


Heisenberg machines with programmable spin circuits

Saleh Bunaiyan^{1,2,*}, Supriyo Datta³, and Kerem Y. Camsari²

¹*Electrical Engineering Department, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia*

²*Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, California 93106, USA*

³*Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA*

 (Received 3 December 2023; revised 22 April 2024; accepted 6 June 2024; published 8 July 2024)

We show that we can harness two recent experimental developments to build a compact hardware emulator for the classical Heisenberg model in statistical physics. The first is the demonstration of spin-diffusion lengths in excess of microns in graphene even at room temperature. The second is the demonstration of low-barrier magnets (LBMs) whose magnetization can fluctuate rapidly even at sub-nanosecond rates. Using experimentally benchmarked circuit models, we show that an array of LBMs driven by an external current source has a steady-state distribution corresponding to a classical system with an energy function of the form $E = -(1/2) \sum_{i,j} J_{ij} (\hat{m}_i \cdot \hat{m}_j)$. This may seem surprising for a nonequilibrium system, but we show that it can be justified by a Lyapunov function corresponding to a system of coupled Landau–Lifshitz–Gilbert (LLG) equations. The Lyapunov function we construct describes LBMs interacting through the spin currents they inject into the spin-neutral substrate. We suggest ways to tune the coupling coefficient J_{ij} so that it can be used as a hardware solver for optimization problems involving continuous variables represented by vector magnetizations, similar to the role of the Ising model in solving optimization problems with binary variables. Finally, we train a Heisenberg XOR gate based on a network of four coupled stochastic LLG equations, illustrating the concept of probabilistic computing with a programmable Heisenberg model.

DOI: [10.1103/PhysRevApplied.22.014014](https://doi.org/10.1103/PhysRevApplied.22.014014)

I. INTRODUCTION

With the slowing down of Moore’s law, there is tremendous interest in unconventional computing approaches. Prominent among these approaches are the energy-based models (EBMs) inspired by statistical physics, in which the problem is mapped to an energy function, such that the solution corresponds to the minimum-energy state [1–3], which can then be identified using powerful optimization algorithms.

A particularly attractive way to find the minimum-energy state of an energy function becomes possible if we can design a physical system whose natural physics makes it relax to this state. Such a system could be harnessed to solve this class of problem orders of magnitude faster and more efficiently than an algorithm implemented using conventional transistors. The main reason behind this gap is the extensive cost of designing stochastic spins in deterministic hardware where it takes tens of

thousands of transistors to emulate the necessary tunable randomness [4].

An interesting EBM is the classical Heisenberg model

$$E = -\frac{1}{2} \sum_{i,j} J_{ij} (\hat{m}_i \cdot \hat{m}_j), \quad (1)$$

where $\hat{m}_{i,j}$ are unit vectors in three dimensions, which is different from the Ising model

$$E_{\text{Ising}} = -\frac{1}{2} \sum_{i,j} J_{ij} s_i s_j, \quad (2)$$

where s_i and s_j are variables with only two allowed values, ± 1 . Hopfield networks and Boltzmann machines based on the Ising model (2) have generated tremendous recent excitement and even mapped to physical systems [5,6]. The modern Hopfield network [7–9], based on the classical Heisenberg model (1), shows promise in machine learning applications and it should be of great interest to map it to an energy-efficient physical system.

In this work, we propose and establish the feasibility of mapping Eq. (1) to a physical system consisting of an array

*Contact author: saleh_bunaiyan@ucsb.edu

of low-barrier magnets (LBMs) interacting via spin currents through a spin-neutral channel (e.g., Cu, graphene), as shown in Fig. 1. The system is similar to the two-dimensional graphene channels with long spin-diffusion length (λ_s) used to demonstrate room-temperature spin logic [10–13], but with one key difference. Here, the magnets are not the usual stable magnets, but LBMs similar to those used to represent Ising spins or probabilistic bits (p-bits) [14,15], which have been shown to fluctuate with gigahertz rates [16–18]. A charge current is driven through each LBM i by an external source, and the associated spin current diffuses through the substrate and exerts a spin torque on a neighboring LBM j , leading to an effective interaction term J_{ij} between them.

Note that this is a particularly compact physical realization of Eq. (1) compared to the reported realizations of the Ising model [Eq. (2)] using p-bits [5,19,20], which require separate hardware to implement “synapses” that sense the state of each p-bit and drive neighboring p-bits

accordingly. The term “synapses” here is used to refer to the coupling strength (weights) between neurons, where we assume that synapses have fixed values. Hence, the structure proposed here (Fig. 1) integrates both the neuron and the synapse at once, thus making it possible to implement relatively fast synapses that can utilize the nanosecond fluctuations that have been demonstrated [16–18] and could be designed to be even faster. The only additional hardware needed here is a final readout of the output magnets, rather than repeated readouts and reinjections from intermediate magnets.

The central result of this paper is to establish that the structure in Fig. 1 indeed minimizes a classical Heisenberg energy function of the form in Eq. (1). This is not at all obvious, since our structure is *not* at equilibrium and is not expected necessarily to obey a Boltzmann law $\propto \exp(-E/k_B T)$, with $k_B T$ being the thermal energy. What we show theoretically is that an energy function of the form (1) constitutes a Lyapunov function that is minimized by the LBM dynamical equations. Our numerical simulations further indicate that different configurations also follow a Boltzmann law at least approximately.

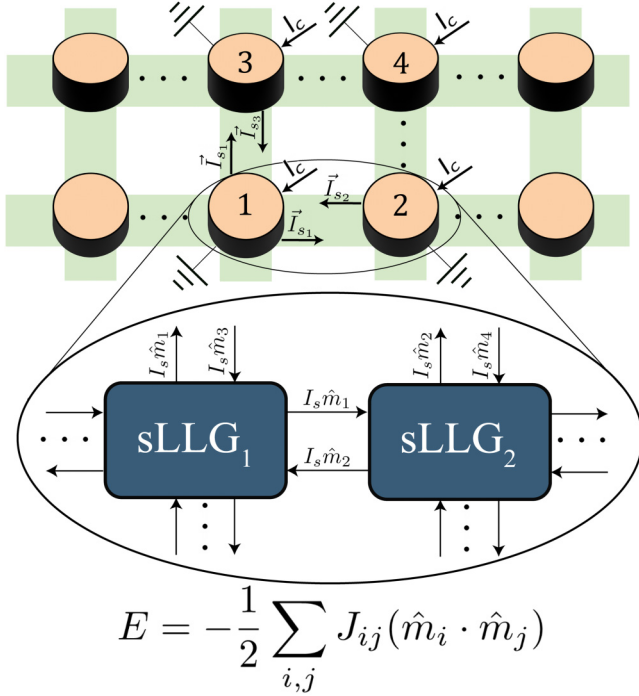


FIG. 1. Heisenberg machines with programmable spin circuits. LBMs and spin-neutral channels form a Heisenberg array, where an external charge current is injected into the LBMs, thereby inducing a spin current inside the channel. Magnets are assumed to have zero energy barrier with no effective anisotropy. The proposed array is analogous to a system of coupled stochastic Landau–Lifshitz–Gilbert (sLLG) equations, where each sLLG sends a spin-current vector \vec{I}_s to its neighbors, where the spin current is defined as a function of the magnet magnetization $\vec{I}_{si} = I_s \hat{m}_i$. This coupled LBM system minimizes the Heisenberg Hamiltonian with continuous spin states, functioning as a Heisenberg machine, analogous to an Ising machine.

II. LYAPUNOV FUNCTIONS

To model a structure like Fig. 1 with N LBMs, we start from a set of N coupled Landau–Lifshitz–Gilbert (LLG) equations using experimentally benchmarked physical parameters from Refs. [13,21,22], which were carefully derived from the work of Bauer, Brataas and Kelly on magnetoelectric circuit theory [23]. These were later turned into explicit circuit models that are modularly simulated in standard circuit simulators [21,24,25].

Throughout this work, we focus on LBMs with no shape or easy-plane anisotropy, which can be practically built by reducing the energy barriers of magnets with perpendicular magnetic anisotropy (PMA). Similar to other dissipative dynamical systems [26], we seek to establish a pseudoenergy function that is minimized over the time evolution of the system. We first show that a single magnet (\hat{m}_1) with an injected spin current \vec{I}_s , coming from another fixed magnet \hat{m}_2 such that $\vec{I}_s = I_s \hat{m}_2$, minimizes the function

$$E = -\left(\frac{I_s}{I_0}\right) \hat{m}_1 \cdot \hat{m}_2 = -J_{12} \hat{m}_1 \cdot \hat{m}_2, \quad (3)$$

where J_{12} is the dimensionless and symmetric coupling coefficient ($J_{12} = J_{21}$) defined as I_s/I_0 . Here I_0 is a constant given by $2q\alpha k_B T/\hbar$, where q is the electron charge, \hbar is the reduced Planck’s constant, and α is the damping coefficient of the nanomagnet. Next, we show how I_0 can be formally derived from the Fokker-Planck equation (FPE).

A. Fokker-Planck equation for coupled LBMs

Suppose that the spin current \vec{I}_s injected into \hat{m}_1 is polarized in the direction of another fixed magnet, \hat{m}_2 , such that $\vec{I}_s = I_s \hat{m}_2$. We will assume $\hat{m}_2 = \hat{z}$ without loss of generality. Any arbitrary direction would produce the same results, but, for the sake of simplifying the mathematics, we choose that direction.

