Infomorphic Computing Enabled by Spintronic Devices Toward Localized and Interpretable Machine Learning

Weisheng Zhao

weisheng.zhao@buaa.edu.cn

Beihang University https://orcid.org/0000-0001-8088-0404

Ziyi Teng

Beihang University

Jiaxuan Chen

Beihang University

Yining Zhao

Beihang University

Yihao Zhang

Beihang University

Huijun Li

Beihang University

Zhaochun Liu

Beihang University

Zhihua Xiao

The Hong Kong University of Science and Technology

Qiming Shao

Hong Kong University of Science and Technology https://orcid.org/0000-0003-2613-3031

Yuxuan Yao

Beihang University https://orcid.org/0000-0003-0478-1601

Hongchao Zhang

Truth Memory Corporation

Hongxi Liu

Truth Memory Corporation https://orcid.org/0009-0004-1562-2539

Shouzhong Peng

Beihang University

Sai Li

Beihang University

Article

Keywords:

Posted Date: June 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9863940/v1

Additional Declarations: There is **NO** Competing Interest.

Infomorphic Computing Enabled by Spintronic Devices Toward Localized and Interpretable Machine Learning

Ziyi Teng[1,2,*], Jiaxuan Chen[1,2,*], Yining Zhao[1,2], Yihao Zhang[1,2], Huijun Li[1,2], Zhaochun Liu[1,2], Zhihua Xiao[3], Qiming Shao[3], Yuxuan Yao[1,2], Hongchao Zhang[4], Hongxi Liu[4], Shouzhong Peng[1,2,#], Sai Li[1,2,#], Weisheng Zhao[1,2,#]

[1]School of Integrated Circuit Science and Engineering, Beihang University, Beijing 100191, China.

[2]National Key Lab of Spintronics, Institute of International Innovation, Beihang University, Hangzhou, 311115, China.

[3]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.

[4]Truth Memory Corporation, Beijing, 100083, China.

[*]These authors contributed equally: Ziyi Teng, Jiaxuan Chen

[#]Corresponding author. Email: shouzhong.peng@buaa.edu.cn, saili@buaa.edu.cn, weisheng.zhao@buaa.edu.cn

Abstract

Infomorphic computing, grounded in partial information decomposition, offers a unified and mathematically interpretable framework for neuron-autonomous local learning. Its neuronal units operate through dual-input, nonlinear and tunably stochastic dynamics, inherently eliminating the cross-layer error propagation and associated energy overhead of backpropagation. However, the lack of hardware primitives that natively embody these dynamics remains a critical bottleneck, preventing its intrinsic algorithmic advantages from translating into practical hardware efficiency. Here, we demonstrate, to our knowledge, the first physical infomorphic neuron using CMOS-compatible spintronic devices, specifically voltage-gated spin-orbit-torque magnetic tunnel junctions (VGSOT-MTJs). We show that the device physics maps natively onto the core information-theoretic operation, establishing what we define as Infomorphic Nonlinear Probabilistic Activation (INPA): the SOT current and gate voltage encode receptive (R) and contextual (C) signals, respectively, jointly modulating the MTJ switching probability to yield a probabilistic binary output. Using this native primitive, we build spintronic infomorphic neural networks capable of supervised learning, unsupervised learning, and associative memory tasks with fully local updates, requiring only ~0.5% of the update computation compared to backpropagation method in the same task. Each INPA operation consumes ~15.86 pJ within 1.6 ns using the VGSOT-MTJ device. At the system level, our laboratory prototype already achieves ~400× higher energy efficiency than GPU-based execution, with projected gains reaching up to six orders of magnitude upon spin-CMOS ASIC integration. This work establishes a scalable hardware foundation for infomorphic computing, positioning spintronics as a key

enabler for next-generation energy-efficient machine intelligence with clear mathematical interpretability and minimal resource overhead.

Introduction

Backpropagation-trained deep networks have fundamentally reshaped scientific discovery and industrial applications[1–6], yet their future scaling is increasingly constrained by energy consumption, data movement, and limited interpretability. As foundation models scale toward the trillion-parameter regime, they are increasingly constrained by two major barriers. First, the "energy wall": training frontier-scale models already requires electricity at the GWh scale[7], and the high energy consumption and costs have exceeded the limits of most institutions[8]. This inefficiency stems from the von Neumann bottleneck, where repeated movement of synaptic weights can consume orders of magnitude more energy than the computation itself.[9,10] Second, the "interpretability wall": standard deep networks operate as opaque "black boxes", where the global minimization of error offers no insight into the internal logic of decision-making[11,12]. Consequently, merely accelerating current architectures is insufficient; overcoming these bottlenecks requires a fundamental paradigm shift toward hardware-native learning rules that are intrinsically local, physically efficient, and mathematically transparent[13].

Central to this paradigm is local learning, where each unit updates its parameters using locally available signals, minimizing global coordination and aligning naturally with parallel and distributed physical substrates[9,14–18]. Yet local learning has not become a general-purpose alternative to backpropagation. Classical biologically inspired rules (e.g., Hebbian learning[19] and its variants), while plausible, consistently fail to scale to deep networks with competitive performance. Conversely, modern alternatives, such as contrastive learning[20,21], local information maximization[22] and predictive coding[23], often rely on task-specific engineering, requiring auxiliary projection heads, negative sampling strategies, or complex error-routing circuits that complicate physical realizability and limit generalization. A remaining challenge is to identify a principled and hardware-compatible local objective that can be evaluated using only neuron-local signals while preserving interpretability.

This gap points to a more fundamental question in machine intelligence: whether learning objectives can be specified at the level of individual computational units, grounded in information-theoretic first principles, and implemented without global error propagation. Infomorphic computing has recently emerged as a rigorous framework that addresses this question by defining neuron-level objectives directly from local input-output statistics.[24]. In this framework, the Infomorphic Neuron (IN) serves as the fundamental information-processing unit (Fig. 1b). Each neuron produces a stochastic output Y whose activation probability is nonlinearly modulated by its two local inputs R and C, and the neuron's output statistics reflect the interaction between the input signals. This input-output relationship defines a joint distribution $\mathrm{P}(\mathrm{R},\mathrm{C},\mathrm{Y})$ entirely from local signals, which is then decomposed by Partial Information Decomposition (PID)[25] into five information atoms describing distinct contributions of the inputs to the neuronal response.[26,27] These atoms are combined into a neuron-level goal function G, and each neuron independently maximizes its own G using only local signals, thereby enabling fully local weight updates without cross-layer error

propagation (Fig. 1d). This PID-based formulation provides a first-principles information-theoretic description of neuron-level learning objectives and links biological notions of local adaptation to scalable artificial networks[28].

