

## Exploiting chaotic dynamics as deep neural networks

Shuhong Liu<sup>1,\*</sup>, Nozomi Akashi<sup>2</sup>, Qingyao Huang<sup>1</sup>, Yasuo Kuniyoshi<sup>1</sup> and Kohei Nakajima<sup>1,†</sup><sup>1</sup>Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan<sup>2</sup>Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

(Received 25 November 2024; accepted 3 June 2025; published 8 July 2025)

Chaos presents complex dynamics arising from nonlinearity and a sensitivity to initial states. These characteristics suggest a depth of expressivity that underscores their potential for advanced computational applications. However, strategies to effectively exploit chaotic dynamics for information processing have largely remained elusive. In this study, we reveal that the essence of chaos can be found in various state-of-the-art deep neural networks. Drawing inspiration from this revelation, we propose a novel method that directly leverages chaotic dynamics for deep learning architectures. Our approach is systematically evaluated across distinct chaotic systems. In all instances, our framework presents superior results to conventional deep neural networks in terms of accuracy, convergence speed, and efficiency. Furthermore, we find an active role of transient chaos formation in our framework. Collectively, this study opens up a new path for the integration of chaos, which has long been overlooked in information processing, and provides insights into the potential fusion of chaotic dynamics within the domains of machine learning and neuromorphic computation.

DOI: [10.1103/PhysRevResearch.7.033031](https://doi.org/10.1103/PhysRevResearch.7.033031)

## I. INTRODUCTION

Chaos is ubiquitous in nature. It is characterized as a bounded and deterministic system with a high sensitivity to small initial state perturbations. As time passes, small perturbations in a system can yield exponentially expanding and aperiodic outcomes. Chaotic behaviors are prevalent across domains, from the intricacies of atmospheric dynamics to fluid mechanics [1]. Chaos is also evident in engineered systems, from the simple double-pendulum system to complex neuromorphic devices [2]. Additionally, numerous studies have illuminated the presence of chaotic dynamics in biological neural systems [3,4], evident in areas from electroencephalogram signal analysis [5] to the dynamical behaviors observed in resting-state brain activity [6].

The adaptability, diversity, and energy efficiency inherent in biological neural systems offer compelling blueprints for novel computational possibilities. Drawing inspiration from these, recent advancements in neuromorphic computing offer the potential to overcome the speed limit imposed by the von Neumann bottleneck [7]. Various neuromorphic devices have been proposed, employing distinct modalities, such as spintronics, photonics, quantum, and nanomaterials [8]. A particular benefit lies in the inherent dynamics of these physical systems because they facilitate proactive state evolution

without requiring additional computation overheads. Such natural evolution of physical components over time ensures computation with minimal external intervention, endowing neuromorphic devices with exceptional computation and energy efficiency.

Despite the appeal of neuromorphic systems, integrating them with chaotic dynamics has posed challenges. Reservoir computing (RC) [9–11] has achieved significant success in implementing neuromorphic systems. RC employs inherent dynamics, termed the “reservoir,” as a recurrent neural network (RNN), enabling efficient hardware implementation and training processes. However, an important condition for achieving reliable performance in RC is the echo state property (ESP) [9]. This property guarantees that the reservoir’s internal state eventually attains independence from its initial conditions, ensuring reproducible computations. This demand straightforwardly contradicts the intrinsic property of chaos, which typically exhibits sensitivity to initial conditions, thereby making the integration of chaotic reservoirs less explored. Recent discoveries have provided evidence of the expansion behavior, a crucial characteristic of chaos, in artificial deep neural networks (DNNs) [12–20], suggesting a profound interplay between chaos and information processing. However, these studies have either focused on the suppression of chaos or utilized the bifurcation at the edge of chaos, where the potential utility of chaos remains unclear.