For this system, the following FPE can be written at steady state [27]:

$$\left(\frac{\alpha \gamma k_B T}{(1 + \alpha^2) M_s \mathcal{V}} \right) \nabla^2 P(\theta_1, \phi_1) - \nabla \cdot \left(P(\theta_1, \phi_1) \frac{d\hat{m}_1}{dt} \right) = 0, \quad (4)$$

where γ is the gyromagnetic ratio of the electron, M_s is the saturation magnetization of the magnet, and \mathcal{V} is the volume of the magnetic body. The FPE derivation discussed in the seminal work of Brown [27] naturally does not consider the effect of the spin-transfer torque since this phenomenon was discovered later. Here, we use a modified FPE that expands Brown's work to include the spin-transfer-torque effect, following the approach discussed in Ref. [28].

At steady state, the magnetization \hat{m}_1 follows a Boltzmann-like law according to the energy, E , defined by Eq. (3):

$$P(\theta_1, \phi_1) = \frac{1}{Z} \exp(-E). \quad (5)$$

In turn, $d\hat{m}_1/dt$ can be obtained from the deterministic LLG equation [29] for \hat{m}_1 :

$$(1 + \alpha^2) \frac{d\hat{m}_1}{dt} = \frac{\hat{m}_1 \times (\vec{I}_s \times \hat{m}_1)}{qN_s} + \frac{\alpha(\hat{m}_1 \times \vec{I}_s)}{qN_s}, \quad (6)$$

where N_s is the number of spins in the magnetic volume, i.e., $N_s = M_s \mathcal{V} / \mu_B$, with μ_B being the Bohr magneton. Direct substitution of Eq. (5) into Eq. (4) shows, after several steps of tedious algebra, that the solution satisfies the FPE. Next, we show that dE/dt is always negative for coupled LBMs.

B. Lyapunov functions: two-magnet system

Armed with our result in Eq. (3), we consider the Lyapunov function for two coupled LBMs ($N = 2$). The main idea is to find an "energy" whose rate of change is nonpositive:

$$\frac{dE}{dt} = \vec{\nabla}_{\hat{m}_1} E \cdot \frac{d\hat{m}_1}{dt} \leq 0. \quad (7)$$

Consider the same scenario from which we started in Eq. (4): magnet \hat{m}_1 is receiving a spin current, \vec{I}_s , polarized in the direction of another fixed magnet, \hat{m}_2 , such that $\vec{I}_s =$

$I_s \hat{m}_2 = (I_0) J_{12} \hat{m}_2$. Here J_{12} is the dimensionless coupling between magnet 2 and magnet 1. For this system, define the Lyapunov function, E , as

$$E = - \left(\frac{I_s}{I_0} \right) \hat{m}_1 \cdot \hat{m}_2 \quad \text{with} \quad \vec{\nabla}_{\hat{m}_1} E = - \left(\frac{I_s}{I_0} \right) \hat{m}_2 = - \frac{\vec{I}_s}{I_0}. \quad (8)$$

From the LLG equation for \hat{m}_1 [Eq. (6)], we obtain

$$C \frac{d\hat{m}_1}{dt} = \vec{I}_s - \hat{m}_1 (\hat{m}_1 \cdot \vec{I}_s) + \alpha (\hat{m}_1 \times \vec{I}_s), \quad (9)$$

where we defined $C = qN_s(1 + \alpha^2) > 0$. The combination of Eqs. (9) and (8) using Eq. (7) yields

$$C \frac{dE}{dt} = - \frac{\vec{I}_s}{I_0} \cdot [\vec{I}_s - \hat{m}_1 (\hat{m}_1 \cdot \vec{I}_s) + \alpha (\hat{m}_1 \times \vec{I}_s)]. \quad (10)$$

Finally, by simplifying Eq. (10), we can show that the condition for energy minimization is satisfied:

$$C \frac{dE}{dt} = - \frac{[\vec{I}_s \cdot \vec{I}_s - (\hat{m}_1 \cdot \vec{I}_s)^2]}{I_0} \leq 0. \quad (11)$$

C. Lyapunov function: N -magnet system

Using the same approach, we extend this result to a system of N coupled LBMs, described by N coupled LLG equations:

$$(1 + \alpha^2) \frac{d\hat{m}_i}{dt} = \frac{\hat{m}_i \times (I_0 \vec{I}_i \times \hat{m}_i)}{qN_s} + \frac{\alpha(\hat{m}_i \times I_0 \vec{I}_i)}{qN_s}, \quad (12)$$

with the input to the i th magnet defined as $\vec{I}_i = \sum_j J_{ij} \hat{m}_j$. Given how the Lyapunov function for the two-magnet system can be formally derived to follow a Boltzmann-like equation, we posit the following Lyapunov function for the N -magnet system:

$$E = - \frac{1}{2} \sum_{ij} J_{ij} (\hat{m}_i \cdot \hat{m}_j) = - \frac{1}{2} \sum_{ij} \frac{I_{sij}}{I_0} (\hat{m}_i \cdot \hat{m}_j). \quad (13)$$

Assuming the reciprocity of the interaction strengths ($J_{ij} = J_{ji}$):

$$\vec{\nabla}_{\hat{m}_i} E = - \sum_j J_{ij} \hat{m}_j = - \vec{I}_i. \quad (14)$$

The coupled LLG equations for N magnets can be written as

$$C \frac{d\hat{m}_i}{dt} = I_0 [\vec{I}_i - \hat{m}_i(\hat{m}_i \cdot \vec{I}_i) + \alpha(\hat{m}_i \times \vec{I}_i)]. \quad (15)$$

The combination of Eqs. (15) and (14) with Eq. (7) yields

$$C \frac{dE}{dt} = \sum_i -\{\vec{I}_i \cdot I_0[\vec{I}_i - \hat{m}_i(\hat{m}_i \cdot \vec{I}_i) + \alpha(\hat{m}_i \times \vec{I}_i)]\}. \quad (16)$$

By noting that $\sum_i \vec{I}_i \cdot \alpha(\hat{m}_i \times \vec{I}_i) = 0$, the expression can be further simplified to

$$C \frac{dE}{dt} = \sum_i -I_0 [\vec{I}_i \cdot \vec{I}_i - (\hat{m}_i \cdot \vec{I}_i)^2] \leq 0, \quad (17)$$

which shows that the condition for energy minimization is also satisfied for N magnets.

The above discussion implies that $dE/dt \leq 0$, where E is now the classical Heisenberg model described in Eq. (1). As a result, in the *deterministic limit*, the system dynamics tends to minimize the energy of a classical Heisenberg model whose parameters J_{ij} can be programmed. As we will show later, in the presence of uncorrelated white noise (see the stochastic LLG equation in Appendix A) and transitions between all states with finite probability (i.e., ergodicity), the tendency to minimize energy leads the system to sample from the Boltzmann distribution with

occasional energy-increasing state transitions enabled by thermal noise. This description of a nonequilibrium system with an equilibrium model is reminiscent of spin currents interacting with PMA magnets [29].

III. NUMERICAL SIMULATION

In this section we will present numerical results for two examples, i.e., a system of two LBMs and a frustrated system of three LBMs, indicating that the results follow the Boltzmann law, where states are sampled according to $\rho(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_N) \propto \exp[-E(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_N)]$.

Simulations were carried out on a standard circuit simulator (HSPICE) using four-component spin circuits [21, 30]. The full circuits used for simulating the two examples are shown in Figs. 2 and 3. The used modules are of two categories: transport and magnetism. Transport through the LBMs into the spin-neutral channel is characterized by an FM-NM interface module, which defines the interface between the ferromagnet (FM) and the normal metal (NM) [21,23]. The magnetic fluctuations of the LBMs are described by the stochastic Landau–Lifshitz–Gilbert module, carefully benchmarked against corresponding FPEs for monodomain magnets [29,30]. The spin-neutral channel between LBMs is solely described by the transport module NM. All transport modules are represented by 4×4 matrices describing the interactions between charge and spins in the z, x, y directions. Further details on the spin-circuit modules, sLLG, and thermal noise are discussed in

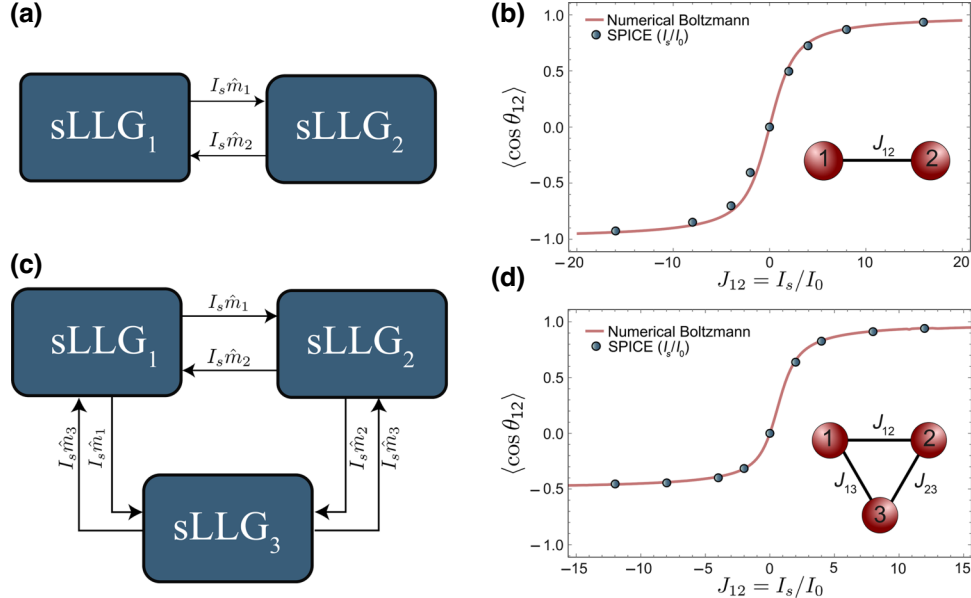


FIG. 2. Coupled LBMs with pure spin currents. (a) Two coupled sLLG equations, where each sLLG sends a spin current along its magnetization direction $\vec{I}_{si} = I_s \hat{m}_i$ to the other sLLG. (b) The two coupled sLLG equations are contacted to the Boltzmann numerical solution by the predefined constant I_0 . (c) Three coupled sLLG equations in a frustrated configuration. (d) The three coupled sLLG equations also match with the Boltzmann law. Note that I_0 is configuration-independent.

