Neuroscience-inspired information-integration system based on stochastic magnetic tunnel junctions

Li Zhao,[1,2,3,‡] Meiting Zhang,[4,‡] Yajun Zhang,[1,2,3] Yuanyuan Mi,[5,*] Zhe Yuan,[2,3,†] and Ke Xia[6]

[1] *Center for Advanced Quantum Studies and Department of Physics, Beijing Normal University, Beijing 100875, China*

[2] *Institute for Nanoelectric Devices and Quantum Computing, Fudan University, Shanghai 200433, China*

[3] *Interdisciplinary Center for Theoretical Physics and Information Sciences (ICTPIS), Fudan University, Shanghai 200433, China*

[4] *Center for Neurointelligence, School of Medicine, Chongqing University, Chongqing 400044, China*

[5] *Department of Psychological and Cognitive Sciences, Tsinghua University, Beijing, 100084, China*

[6] *School of Physics, Southeast University, Nanjing 211189, China*

(Received 26 August 2023; revised 26 March 2024; accepted 17 May 2024; published 17 June 2024)

Effective integration of multiple sensory information in the brain is essential to achieve accurate information perception and processing despite the noise and imperfections in biological neural systems, providing strong inspiration for improving computational accuracy in neuromorphic computing. In the future development of brainlike chips, e.g., biomimetic robots, multisensory-information integration is an essential capability for recognizing the external circumstances. Here, based on the computational model proposed by neuroscientists, we propose a hardware implementation scheme of the decentralized multisensory information-integration system using spintronic devices, in which magnetic tunnel junctions are employed as artificial neurons with stochastic dynamics. Using a one-dimensional continuous variable (orientation, head direction, etc.) as a typical example, we demonstrate that the input information from noisy cues is extracted with high accuracy after integration. This spintronic neuromorphic system exhibits remarkable tolerance for both nonuniform devices and the malfunction of different modules. The computational advantages and robustness of the spintronics-based information-integration system provide an important basis for the development of hardware neuromorphic platforms.

DOI: 10.1103/PhysRevApplied.21.064040

I. INTRODUCTION

Digital computers nowadays rely on the accurate processing of data, where every bit must be correctly computed, transferred, and stored. Ensuring accuracy in storage and computation consumes additional energy [1]. In contrast, biological neural systems contain significant noise in the information, stemming either from the sensory organs as noisy input signals or during the neural computation due to the stochastic dynamics of the neurons and synapses [2–6]. Despite this noise, the neural system can achieve sophisticated cognitive and behavioral tasks. For example, humans can distinguish orientation differences of less than 1° [7], and orientation is a typical continuous variable in object recognition.

Stimulus information cannot be accurately encoded in the activity of a single neuron due to noise from input signals and the intrinsic network topology. Neural circuits generally adopt a population coding strategy to represent external stimuli [8–10], as shown in Fig. 1(a). They integrate relevant sensory cues from other circuits optimally to improve perception (i.e., multisensory-information integration) [12–15]. The continuous-attractor neural network (CANN) is a canonical model for representing continuous features through population coding, such as orientation [16], movement direction [17], head direction [18], spatial locations [19], and different views of objects [20,21]. Computations using CANNs yield consistent results with neural experiments [22–25]. Coupling several CANNs can enable Bayesian optimization of multiple information integration [26–29]. In addition, multisensory-information integration is a necessary function for future intelligent systems, such as biomimetic robots, to accurately perceive the external circumstances [30].

In a CANN, every individual neuron has its own preference, and its responsive tuning curve to external stimuli is bell shaped and centered at its corresponding preference [8,19,31]; see Fig. 1(a). The stochastic dynamic properties of such neurons in population coding [10,32] are perfectly reproduced by a magnetic tunnel junction (MTJ) in the

*Corresponding author: miyuanyuan@tsinghua.edu.cn

†Corresponding author: yuanz@fudan.edu.cn

‡L. Zhao and M. Zhang contributed equally to this work.