Prompted by these insights, our study dives deep into the heart of the challenge: Can chaotic dynamics truly be assets in DNNs? Our research systematically examines the effectiveness of direct integration of chaotic dynamics in information processing, challenging conventional notions that view such dynamics as noncontributive. For DNNs, our investigation centers on revealing the existence of expansion properties. Our observations reveal the essence of chaos in

\*Contact author: [s-liu@isi.imi.i.u-tokyo.ac.jp](mailto:s-liu@isi.imi.i.u-tokyo.ac.jp)†Contact author: [k-nakajima@isi.imi.i.u-tokyo.ac.jp](mailto:k-nakajima@isi.imi.i.u-tokyo.ac.jp)

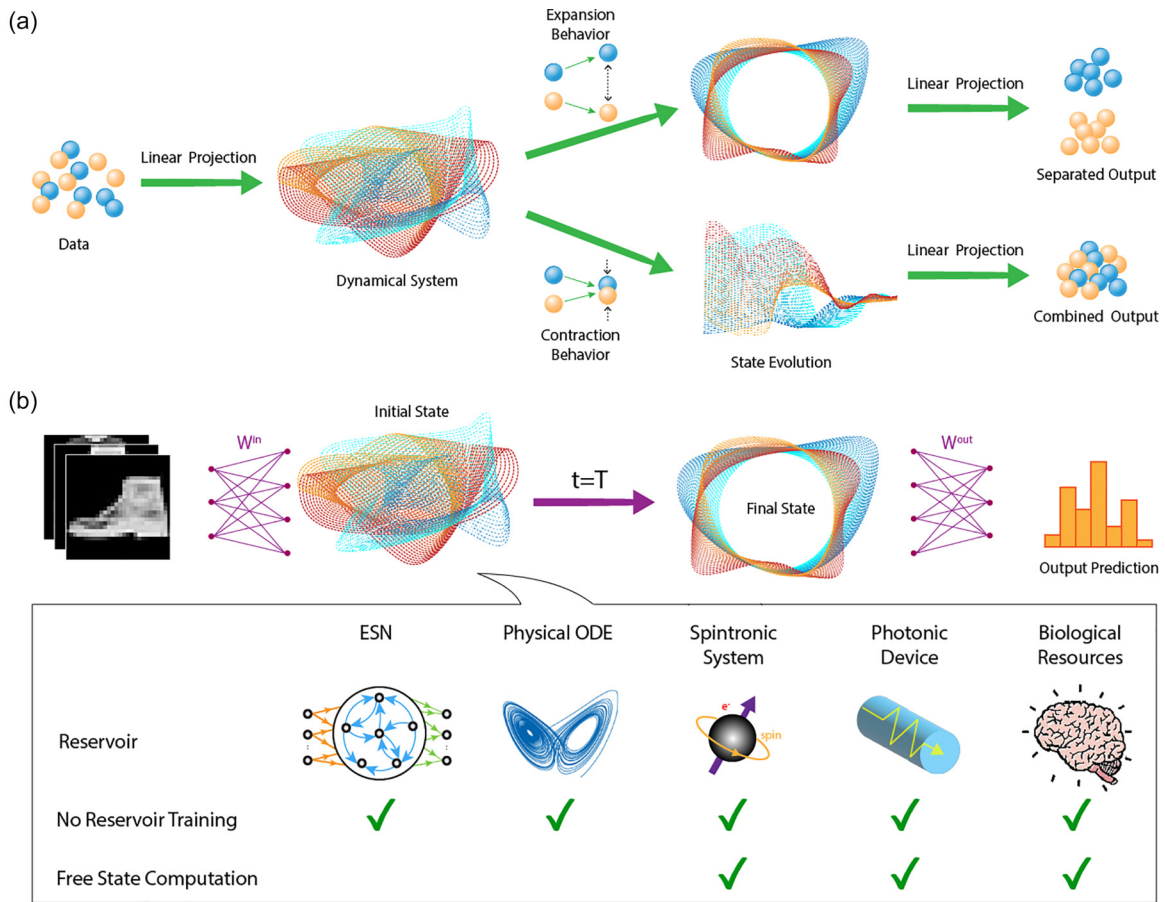


FIG. 1. (a) Illustration of expansion behavior in the dynamical system. (b) Conceptual design of the proposed framework utilizing chaotic dynamics for the purpose of image classification. The terms  $W_{in}$  and  $W_{out}$  denote the linear transformation of the state space, while  $T$  signifies the duration of self-evolution in the employed dynamical system.

state-of-the-art DNNs and illuminate its pivotal role in input separation. With this finding, we then propose a straightforward but effective framework that utilizes chaotic dynamics directly as computational media. This strategy requires many fewer trainable weights while exhibiting rapid convergence. The implications of our approach are particularly profound for the future of neuromorphic devices. To underscore its effectiveness, our framework is assessed across diverse chaotic dynamics—from discrete to continuous systems and from spatiotemporal chaos to intricate neuromorphic systems based on spin-torque oscillators.