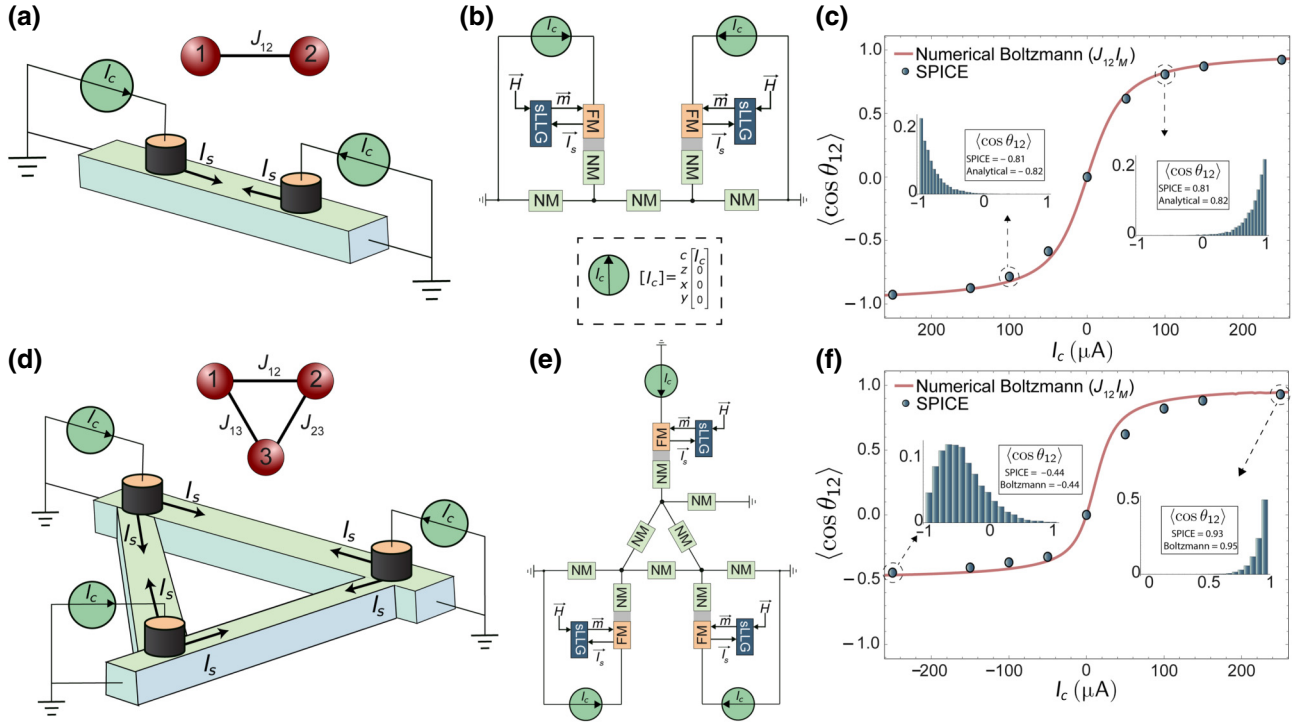


FIG. 3. Simulated hardware for the Heisenberg emulator. (a) Two-LBM system that emulates a Heisenberg model of two spins. (b) Simulated configuration in HSPICE using benchmarked circuit models. Transport physics from the LBM to the channel and from the channel to the ground are characterized by FM-NM and NM modules, respectively. Magnetization dynamics of the LBM are described by the sLLG modules. All circuits are described by four components: charge and spins (z, x, y). (c) The Boltzmann solution is related to the proposed hardware through I_M , which takes the charge-to-spin-current conversions into account. The relation between spin current and interaction strength is given by $J_{12} = I_s/I_0$. (d) Three-LBM system that emulates a Heisenberg model of three spins in a frustrated configuration. (e) Simulated configuration in HSPICE. (f) The Boltzmann solution for the three-magnet system. We use the same I_M value as in the two-magnet system (even though a new charge-to-spin mapping may need to be measured or calculated; see Appendix B). The insets in panels (c) and (f) show the probability distributions of the magnets' correlation $\cos \theta_{12}$ at a given input I_c , which can be analytically derived for two LBMs with the aid of Eq. (20).

Appendix A. The physical parameters we used are reported in Appendix E.

To show that a system of coupled LBMs samples from the classical Heisenberg model, we start with a pure sLLG network that resembles the system of coupled LBMs. All numerical examples are based on the sLLG equation with the corresponding FPE given by Eq. (4).

A. Coupled LBMs without transport

In Fig. 2, we numerically study a system of two coupled sLLG equations, and we compare its response with the numerical solution of the Heisenberg Hamiltonian based on the Boltzmann law. We study the system by observing the correlation between the two coupled LBMs, which can be described by the cosine of the relative angle between their magnetization vectors $\langle \cos \theta_{12} \rangle$. In the two configurations of Fig. 2, we compute correlations between the magnets purely based on spin currents; for the full proposed hardware, more complex analysis is needed.

If the two LBMs are coupled by J_{ij} , their correlation can be computed from a Boltzmann-like equation using Eq. (1):

$$\langle \cos \theta_{12} \rangle = \int_{S_1} \int_{S_2} \frac{1}{Z} \cos \theta_{12} \exp(-E) dS_1 dS_2. \quad (18)$$

Here $dS_i = \sin \theta_i d\theta_i d\phi_i$ describes integration on the surface of the unit sphere and Z is the partition function. For any given symmetric coupling J_{ij} between LBMs 1 and 2, we can evaluate Eq. (18) numerically. Note that for two-magnet system the average correlation can also be obtained through solving for the average magnetization $\langle m_z \rangle$ for a single magnet,

$$\langle m_z \rangle = \frac{\int_{\theta=0}^{\theta=\pi} \int_{\phi=0}^{\phi=2\pi} \frac{1}{Z} \cos \theta \exp\left(\frac{I_s}{I_0} \cos \theta\right) \sin \theta d\theta d\phi}{\int_{\theta=0}^{\theta=\pi} \int_{\phi=0}^{\phi=2\pi} \frac{1}{Z} \exp\left(\frac{I_s}{I_0} \cos \theta\right) \sin \theta d\theta d\phi}, \quad (19)$$

which is found to be the Langevin function

$$\langle m_z \rangle = \langle \cos \theta \rangle = \coth \left(\frac{I_s}{I_0} \right) - \frac{I_0}{I_s}. \quad (20)$$

The Langevin function exactly overlaps with the Boltzmann numerical solution of $\langle \cos \theta_{12} \rangle$ shown in Fig. 2(b). This can be viewed as measuring the relative angle between the two magnets when one of the magnets is fixed to $+\hat{z}$ and the other magnet is free to move. This trick works due to the spherical symmetry of the system, discussed in more detail in Appendix B.

We then study a frustrated system of three magnets, with the network configuration shown in Figs. 2(c) and 2(d), described by the Boltzmann law as

$$\langle \cos \theta_{12} \rangle = \int_{S_1} \int_{S_2} \int_{S_3} \frac{1}{Z} \cos \theta_{12} \exp(-E) dS_1 dS_2 dS_3. \quad (21)$$

Our numerical results in Fig. 2 indicate that the two- and three-magnet systems are well described by the Boltzmann law, indicating that the network is actually sampling from the classical Heisenberg model. Moreover, it is important to note the generality of the constant I_0 , which could guide the experimental realization of these networks.

B. Programmable spin-circuits

The proposed physical structures of the two studied systems are described in Figs. 3(a) and 3(d). All magnets were chosen to be identical and receive the same input current I_c . In our setup, the charge current I_c flows to a nearby ground injecting pure spin currents in all directions, as commonly done in nonlocal spin valves [13,31].

To make contact between the circuit shown in Fig. 3(a) and Eq. (18), we need the same normalizing parameter $I_0 = I_{sij}/J_{ij}$, where I_{sij} is the component of the spin current along magnet j incident to magnet i . In addition to that, we need another configuration-dependent parameter relating the injected charge current to the spin currents. We define a single parameter I_M that combines the two such that $J_{ij} = I_c/I_M$. For the two-magnet system, we can find I_{sij} analytically by defining a new basis: $(\hat{m}_1, \hat{m}_2, \hat{m}_1 \times \hat{m}_2)$ similar to the approach used in Ref. [32]. By a clever coordinate transformation, we exactly solve for I_M in the two-magnet system in Appendix B. Numerical results from our spin circuits match those obtained from our analytical solution, as shown in Fig. 3, where the analytical solution was found by substituting $J_{12} = I_c/I_M$ in Eq. (20). Note that, in general systems ($N > 2$), I_M may depend on the geometry of the channel and it may need to be found by experimental calibration in different systems.

The frustrated system in Fig. 3(d) can also be described by Eq. (21). All magnets have identical charge inputs

(swept from negative to positive values of I_c , where positive electron current is measured from the magnet into the channel), leading to uniform $J_{ij} = \pm J_0$ ferromagnetic and antiferromagnetic interactions. At all currents we observe identical correlations, i.e., $\langle \theta_{12} \rangle = \langle \theta_{13} \rangle = \langle \theta_{23} \rangle$, and thus we only report $\langle \cos \theta_{12} \rangle$ without loss of generality in Fig. 3. Similar to the two-magnet system, positive currents lead to ferromagnetic correlation. However, negative currents result in a saturated correlation of $\cos \theta_{12} = -0.5$, where each magnet (on average) has a 120° separation between its neighbors [Fig. 3(f)], reminiscent of frustrated magnets realizing XY models [2].

