassan, K. Y. Camsari, and S. Datta, "Accelerating machine learning using stochastic embedded mtj," 2018, unpublished.

[44] S. Nasrin, J. L. Drobitch, S. Bandyopadhyay, and A. R. Trivedi, "Low power restricted boltzmann machine using mixed-mode magneto-tunneling junctions," *IEEE Electron Device Letters*, vol. 40, no. 2, pp. 345–348, 2019.

[45] D. Heckerman, A. Mamdani, and M. P. Wellman, "Real-world applications of bayesian networks," *Communications of the ACM*, vol. 38, no. 3, pp. 24–26, 1995.

[46] R. Faria, K. Y. Camsari, and S. Datta, "Implementing bayesian networks with embedded stochastic mram," *AIP Advances*, vol. 8, no. 4, p. 045101, 2018.

[47] T. Albash and D. A. Lidar, "Adiabatic quantum computation," *Rev. Mod. Phys.*, vol. 90, p. 015002, Jan 2018.

[48] R. Martoňák, G. E. Santoro, and E. Tosatti, "Quantum annealing of the traveling-salesman problem," *Physical Review E*, vol. 70, no. 5, p. 057701, 2004.

[49] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, "Intrinsic optimization using stochastic nanomagnets," *Scientific Reports*, vol. 7, p. 44370, 2017. [Online]. Available: http://doi.org/10.1038/srep44370

[50] A. K. Chandra, A. Das, and B. K. Chakrabarti, *Quantum quenching, annealing and computation*. Springer Science & Business Media, 2010, vol. 802.

[51] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara *et al.*, "A fully programmable 100-spin coherent ising machine with all-to-all connections," *Science*, vol. 354, no. 6312, pp. 614–617, 2016.

[52] C. Yoshimura, M. Hayashi, T. Okuyama, and M. Yamaoka, "Fpga-based annealing processor for ising model," in *2016 Fourth International Symposium on Computing and Networking (CANDAR)*. IEEE, 2016, pp. 436–442.

[53] Y. Shim, A. Jaiswal, and K. Roy, "Ising computation based combinatorial optimization using spin-hall effect (she) induced stochastic magnetization reversal," *Journal of Applied Physics*, vol. 121, no. 19, p. 193902, 2017.

[54] T. Wang and J. Roychowdhury, "Oscillator-based ising machine," *arXiv preprint arXiv:1709.08102*, 2017.

[55] T. Okuyama, M. Hayashi, and M. Yamaoka, "An ising computer based on simulated quantum annealing by path integral monte carlo method," in *Rebooting Computing (ICRC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[56] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, "Evidence for quantum annealing with more than one hundred qubits," *Nature Physics*, vol. 10, no. 3, p. 218, 2014.

[57] K. Y. Camsari, S. Chowdhury, and S. Datta, "Scaled quantum circuits emulated with room temperature p-bits," *arXiv preprint arXiv:1810.07144*, 2018.

[58] B. Heim, T. F. Rønnow, S. V. Isakov, and M. Troyer, "Quantum versus classical annealing of ising spin glasses," *Science*, vol. 348, no. 6231, pp. 215–217, 2015.

[59] K. Kechedzhi and V. N. Smelyanskiy, "Open-system quantum annealing in mean-field models with exponential degeneracy," *Physical Review X*, vol. 6, no. 2, p. 021028, 2016.

[60] F. L. Traversa and M. Di Ventra, "Polynomial-time solution of prime factorization and np-complete problems with digital memcomputing machines," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, no. 2, p. 023107, 2017.

[61] M. Di Ventra and F. L. Traversa, "Perspective: Memcomputing: Leveraging memory and physics to compute efficiently," *Journal of Applied Physics*, vol. 123, no. 18, p. 180901, 2018.

[62] A. Z. Pervaiz, L. A. Ghantasala, K. Y. Camsari, and S. Datta, "Hardware emulation of stochastic p-bits for invertible logic," *Scientific Reports*, vol. 7, p. 10994, Sep. 2017.