## II. RESULTS

Given the lack of (i) a general assessment of chaotic properties in DNNs and (ii) efficient methods for utilizing chaotic dynamics for information processing, our study is carried out in two parts. Initially, we analyze the behavior of conventional DNNs from the viewpoint of dynamical systems, focusing on examining the emergence of the expansion property in various tasks. Subsequently, we propose a framework for exploiting chaotic dynamics as the computational medium and conduct experiments to verify its effectiveness. Figure 1(a) depicts the fundamental characteristics of the expansion property. In this process, input data from closely situated positions clearly separate through diverging paths as the state evolves. Figure 1(b)

outlines the conceptual design of our proposed framework. This design capitalizes on the intrinsic dynamics of the integrated system, wherein only two layers of linear projection can be trained. This approach, especially when incorporating physical systems, significantly enhances energy efficiency by leveraging natural state progression.

### A. Expansion property in deep neural networks

To reveal the presence of expansion property in neural networks, we chose a variety of state-of-the-art architectures, including the multilayer perceptron (MLP), the convolutional neural network (CNN), the convolutional autoencoder (AE), and the transformer-based language model known as bidirectional encoder representations from transformers (BERT), and applied them to tasks across distinct domains. We analyzed each model from two perspectives—its overall model and its layerwise dynamics—by assessing their transient dynamics using the finite-time maximum Lyapunov exponent (FTMLE), which can be derived from a general form of Lyapunov exponent

$$\lambda[T](\mathbf{x}_i) = \max_{\delta \mathbf{x}_i} \frac{1}{T} \log \frac{\|\delta \mathbf{x}_{i+T}\|}{\|\delta \mathbf{x}_i\|} \quad (1)$$

by using a finite  $T$  as

$$\lambda_T^{\text{FTMLE}}(\mathbf{x}_i) = \lambda[T](\mathbf{x}_i), \quad (2)$$

where  $x_i$  represents the current state at time  $i$ ,  $\delta x_i$  indicates the perturbation applied on  $x_i$ , and  $T$  is typically the time of iteration. In the context of evaluating FTMLE in DNNs,  $T$  denotes the number of propagated layers. The perturbations  $\delta x_i$  are introduced to the original input data, rather than to arbitrary values, to reflect the system's transient dynamics. The maximum Lyapunov exponent (MLE), as outlined in Eq. (1) in Sec. II A in the Supplemental Material [21], characterizes the global dynamics of the system as  $T$  tends toward infinity. Conversely, the local Lyapunov exponent (LLE), detailed in Eq. (2) in Sec. II A in the Supplemental Material [21], describes the layer's transient dynamics by setting  $T = 1$ :

$$\lambda[T](\mathbf{x}_i) = \max_{\delta \mathbf{x}_i} \frac{1}{T} \log \frac{\|\delta \mathbf{x}_{i+T}\|}{\|\delta \mathbf{x}_i\|}, \quad (3)$$

$$\lambda_1^{\text{LLE}}(\mathbf{x}_i) = \lambda[1](\mathbf{x}_i), \quad (4)$$

$$\lambda_T^{\text{FTMLE}}(\mathbf{x}_i) = \lambda[T](\mathbf{x}_i), \quad (5)$$

$$\lambda_\infty^{\text{MLE}} = \lim_{T \rightarrow \infty} \lambda_T^{\text{FTMLE}}(\mathbf{x}_0). \quad (6)$$

### 1. Multilayer perceptron

Comparable experiments evaluating the FTMLE in shallow MLPs have been conducted in previous research [22,23]. We expanded upon their analyses by tracking the FTMLE values of individual samples across each hidden layer. The MLP was trained for multiclass classification tasks. Specifically, we selected a two-dimensional input state space, enabling us to monitor the change in FTMLE values along the propagation through hidden layers. As depicted in Fig. 2(a), the trained MLP overall exhibited negative FTMLE values, with the first three layers predominantly displaying positive values. In contrast, the later layers, which were responsible for finer-grained classification, tended to exhibit negative FTMLE values, indicating a contraction behavior. In addition, the FTMLE heat map shows the formation of the decision boundaries in hidden states, wherein broader linear boundaries progressively transformed into more specific nonlinear boundaries. Regions with positive FTMLE values suggested an expansion behavior of features closely associated with the decision-making process. Additional experiments carried out on various datasets are provided in Sec. IB in the Supplemental Material [21]. The comparison of these results to the FTMLE analysis of untrained MLPs, as detailed in Sec. IC in the Supplemental Material [21], reveals a notable observation: the emergence of positive FTMLE values. This finding suggests that expansion properties are employed during training to facilitate classification via the separation of samples.