IV. POWER ESTIMATION

Since the proposed hardware comprises both the neuron and the synapse, the power consumption is expected to be an order of magnitude less than that of possible digital implementations. There are no detailed estimations of the digital footprint for continuous stochastic neurons. However, transistor-level projections indicate that tens of thousands of transistors operating at tens of microwatts are needed even for *binary* stochastic neurons [4,33].

To generate significant correlations, our systems presented in Fig. 3 consume around 180 and 270 nW for the two- and three-LBM systems, about 100 nW per LBM terminal. We measure this power as the average at the two extremes of input charge currents for the ferromagnetic and the antiferromagnetic cases. These numbers can be understood by relating the necessary spin currents to charge currents. Spin currents need to be around $\approx \pm 15I_0 \approx 1.88 \mu\text{A}$ to create large negative or positive correlations. With interface polarizations of $P \approx 0.1$, interface and side-channel resistances of $R_{\text{int}} \approx 1 \Omega$ and $R_{\text{side}} \approx 0.5 \Omega$, a spin-diffusion length of 400 nm along 200 nm channels, and resistive division factors diverting the spin currents, charge currents of around $I_c = 250 \mu\text{A}$ are needed (see Appendix B). These charge currents lead to $I_c^2 R$ losses of around 100 nW per LBM arm. Considering the 10–20 μW estimations for binary stochastic neurons for the simpler Ising model, the proposed emulator should be at least three to four orders of magnitude more energy-efficient over digital implementations of the classical Heisenberg model.

V. PROGRAMMABILITY

Another practical aspect is the tunability of the interaction strengths, J_{ij} . In the proposed hardware, these interactions are described by the amount of spin current received from the other LBMs, which can be tuned through many parameters (e.g., input current, channel material, etc.). In Fig. 4, we analyze the programmability of J_{ij} for a channel by tuning the channel length L_{ch} . The spin-diffusion length for Cu is around 400 nm at room temperature [34]; the results show a reasonable tunability range for J_{ij} . Recently,

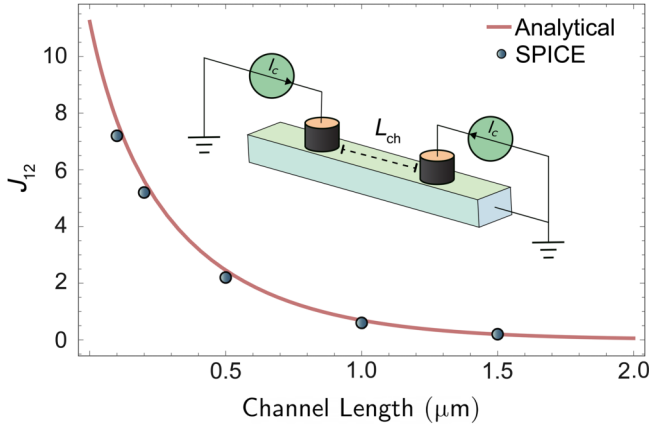


FIG. 4. Programmability of the interaction strength J_{ij} . At fixed charge current I_c , the NM channel between LBMs encodes the magnitude of the coupling strength between magnets. Extending the channel length L_{ch} reduces the correlation between the magnets due to more spins being neutralized. The polarity of J_{ij} is controlled through the direction of the input current I_c . The analytical analysis is provided in Appendix B. We assume the channel to be Cu with 400-nm spin-diffusion length.

it was reported that graphene can have a spin-diffusion length of up to 26 μm at room temperature [12]. Having longer spin-diffusion length relaxes the constraints on tuning the interaction strength J_{ij} and we can achieve higher values of J_{ij} ; both are critical for machine learning applications, such as Boltzmann machines [35,36]. Throughout, we assumed configurations where all charge currents have the same sign and magnitude, leading to J_{ij} with the same sign.

However, systems with frustration involve positive and negative J_{ij} . One way to implement $\pm J_{ij}$ in one configuration is by using auxiliary magnets. We set the network to have all negative weights, and, through graph embedding via auxiliary nodes (Fig. 5), a mix of positive and negative

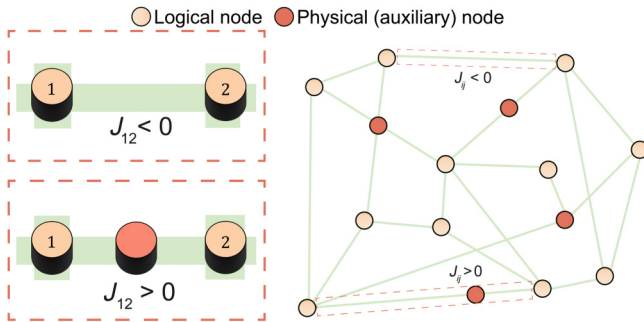


FIG. 5. Programming \pm weights. Auxiliary magnets are introduced to choose the polarity of the interaction strength J_{ij} , where the input current is fixed such that all magnets have negative correlation ($J_{ij} < 0$). The magnitude of J_{ij} is tuned by the channel length between any two spins (\hat{m}_i, \hat{m}_j).

weights can be obtained, where the role of auxiliary magnets is to have a double-negative effect such that $-(-J_{ij}) = +J_{ij}$. Keeping in mind that to ensure symmetric J_{ij} values, we inject the same charge current over all LBMs.

From a practical point of view, our programming approach is readily applicable for inference (or sampling) problems, where J_{ij} are fixed and do not need to change. This would also include optimization approaches such as parallel tempering. For optimization problems, where the system needs to be annealed by varying the temperature, e.g., simulated annealing, we can tune the system temperature for a given set of J_{ij} by gradually increasing the injected charge current I_c from a low to a high value. For machine learning applications where the system needs to update J_{ij} at each training step, using the same programmability approach would be challenging. One possibility for training the system without rebuilding the circuit at each training step is to introduce multiple ferromagnetic contacts that can inject currents with varying distances between electrodes.

VI. READOUT OF LBMS

An important practical consideration is the ability to read out the correlations induced in LBMs. Advances in modern spintronic capabilities offer various possibilities. We briefly consider three approaches that could be used.

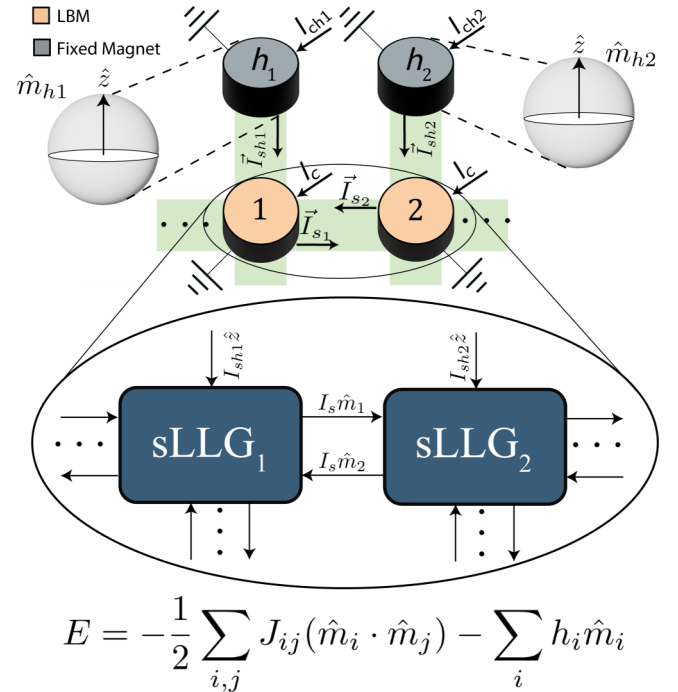


FIG. 6. Heisenberg machine with bias terms. For an LBM array, bias terms h_i can be introduced by stable magnets with a fixed magnetization direction \hat{m}_{hi} producing fixed spin current I_{shi} to incident magnets.

First, an experimentally demonstrated approach is using nearby ferromagnets for potentiometric readout, e.g., as commonly used for topological insulators (TIs) [37,38]. Secondly, inverse spin Hall effects in heavy metals and TIs can be used via terminals placed near the LBMs [39,40]. Finally, magnetic tunnel junctions (MTJs) built on top of the LBMs could be used for readout as commonly used in spin-orbit-torque-based structures [41]. This approach may require isolated and separate terminals for the MTJ readout and charge injection over the LBM similar to those in all-spin logic devices; this is potentially more challenging from a fabrication point of view [42].

In the schemes we proposed, readout of the full magnetization vector is not possible, and the methods we suggest can only read the projection of the magnet state along a specified direction. Even with this loss of information, the continuous nature of spins is preserved and the internal dynamical evolution of the system relies fully on the three-dimensional (3D) magnetization vectors.

VII. TRAINING HEISENBERG MACHINES

The energy model shown in Eq. (1) did not include bias terms, for simplicity. To our system of LBMs, bias terms h_i can be introduced using stable magnets with a fixed

magnetization, as shown in Fig. 6, i.e.,

$$E = -\frac{1}{2} \sum_{ij} J_{ij} (\hat{m}_i \cdot \hat{m}_j) - \sum_i h_i \hat{m}_i. \quad (22)$$

More details about Eq. (22), and how it breaks the zero net magnetization that was caused by the symmetry of Eq. (1), are provided in Appendix C. These bias terms make the energy model completely general, allowing us to construct networks where arbitrary correlations between the LBMs can be designed.

To show this, we train a network of LBMs using a continuous generalization of the contrastive divergence algorithm [43,44] (see Appendix D for details), where a Heisenberg XOR gate based on four coupled LBMs is obtained with three XOR spins and an auxiliary spin (Fig. 7). For this example, we used coupled sLLG equations without our full transport models for simplicity. Further, by linearly scaling the weights and biases with the dimensionless inverse algorithmic temperature β , we perform an annealing of the system energy. In practice, this could be done, for example, by increasing the charge current injected into the LBMs. Note that β describes only a universal scaling factor of the weights and biases; it is detached from the actual temperature T that determines the physical parameters I_0 .