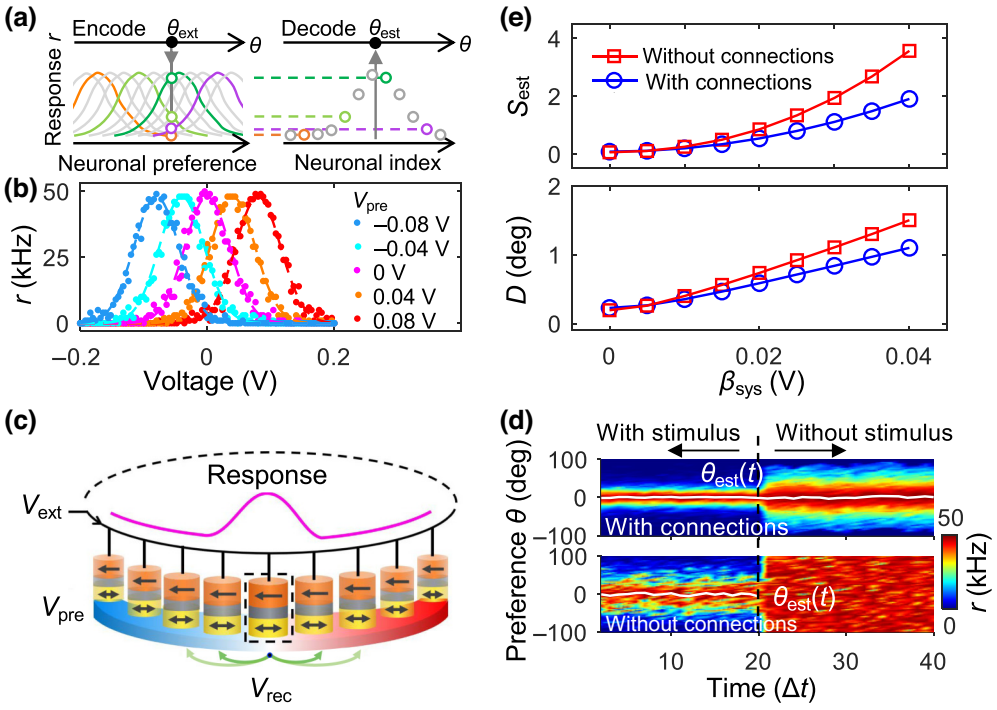

FIG. 1. (a) Schematic illustration of population coding. Lines in the left panel indicate the neurons' tuning curves, which are bell shaped and centered at their preferences. One-dimensional variable θ_{ext} induces activity in a group of neurons, resulting in a bumplike population activity (right panel). Stimulus value θ_{est} is then decoded based on the neural population activity [11]. (b) Switching frequency of the MTJs with different preferences as functions of the input voltage. Frequency is calculated by sampling 100 states within 1 ms. Dashed lines are fitted using a Gaussian function. (c) Schematic of the proposed architecture of a MTJ-based CANN. Every MTJ in the CANN is subjected to a unique local bias to represent the neuron with its own preferred feature, as indicated by the substrate at the bottom. Green arrows denote excitatory connections for the MTJ in the dashed frame, and the connections have translationally invariant symmetry for all the MTJs. Solid magenta line schematically illustrates the bell-shaped response bump of the network under an external stimulus. Inhibitory connections are not explicitly shown. (d) Population activities of the network with and without connections. Solid white lines represent the estimate decoded from the network response. Stimulus is input to the network at $t < 20\Delta t$. (e) Deviation of the network estimate from the real stimulus and the variance of the network estimate as functions of the system's noise strength.

superparamagnetic regime [33,34], which can serve as the fundamental element for the hardware implementation of population coding [35]. In particular, the switching frequency of a MTJ driven by thermal perturbation can be simply tuned by an external electric and/or magnetic bias to shift its preference [36]. Then a group of MTJs with various biases naturally encode a stimulus of continuous features [37,38], as shown in Fig. 1(b). Thanks to mature fabrication techniques, such MTJs can be optimized to reduce their switching time down to nanoseconds [39]. They are particularly suitable for application in future nanodevices [40–44] with fast operation speed, low cost, and low energy consumption [45–47]. There was a pioneering study to realize population coding with superparamagnetic MTJs [36], where the summation of the weighted activity of several MTJs with shifted tuning curves was employed to represent nonlinear functions, and the realization of this approach was dependent on MTJs with a smaller area and lower energy consumption than those based entirely on CMOS devices. However, the computational advantages of population coding in a network consisting of interconnected MTJs have not yet been fully exploited, partly because of the challenges in developing and applying suitable neuromorphic algorithms [48] to let MTJ-based networks have designed functions and good performance.

Based on recent progress in experimental and computational neurosciences for the application of population coding and multiple sensory integration in information processing [12,26,28,29], we propose a spintronic realization scheme of several coupled CANNs with high tolerance to system noise and input noise. The scheme employs the unique dynamic characteristics of MTJs and provides an ideal implementation for the computational neuroscience model [28,29]. As a proof of concept, a decentralized model for multisensory-information integration is implemented with MTJs, which have stochastic dynamics driven by thermal fluctuations. Through multiscale simulations, we demonstrate its capability to accurately extract

one-dimensional orientation by integrating multiple noisy input signals. This bioinspired computing scheme is particularly robust against hardware variability and the different malfunctions of modules in the network.

II. PROPOSAL FOR THE MTJ-BASED CANN

We design a magnetic CANN consisting of a group of MTJs, the preferences of which are uniformly distributed in a one-dimensional receptive field, as sketched in Fig. 1(c). The MTJs are connected via translationally invariant excitatory connections, i.e., the connection strength between two neurons only depends on the difference between their preferred stimuli [27]; see Eq. (A4) in the Appendix. Such translationally invariant connections can be realized through peripheral circuits [49]. Synaptic strengths are globally scaled by the parameter J_{rec} and global inhibition is introduced to maintain the balance of excitatory and inhibitory dynamics in the network; see the Appendix.

A MTJ consists of two ferromagnetic layers, such as Co-Fe-B, that are separated by an insulator spacer like MgO. While the magnetization of one ferromagnetic layer is fixed by, e.g., an exchange bias due to an adjacent antiferromagnetic material, the other magnetic layer can flip its magnetization under the superparamagnetic regime via thermal excitation. When the two magnetizations are parallel and antiparallel, the MTJ exhibits low and high resistance, respectively. For a single MTJ under an electric voltage (V), the thermally induced stochastic switching frequency (denoted as r hereafter to mimic the firing rate of a single neuron) is determined by the temperature and energy barrier between the parallel and antiparallel states, as described by the generalized Néel-Brown theory [33,50]. A bell-shaped tuning curve of the MTJ follows the relation $r \propto f\,[(V - V_{\text{pre}})^2]$, where f is a nonlinear function, as described in the Supplemental Material [51]. V_{pre} is the preferred value determined by an electrical or magnetic bias on the free layer and is uniformly distributed across the whole receptive field; see the Appendix.

The CANN receives sensory information with considerable noise from a local sensor, which is converted into a global input voltage (V_{ext}) for all the MTJs in the network. This global input, used to encode the continuous external stimulus, leverages the distinctive dynamic properties of MTJs and significantly simplifies the input process in the conventional neuroscience model. The total voltage on a MTJ (V_{tot}) becomes a superposition of the external stimulus (V_{ext}), the local bias (V_{pre}), and the recurrent interactions from other MTJs (V_{rec}). The input noise and system noise are superimposed on V_{ext} and V_{rec}, respectively, both modeled by independent random noise. Under V_{tot}, the response of every MTJ is updated by counting the switching frequency in a fixed time interval. This frequency is then transferred to other MTJs via the connections as an electrical voltage, according to the inverse function, f^{-1}, as described in the Supplemental Material [51]. Thus, the computation is iteratively performed, and the CANN generates bump activity to encode the external stimulus. Information is then extracted by decoding the response bump, taking the center of mass of the bump in the past 10 time steps for network dynamics as the instantaneous estimate [59,60], which is equivalent to the maximum likelihood estimate [4].