[63] S. C. Smithson, N. Onizawa, B. H. Meyer, W. J. Gross, and T. Hanyu, "Efficient cmos invertible logic using stochastic computing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2019.

[64] P. Debashis, R. Faria, K. Y. Camsari, and Z. Chen, "Design of stochastic nanomagnets for probabilistic spin logic," *IEEE Magnetics Letters*, vol. 9, pp. 1–5, 2018. [Online]. Available: https://doi.org/10.1109/LMAG.2018.2860547

[65] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nature nanotechnology*, vol. 10, no. 3, p. 187, 2015.

[66] Z. Luo, C. Xiong, X. Zhang, Z.-G. Guo, J. Cai, and X. Zhang, "Extremely large magnetoresistance at low magnetic field by coupling the nonlinear transport effect and the anomalous hall effect," *Advanced Materials*, vol. 28, no. 14, pp. 2760–2764, 2016.

[67] L. Lopez-Diaz, L. Torres, and E. Moro, "Transition from ferromagnetism to superparamagnetism on the nanosecond time scale," *Physical Review B*, vol. 65, no. 22, p. 224406, 2002.

[68] D. Vodenicarevic, N. Locatelli, A. Mizrahi, J. S. Friedman, A. F. Vincent, M. Romera, A. Fukushima, K. Yakushiji, H. Kubota, S. Yuasa *et al.*, "Low-energy truly random number generation with superparamagnetic tunnel junctions for unconventional computing," *Physical Review Applied*, vol. 8, no. 5, p. 054045, 2017. [Online]. Available: https://doi.org/10.1103/PhysRevApplied.8.054045

[69] R. P. Cowburn, D. K. Koltsov, A. O. Adeyeye, M. E. Welland, and D. M. Tricker, "Single-domain circular nanomagnets," *Physical Review Letters*, vol. 83, no. 5, p. 1042, 1999.

[70] P. Debashis, R. Faria, K. Y. Camsari, J. Appenzeller, S. Datta, and Z. Chen, "Experimental demonstration of nanomagnet networks as hardware for Ising computing," in *2016 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2016, pp. 34.3.1–34.3.4.

[71] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, "A perpendicular-anisotropy cofeb–mgo magnetic tunnel junction," *Nature materials*, vol. 9, no. 9, p. 721, 2010.

[72] B. Behin-Aein, A. Sarkar, and S. Datta, "Modeling circuits with spins and magnets for all-spin logic," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*. IEEE, 2012, pp. 36–40.

[73] B. Behin-Aein, "Computing multi-magnet based devices and methods for solution of optimization problems," Apr. 15 2014, uS Patent 8,698,517.

[74] A. F. Vincent, J. Larroque, N. Locatelli, N. B. Romdhane, O. Bichler, C. Gamrat, W. S. Zhao, J. Klein, S. Galdin-Retailleau, and D. Querlioz, "Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 2, pp. 166–174, April 2015.

[75] B. Behin-Aein, V. Diep, and S. Datta, "A building block for hardware belief networks," *Scientific Reports*, vol. 6, p. 29893, Jul. 2016.

[76] A. Sengupta, M. Parsa, B. Han, and K. Roy, "Probabilistic deep spiking neural systems enabled by magnetic tunnel junction," *IEEE Transactions on Electron Devices*, vol. 63, no. 7, pp. 2963–2970, 2016.

[77] Y. Lv and J.-P. Wang, "A single magnetic-tunnel-junction stochastic computing unit," in *Electron Devices Meeting (IEDM), 2017 IEEE International*. IEEE, 2017, pp. 36–2.

[78] C. H. Bennett, "The thermodynamics of computation—a review," *International Journal of Theoretical Physics*, vol. 21, no. 12, pp. 905–940, 1982.

[79] I. M. Miron, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, S. Pizzini, J. Vogel, and P. Gambardella, "Current-driven spin torque induced by the rashba effect in a ferromagnetic metal layer," *Nature materials*, vol. 9, no. 3, p. 230, 2010.

[80] L. Liu, C.-F. Pai, Y. Li, H. Tseng, D. Ralph, and R. Buhrman, "Spin-torque switching with the giant spin hall effect of tantalum," *Science*, vol. 336, no. 6081, pp. 555–558, 2012.