### 2. Convolutional neural network

We employed a pretrained ResNet50 model [24] that had been trained on the ImageNet dataset [25] and conducted our analyses using a subsampling from its validation set. As depicted in Fig. 2(b), our analysis of ResNet50 revealed an overall positive FTMLE distribution. Specifically, the initial max-pooling layer and the subsequent four bottleneck layers exhibited positive FTMLE values, implying that the expansion property was primarily harnessed for both low-level and high-level feature extraction purposes. Conversely, in the final

average-pooling layer, where feature contraction occurred to generate predictions, we observed negative FTMLE values. In comparison to the untrained CNNs described in Sec. IC in the Supplemental Material [21], for which certain layers displayed positive FTMLE values possibly due to the network's structure, we noticed that the training process enhanced expansion behavior. This enhancement was evidenced by the remarkably more positive post-training FTMLE distributions. These findings suggest that the network effectively utilized the expansion property during the feature extraction stage.

### 3. Autoencoder

In addition to classification tasks, our study also examined image generation as a regression task. A pretrained AE based on ResNet-18 was loaded on the CIFAR10 dataset [26] for this purpose. The task required that output images retain the same dimensions as the input images while displaying feature compression in the encoder and decompression in the decoder. As depicted in Fig. 2(c), the overall network exhibited expansion property; however, when examined separately, the encoder and decoder clearly aligned with the task requirements by exhibiting contraction and expansion behaviors, respectively. When contrasted with the untrained AEs that displayed purely negative overall model FTMLE distributions, as shown in Sec. IC in the Supplemental Material [21], it is evident that the training process endowed the model with the expansion property, enabling it to effectively encode and decode input images.

### 4. Bidirectional encoder representations from transformers (BERT)

BERT [27] serves as a promising backbone network for learning feature representations in natural language processing tasks. For computational efficiency, we used BERT-mini [28], a more compact variant of BERT equipped with four layers of transformer encoder and an embedding size of 256. To investigate the internal dynamics of feature propagation across transformer layers in BERT-mini, we conducted our analysis using the IMDb benchmark [29].

As illustrated in Fig. 2(d), the overall model and its sublayers exhibited positive FTMLE values. In particular, each sublayer showed dominant expansion behavior, suggesting that the language token encoding process requires the use of the expansion property for effective language representation learning. Compared to the evaluation of the untrained BERT-mini model detailed in Sec. IC in the Supplemental Material [21], the training process demonstrated a significant enhancement in the model's expansion behavior. Section I.D in the Supplemental Material provides a more in-depth analysis of the relationship between the FTMLE distribution and the input token density. We therefore conclude that input sentences with richer content tokens (fewer padding tokens) are more likely to expand, thus preserving their information during layer propagation. In contrast, input sentences with fewer content tokens (more padding tokens) are more inclined to contract as they contain less information.

Our experiments revealed that the expansion property was consistently observed across various state-of-the-art DNNs. Furthermore, in contrast to the randomly initialized DNNs