Before performing information integration, we first demonstrate the capability of information encoding using such a spintronic CANN, where recurrent connections in the network are essential for the accuracy and retention of encoded data. The dynamics of a CANN with 100 MTJs is performed with and without recurrent connections. Detailed dynamic equations are provided in the Appendix. Since the CANN is translationally invariant, the response to an arbitrary stimulus (V_{ext}) corresponds to the same bump centered at the corresponding MTJ with $V_{\text{pre}} = V_{\text{ext}}$. Therefore, we only consider in this work a constant external stimulus of the orientation, $\theta_{\text{ext}} = 0^\circ$, without loss of generality. Then the input voltage, V_{ext}, is applied to the network without input noise ($\beta_{\text{inp}} = 0$) for a finite period, $t \leq 20\Delta t$, and the firing rates (switching frequencies) of the most active MTJs in the CANN are plotted in Fig. 1(d).

With the input signal, both the MTJs without connections and the CANN with connections exhibit a bump-shaped response centered at the MTJs with a preference of 0°. The white lines represent the decoded estimate, i.e., the center of mass of the network response. The CANN with connections results in much smaller uncertainty than the unconnected MTJs, indicating that the recurrent connections lead to more accurate information encoding.

After turning the stimulus off at $t = 20\Delta t$, all the MTJs without connections reach the highest firing rate, and the encoded information disappears. In contrast, the CANN with connections still holds a bell-shaped response, with its center near the previous external stimulus, $\theta_{\text{ext}} = 0^\circ$, indicating that the network has memorized the information. The retention of the response bump is attributed to the connections in the CANN, which determine the energy landscape of the dynamic network. Specifically, a series of local minima, referred to as attractors, are continuously distributed over the feature space due to the translationally invariant connections, in contrast to the discrete attractors in the Hopfield network [44]. When the CANN receives an external input, it evolves to the corresponding attractor state and remains at this steady state after the input is removed. Such an inertialike effect not only helps to reduce the influence of system noise in the network, but also holds the stimulus information for other cognitive functions, such as decision making and spatial navigation. By repeating 20 runs with 400 trials each, we calculated

the average deviation of the network estimate, θ_{est}, from the real stimulus, θ_{ext}, $D = |\theta_{est} - \theta_{ext}|$, and the variance of θ_{est} (denoted by S_{est}) as a function of system noise strength, β_{sys}, as shown in Fig. 1(e). Both the deviation and variance of the CANN with connections are smaller than those without connections. Therefore, although independent MTJs are applicable for the population coding strategy, the connections in the CANN essentially provide the unique advantages of the network. It is worth noting that a single CANN can barely reduce the input noise because such noise simultaneously applies to all the MTJs in the CANN, resulting in a global shift in the network estimate. In biological neural systems, information received from different senses is integrated to achieve high accuracy, where the independent input noise for different senses can be effectively reduced.

III. INFORMATION INTEGRATION BY TWO COUPLED MTJ-BASED CANNS

Inspired by multiple sensory-information integration in the biological neural system, we design a decentralized information-integration model consisting of $L(L \geq 2)$ CANNs, which are reciprocally connected, as sketched in Fig. 2(a). The connection strengths between different networks also depend on the difference in the preferences of the corresponding MTJs and are scaled by another parameter, J_{rep}. The functional task of the coupled CANNs is to reduce the uncertainties of input cues to extract precise information. In the following, we demonstrate the functionality and robustness in hardware implementation with MTJs via numerical simulations. In particular, the MTJ-based networks show a better performance and require less-critical network parameters than the computational neuroscience model, providing great flexibility for device fabrication and applications.

First, we consider two CANNs, each consisting of 100 MTJs, which receive a fixed external stimulus corresponding to $\theta_{ext} = 0^\circ$. Independent white Gaussian noise with strengths of $\beta_{sys} = 0.015$ V and $\beta_{inp} = 0.005$ V are superposed on the interior and input signals, respectively. To examine the effect of information integration, we consider three strategies. The first two strategies correspond to the external stimulus applied separately to the first network (denoted as Net 1) and the second network (denoted as Net 2). In the third strategy, we let both networks receive their external inputs simultaneously. The three strategies are schematically illustrated in Fig. 2(b), and the firing rates (switching frequencies) are shown in Fig. 2(c). When only a single network receives the external stimulus, the other network is excited indirectly by the internetwork connections, and thus, exhibits a wider response. If both networks receive their external inputs simultaneously, the bandwidth of the firing rate slightly decreases.

Owing to the reciprocal structure of the two CANNs, we investigate the network's estimate, θ_{est}, under the three external input strategies by decoding the response of the first network (i.e., Net 1). For every external input strategy, a run with 400 trials is performed and the calculated values of θ_{est} are shown as scattered dots in Fig. 2(d). The corresponding distribution is plotted on the right side. Then we repeat 20 runs with 400 trials each and plot the average distributions of θ_{est} in Fig. 2(e) for the three strategies. All three curves exhibit a bell-shaped distribution centered at $\theta_{ext} = 0^\circ$, as expected. Moreover, the integrated information with two simultaneous sensory inputs has a narrower distribution than the other cases with only one sensory input. The variances of θ_{est} are plotted in Fig. 2(f) for a quantitative comparison, where the low green column manifests the accuracy of the integrated sensory information.