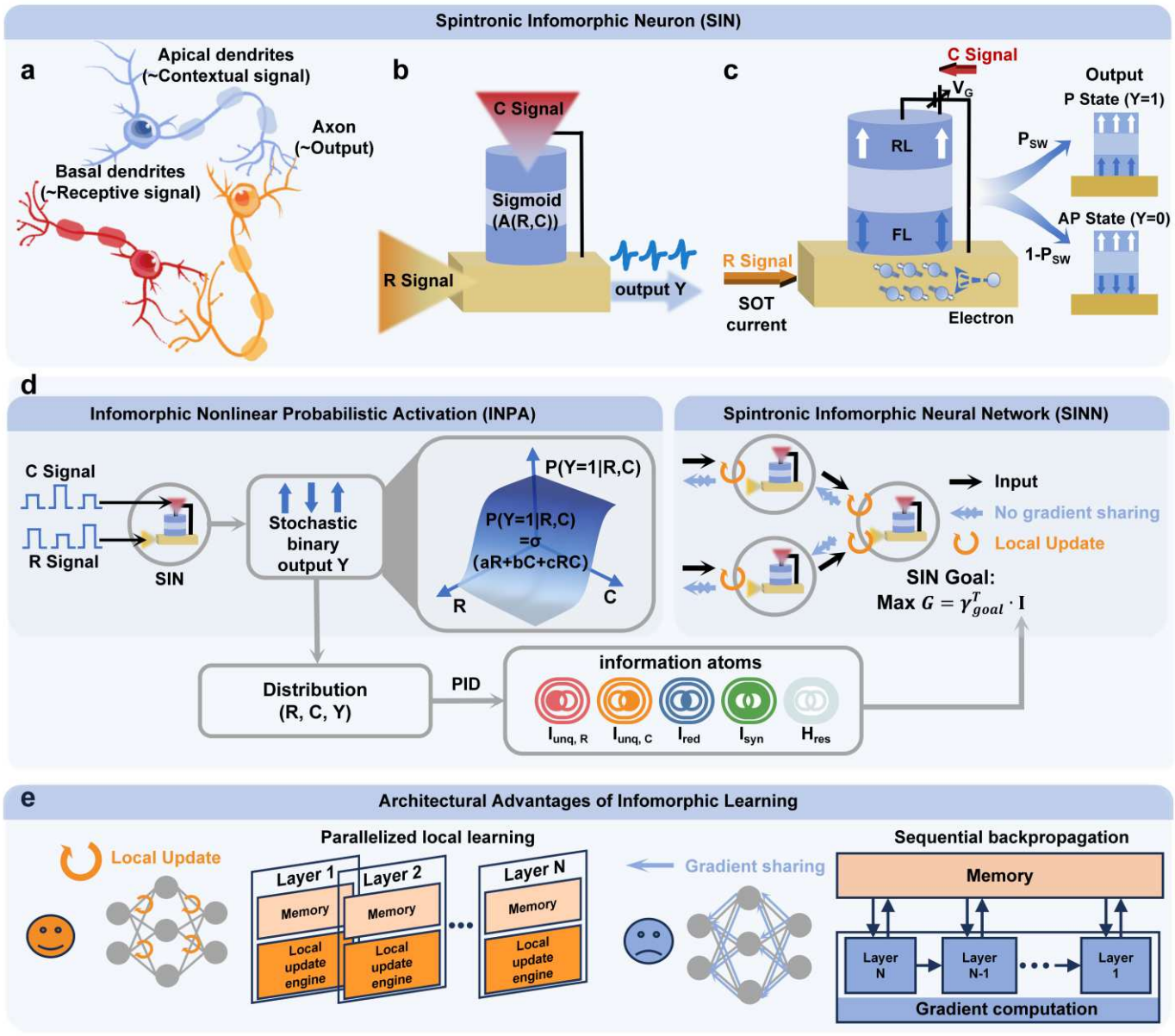

Fig. 1 | Spintronic infomorphic neuron and network architecture. a, Schematic of a cortical pyramidal neuron, where basal dendrites receive R signals that primarily drive neuronal output, while apical dendrites convey C signals that modulate neuronal response. **b,** Conceptual model of an infomorphic neuron. R and C signals jointly determine the probability of a binary neuronal output through a nonlinear activation function followed by a sigmoid transformation. **c,** Spintronic realization of an infomorphic neuron using a VGSOT-MTJ. The R signal is encoded by the spin-orbit-torque current, while the C signal is applied through the gate voltage, together modulating the stochastic switching probability of the MTJ and generating a probabilistic binary output. **d,** INPA and information-based local learning. Each

spintronic SIN produces a stochastic binary output, and the joint distribution of R signal, C signal and output is decomposed into unique ($I_{unq,R}, I_{unq,C}$), redundant (I_{red}), synergistic (I_{syn}) contributions and residual entropy (H_{res}) via PID, which define a neuron-local learning objective. **e,** SINN with intrinsically local and parallel updates. Each SIN optimizes its own local objective without global gradient sharing, eliminating the need for global backpropagation and enabling fast and power-efficient infomorphic learning.

Despite its theoretical elegance, the physical implementation of the IN remains challenging because the model imposes stringent structural and functional requirements. Structurally, the IN is inspired by the compartmental organization of cortical pyramidal neurons[29] (Fig. 1a), separately integrating two distinct classes of input signals: basal dendrites that integrate receptive (feedforward) signals, and apical dendrites that convey contextual signals[30], including feedback[31,32], labels[19], or rewards[33], which modulate the learning process[34,35]. Functionally, the IN requires two non-trivial computational operations. First, it requires nonlinear information integration, in which R and C inputs are combined through multiplicative interaction terms, rather than through a purely additive weighted sum. Second, it requires probabilistic activation, in which the output is mapped through a sigmoid function and sampled via a Bernoulli process, yielding intrinsically stochastic binary responses (Fig. 1b). These requirements can be implemented in CMOS, but typically at the cost of additional peripheral circuits for nonlinear computation, stochastic sampling, and signal routing, which can increase system overhead. The resulting peripheral overhead and data movement can outweigh the cost of the core computation itself[36–38]. This mismatch creates a significant “implementation gap”. As a result, infomorphic local learning has so far been limited to theory, hindering its scalable physical realization and practical deployment.

In this work, we demonstrate that the voltage-gated spin-orbit-torque magnetic tunnel junction (VGSOT-MTJ) provides a physical substrate suited to the structural and functional requirements of the IN, combining two independent control ports with intrinsically stochastic MTJ switching dynamics[39–41]. We experimentally realize a Spintronic Infomorphic Neuron (SIN), in which the SOT current encodes the R signal and the gate voltage (V_{G}) encodes the C signal. These two inputs jointly modulate the probabilistic switching dynamics of the free-layer magnetization, producing a nonlinear, continuously tunable switching probability that directly defines the binary probabilistic neuronal output Y (Fig. 1c). We term this device-level behavior Infomorphic Nonlinear Probabilistic Activation (INPA), which serves as the core functional primitive of the SIN. The joint probability distribution $p(R, C, Y)$ provides the statistical basis for PID, enabling local learning objectives to be constructed from five information atoms that drive weight updates (Fig. 1d). Building on this physical primitive, we construct spintronic infomorphic neural networks (SINNs) that enables fully parallel local updates without cross-layer gradient backpropagation and demonstrate supervised learning, unsupervised learning, and associative memory tasks (Fig. 1e; Supplementary Fig. 4 for the local-update workflow comparison with conventional backpropagation). This work establishes the first device realization of infomorphic

computing, bridging information-theoretic learning principles with device physics and laying a hardware foundation for high-performance infomorphic chips.

Results

Device characterization and probabilistic switching

The SIN device primitive requires both reliable binary-state readout and probabilistic switching controlled by two independent terminals. We therefore characterize the VGSOT-MTJ that serves as the device primitive of the SIN. The device employs a synthetic antiferromagnetic (SAF)/CoFeB/MgO/CoFeB/W stack (Fig. 2a) and exhibits clear perpendicular magnetic anisotropy, as confirmed by its magnetic hysteresis loop. A tunneling magnetoresistance (TMR) ratio of 119% (Fig. 2b) ensures robust electrical discrimination between the two magnetic states. Voltage-gated SOT switching under various gate voltages (Fig. 2c) shows that applying a positive gate voltage reduces the critical switching voltage by approximately 25% at V_G = 0.5 V, confirming that the voltage-controlled magnetic anisotropy (VCMA) effect modulates the energy barrier for SOT-driven magnetization reversal.