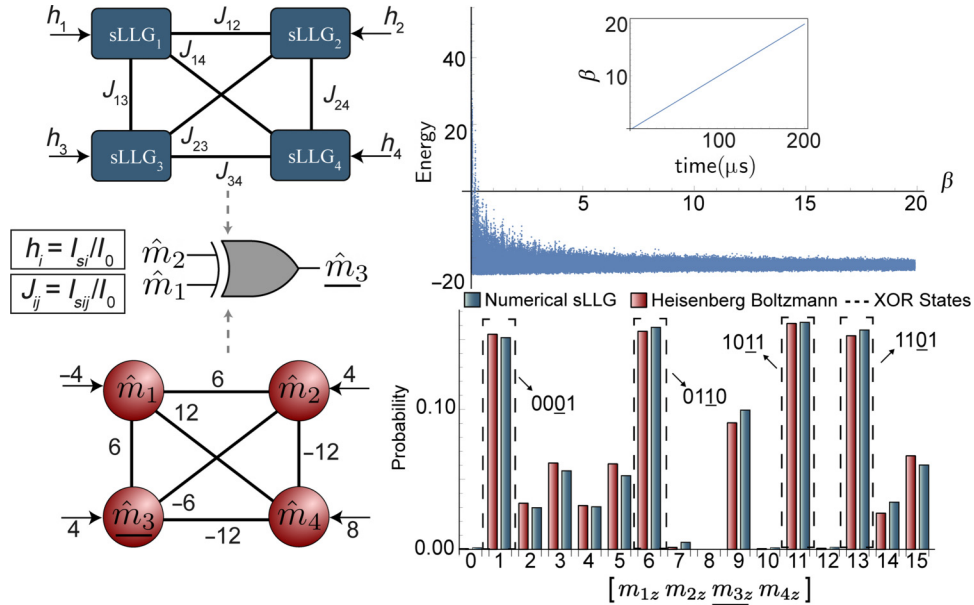


FIG. 7. Heisenberg XOR gate. Training four coupled LBMs modeled by sLLGs (\hat{m}_4 is an auxiliary spin). The interaction strength $J_{ij} = I_{sij}/I_0$ and bias terms $h_i = I_{si}/I_0$ (shown in the figure) are learned to implement a Heisenberg XOR gate; continuous spins are binarized by thresholding at zero. The network was annealed with a linear profile of the dimensionless inverse algorithmic temperature β , where β linearly scales the weights and the biases in the network. In physical implementations, β can be adjusted by changing the injected charge currents through each LBM. Results show saturation around the ground state of the network. The truth table of the XOR was verified numerically by the network response, with the correct states being {1, 6, 11, 13}, matching the expected response obtained by the Boltzmann law. For the histogram producing the truth table, the network was simulated at a constant $\beta = 4$. The value $I_0 \approx 125$ nA is fixed for all sLLG equations.

Our annealing result aligns well with the ground state; the error in the network minimum state is around 0.01% after 200 μs . We also obtain the full truth table for the XOR gate by sampling the network output at fixed β for 500 μs . To generate Boolean output values, continuous spins were binarized using thresholding at the zero point. No thresholding was performed during the simulation; however, all spins were binarized post-simulation. Figure 7 shows that the network response is identical to that obtained by the Boltzmann law, and in agreement with a probabilistic XOR gate operation. Further details are provided in Appendix D. In this example, we showed how the 3D dynamics of the Heisenberg model can be adjusted to build a probabilistic XOR gate with *Boolean* output states. The full continuous nature of our spins could be useful in the areas of stochastic computing where arithmetic operations such as multiplication, division, and factorization may be simplified by continuous stochastic neurons [45].

VIII. CONCLUSION

This work presents a programmable hardware platform to emulate the classical Heisenberg model using nonlocal spin valves. With analytical and numerical results, we establish that the physics of low-barrier nanomagnets with perpendicular magnetic anisotropy leads to a coupled system whose steady-state behavior is described by a Boltzmann factor, $\propto \exp(-\beta H)$, where H is the 3D classical Heisenberg Hamiltonian. With the tremendous current interest in building programmable computing hardware for the Ising model, the compact and energy-efficient realization of the classical Heisenberg model whose classical emulation would be much more costly than equivalent

Ising systems could be useful for a number of applications. These include training modern Hopfield networks to solve continuous optimization problems. We leave natural extensions of the concept to in-plane magnetic anisotropy magnets realizing the classical *XY* model for future investigation. We envision that spin-circuit networks with LBMs can be extended beyond conservative systems described by an energy, such as Bayesian (belief) networks with asymmetric network connections [36,46,47].

The spin-circuit codes used in this study are available on GitHub [48].

ACKNOWLEDGMENTS

We acknowledge support from ONR-MURI Grant No. N000142312708, OptNet: Optimization with p-Bit Networks. The authors are grateful to Shun Kanai, Saroj Dash, Punyashloka Debashis, and Zhihong Chen for fruitful discussions.

APPENDIX A: SPIN-CIRCUIT MODULES

Spin circuits [21,23–25] provide a generalization of ordinary charge circuits where each node in the circuit is represented by a four-dimensional voltage, $[V_c V_z V_x V_y]^T$, corresponding to three spin components (z, x, y) and one charge component. Figure 8 shows the details of the spin circuit for the two-LBM system considered in the main text. For the normal metal (NM) module, the series conductance G_{se} and the shunt conductance G_{sh} are defined as

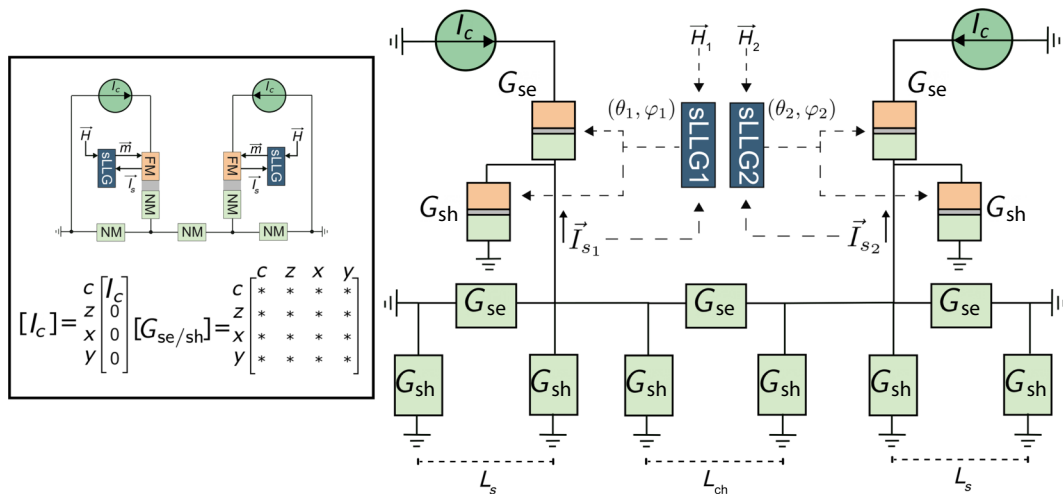


FIG. 8. Spin-circuit modules. The LBMs and the spin-neutral channel are described by a combination of series G_{se} and shunt G_{sh} conductances, where each conductance is a 4×4 matrix that takes into consideration the interaction between charges and spins in the z, x, y directions. The sLLG module solves the stochastic LLG equation as a function of the received spin current \vec{I}_s and thermal noise.

$$G_{\text{se}} = \left[\begin{array}{c|cccc} & c & z & x & y \\ \hline c & G_c & 0 & 0 & 0 \\ z & 0 & G_s & 0 & 0 \\ x & 0 & 0 & G_s & 0 \\ y & 0 & 0 & 0 & G_s \end{array} \right], \quad G_{\text{sh}} = \left[\begin{array}{c|cccc} & c & z & x & y \\ \hline c & 0 & 0 & 0 & 0 \\ z & 0 & G'_s & 0 & 0 \\ x & 0 & 0 & G'_s & 0 \\ y & 0 & 0 & 0 & G'_s \end{array} \right],$$

where we define $G_c = A_{\text{NM}}/(\rho_{\text{NM}}L)$, $G_s = A_{\text{NM}}/(\rho_{\text{NM}}\lambda_s)\text{csch}(L/\lambda_s)$, and $G'_s = A_{\text{NM}}/(\rho_{\text{NM}}\lambda_s)\tanh(L/2\lambda_s)$. Here A_{NM} denotes the NM's area, ρ_{NM} is the NM's resistivity, L is the NM's length, and λ_s is the NM's spin-diffusion length. Similarly, the shunt and series conductance for the FM-NM interface, if the magnet is pointing in the $+\hat{z}$ direction, are defined as

$$G_{\text{se}} = \left[\begin{array}{c|cccc} & c & z & x & y \\ \hline c & G & PG & 0 & 0 \\ z & PG & G & 0 & 0 \\ x & 0 & 0 & 0 & 0 \\ y & 0 & 0 & 0 & 0 \end{array} \right], \quad G_{\text{sh}} = \left[\begin{array}{c|cccc} & c & z & x & y \\ \hline c & 0 & 0 & 0 & 0 \\ z & 0 & 0 & 0 & 0 \\ x & 0 & 0 & aG & bG \\ y & 0 & 0 & -bG & aG \end{array} \right], \quad (\text{A1})$$

where G is the interface conductance, P is the interface polarization, and a and b are the real and imaginary coefficients of the spin-mixing conductance, respectively. The interface conductance can be rotated via $G_{\{\text{sh,se}\}} = [U_R]^T [G_{\{\text{sh,se}\}}] [U_R]$, where the rotation matrix $[U_R]$ is given by

$$\left[\begin{array}{c|cccc} & c & z & x & y \\ \hline c & 1 & 0 & 0 & 0 \\ z & 0 & \cos \theta & \sin \theta \cos \phi & \sin \theta \sin \phi \\ x & 0 & -\sin \theta \cos \phi & \cos \theta + \sin^2 \phi (1 - \cos \theta) & -\sin \phi \cos \phi (1 - \cos \theta) \\ y & 0 & -\sin \theta \sin \phi & \sin \phi \cos \phi (1 - \cos \theta) & \cos \theta + \cos^2 \phi (1 - \cos \theta) \end{array} \right].$$