As a benchmark, we employ the Bayesian inference to estimate the variance for comparison. To do so, we consider the first two strategies shown in Fig. 2(b) and obtain θ^1_{est} and θ^2_{est} by decoding the output of the first network. Following the Bayes theorem, we calculate the variance of Bayesian inference across 400 trials, $S_B = [S(\theta^1_{est})^{-1} + S(\theta^2_{est})^{-1}]^{-1}$ [61]; see Eq. (A9) in the Appendix. After averaging the results from 20 runs with 400 trials for every run, we find the average variance using the Bayesian inference is larger than that from network integration. This quantitative comparison demonstrates the remarkable accuracy of integrating multiple sensory information using the reciprocally connected magnetic CANNs.

The network with superparamagnetic MTJs for sensory-information integration has an excellent performance across a wide range of network parameters. This includes the voltage boundary of neuron preference, strengths of synaptic connections, the threshold of the firing rate, the number of neurons, and the sampling rate of MTJs, as systematically examined and documented in the Supplemental Material [51]. In addition, the synaptic connections can be truncated within just a few neighboring MTJs, while the performance of the network is kept the same; see the Supplemental Material [51]. The networks work reasonably well in the presence of input and system noise, as well as the inhomogeneity of superparamagnetic MTJs because of the population coding strategy of CANNs [62]; see the Supplemental Material [51]. These advantages are very helpful for reducing complexity and energy consumption in future applications.

Our neuromorphic model implemented by MTJs exhibits more advantages than the computational neuroscience model [26,28,29]. For example, the achieved integration of multisensory information does not require critically fixed J_{rep} and J_{rec} (see the Supplemental Material [51]), in sharp contrast to the original computational model [26]. This is because the tuning curve of a neuron

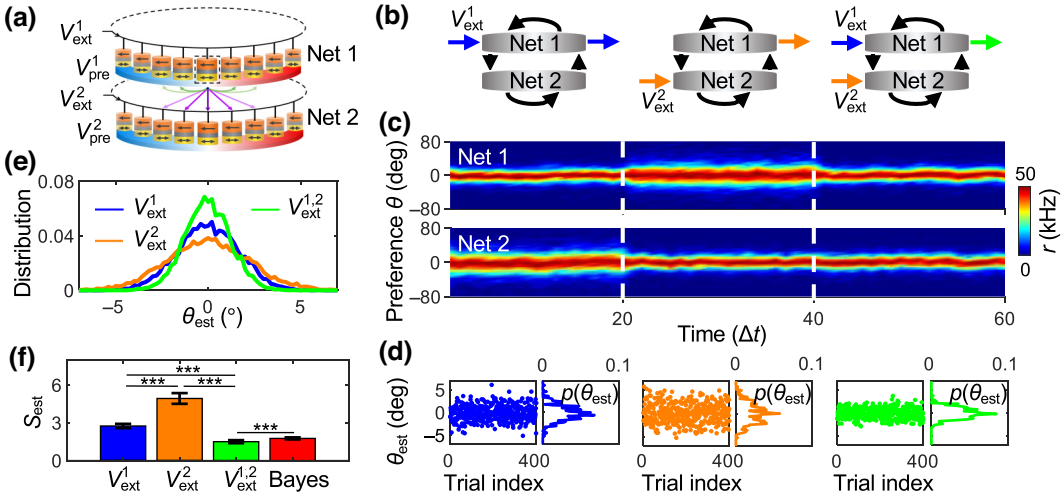

FIG. 2. (a) Structure of two coupled CANNs based on MTJs. Two CANNs have the same structure as in Fig. 1(c). Each CANN receives an independent voltage as the input sensory signal and communicates through the excitatory connections between networks denoted by purple arrows. (b) Schematic diagram of the three external stimulus input conditions: (1) only V^1_{ext} is applied to network 1 (left); (2) only V^2_{ext} is applied to network 2 (center); (3) V^1_{ext} and V^2_{ext} are simultaneously applied to the two networks (right). (c) Population activities of two networks for the three stimulus conditions in a temporal order. (d) Decoded θ_{est} from the firing rate in network 1 for 400 trials, while the right curves illustrate the distribution of the 400 estimated values. (e) Average distribution of extracted θ_{est}. (f) Mean variance obtained over 20 runs. Error bars indicate the standard deviation of the variances over 20 runs. Bayesian inference is plotted for comparison. Three asterisks indicate the significant difference (***, $p < 0.001$).

in the computational model relies sensitively on the connections in the network. However, a MTJ naturally has a bumplike response function for the input voltage, even if it is not connected to any others. Such an intrinsic dynamic property offers a large flexibility in choosing appropriate network parameters to achieve a better performance and greatly reduces the difficulty in hardware implementation of brain-inspired computing.

IV. NETWORK WITH THREE INTERCONNECTED MTJ-BASED CANNS: ROBUSTNESS AGAINST MODULE BREAKDOWN

The above network can be generalized to contain three or more CANNs that are mutually connected. Here, we take a network with three CANNs as an example, sketched in the inset of Fig. 3, to verify its functionality in the presence of module breakdown. With a single external input (V^i_{ext}, $i = 1, 2, 3$), the variance of the network estimate obtained by decoding the response of the first network is plotted as blue, orange, and purple bars, respectively. When the three CANNs simultaneously receive V^i_{ext}, the variance is much smaller (the green bar) than those with only one of the CANNs receiving the external input. The performance of information integration by three CANNs is even better than that of the standard Bayesian inference (the red bar).