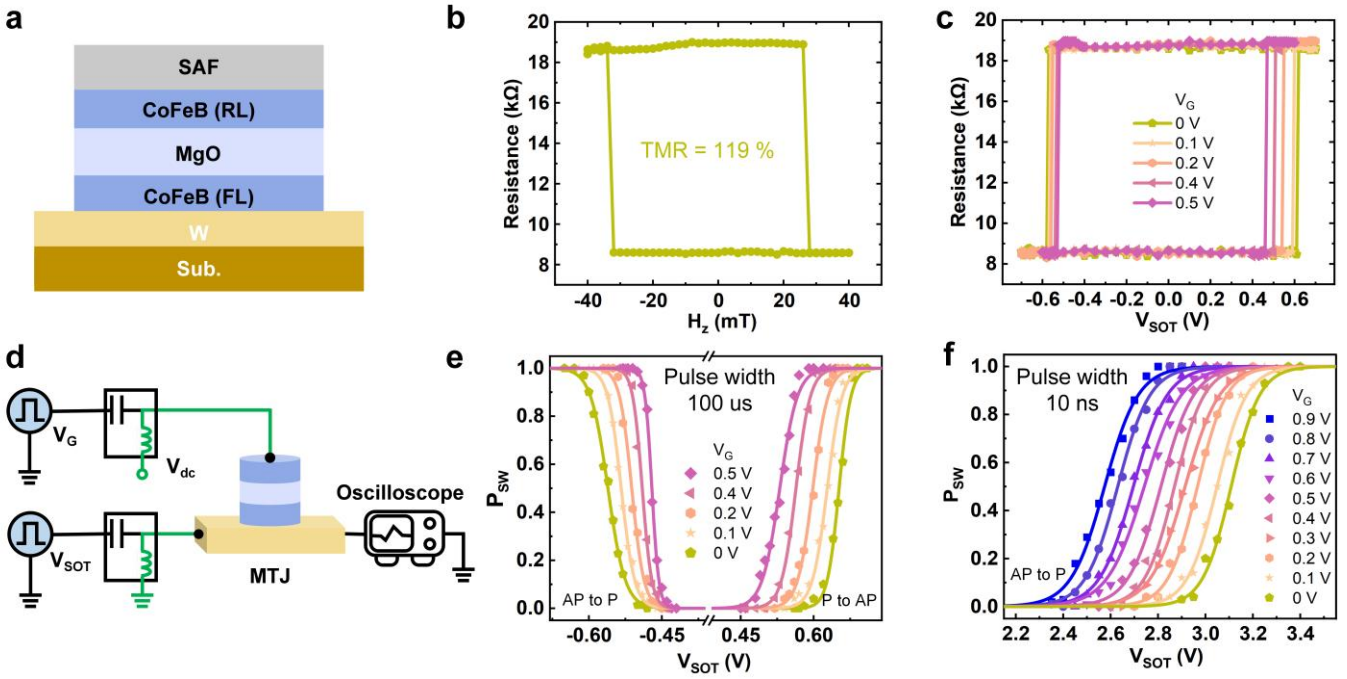

Fig. 2 | Device characterization and probabilistic switching behavior of the VGSOT–MTJ used as the SIN device primitive. a, Schematic of the film stack of the fabricated VGSOT–MTJ device. **b**, TMR measurement under perpendicular magnetic fields, showing a TMR of 119%. **c**, VGSOT switching under different gate voltages V_G, demonstrating gate-controlled modulation of the critical SOT switching voltage V_{SOT}. **d**, Experimental setup used for probabilistic switching measurements. **e**, Magnetization switching probability P_{SW} as a function of V_{SOT} for various V_G, measured with 100 μs pulses and averaged over 100 switching cycles. **f**, Magnetization switching probability P_{SW} as a function of V_{SOT} for different V_G, measured with 10 ns short pulses and averaged over 100 switching cycles.

To evaluate the stochastic switching behavior, we measure switching probabilities under pulsed excitation (Fig. 2d). A switching event induced by a single pulse is encoded as "1" (high state), whereas no switching is encoded as "0" (low state). Thus, each pulse directly yields a binary stochastic output at the device level, consistent with a Bernoulli process. For long pulses (100 μs), we characterize both antiparallel-to-parallel (AP-to-P) and parallel-to-antiparallel (P-to-AP) transitions and estimate the switching probability P_{SW} by averaging over 100 cycles (Fig. 2e), confirming well-defined probabilistic switching in both magnetization-reversal directions. We then examine operation in the nanosecond regime using 10 ns pulses (Fig. 2f). Although both switching directions exhibit well-defined probabilistic behavior under long pulses, the SIN defines its binary neuronal output using the AP-to-P switching branch.

We therefore focus on this transition for the subsequent extraction of the activation function. As shown in Fig. 2f, the P_{SW} measured with 10 ns pulses across a series of gate voltages is well described by sigmoid functions and spans the full probability range from 0 to 1. Under a given input condition, each pulse therefore produces a binary stochastic output with a continuously tunable Bernoulli probability, providing the device-level basis for the subsequent activation-function extraction. The detailed measurement procedure for each probability data point is provided in Supplementary Fig. 1.

Device-to-device variation is a critical factor in determining overall network reliability and learning accuracy, so we extended our characterization to an array of fabricated devices. Supplementary Fig. 2 presents the probabilistic switching behavior of six devices randomly selected from the fabricated VGSOT-MTJ array. All devices exhibit highly consistent sigmoidal $P_{\mathrm{SW}} - V_{\mathrm{SOT}}$ curves and minimal device-to-device variation, with an average P_{SW} standard deviation of $\sigma_{SW} = 0.031$.

Together, these results show that the VGSOT-MTJ combines reliable binary readout, gate-tunable probabilistic switching and sigmoid-like switching statistics in the nanosecond regime, thereby establishing the device basis for implementing the SIN.

VGSOT-MTJ-based SIN module design

The core operation of the SIN is the INPA, a dual-input probabilistic operation that combines nonlinear input integration with stochastic binary output generation. Here we use supervised learning on the MNIST handwritten-digit dataset as a representative example to illustrate the mapping between the abstract INPA process and its device-level implementation (Fig. 3a). In this case, the pixel intensity of a digit image (e.g., "7") serves as the receptive signal R, whereas its label information provides the contextual signal C. These two inputs jointly determine the neuronal activation $A(R, C)$, which is transformed by a sigmoid function $\sigma(A)$ into an output probability. A Bernoulli process then generates a binary output $Y \in \{0,1\}$, completing the INPA of the infomorphic neuron.

As shown in Fig. 3b, the VGSOT-MTJ directly implements INPA in hardware. The two inputs play distinct physical roles: the SOT pulse provides the direct drive for magnetization reversal and thus serves as the receptive pathway, whereas the gate voltage primarily modulates the anisotropy energy barrier through the VCMA effect, thereby reshaping the switching-probability landscape and serving as the contextual pathway. Their cooperative action

determines the switching probability between the P and AP states, which are assigned to Y=1 and Y=0, respectively. Together, these two physical inputs realize a nonlinear probabilistic activation, while the intrinsic stochasticity of magnetization dynamics produces the binary neuronal output.