In our numerical modules, the sLLG module provides the instantaneous magnetization directions for all LBMs $\hat{m}_i(\theta_i, \phi_i)$, and the circuit simulator rotates the conductances based on the new magnetization. On the other hand, the magnetizations are updated by solving the stochastic LLG based on the received spin currents \vec{I}_s along with the thermal noise that enters the effective fields \vec{H} :

$$(1 + \alpha^2) \frac{d\hat{m}_1}{dt} = -|\gamma|\hat{m}_1 \times \vec{H} - \alpha|\gamma|\hat{m}_1 \times (\hat{m}_1 \times \vec{H}) + \frac{\hat{m}_1 \times (\vec{I}_s \times \hat{m}_1)}{qN_s} + \frac{\alpha(\hat{m}_1 \times \vec{I}_s)}{qN_s}. \quad (\text{A2})$$

For LBMs with no effective anisotropy, the magnetic field \vec{H} is fully characterized by the thermal noise defined as the 3D uncorrelated magnetic fluctuation \vec{H}_n in the (z, x, y) directions:

$$\text{Var}(H_n^{z,x,y}) = \frac{2\alpha kT}{|\gamma|M_s\mathcal{V}}, \quad \mathbb{E}[H_n^{z,x,y}] = 0. \quad (\text{A3})$$

For these magnets, in the absence of the thermal noise, the sLLG equation (A2) will be reduced to Eq. (6).

The self-consistency between magnetism and transport is well defined because electronic timescales are much faster than magnetization dynamics. Therefore, at each discrete time point, a lumped circuit module for the transport can be defined based on the new magnetizations.

APPENDIX B: MAPPING CHARGE TO SPIN

Finding the exact analytical expression for the spin-to-charge mapping is challenging, but, in the case of a two-magnet system, this can be done analytically by a clever coordinate transformation and heavy algebraic manipulation. We first define a reference frame (z', x', y') , where the $+\hat{z}'$ axis always coincides with the fluctuating direction of the magnet \hat{m}_1 . We use this new coordinate system to transform the (z, x, y) coordinates of the channels and the interface matrix of the second magnet. Since the channels we consider in this paper are isotropic in spin (in the absence of any spin-orbit torques or directional spin relaxation), the channel conductances are unaffected by this transformation.

For the second magnet, we have $\hat{m}'_2 = R \hat{m}_2$, with

$$R = \begin{bmatrix} z & \cos \theta_1 & \sin \theta_1 \cos \phi_1 & \sin \theta_1 \sin \phi_1 \\ x & -\cos \varphi \sin \theta_1 & \cos \varphi \cos \theta_1 \cos \phi_1 - \sin \varphi \sin \phi_1 & \cos \varphi \cos \theta_1 \sin \phi_1 + \sin \varphi \cos \phi_1 \\ y & \sin \varphi \sin \theta_1 & -\cos \varphi \sin \phi_1 - \sin \varphi \cos \theta_1 \cos \phi_1 & \cos \varphi \cos \phi_1 - \sin \varphi \cos \theta_1 \sin \phi_1 \end{bmatrix}, \quad (\text{B1})$$

where the angle φ describes the applied rotation around the (x, y) plane. Note that R is described in (z', x', y') rather than (x', y', z') , since this is how we express the channel and interface matrices in the spin-circuit modules described in Eq. (A1). The new coordinates simplify the system for the purpose of finding the charge-to-spin conversion ratio, since now the coupled system reduces to the simpler configuration where one magnet (\hat{m}_1) is fixed to the $+\hat{z}'$ direction and injects a spin current of the form $\vec{I}_s = I_s \hat{z}'$.

Another key point that needs to be considered is extracting the component of incident spin currents along the direction of transmitting magnets. As such, we need to decompose incident spin currents (such as \vec{I}_{s2} in Fig. 8) into their constituents. For this purpose, we choose a commonly used nonorthogonal basis for the two-magnet system [32] ($\hat{m}_1, \hat{m}_2, \hat{m}_1 \times \hat{m}_2$). This allows us to clearly resolve the individual contributions of each magnet to incident spin currents. This new basis can be described by the transformation matrix A that turns (z, x, y) to $(\hat{m}_1, \hat{m}_2, \hat{m}_1 \times \hat{m}_2)$:

$$A = \begin{bmatrix} z & \cos \theta_1 & \cos \theta_2 & \sin \theta_1 \sin \theta_2 \sin(\phi_2 - \phi_1) \\ x & \sin \theta_1 \cos \phi_1 & \sin \theta_2 \cos \phi_2 & \cos \theta_2 \sin \theta_1 \sin \phi_1 - \cos \theta_1 \sin \theta_2 \sin \phi_2 \\ y & \sin \theta_1 \sin \phi_1 & \sin \theta_2 \sin \phi_2 & \sin \theta_2 \cos \theta_1 \cos \phi_2 - \sin \theta_1 \cos \theta_2 \cos \phi_1 \end{bmatrix}. \quad (\text{B2})$$

Then the spin-current component of interest is described by

$$\vec{I}_{s2} = (RA)^{-1} \vec{I}_{s2}(z', x', y') = I_{sm1} \hat{m}'_1 + I_{sm2} \hat{m}'_2 + I_{s\perp} (\hat{m}'_1 \times \hat{m}'_2). \quad (\text{B3})$$

It may seem that, in order to find an analytical expression for \vec{I}_{s2} , we must know (θ_2, ϕ_2) , since the conductance matrices for the second magnet are a function of (θ_2, ϕ_2) . However, when we assume $a = 1$, the spin-current component along the direction of \hat{m}_1 , described by I_{sm1} , is independent of the instantaneous direction of \hat{m}_2 . This allows us to calculate I_{sm1} , but in our new coordinate system (where $\hat{m}_1 = \hat{z}'$). The spin-current I_{s2} also needs to be expressed in (z', y', x') .

After these basis transformations and tedious algebra using standard circuit theory, we solve the two-magnet system and arrive at the analytical expression (keeping only leading-order terms for P , since typically $P^2 \ll 1$):

$$\frac{I_{sm1}}{I_c} = \frac{PR_{\text{int}}R_{\text{sp}}}{R_{\text{int}}^2 \text{csch}^2\left(\frac{L_s}{\lambda_s}\right) \sinh\left(\frac{L_{\text{ch}} + 2L_s}{\lambda_s}\right) + 2R_{\text{int}}R_{\text{sp}} \text{csch}\left(\frac{L_s}{\lambda_s}\right) \sinh\left(\frac{L_{\text{ch}} + L_s}{\lambda_s}\right) + R_{\text{sp}}^2 \sinh\left(\frac{L_{\text{ch}}}{\lambda_s}\right)}. \quad (\text{B4})$$

Here R_{sp} is defined as the resistance of a block of NM with length λ_s , i.e., $R_{\text{sp}} = \rho \lambda_s / A_{\text{NM}}$; R_{int} is the interface resistance, $1/G$; and L_{ch} and L_s are the side- and middle-channel lengths of the normal metal.

The total mapping factor for the proposed hardware then becomes $I_M = I_0(I_c/I_{sm1})$. We used I_M in Fig. 3 to relate our dimensionless Boltzmann models to our full numerical results and have obtained agreement. We also used I_M in Fig. 4 where the interaction strength $J_{12} = I_c/I_M$ was analyzed by sweeping L_{ch} .

It is instructive to consider the limits of Eq. (B4), assuming no spin relaxation, $\lambda_s \rightarrow \infty$:

$$\frac{I_{sm1}}{I_c} = P \left(\frac{R_{\text{side}}}{R_{\text{int}} + R_{\text{side}}} \right) \left(\frac{R_{\text{side}} \parallel R_{\text{int}}}{2(R_{\text{side}} \parallel R_{\text{int}}) + R_{\text{ch}}} \right), \quad (\text{B5})$$

where R_{side} and R_{ch} are the charge resistances of the side ($\rho L_s / A_{\text{NM}}$) and middle ($\rho L_{\text{ch}} / A_{\text{NM}}$) channels, respectively.

In our power calculation for significant saturation, we let $I_c = 15I_0(I_c/I_{sm1}) = 15I_M$ to obtain the $I_c^2 R$ dissipation, $R = R_{\text{side}} + R_{\text{int}}$; note that the term I_c/I_{sm1} has a fixed value for this two-LBM configuration. Pure spin-neutral channels can further optimize the power consumption of our proposed device, graphene being a good example [12,13].

Finally, we note that in the above analysis we added a special choice of the mixing conductance ($a = 1$), which leads to the (θ_2, ϕ_2) independence of I_{sm1} . For arbitrary choices of the mixing conductance, one needs to solve a

system of self-consistent equations for every new input I_c such that (θ_2, ϕ_2) and \vec{I}_{s2} agree.