Next, we test the functionality of the network when the module does not work properly. Three possible problems are specifically introduced, namely, the external input signal for the third CANN is blocked (B1), the reciprocal signal transmitted towards the third CANN is blocked

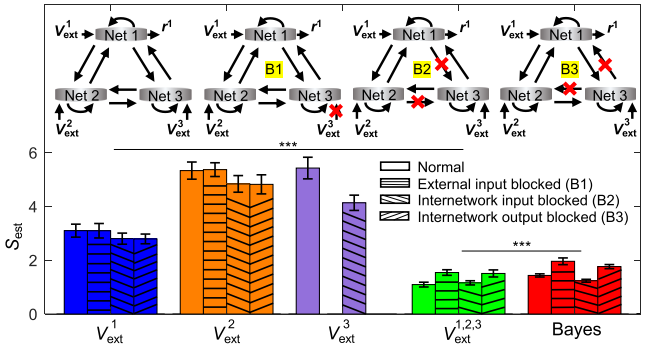

FIG. 3. Information integration by three networks that are reciprocally connected with one another. Top inset illustrates the system consisting of three CANNs. From left to right, normal; external stimulus for network 3 not received (B1); input communication to network 3 blocked (B2); output communication from networks blocked (B3). Bottom histograms show the calculated average variance from network 1 under the four input conditions: networks 1–3 receive the external stimuli, respectively, and all three networks receive the input simultaneously. Error bars represent the standard deviation over 20 runs, while every run contains 400 independent trials. Variance of Bayesian inference is plotted for comparison. Note that with B1 or B3 and only network 3 receiving the external stimulus, the system does not respond. Three asterisks indicate the significant difference (***, $p < 0.001$).

(B2), and the reciprocal signal transmitted from the third CANN is blocked (B3); see the inset of Fig. 3. The variances integrated by the normal and faulty networks with different external inputs are plotted at the bottom of Fig. 3. These module breakdowns may affect the integrated information in cases with a single external input. Nevertheless, the variance is much less changed with the three simultaneous inputs, as shown by the green bars. In addition, the variance from network integration is noticeably smaller than that from the Bayesian inference. It is worth noting that, for a single input, the variances under the breakdown of B2 and B3 are smaller than the normal case, which can be understood as follows. For example, under B2 or B3, and only the first or the second network receiving the external input, the system noise of the third network does not influence the estimate by the first network due to blocked communication. Thus, the variance is slightly lowered. When only the third CANN receives the external input under B2, the system noise of the other two networks transferred to the third network is filtered owing to the communication breakdown. Consequently, the output information from Net 3 is more accurate, resulting in a smaller variance (the right purple column). However, if three networks simultaneously receive external inputs, the smallest variance is only achieved for the normal case, as shown by the green columns.

The robustness of this network against module breakdown is attributed to its decentralized structure. Every CANN first makes a local estimate with its own population coding strategy, and communication between CANNs via the reciprocal connections further help to achieve a more accurate global estimate. This is significantly different from the conventional centralized computing structure, where a breakdown in the central computing unit must lead to the whole system failing.

V. CONCLUSIONS

We develop a population coding scheme implemented by superparamagnetic MTJs, which are used to construct the CANN. The recurrent connections within the CANN play an essential role in the accuracy and retention of encoded information. The mutually connected CANNs are designed as a decentralized neural network for integrating sensory information. Numerical simulations demonstrate that this spintronics-based hardware network has an excellent capability to integrate noisy sensory information and does not sensitively depend on the network parameters. This decentralized network is robust against various breakdowns of sensory and communicative modules. Since multisensory-information integration is a necessary and crucial unit in future robotics and artificial intelligence as a basis for decision making, our findings provide an important reference for the development of neuromorphic computing hardware.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grants No. 12174028, No. 11734004, and No. T2122016) and the National Science and Technology Innovation 2030 Major Program (Grant No. 2021ZD0203700/2021ZD0203705).

APPENDIX

Dynamics of the MTJ-based CANNs

The decentralized model for multisensory-information integration in this work consists of $L(L \geq 2)$ one-dimensional CANNs. In every CANN, a group of MTJs are aligned in a ring according to their preference with periodic boundary conditions. The preference of the ith MTJ in the lth CANN ($V^l_{\mathrm{pre},i}, i = 1, \ldots, N, l = 1, \ldots, L$, where N is the size of each CANN) is equally distributed in the range of $[-V_{\mathrm{bound}}, +V_{\mathrm{bound}})$, where V_{bound} is the voltage boundary of neuron preference. To mimic the multisensory information-integration mechanism, we map the preferred value of MTJ at $V^l_{\mathrm{pre},i}$ to the continuous-orientation-feature space (θ^l_i) in the range of $[-\pi, +\pi)$, i.e., $\theta^l_i = \pi V^l_{\mathrm{pre},i}/V_{\mathrm{bound}}$.

The dynamics of each MTJ is determined by the total applied voltage, $V^l_i(t)$, including its preference voltage, $V^l_{\mathrm{pre},i}(t)$; the interactions from other neurons, $V^l_{\mathrm{rec},i}(t)$ (where the system noise is implicitly included); and the external input signal, $V^l_{\mathrm{ext},i}(t)$ (where the input noise is superposed), namely,

$$V^l_{\mathrm{tot},i}(t) = -V^l_{\mathrm{pre},i}(t) + V^l_{\mathrm{rec},i}(t) + V^l_{\mathrm{ext},i}(t). \tag{A1}$$

Here, $V^l_{\mathrm{rec},i}(t)$ includes both the recurrent interactions from neurons in the same CANN and the reciprocal interactions from neurons in other CANNs, which are given by

$$V^l_{\mathrm{rec},i}(t) = \frac{\sum_m \sum_j J^{lm}_{ij}[V^m_j(t-\Delta t) + V^m_{\mathrm{sys},j}(t)]}{1 + k\sqrt{\sum_i \left\langle \sum_m \sum_j J^{lm}_{ij}[V^m_j(t-\Delta t) + V^m_{\mathrm{sys},j}(t)] \right\rangle^2}}. \tag{A2}$$

$V^l_{\mathrm{rec},i}(t)$ increases with the recurrent inputs from other neurons and then saturates gradually due to normalization by the overall network activity. This normalization operation could be implemented through global shunting inhibition by inhibitory neurons [49], where the parameter k controls the inhibition strength. $V^m_j(t-\Delta t)$ is calculated by $f^{-1}[r^m_j(t-\Delta t)]$, where $r^m_j(t-\Delta t)$ is the switching frequency of the j th MTJ in the mth CANN at $t - \Delta t$.