[81] L. You, O. Lee, D. Bhowmik, D. Labanowski, J. Hong, J. Bokor, and S. Salahuddin, "Switching of perpendicularly polarized nanomagnets with spin orbit torque without an external magnetic field by engineering a tilted anisotropy," *Proceedings of the National Academy of Sciences*, vol. 112, no. 33, pp. 10 310–10 315, 2015.

[82] A. van den Brink, G. Vermijs, A. Solignac, J. Koo, J. T. Kohlhepp, H. J. Swagten, and B. Koopmans, "Field-free magnetization reversal by spin-hall effect and exchange bias," *Nature communications*, vol. 7, p. 10854, 2016.

[83] Y.-W. Oh, S.-h. C. Baek, Y. Kim, H. Y. Lee, K.-D. Lee, C.-G. Yang, E.-S. Park, K.-S. Lee, K.-W. Kim, G. Go *et al.*, "Field-free switching of perpendicular magnetization through spin–orbit torque in antiferromagnet/ferromagnet/oxide structures," *Nature nanotechnology*, vol. 11, no. 10, p. 878, 2016.

[84] P. Debashis and Z. Chen, "Experimental demonstration of a spin logic device with deterministic and stochastic mode of operation," *Scientific reports*, vol. 8, no. 1, p. 11405, 2018.

[85] M. Hu, J. P. Strachan, Z. Li, E. M. Grafals, N. Davila, C. Graves, S. Lam, N. Ge, J. J. Yang, and R. S. Williams, "Dot-product engine for neuromorphic computing: programming 1t1m crossbar to accelerate matrix-vector multiplication," in *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE*. IEEE, 2016, pp. 1–6.

[86] V. Ostwal, P. Debashis, R. Faria, Z. Chen, and J. Appenzeller, "Spin-torque devices with hard axis initialization as stochastic binary neurons," *Scientific reports*, vol. 8, no. 1, p. 16689, 2018. [Online]. Available: https://doi.org/10.1038/s41598-018-34996-2

[87] V. Ostwal and J. Appenzeller, "Spin–orbit torque-controlled magnetic tunnel junction with low thermal stability for tunable random number generation," *IEEE Magnetics Letters*, vol. 10, pp. 1–5, 2019.

[88] N. X. Sun and G. Srinivasan, "Voltage control of magnetism in multiferroic heterostructures and devices," in *Spin*, vol. 2, no. 03. World Scientific, 2012, p. 1240004.

[89] A. K. Biswas, H. Ahmad, J. Atulasimha, and S. Bandyopadhyay, "Experimental demonstration of complete 180 ° reversal of magnetization in isolated co nanomagnets on a pmn–pt substrate with voltage generated strain," *Nano letters*, vol. 17, no. 6, pp. 3478–3484, 2017.

[90] P. B. Meisenheimer, S. Novakov, N. M. Vu, and J. T. Heron, "Perspective: Magnetoelectric switching in thin film multiferroic heterostructures," *Journal of Applied Physics*, vol. 123, no. 24, p. 240901, 2018.

[91] F. Wang, L. Luo, D. Zhou, X. Zhao, and H. Luo, "Complete set of elastic, dielectric, and piezoelectric constants of orthorhombic 0.71 pb (mg 1/ 3 nb 2/ 3) o 3–0.29 pb ti o 3 single crystal," *Applied physics letters*, vol. 90, no. 21, p. 212903, 2007.

[92] K. Roy, S. Bandyopadhyay, and J. Atulasimha, "Binary switching in a 'symmetric'potential landscape," *Scientific reports*, vol. 3, p. 3038, 2013.

[93] K. Munira, S. C. Pandey, W. Kula, and G. S. Sandhu, "Voltage-controlled magnetization switching in mrams in conjunction with spin-transfer torque and applied magnetic field," *Journal of Applied Physics*, vol. 120, no. 20, p. 203902, 2016.