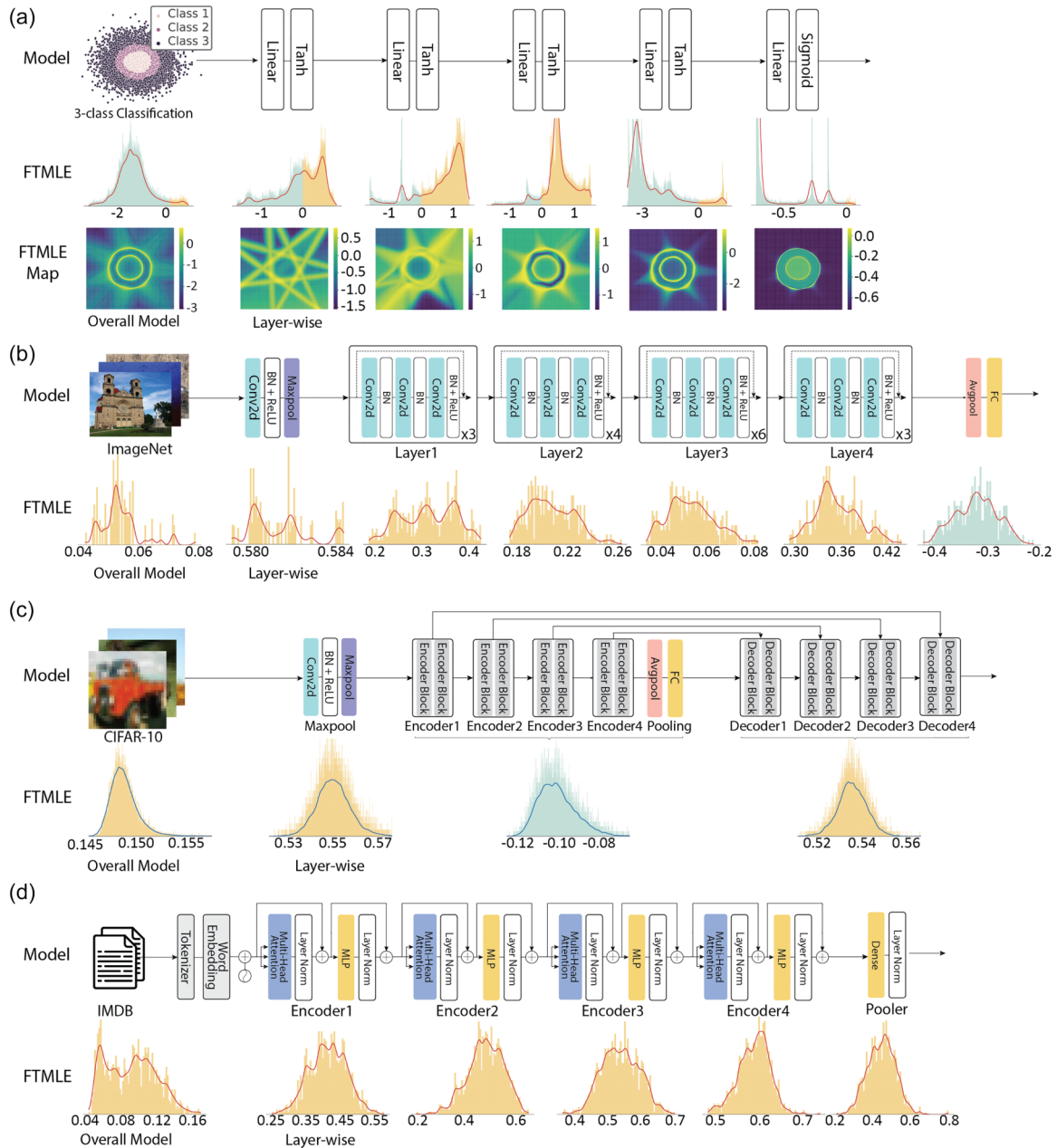


FIG. 2. The FTMLE values with respect to input samples, showing overall and layerwise distributions. Blue areas indicate contraction zones; orange areas indicate expansion. (a) MLP on a synthetic three-class classification dataset; test data uniformly sampled in 2D space. (b) ResNet50 on ImageNet validation set. (c) Autoencoder on CIFAR10 test set. (d) BERT-mini on IMDb test set.

before training, which demonstrated either no or only weak expansion behavior, the DNNs after training exhibited strong expansion behavior. This indicates that the training process equipped or reinforced the models with this capability for feature expansion. These observations lend strong support to the notion that expansion maps are utilized in DNNs to facilitate information propagation across layers.

## B. Exploiting chaotic systems as computation media

Driven by the above observation, we propose a framework that incorporates chaotic systems into neural networks as backbone models. To harness the intrinsic expansion property of chaotic systems, we simply introduced linear

transformation layers before and after the dynamical system. These two linear layers could be tuned by leveraging back-propagation, while the intermediate system remains fixed to preserve its intrinsic dynamics. To show the generalization capability of our proposed method, we chose a variety of chaotic systems as the computational medium, encompassing both discrete and continuous types.