The system noise voltage, $V^m_{\mathrm{sys},j}(t)$, in Eq. (A2) is given by

$$V^m_{\mathrm{sys},j}(t) = \beta^m_{\mathrm{sys}}\xi^m_{\mathrm{sys},j}(t), \tag{A3}$$

where $\xi^m_{\mathrm{sys},j}(t)$ is Gaussian white noise with zero mean and unit variance, and β^m_{sys} corresponds to its strength.

The interaction strength (J^{lm}_{ij}) from the jth MTJ in the mth CANN to the ith MTJ in the lth CANN is set to be

$$J^{lm}_{ij} = \begin{cases} \tilde{J}^{lm}\exp\left[-\left(\frac{\theta^l_i-\theta^m_j}{\sqrt{2}\sigma^{lm}}\right)^2\right], & \text{if } |\theta^l_i-\theta^m_j| < \theta_{\mathrm{th}}, \\ 0, & \text{else}, \end{cases} \tag{A4}$$

where $\tilde{J}^{lm}$ and σ^{lm} denote the absolute interactive strength and range between MTJs, and $\tilde{J}^{lm} = J_{\mathrm{rec}}$ and $\sigma^{lm} = \sigma_{\mathrm{rec}}$ for $l = m$ and $\tilde{J}^{lm} = J_{\mathrm{rep}}$ and $\sigma^{lm} = \sigma_{\mathrm{rep}}$ otherwise. The connection J^{lm}_{ij} is a function of $\theta^l_i - \theta^m_j$ and is, therefore, spatially translation invariant, so that each CANN enables a family of continuous attractors. Notably, we only consider short-range connections within the condition of $|\theta^l_i - \theta^m_j| < \theta_{\mathrm{th}}$ to save computing power consumption; see the Supplemental Material [51].

The external inputs, $V^l_{\mathrm{ext},i}(t)$, are set to be a constant, α^l_{ext}, due to the computing characteristic superposed by input noise,

$$V^l_{\mathrm{ext},i}(t) = \alpha^l_{\mathrm{ext}} + \beta^l_{\mathrm{inp}}\xi^l_{\mathrm{inp}}(t), \tag{A5}$$

where $\alpha^l_{\mathrm{ext}} \in [-V_{\mathrm{bound}}, +V_{\mathrm{bound}})$ refers to the orientation of the external input in the lth layer ($\theta^l_{\mathrm{ext}} = \pi\alpha^l_{\mathrm{ext}}/V_{\mathrm{bound}}$). $\xi^l_{\mathrm{inp}}(t)$ is Gaussian white noise with zero mean and unit variance, and β^l_{inp} is the corresponding input noise strength. The firing rate of the ith MTJ in the lth layer at time t, $r^l_i(t)$, is calculated using the generalized Néel-Brown theory, as described in the Supplemental Material [51].

Variance determined using Bayesian inference

We consider a single external stimulus θ_{ext}, generating two sensory cues (V^1_{ext} and V^2_{ext}), where the noise is assumed to be independent for the two cues. According to Bayes theorem, the posterior distribution, $p(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}}, V^2_{\mathrm{ext}})$, satisfies the following relationship:

$$p(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}}, V^2_{\mathrm{ext}}) \propto p(V^1_{\mathrm{ext}}|\theta_{\mathrm{ext}})p(V^2_{\mathrm{ext}}|\theta_{\mathrm{ext}})p(\theta_{\mathrm{ext}}). \tag{A6}$$

Here, $p(V^l_{\mathrm{ext}}|\theta_{\mathrm{ext}})$ $(l = 1, \mathrm{or}\, 2)$ represents the likelihood function modeled by a Gaussian distribution, and it indicates the probability of a particular value of V^l_{ext} generated by the given stimulus θ_{ext}. $p(\theta_{\mathrm{ext}})$ *a priori* specifies the probability of the presence of θ_{ext} and is assumed to be a uniform distribution.

Applying Bayes theorem again, the likelihood functions, $p(V^l_{\mathrm{ext}}|\theta_{\mathrm{ext}})$, can be rewritten as

$$p(V^l_{\mathrm{ext}}|\theta_{\mathrm{ext}}) \propto p(\theta_{\mathrm{ext}}|V^l_{\mathrm{ext}})p(V^l_{\mathrm{ext}}). \tag{A7}$$

Because of the uniform distribution of stimuli θ_{ext}, $p(V^l_{\mathrm{ext}})$ is also uniform. Consequently, $p(\theta_{\mathrm{ext}}|V^l_{\mathrm{ext}})$ is a Gaussian function following the likelihood functions. Substituting Eq. (A7) into Eq. (A6) and omitting the uniform probabilities of $p(\theta_{\mathrm{ext}})$ and $p(V^l_{\mathrm{ext}})$, we have

$$p(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}}, V^2_{\mathrm{ext}}) \propto p(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}})p(\theta_{\mathrm{ext}}|V^2_{\mathrm{ext}}). \tag{A8}$$

Since both $p(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}})$ and $p(\theta_{\mathrm{ext}}|V^2_{\mathrm{ext}})$ are Gaussians, their product, $p(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}}, V^2_{\mathrm{ext}})$, must be a Gaussian function, and its variance can be directly computed from the variances of the other two Gaussians as

$$S(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}}, V^2_{\mathrm{ext}})^{-1} = S(\theta_{\mathrm{ext}}|V^1_{\mathrm{ext}})^{-1} + S(\theta_{\mathrm{ext}}|V^2_{\mathrm{ext}})^{-1}. \tag{A9}$$

[1] K. Lin, A. Kansal, D. Lymberopoulos, and F. Zhao, in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, MobiSys'10* (Association for Computing Machinery Press, New York, 2010) pp. 285–298.

[2] G. Deco, E. T. Rolls, and R. Romo, Stochastic dynamics as a principle of brain function, Prog. Neurobiol. **88**, 1 (2009).