[94] T. Shen, V. Ostwal, K. Y. Camsari, and J. Appenzeller, "Demonstration of a strain-mediated magnetoelectric write and read unit in a co60fe20b20/pb (mg1/3nb2/3) 0.7 ti0. 3o3 heterostructure," *arXiv preprint arXiv:1901.01368*, 2019.

[95] K. Y. Camsari, R. Faria, O. Hassan, B. M. Sutton, and S. Datta, "Equivalent circuit for magnetoelectric read and write operations," *Physical Review Applied*, vol. 9, no. 4, p. 044020, 2018.

[96] A. Klimov, N. Tiercelin, Y. Dusch, S. Giordano, T. Mathurin, P. Pernod, V. Preobrazhensky, A. Churbanov, and S. Nikitov, "Magnetoelectric write and read operations in a stress-mediated multiferroic memory cell," *Applied Physics Letters*, vol. 110, no. 22, p. 222401, May 2017.

[97] K. Y. Camsari, S. Salahuddin, and S. Datta, "Implementing p-bits With Embedded MTJ," *IEEE Electron Device Letters*, vol. 38, no. 12, pp. 1767–1770, Dec. 2017.

[98] W. T. Coffey and Y. P. Kalmykov, "Thermal fluctuations of magnetic nanoparticles: Fifty years after brown," *Journal of Applied Physics*, vol. 112, no. 12, p. 121301, 2012.

[99] W. F. Brown Jr, "Thermal fluctuations of a single-domain particle," *Physical Review*, vol. 130, no. 5, p. 1677, 1963.

[100] U. Atxitia, T. Birk, S. Selzer, and U. Nowak, "Superparamagnetic limit of antiferromagnetic nanoparticles," *arXiv preprint arXiv:1808.07665*, 2018.

[101] J. Kaiser, A. Rustagi, K. Y. Camsari, J. Z. Sun, S. Datta, and P. Upadhyaya, "Ultrafast fluctuations in low-barrier magnets," *arXiv preprint arXiv:1902.03312*, 2019.

[102] O. Hassan, R. Faria, K. Y. Camsari, J. Z. Sun, and S. Datta, "Low-barrier magnet design for efficient hardware binary stochastic neurons," *IEEE Magnetics Letters*, vol. 10, pp. 1–5, 2019.

[103] M. A. Abeed and S. Bandyopadhyay, "Low barrier nanomagnet design for binary stochastic neurons: Design challenges for real nanomagnets with fabrication defects," *IEEE Magnetics Letters*, 2019.

[104] C. M. Liyanagedera, A. Sengupta, A. Jaiswal, and K. Roy, "Stochastic spiking neural networks enabled by magnetic tunnel junctions: From nontelegraphic to telegraphic switching regimes," *Physical Review Applied*, vol. 8, no. 6, p. 064017, 2017.

[105] C. Lin, S. Kang, Y. Wang, K. Lee, X. Zhu, W. Chen, X. Li, W. Hsu, Y. Kao, M. Liu *et al.*, "45nm low power cmos logic compatible embedded stt mram utilizing a reverse-connection 1t/1mtj cell," in *Electron Devices Meeting (IEDM), 2009 IEEE International*. IEEE, 2009, pp. 1–4.

[106] Y. Song, J. Lee, H. Shin, K. Lee, K. Suh, J. Kang, S. Pyo, H. Jung, S. Hwang, G. Koh *et al.*, "Highly functional and reliable 8mb stt-mram embedded in 28nm logic," in *Electron Devices Meeting (IEDM), 2016 IEEE International*. IEEE, 2016, pp. 27–2.

[107] D. Shum, D. Houssameddine, S. T. Woo, Y. S. You, J. Wong, K. W. Wong, C. C. Wang, K. H. Lee, K. Yamane, V. B. Naik, C. S. Seet, T. Tahmasebi, C. Hai, H. W. Yang, N. Thiyagarajah, R. Chao, J. W. Ting, N. L. Chung, T. Ling, T. H. Chan, S. Y. Siah, R. Nair, S. Deshpande, R. Whig, K. Nagel, S. Aggarwal, M. DeHerrera, J. Janesky, M. Lin, H. J. Chia, M. Hossain, H. Lu, S. Ikegawa, F. B. Mancoff, G. Shimon, J. M. Slaughter, J. J. Sun, M. Tran, S. M. Alam, and T. Andre, "Cmos-embedded stt-mram arrays in 2x nm nodes for gp-mcu applications," in *2017 Symposium on VLSI Technology*, June 2017, pp. T208–T209.