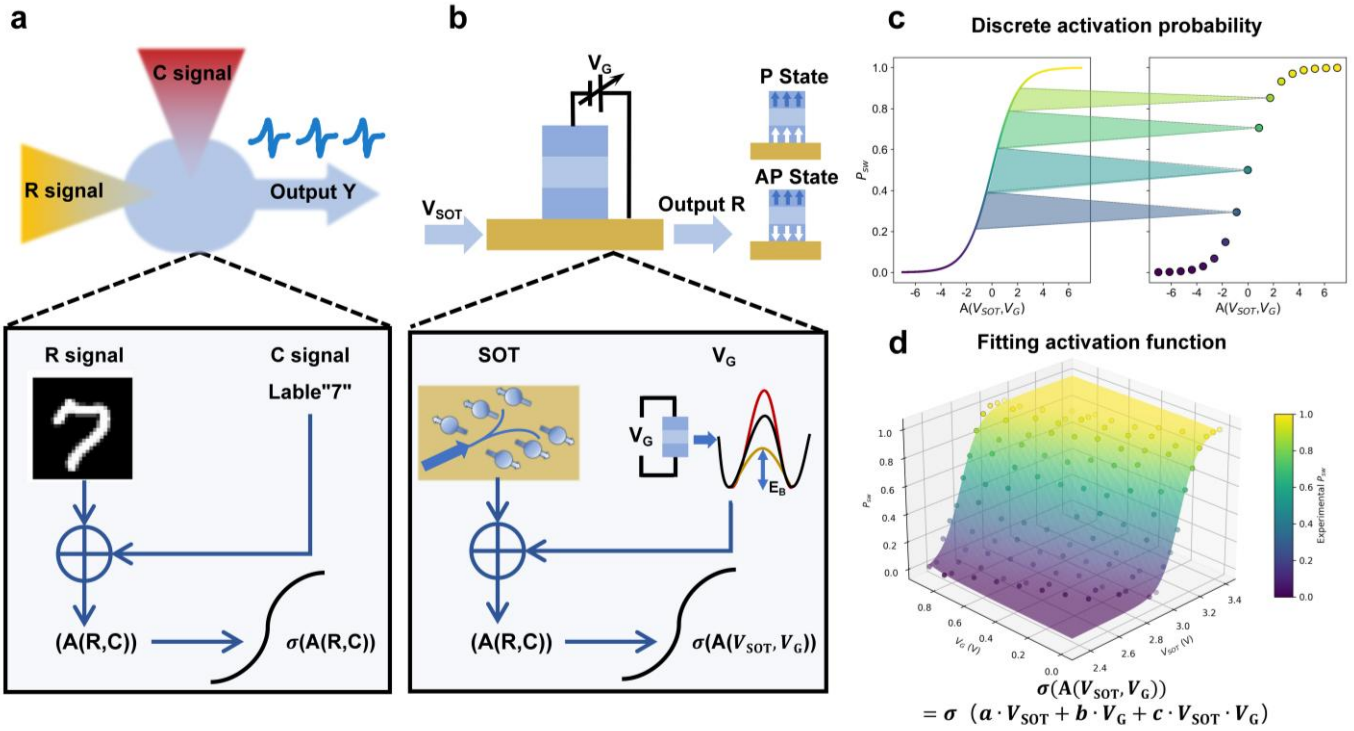

Fig. 3 | Working principle of the SIN. a, Schematic of an infomorphic neuron receiving R and C inputs, which are integrated through an activation function $A(R,C)$ and transformed by a sigmoid function to generate a spike probability. **b**, Physical implementation using a VGSOT-MTJ. The SOT pulse encodes the R signal and V_{G} serves as C signal. During switching, V_{G} tunes the magnetic anisotropy energy barrier, thereby modulating the stochastic switching probability of the MTJ. **c**, Discrete representation of the nonlinear activation probability in the SIN. Owing to finite switching statistics, the activation probability corresponding to a given $(V_{\mathrm{SOT}}, V_{\mathrm{G}})$ input pair is represented by discrete probability levels. **d**, Fitting of the neuron activation function. Activation probabilities obtained under different $(V_{\mathrm{SOT}}, V_{\mathrm{G}})$ conditions are fitted to yield the effective activation function of the SIN, described by $\sigma(A(V_{\mathrm{SOT}}, V_{\mathrm{G}}))$.

For a given input pair, the SIN produces a single binary output in one operation, with the probability of Y=1 determined by the corresponding switching probability. To estimate this probability experimentally, we repeat the switching measurement under identical $(V_{\mathrm{SOT}}, V_{\mathrm{G}})$ conditions and calculate the fraction of Y=1 events. Because this estimate is based on a finite number of trials, the sampled activation probabilities appear as discrete levels, as shown

in Fig. 3c. This repeated sampling is used only for device characterization and activation-function extraction; in operation, each input elicits a single stochastic binary output.

By repeatedly sampling under different (V_{SOT}, V_{G}) conditions, we obtain the switching-probability distribution P_{SW} (V_{SOT}, V_{G}). Fitting the experimental data yields the activation function $\sigma(A(V_{\mathrm{SOT}}, V_{\mathrm{G}}))$ of the SIN, expressed as:

$$P_{\mathrm{SW}}(Y = 1|V_{\mathrm{SOT}}, V_{\mathrm{G}}) = \sigma(aV_{\mathrm{SOT}} + bV_{\mathrm{G}} + cV_{\mathrm{SOT}}V_{\mathrm{G}}),$$

where the cross term $V_{\mathrm{SOT}}V_{\mathrm{G}}$ captures the nonlinear coupling between the two physical inputs. The fitted three-dimensional activation landscape (Fig. 3d; Supplementary Fig. 3) shows continuous dual-input control of the output probability across the full activation range. This experiment therefore extracts the neuron activation function directly from device statistics, yielding a hardware-implemented probabilistic activation in which both nonlinear input integration and stochastic sampling are performed intrinsically by the device, without auxiliary random-number generators or external nonlinear circuits. Taken together, these results show that the VGSOT-MTJ fulfills the key structural and functional requirements of the IN, thereby realizing a SIN and providing the physical neuron module for constructing SINN.

Supervised learning with SINN

Supervised learning was implemented using a single-layer SINN comprising ten neurons trained on the MNIST handwritten digit dataset (Fig. 4a–e). Image pixels were applied as R inputs with all-to-all connectivity to the SINs, providing the driving signals, while digit labels were introduced during training as C inputs through one-to-one connectivity, serving as modulatory signals. This architectural asymmetry enables each neuron to maximize the redundant information $I_{redundant}$ shared between image and label inputs, thereby promoting specialization toward a specific digit class.

During inference, only the R inputs are applied and C inputs are removed. Classification is performed by evaluating the spike rates of all ten neurons over a fixed temporal window of 10 timesteps and selecting the neuron with the highest firing rate. As illustrated in Fig. 4b for the representative case of class 0, neurons initially exhibit comparable spike rates, indicating poor class separability. With continued training, the neuron corresponding to class 0 progressively develops a dominant response, demonstrating successful emergence of class-specific representations. The evolution of PID information quantities during training is summarized in Fig. 4c. The redundant information (I_{red}) increases monotonically, while the residual entropy (H_{res}) decreases, consistent with the PID objective: as the shared information between image and label inputs increases, the network develops increasingly selective responses to individual digit classes.

To validate the physical feasibility of INPA, we implemented a proof-of-concept hardware platform based on a single VGSOT-MTJ device with programmable write and read operations (Supplementary Fig. 5 and Supplementary Note 6). The INPA characteristics of VGSOT-MTJ-based SINs used in Fig. 4d were extracted directly from this hardware platform, whereas network-level parallelism was implemented through time multiplexing.

Figure 4d compares the classification performance of infomorphic networks constructed using three neuron models: a software-based infomorphic neuron, a VGSOT-MTJ-based SIN with the full activation function (including the coupling

term $c \cdot V_{\mathrm{SOT}} \cdot V_{\mathrm{G}}$), and a VGSOT-MTJ-based SIN with the coupling term removed ($c = 0$). The full-model VGSOT-MTJ-based SINN achieves a classification accuracy of 88.13%, comparable to that of the software implementation. In contrast, the network using the decoupled activation function ($c = 0$) exhibits degraded performance, indicating that the coupling term is important for preserving the dual-input interaction required for INPA. The normalized final synaptic weight matrices of the three networks are shown in Fig. 4e. The full-model VGSOT-MTJ-based SINN exhibits weight patterns closely matching those obtained from the software model, whereas the $c = 0$ SINN shows pronounced deviations. These results confirm that the multiplicative interaction between V_{SOT} and V_{G} in the VGSOT-MTJ is critical for implementing INPA and translating infomorphic learning principles into physical hardware.