[3] T. J. Sejnowski, On the stochastic dynamics of neuronal interaction, Biol. Cybern. **22**, 203 (1976).

[4] W. Maass and A. M. Zador, in *Advances in Neural Information Processing Systems*, edited by M. Jordan, M. Kearns, and S. Solla (MIT Press, Cambridge, 1997), Vol. 10.

[5] A. Destexhe and D. Contreras, Neuronal computations with stochastic network states, Science **314**, 85 (2006).

[6] S. Habenschuss, Z. Jonke, and W. Maass, Stochastic computations in cortical microcircuit models, PLoS Comput. Biol. **9**, e1003311 (2013).

[7] S. Ding, C. J. Cueva, M. Tsodyks, and N. Qian, Visual perception as retrospective Bayesian decoding from high- to low-level features, Proc. Natl. Acad. Sci. U. S. A. **114**, E9115 (2017).

[8] B. B. Averbeck, P. E. Latham, and A. Pouget, Neural correlations, population coding and computation, Nat. Rev. Neurosci. **7**, 358 (2006).

[9] R. S. Zemel and P. Dayan, in *Advances in Neural Information Processing Systems*, edited by M. Kearns, S. Solla, and D. Cohn (MIT Press, Cambridge, 1998), Vol. 11.

[10] A. Pouget, P. Dayan, and R. Zemel, Information processing with population codes, Nat. Rev. Neurosci. **1**, 125 (2000).

[11] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, Neuronal population coding of movement direction, Science **233**, 1416 (1986).

[12] Y. Gu, D. E. Angelaki, and G. C. Deangelis, Neural correlates of multisensory cue integration in macaque MSTd, Nat. Neurosci. **11**, 1201 (2008).

[13] M. O. Ernst and M. S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion, Nature **415**, 429 (2002).
[14] D. E. Angelaki, G. Yong, and G. C. Deangelis, Multisensory integration: Psychophysics, neurophysiology, and computation, Curr. Opin. Neurobiol. **19**, 452 (2009).
[15] K. Cai and J. Shen, in *2011 International Conference on System science, Engineering Design and Manufacturing Informatization* (Guiyang, China, 2011) pp. 352–355.
[16] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Theory of orientation tuning in visual cortex, Proc. Nati. Acad. Sci. U. S. A. **92**, 3844 (1995).
[17] A. P. Georgopoulos, M. Taira, and A. Lukashin, Cognitive neurophysiology of the motor cortex, Science **260**, 47 (1993).
[18] K. Zhang, Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory, J. Neurosci. **16**, 2112 (1996).
[19] A. Samsonovich and B. L. McNaughton, Path integration and cognitive mapping in a continuous attractor neural network mode, J. Neurosci. **17**, 5900 (1997).
[20] N. K. Logothetis, J. Pauls, and T. Poggio, Shape representation in the inferior temporal cortex of monkeys, Curr. Biol. **5**, 552 (1995).
[21] N. K. Logothetis, J. Pauls, and T. Poggio, Spatial reference frames for object recognition: Tuning for rotations in depth, in AI Memos, No. 1533. (MIT Press, Cambridge, 1995). Available at http://hdl.handle.net/1721.1/6637
[22] C. C. A. Fung, K. Y. M. Wong, H. Mao, and S. Wu, Fluctuation-response relation unifies dynamical behaviors in neural fields, Phys. Rev. E. **92**, 022801 (2015).
[23] Y. Y. Mi, C. C. A. Fung, K. Y. M. Wong, and S. Wu, in *Advances in Neural Information Processing Systems*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (MIT Press, Cambridge, 2014), Vol. 27.
[24] H. T. Blair and P. E. Sharp, Anticipatory head direction signals in anterior thalamus: Evidence for a thalamocortical circuit that integrates angular head motion to compute head direction, J. Neurosci. **15**, 6260 (1995).
[25] J. S. Taube and R. U. Muller, Comparisons of head direction cell activity in the postsubiculum and anterior thalamus of freely moving rats, Hippocampus **8**, 87 (1998).
[26] W. H. Zhang, A. Chen, M. J. Rasch, and S. Wu, Decentralized multisensory information integration in neural systems, J. Neurosci. **36**, 532 (2016).
[27] S. Wu, K. Y. M. Wong, C. C. A. Fung, Y. Y. Mi, and W. H. Zhang, Continuous attractor neural networks: Candidate of a canonical model for neural information representation, F1000Res. **5**, F1000 Faculty Rev-156 (2016).
[28] W. H. Zhang, H. Wang, A. Chen, Y. Gu, T. S. Lee, K. Y. M. Wong, and S. Wu, Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation, eLife **8**, e43753 (2019).
[29] W. H. Zhang, H. Wang, K. Y. M. Wong, and S. Wu, in *Advances in Neural Information Processing Systems*, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (MIT Press, Cambridge, 2016), Vol. 29.
[30] T. Zeng, F. Tang, D. Ji, and B. Si, NeuroBayesSLAM: Neurobiologically inspired Bayesian integration of multisensory information for robot navigation, Neural Networks **126**, 21 (2020).
[31] S. Deneve, P. E. Latham, and A. Pouget, Reading population codes: A neural implementation of ideal observers, Nat. Neurosci. **2**, 740 (1999).
[32] W. J. Ma, J. M. Beck, P. E. Latham, and A. Pouget, Bayesian inference with probabilistic population codes, Nat. Neurosci. **9**, 1432 (2006).
[33] Z. Li and S. Zhang, Thermally assisted magnetization reversal in the presence of a spin-transfer torque, Phys. Rev. B **69**, 134416 (2004).
[34] W. Rippard, R. Heindl, M. Pufall, S. Russek, and A. Kos, Thermal relaxation rates of magnetic nanoparticles in the presence of magnetic fields and spin-transfer effects, Phys. Rev. B **84**, 064439 (2011).
[35] H. Tu, L. Zhang, Y. Luo, W. Lv, T. Lei, J. Cai, B. Fang, G. Finocchio, L. Bian, S. Li, B. Zhang, and Z. Zeng, Neural-like population coding based on spin-torque diode, Appl. Phys. Lett. **122**, 122402 (2023).
[36] A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yuasa, J. Grollier, and D. Querlioz, Neural-like computing with populations of superparamagnetic basis functions, Nat. Commun. **9**, 1533 (2018).
[37] J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, Neuromorphic spintronics, Nat. Electron. **3**, 360 (2020).
[38] J. Grollier, D. Querlioz, and M. D. Stiles, Spintronic nanodevices for bioinspired computing, Proc. IEEE Inst. Electron. Eng. **104**, 2024 (2016).
[39] 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).
[40] K. Y. Camsari, B. M. Sutton, and S. Datta, *p*-Bits for probabilistic spin logic, Appl. Phys. Rev. **6**, 011305 (2019).
[41] A. Sengupta and K. Roy, Short-term plasticity and long-term potentiation in magnetic tunnel junctions: Towards volatile synapses, Phys. Rev. Appl. **5**, 024012 (2016).
[42] S. Jung, H. Lee, S. Myung, H. Kim, S. K. Yoon, S.-W. Kwon, Y. Ju, M. Kim, W. Yi, S. Han, B. Kwon, B. Seo, K. Lee, G.-H. Koh, K. Lee, Y. Song, C. Choi, D. Ham, and S. J. Kim, A crossbar array of magnetoresistive memory devices for in-memory computing, Nature **601**, 211 (2022).
[43] X. Lan, Y. Cao, X. Liu, K. Xu, C. Liu, H. Zheng, and K. Wang, Gradient descent on multilevel spin-orbit synapses with tunable variations, Adv. Intell. Syst. **3**, 2000182 (2021).
[44] W. C. Yu, J. Xiao, and G. E. W. Bauer, Hopfield neural network in magnetic textures with intrinsic Hebbian learning, Phys. Rev. B **104**, L180405 (2021).
[45] R. Zand, K. Y. Camsari, S. D. Pyle, I. Ahmed, C. H. Kim, and R. F. DeMara, in *Proceeding of the 2018 Great Lakes Symposium on VLSI, GLSVLSI'18* (Association for Computing Machinery Press, New York, 2018) pp. 15–20.
[46] D. Vodenicarevic, N. Locatelli, A. Mizrahi, J. S. Friedman, A. F. Vincent, M. Romera, A. Fukushima, K. Yakushiji, H. Kubota, S. Yuasa, S. Tiwari, J. Grollier, and D. Querlioz, Low-energy truly random number generation with superparamagnetic tunnel junctions for unconventional computing, Phys. Rev. Appl. **8**, 054045 (2017).