### 1. Feedforward echo state network (FFESN)

The echo state network (ESN) emerges as a distinctive subcategory of RNNs and is distinguishable by its nontrainable recurrent layer [9]. Figure 3(a) illustrates the conventional network architecture of ESN designated for time-series data,

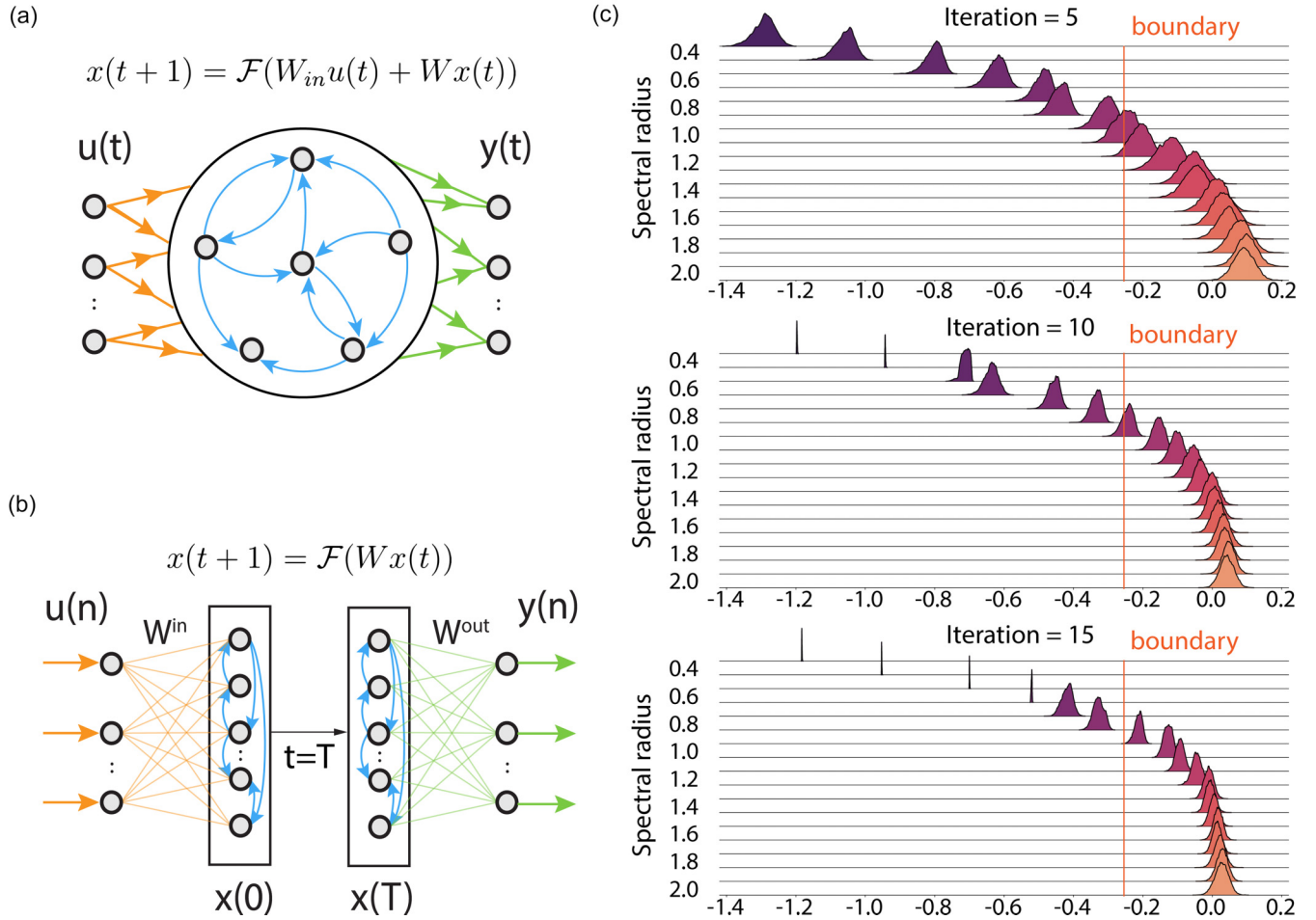


FIG. 3. The architecture and analysis of the FFESN. Training and evaluation were conducted on the MNIST dataset. (a) Traditional ESN framework. (b) The FFESN model takes a linearly transformed input as the initial state. The forthcoming state depends solely on the reservoir's present state. (c) FTMLE analysis of the FFESN, considering a spectral radius  $\rho$  varying from 0.3 to 2.0. Iteration steps  $T$  were chosen as 5, 10, and 15, corresponding to the upper, middle, and lower graphs, respectively. Each pinnacle on the ridge signifies an FTMLE distribution for the reservoir at each  $T$ .  $\rho$  under 0.3, exhibiting more negative trends, are excluded.