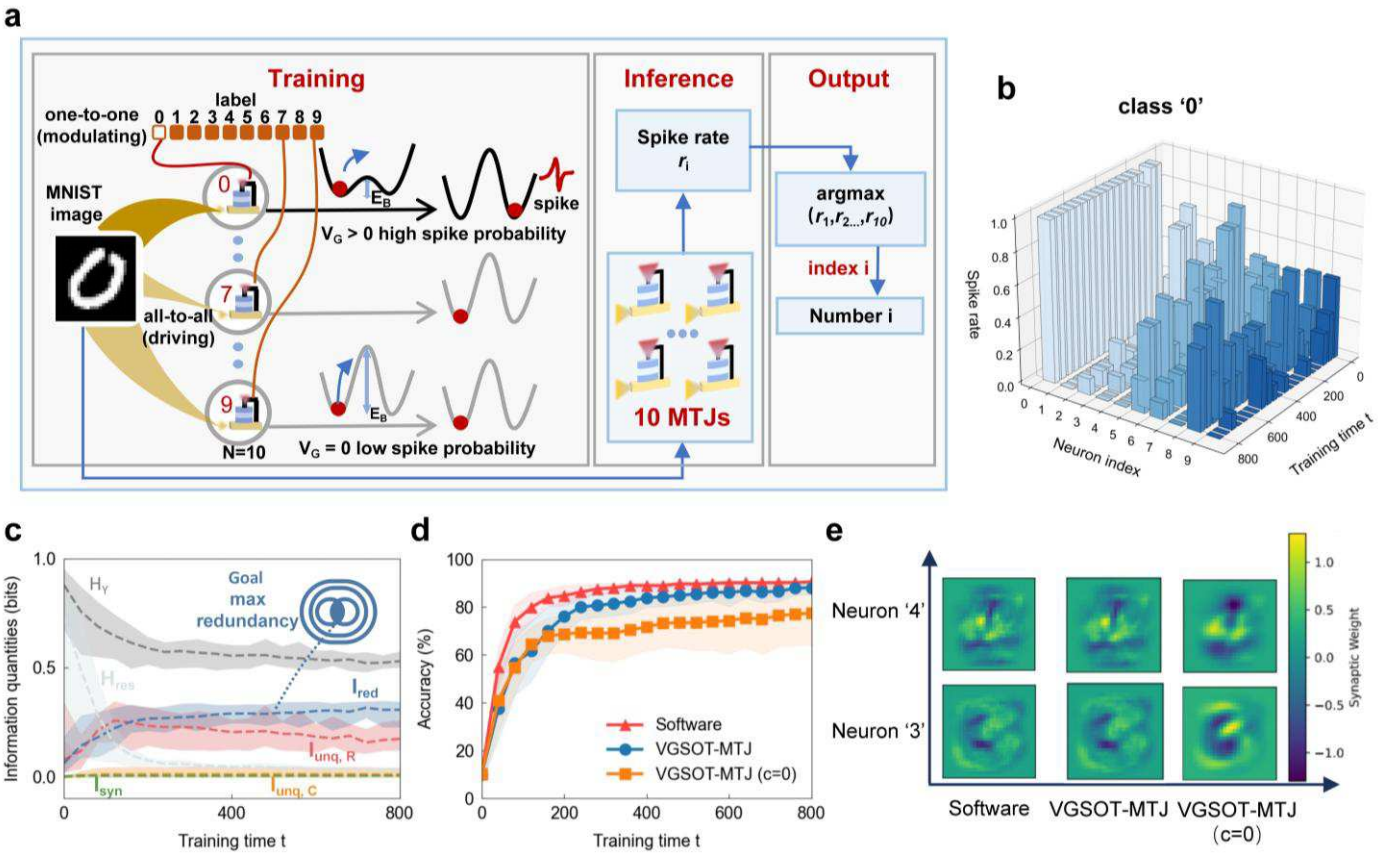

Fig. 4 | Supervised learning with SINN. **a**, Schematic of supervised learning on the MNIST dataset. During training, ten neurons (N=10) receive image pixels as R inputs and digit labels as one-to-one C inputs. During inference, C inputs are removed; classification is performed based on the spike-rate responses of the neurons. **b**, Evolution of spike rates during training for the case of class 0. **c**, Information quantities averaged over all neurons, showing the optimization of redundant information between image input and label signal, based on 20 independent training runs. **d**, MNIST image recognition accuracies of networks constructed using software-based infomorphic neurons, VGSOT-MTJ-based SINs, and VGSOT-MTJ-based SINs with the coupling term removed ($c = 0$). **e,** Learned synaptic weight distributions of the three types of neural networks after training.

Unsupervised learning and associative memory in recurrent SINN

We next examined recurrent SINN in unsupervised learning and associative memory tasks to demonstrate the versatility of the VGSOT-MTJ-based framework (Fig. 5a–h). We first constructed a minimal two-neuron recurrent SINN for unsupervised learning on the breast cancer dataset (Fig. 5a–c). Each neuron received standardized R inputs derived from the 30-dimensional feature vector, and no external C input was provided. Under this setting, the network is driven to maximize the unique information ($I_{unq,R}$) carried by the R inputs, leading to unsupervised feature-level specialization. As shown in Fig. 5b, $I_{unq,R}$ increases during training, whereas the residual entropy (H_{res}) decreases, indicating that the two neurons progressively specialize toward distinct, non-overlapping feature representations. Across 20 independent trials, the network achieves an average classification accuracy of approximately 81.1%, with a maximum accuracy of 92.6% (Fig. 5c), as evaluated under the better-matched neuron-to-class assignment.

We next investigated associative memory using a recurrent SINN comprising 35 VGSOT-MTJ-based SINs (Fig. 5d–h). The dataset consists of 26 capital letters, each represented by a 5×7 binary pixel pattern. In this recurrent configuration, R inputs at time t were generated from the weighted network state at time $t-1$, whereas external letter patterns were applied as C signals through one-to-one pixel-to-neuron mapping. The non-volatile nature of the VGSOT-MTJ enabled direct persistence of binary neuron states between successive timesteps.

During training, N noiseless letter patterns were randomly selected and presented sequentially as contextual cues, where N denotes the number of stored patterns shown in Fig. 5g,h. Synaptic weights were locally updated to maximize the redundant information (I_{red}) between recurrent receptive signals at time $t-1$ and contextual signals at time t. During inference, a noisy letter pattern (α = 0.2, representing randomly flipped pixels) was presented as a contextual cue for one timestep, after which the contextual input was removed and the network evolved freely for 20 timesteps. Recall accuracy was evaluated from the cosine similarity between the final network state and the original noiseless pattern (Fig. 5d,f). As shown in Fig. 5e, I_{red} increases monotonically while H_{res} decreases, indicating the progressive formation of stable associative memory states. Correspondingly, recall accuracy improves with training and decreases as the number of stored patterns increases (Fig. 5f), reflecting the higher recall difficulty under larger memory loads.

To evaluate robustness against device non-idealities, we incorporated experimentally measured device-to-device variations (Supplementary Fig. 2) into the network model, with details provided in Supplementary Note 7. As shown in Fig. 5g, SINN with device variation maintains recall accuracy close to that of ideal-device SINN across different numbers of stored patterns, indicating strong tolerance to hardware variability. Finally, we compared the associative memory performance of the SINN with that of a classical Hopfield neural network (HNN). As shown in Fig. 5h, the

SINN achieves higher recall accuracy and superior noise tolerance, demonstrating the advantage of the SINN framework for robust associative memory.