[47] M. R. Pufall, W. H. Rippard, S. Kaka, S. E. Russek, T. J. Silva, J. Katine, and M. Carey, Large-angle, gigahertz-rate random telegraph switching induced by spin-momentum transfer, Phys. Rev. B **69**, 214409 (2004).
[48] B. Dieny, *et al.*, Opportunities and challenges for spintronics in the microelectronics industry, Nat. Electron. **3**, 446 (2020).
[49] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit, Nature **405**, 947 (2000).
[50] A. F. Vincent, N. Locatelli, J.-O. Klein, W. S. Zhao, S. Galdin-Retailleau, and D. Querlioz, Analytical macrospin modeling of the stochastic switching time of spin-transfer torque devices, IEEE Trans. Electron Devices **62**, 164 (2014).
[51] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.21.064040 for the MTJ dynamics described by the generalized Néel-Brown theory and micromagnetics, the influence of network parameters and size, cutoffs for connections, the tolerance for noise, the diversity of MTJs, and the response time of MTJs; it also includes Refs. [52–58].
[52] J. C. Slonczewski, Current-driven excitation of magnetic multilayers, J. Magn. Magn. Mater. **159**, L1 (1996).
[53] L. Berger, Emission of spin waves by a magnetic multilayer traversed by a current, Phys. Rev. B **54**, 9353 (1996).
[54] W. F. Brown Jr., Thermal fluctuations of a single-domain particle, J. Appl. Phys. **130**, 1677 (1963).
[55] A. Vansteenkiste, J. Leliaert, M. Dvornik, M. Helsen, F. Garcia-Sanchez, and B. V. Waeyenberge, The design and verification of MuMax3, AIP Adv. **4**, 107133 (2014).
[56] L. Lopez-Diaz, L. Torres, and E. Moro, Transition from ferromagnetism to superparamagnetism on the nanosecond time scale, Phys. Rev. B **65**, 224406 (2002).
[57] J. Pei, *et al.*, Towards artificial general intelligence with hybrid Tianjic chip architecture, Nature **572**, 106 (2019).
[58] W. X. Lv, J. L. Cai, H. Y. Tu, L. K. Zhang, R. X. Li, Z. Yuan, G. Finocchio, S. P. Li, X. M. Sun, L. F. Bian, B. S. Zhang, R. Xiong, and Z. M. Zeng, Stochastic artificial synapses based on nanoscale magnetic tunnel junction for neuromorphic applications, Appl. Phys. Lett. **121**, 232406 (2022).
[59] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti, and J. T. Massey, On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex, J. Neurosci. **2**, 1527 (1982).
[60] S. Wu, S.-I. Amari, and H. Nakahara, Population coding and decoding in a neural field: A computational study, Neural Comput. **14**, 999 (2002).
[61] M. O. Ernst, in *Human Body Perception from the Inside Out*, edited by G. Knoblich, M. Grosjean, I. Thornton, and M. Shiffrar (Oxford University Press, Oxford, 2005), pp. 105–131.
[62] Y. Qiao, Y. J. Zhang, and Z. Yuan, Nonuniform magnetic domain-wall synapses enabled by population coding, New J. Phys. **25**, 033031 (2023).