channeling it through internal reservoirs. To perform tasks of nontemporal data, such as image classification, we utilized the feedforward echo state network (FFESN) [30]. This architecture eliminates the requirement for time-series input  $u(t)$ , instead receiving input data via a linear transformation layer that determines the initial state of the ESN iteration layer, as depicted in Fig. 3(b). These linear projection layers, akin to those used in extreme learning machines or random projection networks [31–34], ensure that nonlinearities emerge solely from the intrinsic dynamics of the integrated system rather than from tunable projection weights. The dynamics of the FFESN are measured through transient behaviors, given the restricted iteration step  $T$ . These dynamics are characterized by the spectral radius  $\rho$ , which is the maximum absolute value among the eigenvalues of the recurrent layer's internal weight matrix, predetermined during model initialization. In standard ESNs, chaotic tendencies appear when  $\rho > 1$  [35,36]. For consistency in analysis, we employed the FTMLE to analyze the recurrent layer, aligning our approach with earlier DNN analyses, to probe its transient expansion or contraction properties. Figure 3(c) illustrates the FTMLE assessment of FFESNs initialized with  $\rho$  values spanning 0.3 to 2.0 and

$T$  values selected from 5, 10, and 15. Notably, the transient behaviors of FFESNs exhibited expansion maps when  $\rho > 1$ . When  $T$  became larger, its corresponding FTMLE distribution exhibited values closer to the MLE steered by  $\rho$ .

To directly illustrate the state changes resulting from the internal dynamics of the recurrent layer, we evaluated the trajectories of reservoir states within the recurrent layer for both nonchaotic and chaotic FFESNs. As depicted in Fig. 4(a), the principal component analysis (PCA) diagram of the chaotic FFESN displayed a more distinct separation trajectory compared to the nonchaotic ESN. In the case of a nonchaotic network, its inherent contracting nature led to a significant contraction of the state space throughout the iteration procedure. This contraction resulted in the gradual vanishing of reservoir neuron values, which asymptotically trended toward zero. In contrast, the chaotic network not only retained the variance but also amplified the clarity of separation after certain iterative steps, thus demonstrating the usefulness of the expansion map induced by chaos. To further investigate the effects of  $\rho$  and  $T$ , we computed the average testing accuracy and the convergent epoch as the number of epochs before reaching optimal accuracy. Figure 4(b) displays the average