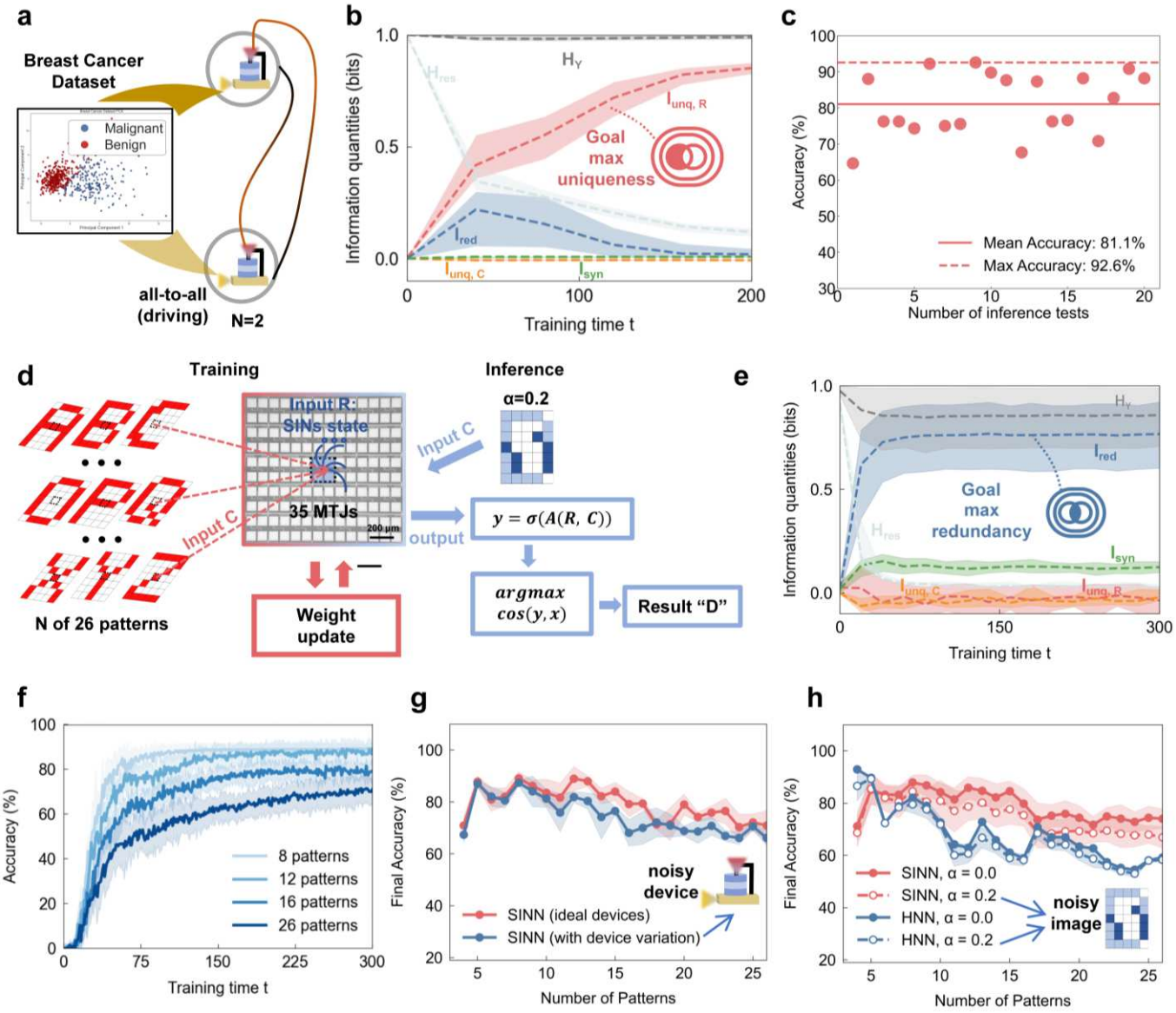

Fig. 5 | Unsupervised learning and associative memory learning with SINN. **a**, Schematic of unsupervised learning performed on a binary breast cancer dataset using two neurons (N = 2). **b**, Information quantities averaged over all neurons during unsupervised learning, based on 20 independent training runs. **c**, Classification accuracies obtained from 20 independent inference tests. **d**, Schematic of associative-memory learning in a recurrent SINN composed of 35 VGSOT–MTJ neurons. **e**, Information quantities averaged over all neurons during associative-memory learning, based on 20 independent training runs. **f**, Training accuracy across 20 independent runs for different numbers of stored patterns. **g**, Final recall accuracy of the SINN with ideal devices and with device-to-device variation

over different numbers of stored patterns. **h**, Final recall accuracy of the SINN compared with a classical HNN under two levels of input image noise.

Discussion

We have demonstrated a proof-of-concept hardware realization of a SIN based on a VGSOT-MTJ, providing a hardware route toward physical implementations of infomorphic computing. By implementing a dual-input, tunably stochastic neuron primitive in hardware, we further constructed the single-layer SINN that support supervised, unsupervised, and associative-memory learning. These results show that the key functional requirements of infomorphic learning can be directly embodied at the device level, rather than being imposed solely through abstract algorithmic design.

This hardware primitive is well suited to local learning in several respects. First, the underlying SOT-driven magnetization dynamics operate on nanosecond timescales, enabling rapid probabilistic state updates. Sub-nanosecond switching was experimentally observed (Supplementary Fig. 6), indicating that the physical neuron can in principle support high-speed on-device operation. Second, because weight updates are governed by local neuronal activity rather than cross-layer error propagation, this framework is naturally compatible with highly parallel learning when extended to larger arrays. Third, the physical realization of probabilistic activation through stochastic VGSOT-MTJ switching provides high intrinsic energy efficiency at the device level. Each INPA update consumes ~15.86 pJ and 1.6 ns in the VGSOT-MTJ devices, while the system-level prototype containing the MCU already achieves ~400 $\times$ higher energy efficiency than the GPU, with projected gains of up to six orders of magnitude upon spin-CMOS ASIC integration (Supplementary Notes 3 and 4).

The high energy efficiency of the SINN extends beyond the device-level INPA operation. At the algorithm-hardware level, infomorphic learning removes the need for global error backpropagation and the repeated cross-layer data movement on which conventional training relies. Our energy analysis, based on a representative Digital Computing-In-Memory baseline (state-of-the-art architecture for backpropagation), shows a ~68 $\times$ energy advantage over backpropagation for the considered supervised-learning task, requiring only ~0.5% of the update computation (Supplementary Note 5). Moreover, the observed robustness to device non-idealities is consistent with the local character of the update rule: perturbations remain confined to individual neurons rather than being recursively amplified across layers[42]. This probabilistic nature of INPA could further reduce sensitivity to device-level fluctuations by embedding stochasticity within the operating principle of the neuron itself.

A key distinction of our work lies in the intrinsic coupling between two physical mechanisms in the VGSOT-MTJ, enabling probabilistic computation to be directly realized at the device level. Unlike prior probabilistic hardware platforms[43–50] and stochastic neuromorphic systems[51,52], where randomness is commonly used for sampling, annealing, or variability emulation[53–57], here, stochasticity is elevated to a core computational primitive. The coupled action of spin-orbit torque and voltage-controlled magnetic anisotropy, independently tunable via two electrical inputs, defines the neuron behavior required to encode the interaction between receptive (R) and contextual (C) inputs, as

prescribed by PID. In this sense, computation is not externally imposed but instead emerges from the underlying CMOS-compatible device physics, enabling practical hardware realization.

This study remains a proof-of-concept demonstration at the device and single-layer network levels. Future work will focus on extending to multilayer architectures, and further peripheral circuits, ultimately paving the way for scalable SINN chips. More broadly, this work demonstrates that device physics can be co-designed with interpretable local learning objectives, offering a potential path toward hardware systems in which learning rules are directly embodied in physical dynamics.