[108] A. Z. Pervaiz, L. A. Ghantasala, K. Y. Camsari, and S. Datta, "Hardware emulation of stochastic p-bits for invertible logic," *Scientific Reports*, vol. 7, p. 10994, Sep. 2017.

[109] B. Sutton, R. Faria, L. A. Ghantasala, K. Y. Camsari, and S. Datta, "Autonomous probabilistic coprocessing with petaflips per second," *arXiv preprint arXiv:1907.09664*, 2019.

[110] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Soc., 2017, vol. 107.

[111] J. Ding, E. Lubetzky, and Y. Peres, "The mixing time evolution of glauber dynamics for the mean-field ising model," *Communications in Mathematical Physics*, vol. 289, no. 2, pp. 725–764, 2009.

[112] S. S. Haykin *et al.*, *Neural networks and learning machines/Simon Haykin.* New York: Prentice Hall,, 2009.

[113] X. Peng, Z. Liao, N. Xu, G. Qin, X. Zhou, D. Suter, and J. Du, "Quantum Adiabatic Algorithm for Factorization and Its Experimental

Implementation," *Physical Review Letters*, vol. 101, no. 22, p. 220405, Nov. 2008.

[114] P. Henelius and S. Girvin, "A statistical mechanics approach to the factorization problem," *arXiv preprint arXiv:1102.1296*, 2011.

[115] S. Jiang, K. A. Britt, A. J. McCaskey, T. S. Humble, and S. Kais, "Quantum annealing for prime factorization," *Scientific Reports (Nature Publisher Group)*, vol. 8, pp. 1–9, 2018.

[116] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, "Integer factorization using stochastic magnetic tunnel junctions," *Nature (in press)*, 2019.

[117] K. L. Pudenz, T. Albash, and D. A. Lidar, "Quantum annealing correction for random ising problems," *Physical Review A*, vol. 91, no. 4, p. 042302, 2015.

[118] T. J. Sejnowski, "Higher-order boltzmann machines," in *AIP Conference Proceedings*, vol. 151, no. 1. AIP, 1986, pp. 398–403.

[119] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[120] D. Heckerman and J. S. Breese, "Causal independence for probability assessment and inference using bayesian networks," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 26, no. 6, pp. 826–831, 1996.

[121] A. Z. Pervaiz, B. M. Sutton, L. A. Ghantasala, and K. Y. Camsari, "Weighted p-bits for fpga implementation of probabilistic circuits," *IEEE transactions on neural networks and learning systems*, 2018.

[122] A. F. Vincent, J. Larroque, N. Locatelli, N. B. Romdhane, O. Bichler, C. Gamrat, W. S. Zhao, J.-O. Klein, S. Galdin-Retailleau, and D. Querlioz, "Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems," *IEEE transactions on biomedical circuits and systems*, vol. 9, no. 2, pp. 166–174, 2015. [Online]. Available: https://doi.org/10.1109/TBCAS.2015.2414423

[123] X. Sun, P. Wang, K. Ni, S. Datta, and S. Yu, "Exploiting hybrid precision for training and inference: A 2t-1fefet based analog synaptic weight cell," in *2018 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2018, pp. 3–1.

[124] W. A. Borders, H. Akima, S. Fukami, S. Moriya, S. Kurihara, Y. Horio, S. Sato, and H. Ohno, "Analogue spin–orbit torque device for artificial-neural-network-based associative memory operation," *Applied Physics Express*, vol. 10, no. 1, p. 013007, 2016.

[125] O. Hassan, K. Y. Camsari, and S. Datta, "Voltage-driven building block for hardware belief networks," *IEEE Design & Test*, vol. 36, no. 3, pp. 15–21, 2019.

[126] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, 2014. [Online]. Available: https://doi.org/10.1126/science.1254642