APPENDIX C: HEISENBERG MACHINE WITH BIAS TERMS

In this section, we discuss how symmetry-breaking bias terms (h_i) can be added to the energy model so that we have

$$E = -\frac{1}{2} \sum_{i,j} J_{ij} (\hat{m}_i \cdot \hat{m}_j) - \sum_i h_i \hat{m}_i. \quad (\text{C1})$$

From a hardware point of view, these biases can be implemented using stable magnets with fixed magnetization states, rather than LBMs. Accordingly, when a charge current is supplied through one of the fixed magnets, the receiving LBM \hat{m}_i will receive a spin current I_{shi} with a fixed \hat{m}_{hi} direction, as we discussed in Fig. 6 in the main text. For subsequent discussions, we assume all fixed magnets are in the $+\hat{z}$ direction, though they can in general be arranged to be fixed in an arbitrary direction.

Next we show numerically that, with bias terms, an array of LBMs can have a finite net magnetization. We examine the bias effect on the frustrated system of three magnets, presented before in the main text [Fig. 3(d)]. Originally, the system had no biases and the net magnetization was

zero due to the \pm symmetry of the system, as shown in Fig. 9. We add bias terms to two of the LBMs at the same inverse temperature, where we observe a finite net magnetization (Fig. 9). In general, desired and arbitrary correlations between constituent magnets can be achieved by a judicious choice or training of these weights, which we discuss next.

APPENDIX D: TRAINING HEISENBERG MACHINES: PROOF OF CONCEPT

In this section, we show how to train Heisenberg machines through an appropriate generalization of the contrastive divergence algorithm, typically used to train Boltzmann machines with binary stochastic neurons. As a representative example, we choose an XOR gate (shown in Fig. 7). To obtain Boolean states, we take measurements along the z axis, and we binarize continuous spins by setting $m_{iz} = 0$ as the thresholding point between 0 and 1.

The thresholded probabilities of any logical state can then be obtained using the Boltzmann law by integrating over the corresponding range for all spins. For example, for a system of two continuous spins, the corresponding probabilities for the four possible states $\{00, 01, 10, 11\}$ are evaluated as follows:

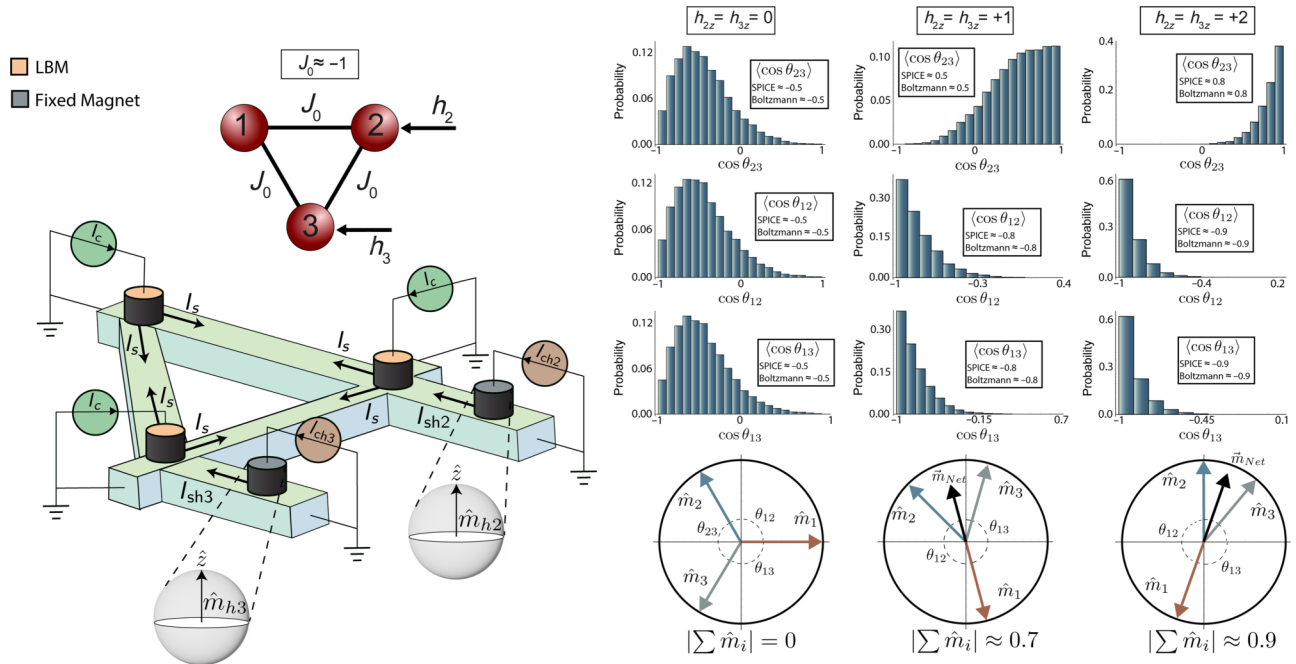


FIG. 9. Heisenberg machines with arbitrary correlations. Bias terms can be introduced by using stable magnets in the network. For this example, the interaction strength J_0 and biases h_i are scaled by a factor of $\beta = 10$. By projecting the magnetization vectors m_i to a unit circle (approximation not exact), we show that, in the absence of biases, the LBM array will have zero net magnetization, while arrays with biases will have finite net magnetization. General correlations between constituent magnets can be achieved this way.

$$P_{ij} = \int_{\phi_1=0}^{\phi_1=2\pi} \int_{\phi_2=0}^{\phi_2=2\pi} \int_{\theta_1=a}^{\theta_1=b} \int_{\theta_2=c}^{\theta_2=d} \frac{1}{Z} \exp(J_{12}\hat{m}_1 \cdot \hat{m}_2 + h_1\hat{m}_1 + h_2\hat{m}_2) \sin \theta_1 \sin \theta_2 d\theta_1 d\theta_2 d\phi_1 d\phi_2, \quad (\text{D1})$$

where the states are read as $[m_{1z}, m_{2z}]$. In this equation, $i, j \in \{0, 1\}$, and (a, b) or $(c, d) \rightarrow (0, \pi/2)$ when i or j is 1, and (a, b) or $(c, d) \rightarrow (\pi/2, \pi)$ when i or j is 0. For the partition function Z , no special adjustment is required; it is evaluated by doing the normal full integral over the four variables such that the summation of all probabilities add to 1.

The update rules for the weights (J_{ij}) and biases (h_i) can be obtained by generalizing the contrastive divergence algorithm commonly used for binary stochastic neurons [43,44]:

$$J_{ij} = J_{ij} + \varepsilon(\langle m_{iz}m_{jz} \rangle_{\text{data}} - \langle m_{iz}m_{jz} \rangle_{\text{model}}) - \varepsilon\lambda J_{ij}, \quad (\text{D2})$$

$$h_i = h_i + \varepsilon(\langle m_{iz} \rangle_{\text{data}} - \langle m_{iz} \rangle_{\text{model}}) - \varepsilon\lambda h_i, \quad (\text{D3})$$

where ε is the learning rate, and λ is the regularization factor. The correlation of the positive phase $\langle m_{iz}m_{jz} \rangle_{\text{data}}$ corresponds to clamping the spins to the XOR truth table, while the negative phase correlation $\langle m_{iz}m_{jz} \rangle_{\text{model}}$ refers to the inference stage. Just as in the typical contrastive divergence algorithm, the correlations can be obtained with probabilistic sampling, though in this case we obtained the correlations exactly using the Boltzmann law at each iteration. At the end of training, we performed minor fine tuning of the weights to get near-integer values. In our XOR example, we defined the correct states to be $\{1, 6, 11, 13\}$, such as (m_{1z}, m_{2z}, m_{3z}) are the two inputs and the output of the XOR gate, respectively. The fourth spin m_{4z} is an auxiliary or hidden spin state.

TABLE I. Physical parameters used in the simulations.

Parameter	Value	Unit
Interface polarization (P)	0.1	—
Gilbert damping coefficient (α)	0.01	—
Saturation magnetization (M_s)	1100×10^3	A/m
Magnet volume (\mathcal{V})	$30 \times 30 \times 2$	nm^3
Interface conductance (G)	1	S
Spin-mixing conductance, real part (aG)	1	—
Spin-mixing conductance, imaginary part (bG)	0	—
NM spin-diffusion length (λ_s)	400	nm
NM resistivity (ρ_{NM})	2.35	$\mu\Omega \text{ cm}$
NM length ($L = L_s = L_{\text{ch}}$)	200	nm
NM area (A_{NM})	1.11×10^{-14}	m^2
Temperature	300	K
Transient time step (SPICE)	10	ps

APPENDIX E: PHYSICAL PARAMETERS

The physical parameters we used in our simulations are reported in Table I, where the parameters of the NM channel are chosen according to experiments performed in metallic nonlocal spin valves [34]. The transient noise (.trannoise) function of HSPICE has been used to solve the stochastic differential equations. This solver has been rigorously benchmarked with exact time-dependent Fokker-Planck equations in Ref. [30].