Materials and Methods

Fabrication and characterization of VGSOT-MTJ devices

To fabricate the top-pinned VGSOT-MTJ devices used in this work, a multilayer stack of W/CoFeB (free layer)/MgO/CoFeB (reference layer)/SAF/Ru/Ta/Ru (from bottom to top, Fig. 2a) was deposited on a thermally oxidized 8-inch Si substrate by magnetron sputtering. Post-deposition annealing was carried out at 350 °C for 20 min in high vacuum to induce strong perpendicular magnetic anisotropy (PMA). The MTJ pillars, with a nominal diameter of 250 nm, were subsequently patterned through standard microfabrication procedures. The TMR ratio was measured at room temperature by the two-probe method. A low testing current I (approximately 10 μA) was applied between the bottom electrode and the top electrode, while a voltage was measured between the same terminals to determine the MTJ resistance.

Electrical measurements

The electrical setup used for switching probability characterization is illustrated in Fig. 2d. Two independent paths were employed: one for the SOT voltage (V_{SOT}) and the other for the VCMA gate voltage (V_G). Each path contained a high-frequency channel and a DC channel, which were decoupled by bias tees. The VCMA path shared its bias tee with the MTJ readout circuit, enabling both pulse application and resistance measurement through the same line. Voltage pulses were generated by an Active Technologies PG-1074 high-frequency pulse generator, and the transmission lines were connected to a 50 Ω impedance-matched Tektronix MSO54 oscilloscope for real-time monitoring of pulse waveform, amplitude, and timing.

For resistance readout, a Keithley 6221 precision current source, and a Keithley 2182A nanovoltmeter, were connected through the DC channel to provide a sensing current and measure the corresponding voltage drop across the MTJ, from which the MTJ resistance was determined. For the switching probability measurements from the high- to low-resistance state, the MTJ was first initialized to the high-resistance (antiparallel) state by applying an external perpendicular magnetic field. During testing, an in-plane magnetic field of 50 mT was applied to assist magnetization switching. Positive V_{SOT} pulses together with V_{G} pulses were then applied to induce switching. After each excitation, the MTJ resistance was measured to determine whether the device had switched to the low-resistance (parallel) state. The same procedure was applied in reverse for the low-to-high resistance switching. At each V_{SOT},, the measurement

was repeated 100 times under identical conditions, and the switching probability was defined as the ratio of successful switching events to the total number of trials.

SIN model

The VGSOT-MTJ was used as the physical neuron primitive in this work. Its switching probability surface, P_{SW} ($Y = 1|V_{\mathrm{SOT}}, V_{\mathrm{G}}$), was experimentally mapped by applying 100 repeated voltage pulses at each ($V_{\mathrm{SOT}}, V_{\mathrm{G}}$) operating point. To obtain the device-level INPA, the measured switching-probability surface was fitted by logistic regression, yielding an effective activation function:

$$A(V_{\mathrm{SOT}}, V_{\mathrm{G}}) = aV_{\mathrm{SOT}} + bV_{\mathrm{G}} + cV_{\mathrm{SOT}}V_{\mathrm{G}},$$

where the coefficients (a, b, c) were obtained by logistic regression. And the corresponding switching (firing) probability was defined as:

$$P_{\mathrm{SW}}(Y = 1|V_{\mathrm{SOT}}, V_{\mathrm{G}}) = \sigma(A(V_{\mathrm{SOT}}, V_{\mathrm{G}})) = \frac{1}{1+e^{-(aV_{\mathrm{SOT}}+bV_{\mathrm{G}}+cV_{\mathrm{SOT}}V_{\mathrm{G}})}},$$

In simulations, neuronal outputs were generated as:

$$Y \sim Bernoulli(P_{\mathrm{SW}}(Y = 1|V_{\mathrm{SOT}}, V_{\mathrm{G}})),$$

thereby directly linking the experimentally measured device statistics to stochastic neuronal outputs. Detailed fitting parameters and goodness-of-fit are provided in Supplementary Fig. 3.

Spintronic infomorphic neural network framework

At the network level, each neuron locally updates its receptive and contextual synaptic weights (w_R, w_C) by stochastic gradient ascent on the PID-based objective:

$$G = \gamma_{goal}^{\mathrm{T}} \cdot I_{\mathrm{PID}},$$

where

$$I_{\mathrm{PID}} = [I_{unq,R}, I_{unq,\mathrm{C}}, I_{red}, I_{syn}, H_{res}]^{\mathrm{T}}$$

Here, γ_{goal} selected the targeted information atom according to the learning paradigm, with I_{red} used for supervised and associative-memory learning, and $I_{unq,R}$ for unsupervised feature extraction. At each forward pass, the INPA-driven stochastic binary output Y, together with the receptive and contextual variables (R, C), are collected into a minibatch, from which the empirical joint distribution $p(R, C, Y)$ was estimated using discretized histograms. The weight updates were then computed as:

$$\Delta w_R = \eta \frac{\partial G}{\partial w_R}, \Delta w_C = \eta \frac{\partial G}{\partial w_C}$$

For each task, receptive and contextual variables were constructed in a task-dependent manner and mapped to ($V_{\mathrm{SOT}}, V_{\mathrm{G}}$) to drive the INPA. During inference, weights were held fixed and task-specific evaluation protocols were applied. Details of the network structure, parameter settings, and histogram configuration are provided in Supplementary Note 1 and Supplementary Table 1; full derivations of the PID-based objective and its gradients are given in Supplementary Note 2.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 62422402, 62374013 and 62404015, the National Key Research and Development Program of China under Grant 2024YFA1410200, Young Elite Scientists Sponsorship Program of the Beijing High Innovation Plan (20250782), Beijing Outstanding Young Scientist Program, the Fundamental Research Funds for the Central Universities, the Nanoelectronics Science and Technology Center at the Hefei Innovation Research Institute of Beihang University.

Data availability

The data that support the plots within this paper and the other findings of this study are available from the corresponding author upon reasonable request.

Code availability

The codes used in this paper are available from the corresponding author upon reasonable request.

References

1. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional Networks. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2261–2269 (IEEE, Honolulu, HI, 2017). doi:10.1109/CVPR.2017.243.

2. Redmon, J. & Farhadi, A. YOLOv3: An Incremental Improvement. Preprint at https://doi.org/10.48550/arXiv.1804.02767 (2018).

3. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017).

4. Brown, T. B. *et al.* Language Models are Few-Shot Learners. Preprint at https://doi.org/10.48550/arXiv.2005.14165 (2020).

5. OpenAI *et al.* GPT-4 Technical Report. Preprint at https://doi.org/10.48550/arXiv.2303.08774 (2024).

6. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533 – 536 (1986).

7. Patterson, D. *et al.* The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. *Computer* **55**, 18–28 (2022).

8. Thompson, N., Greenewald, K., Lee, K. & Manso, G. F. The Computational Limits of Deep Learning. in *Ninth Computing within Limits 2023* (LIMITS, Virtual, 2023). doi:10.21428/bf6fb269.1f033948.

9. Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J. & Hinton, G. Backpropagation and the brain. *Nat Rev Neurosci* **21**, 335–346 (2020).

10. Zou, X., Xu, S., Chen, X., Yan, L. & Han, Y. Breaking the von Neumann bottleneck: architecture-level processing-in-memory technology. *Sci. China Inf. Sci.* **64**, 160404 (2021).

11. Qamar, T. & Bawany, N. Z. Understanding the black-box: towards interpretable and reliable deep learning models.

12. Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22071–22080 (2019).

13. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* **1**, 206–215 (2019).

14. Grollier, J. *et al.* Neuromorphic spintronics. *Nat Electron* **3**, 360–370 (2020).

15. Zenke, F. & Neftci, E. O. Brain-Inspired Learning on Neuromorphic Substrates. *Proc. IEEE* **109**, 935–950 (2021).

16. Bellec, G. *et al.* A solution to the learning dilemma for recurrent networks of spiking neurons. *Nat Commun* **11**, 3625 (2020).

17. Greatorex, H. *et al.* A neuromorphic processor with on-chip learning for beyond-CMOS device integration. *Nat Commun* **16**, 6424 (2025).

18. Payvand, M. *et al.* Self-organization of an inhomogeneous memristive hardware for sequence learning. *Nat Commun* **13**, 5793 (2022).

19. Zhang, T. *et al.* A brain-inspired algorithm that mitigates catastrophic forgetting of artificial and spiking neural networks with low computational cost. *Sci. Adv.* **9**, eadi2947 (2023).

20. Ahamed, M. A., Chen, J. & Imran, A.-A.-Z. Forward-Forward Contrastive Learning. Preprint at https://doi.org/10.48550/arXiv.2305.02927 (2023).

21. Illing, B., Ventura, J., Bellec, G. & Gerstner, W. Local plasticity rules can learn deep representations using self-supervised contrastive predictions. *In Advances in Neural Information Processing Systems (NeurIPS 2021)* (2021).

22. Löwe, S., O'Connor, P. & Veeling, B. S. Putting An End to End-to-End: Gradient-Isolated Learning of Representations. Preprint at https://doi.org/10.48550/arXiv.1905.11786 (2020).

23. Millidge, B., Tschantz, A. & Buckley, C. L. Predictive Coding Approximates Backprop Along Arbitrary Computation Graphs. *Neural Computation* **34**, 1329–1368 (2022).

24. Makkeh, A. *et al.* A general framework for interpretable neural learning based on local information-theoretic goal functions. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2408125122 (2025).

25. Wibral, M., Priesemann, V., Kay, J. W., Lizier, J. T. & Phillips, W. A. Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain and Cognition* **112**, 25–38 (2017).

26. Gutknecht, A. J., Wibral, M. & Makkeh, A. Bits and pieces: understanding information decomposition from part-whole relationships and formal logic. *Proc. R. Soc. A.* **477**, 20210110 (2021).

27. Williams, P. L. & Beer, R. D. Nonnegative Decomposition of Multivariate Information. Preprint at https://doi.org/10.48550/arXiv.1004.2515 (2010).

28. Shu, Y., Hasenstaub, A. & McCormick, D. A. Turning on and off recurrent balanced cortical activity. *Nature* **423**, 288–293 (2003).

29. Körding, K. P. & König, P. Supervised and Unsupervised Learning with Two Sites of Synaptic Integration. *J. Comput. Neurosci.* **11**, 207 – 215 (2001).

30. Park, J. *et al.* Contribution of apical and basal dendrites to orientation encoding in mouse V1 L2/3 pyramidal neurons. *Nat Commun* **10**, 5372 (2019).

31. Renner, A., Sheldon, F., Zlotnik, A., Tao, L. & Sornborger, A. The backpropagation algorithm implemented on spiking neuromorphic hardware. *Nat Commun* **15**, 9691 (2024).

32. Lillicrap, T. P., Cownden, D., Tweed, D. B. & Akerman, C. J. Random synaptic feedback weights support error backpropagation for deep learning. *Nat Commun* **7**, 13276 (2016).

33. Schultz, W. Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology* **80**, 1–27 (1998).

34. Chéreau, R., Williams, L. E., Bawa, T. & Holtmaat, A. Circuit mechanisms for cortical plasticity and learning. *Seminars in Cell & Developmental Biology* **125**, 68–75 (2022).

35. Takahashi, N., Oertner, T. G., Hegemann, P. & Larkum, M. E. Active cortical dendrites modulate perception. *Science* **354**, 1587–1590 (2016).

36. Fan, H. *et al.* High-Performance FPGA-based Accelerator for Bayesian Neural Networks. in *2021 58th ACM/IEEE Design Automation Conference (DAC)* 1063–1068 (IEEE, San Francisco, CA, USA, 2021). doi:10.1109/DAC18074.2021.9586137.

37. Chen, Y.-H., Krishna, T., Emer, J. S. & Sze, V. Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks. *IEEE J. Solid-State Circuits* **52**, 127–138 (2017).

38. Tsmots, I., Skorokhoda, O. & Rabyk, V. Hardware Implementation of Sigmoid Activation Functions using FPGA. in *2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM)* 34–38 (IEEE, Polyana, Ukraine, 2019). doi:10.1109/CADSM.2019.8779253.

39. Liu, Z. *et al.* Robust Field-Free Voltage-Gated Spin-Orbit Torque Switching in IrMn-Based Perpendicular Magnetic Tunnel Junctions. *IEEE Electron Device Lett.* **46**, 745–748 (2025).

40. Li, W. *et al.* Selective Data Writing in IrMn-Based Perpendicular Magnetic Tunnel Junction Array Through Voltage-Gated Spin-Orbit Torque. *IEEE Electron Device Lett.* **45**, 921–924 (2024).

41. Shao, Q. *et al.* Spintronic memristors for computing. *npj Spintronics* **3**, 16 (2025).

42. Artime, O. *et al.* Robustness and resilience of complex networks. *Nat Rev Phys* **6**, 114–131 (2024).

43. Camsari, K. Y., Faria, R., Sutton, B. M. & Datta, S. Stochastic p -Bits for Invertible Logic. *Phys. Rev. X* **7**, 031014 (2017).

44. Zhao, W. *et al.* Ultra-Fast Ising Machine for Combinatorial Optimization using Sub-nanosecond CMOS-integrated Magnetoresistive Random Access Memory. Preprint at https://doi.org/10.21203/rs.3.rs-8350827/v1 (2026).

45. Lin, Y. *et al.* Deep Bayesian active learning using in-memory computing hardware. *Nat Comput Sci* **5**, 27–36 (2024).

46. He, Y., Fang, C., Luo, S. & Liang, G. Many-Body Effects-Based Invertible Logic With a Simple Energy Landscape and High Accuracy. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **9**, 83–91 (2023).

47. He, Y. *et al.* Solving Boolean Satisfiability Problems Using A Hypergraph- based Probabilistic Computer. *Preprint at* (2025).

48. Luo, S., He, Y., Cai, B., Gong, X. & Liang, G. Probabilistic-Bits Based on Ferroelectric Field-Effect Transistors for Probabilistic Computing. *IEEE Electron Device Lett.* **44**, 1356 – 1359 (2023).

49. Duffee, C. *et al.* An integrated-circuit-based probabilistic computer that uses voltage-controlled magnetic tunnel junctions as its entropy source. *Nat Electron* **8**, 784–793 (2025).

50. Aadit, N. A. *et al.* Massively parallel probabilistic computing with sparse Ising machines. *Nat Electron* **5**, 460–468 (2022).

51. Cartiglia, M. *et al.* Stochastic dendrites enable online learning in mixed-signal neuromorphic processing systems. in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)* 476–480 (IEEE, Austin, TX, USA, 2022). doi:10.1109/ISCAS48785.2022.9937833.

52. Cai, J. *et al.* Voltage-Controlled Spintronic Stochastic Neuron Based on a Magnetic Tunnel Junction. *Phys. Rev. Applied* **11**, 034015 (2019).

53. Borders, W. A. *et al.* Integer factorization using stochastic magnetic tunnel junctions. *Nature* **573**, 390–393 (2019).

54. He, Y. *et al.* A hardware demonstration of a universal programmable RRAM-based probabilistic computer for molecular docking. *Nat Commun* **17**, 702 (2025).

55. Zhu, D. *et al.* Thermally Driven Leaky-Integrate-and-Fire Spintronic Neurons With Stray-Field-Enabled Self-Reset for Neuromorphic Computing. *IEEE Electron Device Lett.* **46**, 1437–1440 (2025).

56. Raimondo, E. *et al.* High-Performance and Reliable Probabilistic Ising Machine Based on Simulated Quantum Annealing. *Phys. Rev. X* **15**, 041001 (2025).

57. Raab, K. *et al.* Brownian reservoir computing realized using geometrically confined skyrmion dynamics. *Nat Commun* **13**, 6982 (2022).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportinginformationVFNC.docx