- [1] Y. W. Teh, M. Welling, S. Osindero, and G. E. Hinton, Energy-based models for sparse overcomplete representations, *J. Mach. Learn. Res.* **4**, 1235 (2003).
- [2] N. G. Berloff, M. Silva, K. Kalinin, A. Askitopoulos, J. D. Töpfer, P. Cilibrizzi, W. Langbein, and P. G. Lagoudakis, Realizing the classical XY Hamiltonian in polariton simulators, *Nat. Mater.* **16**, 1120 (2017).
- [3] P. Huembeli, J. M. Arrazola, N. Killoran, M. Mohseni, and P. Wittek, The physics of energy-based models, *Quantum Mach. Intell.* **4**, 1 (2022).
- [4] N. S. Singh, K. Kobayashi, Q. Cao, K. Selcuk, T. Hu, S. Niazi, N. A. Aadit, S. Kanai, H. Ohno, S. Fukami, and K. Y. Camsari, CMOS plus stochastic nanomagnets enabling heterogeneous computers for probabilistic inference and learning, *Nat. Commun.* **15**, 46645 (2024).
- [5] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, Integer factorization using stochastic magnetic tunnel junctions, *Nature* **573**, 390 (2019).
- [6] N. Mohseni, P. L. McMahon, and T. Byrnes, Ising machines as hardware solvers of combinatorial optimization problems, *Nat. Rev. Phys.* **4**, 363 (2022).
- [7] H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, L. Gruber, M. Holzleitner, T. Adler, D. Kreil, M. K. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, in *International Conference on Learning Representations* (Vienna, Austria (Online), 2021).
- [8] D. Krotov, A new frontier for Hopfield networks, *Nat. Rev. Phys.* **5**, 366 (2023).
- [9] M. Widrich, B. Schäfl, M. Pavlović, H. Ramsauer, L. Gruber, M. Holzleitner, J. Brandstetter, G. K. Sandve, V. Greiff, S. Hochreiter, and G. Klambauer, in *Advances in Neural Information Processing Systems*, Vol. 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Curran Associates, 2020), p. 18832.
- [10] R. Ishihara, Y. Ando, S. Lee, R. Ohshima, M. Goto, S. Miwa, Y. Suzuki, H. Koike, and M. Shiraishi, Gate-tunable spin XOR operation in a silicon-based device at room temperature, *Phys. Rev. Appl.* **13**, 044010 (2020).
- [11] J. Panda, M. Ramu, O. Karis, T. Sarkar, and M. V. Kamalakar, Ultimate spin currents in commercial chemical vapor deposited graphene, *ACS Nano* **14**, 12771 (2020).

- [12] T. Bisswanger, Z. Winter, A. Schmidt, F. Volmer, K. Watanabe, T. Taniguchi, C. Stampfer, and B. Beschoten, CVD bilayer graphene spin valves with 26 μm spin diffusion length at room temperature, *Nano Lett.* **22**, 4949 (2022).
- [13] D. Khokhriakov, S. Sayed, A. M. Hoque, B. Karpiak, B. Zhao, S. Datta, and S. P. Dash, Multifunctional spin logic operations in graphene spin circuits, *Phys. Rev. Appl.* **18**, 064063 (2022).
- [14] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, Stochastic p-Bits for Invertible Logic, *Phys. Rev. X* **7**, 031014 (2017).
- [15] K. Y. Camsari, S. Salahuddin, and S. Datta, Implementing p-bits with embedded MTJ, *IEEE Electron Device Lett.* **38**, 1767 (2017).
- [16] K. Hayakawa, S. Kanai, T. Funatsu, J. Igarashi, B. Jinnai, W. A. Borders, H. Ohno, and S. Fukami, Nanosecond Random Telegraph Noise in In-Plane Magnetic Tunnel Junctions, *Phys. Rev. Lett.* **126**, 117202 (2021).
- [17] C. Safranski, J. Kaiser, P. Trouilloud, P. Hashemi, G. Hu, and J. Z. Sun, Demonstration of nanosecond operation in stochastic magnetic tunnel junctions, *Nano Lett.* **21**, 2040 (2021).
- [18] L. Schnitzspan, M. Kläui, and G. Jakob, Nanosecond true-random-number generation with superparamagnetic tunnel junctions: Identification of Joule heating and spin-transfer-torque effects, *Phys. Rev. Appl.* **20**, 024002 (2023).
- [19] J. Kaiser, W. A. Borders, K. Y. Camsari, S. Fukami, H. Ohno, and S. Datta, Hardware-aware in situ learning based on stochastic magnetic tunnel junctions, *Phys. Rev. Appl.* **17**, 014016 (2022).
- [20] A. Grimaldi, K. Selcuk, N. A. Aadit, K. Kobayashi, Q. Cao, S. Chowdhury, G. Finocchio, S. Kanai, H. Ohno, S. Fukami, and K. Y. Camsari, in *2022 International Electron Devices Meeting (IEDM)* (IEEE, San Francisco, CA, USA, 2022), p. 22.4.1.
- [21] K. Y. Camsari, S. Ganguly, and S. Datta, Modular approach to spintronics, *Sci. Rep.* **5**, 10571 (2015).
- [22] S. Sayed, V. Q. Diep, K. Y. Camsari, and S. Datta, Spin funneling for enhanced spin injection into ferromagnets, *Sci. Rep.* **6**, 28868 (2016).
- [23] A. Brataas, G. E. Bauer, and P. J. Kelly, Non-collinear magnetoelectronics, *Phys. Rep.* **427**, 157 (2006).
- [24] S. Srinivasan, V. Diep, B. Behin-Aein, A. Sarkar, and S. Datta, in *Handbook of Spintronics*, edited by Y. Xu, D. Awschalom, and J. Nitta (Springer Netherlands, 2016), p. 1281.
- [25] S. Manipatruni, D. E. Nikonov, and I. A. Young, Modeling and design of spintronic integrated circuits, *IEEE Trans. Circuits Syst. I Regul. Pap.* **59**, 2801 (2012).
- [26] R. Graham and T. Tél, Existence of a Potential for Dissipative Dynamical Systems, *Phys. Rev. Lett.* **52**, 9 (1984).
- [27] W. F. Brown, Jr., Thermal fluctuations of a single-domain particle, *Phys. Rev.* **130**, 1677 (1963).
- [28] Z. Li and S. Zhang, Thermally assisted magnetization reversal in the presence of a spin-transfer torque, *Phys. Rev. B* **69**, 134416 (2004).
- [29] W. H. Butler, T. Mewes, C. K. A. Mewes, P. B. Visscher, W. H. Rippard, S. E. Russek, and R. Heindl, Switching distributions for perpendicular spin-torque devices within the macrospin approximation, *IEEE Trans. Magn.* **48**, 4684 (2012).
- [30] M. M. Torunbalci, P. Upadhyaya, S. A. Bhave, and K. Y. Camsari, Modular compact modeling of MTJ devices, *IEEE Trans. Electron Devices* **65**, 4628 (2018).
- [31] T. Kimura, Y. Otani, and J. Hamrle, Switching Magnetization of a Nanoscale Ferromagnetic Particle using Nonlocal Spin Injection, *Phys. Rev. Lett.* **96**, 037201 (2006).
- [32] D. Datta, B. Behin-Aein, S. Datta, and S. Salahuddin, Voltage asymmetry of spin-transfer torques, *IEEE Trans. Nanotechnol.* **11**, 261 (2011).
- [33] P. Debashis, H. Li, D. Nikonov, and I. Young, Gaussian random number generator with reconfigurable mean and variance using stochastic magnetic tunnel junctions, *IEEE Magn. Lett.* **13**, 1 (2022).
- [34] T. Kimura, T. Sato, and Y. Otani, Temperature Evolution of Spin Relaxation in a NiFe/Cu Lateral Spin Valve, *Phys. Rev. Lett.* **100**, 066602 (2008).
- [35] G. E. Hinton, T. J. Sejnowski, and D. H. Ackley, Boltzmann machines: Constraint satisfaction networks that learn, Technical Report, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA (1984).
- [36] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, A learning algorithm for Boltzmann machines, *Cogn. Sci.* **9**, 147 (1985).
- [37] S. Hong, V. Diep, S. Datta, and Y. P. Chen, Modeling potentiometric measurements in topological insulators including parallel channels, *Phys. Rev. B* **86**, 085131 (2012).
- [38] J. Kim, C. Jang, X. Wang, J. Paglione, S. Hong, J. Lee, H. Choi, and D. Kim, Electrical detection of the surface spin polarization of the candidate topological Kondo insulator SmB_6 , *Phys. Rev. B* **99**, 245148 (2019).
- [39] W. Y. Choi, I. C. Arango, V. T. Pham, D. C. Vaz, H. Yang, I. Groen, C.-C. Lin, E. S. Kabir, K. Oguz, P. Debashis, *et al.* All-electrical spin-to-charge conversion in sputtered $\text{Bi}_x\text{Se}_{1-x}$, *Nano Lett.* **22**, 7992 (2022).
- [40] V. T. Pham, I. Groen, S. Manipatruni, W. Y. Choi, D. E. Nikonov, E. Sagasta, C.-C. Lin, T. A. Gosavi, A. Marty, L. E. Hueso, *et al.* Spin-orbit magnetic state readout in scaled ferromagnetic/heavy metal nanostructures, *Nat. Electron.* **3**, 309 (2020).
- [41] 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* **336**, 555 (2012).
- [42] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, Proposal for an all-spin logic device with built-in memory, *Nat. Nanotechnol.* **5**, 266 (2010).
- [43] G. E. Hinton, Training products of experts by minimizing contrastive divergence, *Neural Comput.* **14**, 1771 (2002).
- [44] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, in *Proceedings of the 24th International Conference on Machine Learning* (Association for Computing Machinery, New York, NY, USA, 2007), p. 473.

- [45] A. Alaghi and J. P. Hayes, Survey of stochastic computing, *ACM Trans. Embedded Comput. Syst. (TECS)* **12**, 1 (2013).
- [46] K.-E. Harabi, T. Hirtzlin, C. Turck, E. Vianello, R. Laurent, J. Droulez, P. Bessière, J.-M. Portal, M. Bocquet, and D. Querlioz, A memristor-based Bayesian machine, *Nat. Electron.* **5**, 88 (2022).
- [47] P. Debashis, V. Ostwal, R. Faria, S. Datta, J. Appenzeller, and Z. Chen, Hardware implementation of Bayesian network building blocks with stochastic spintronic devices, *Sci. Rep.* **10**, 72842 (2020).
- [48] S. Bunaiyan and K. Y. Camsari, Heisenberg machines (HSPICE code), 2023, available online at <https://github.com/OPUSLab/HeisenbergMachines>.