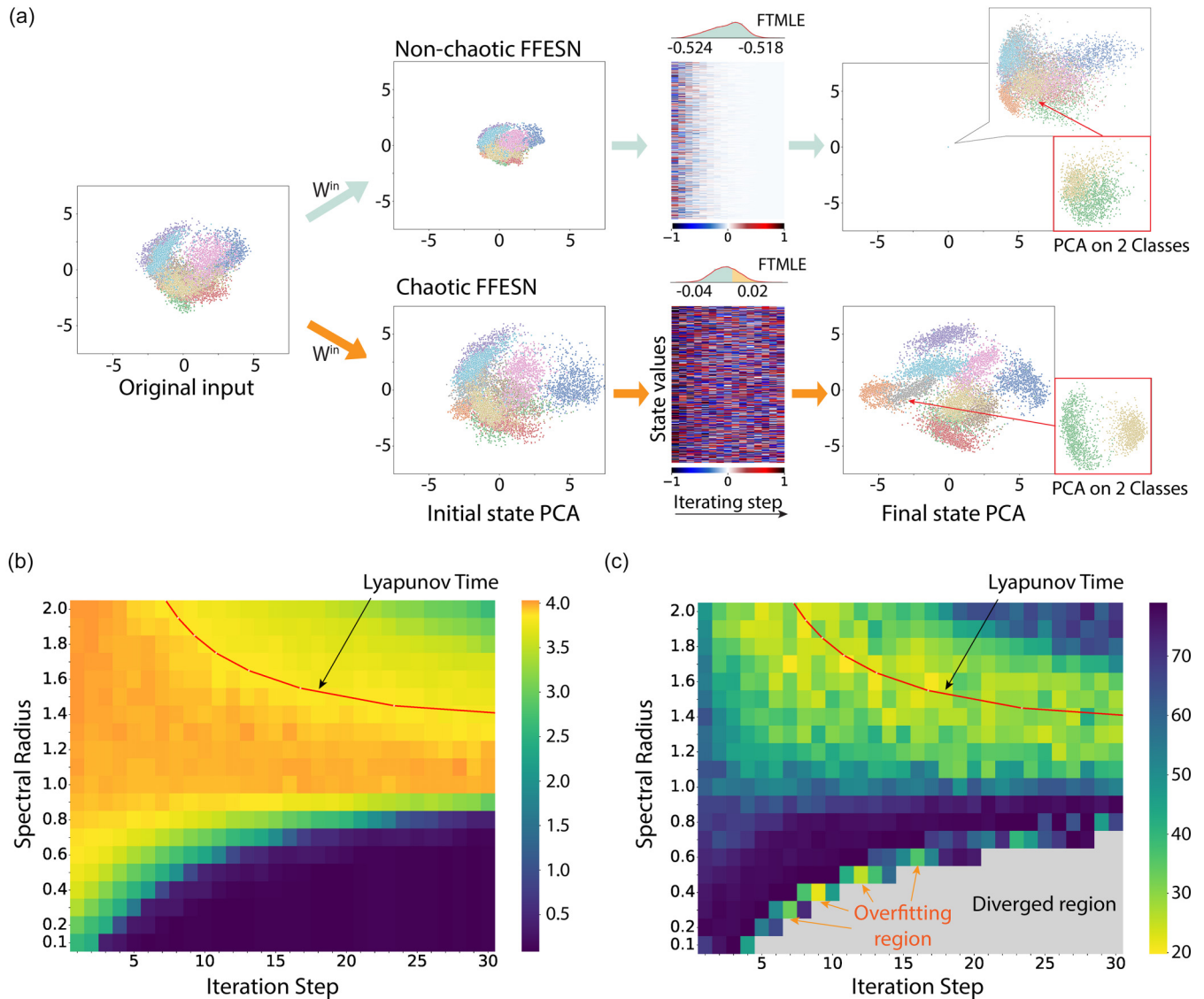


FIG. 4. (a) Comparison of classification capability between FFESNs with  $\rho$  of 0.6 and 1.4 at  $T = 15$ . This includes initial and final reservoir states and neuron state dynamics over time, with neuron numbers on the vertical axis, timestamps on the horizontal, and neuron values color coded. FTMLE distributions are shown for  $T = 15$ , with expansion behavior indicated in orange. PCA on the final state image highlights differences in specific categories. (b) The heat map reflects testing accuracy across  $\rho$  and  $T$  values, using  $|\log(\epsilon)|$  for prediction error shaded in colors, with higher values indicating better accuracy. The red line shows the Lyapunov time, denoting the system's predictability timeframe (c) The heat map displaying the FFESN convergence rate, based on epochs needed for optimal accuracy within a  $5 \times 10^{-4}$  error margin, averaged over five trials. The gray area indicates performance akin to random predictions, and the orange arrows point to overfitting areas.

accuracy heat map; the optimal accuracy was found in the chaotic region where  $\rho > 1$ . Additionally, as  $T$  increased, the accuracy of the chaotic network tended to be preserved, in contrast to the nonchaotic network, for which the accuracy suddenly dropped after a few iterations.

Moreover, the red line represents the Lyapunov time, a key time measure that quantifies how long it takes for small disturbances to expand across the full magnitude of the attractor, offering a predictive time scale for the predictability of the system's behavior. This serves as an insightful indicator, identifying the configuration of  $T$  for which the network achieves its optimal accuracy region. Figure 4(c) displays the average epoch at which convergence occurred during training. We

observed that chaotic FFESNs tended to converge more rapidly when  $T$  aligned with its associated Lyapunov time compared to other regions. This quicker convergence can likely be attributed to more effective weight optimization for linear layers, induced by chaotic iteration layers, throughout the training process. Additionally, with a relatively large  $T$ , the network suffered from overfitting, resulting in lower accuracy or even divergence. An evaluation of both optimal accuracy and convergence speed showed that the chaotic network exhibited a marked superiority over its nonchaotic counterpart. As shown in Sec. V A in the Supplemental Material [21], we conducted an evaluation of FFESN using the Fashion-MNIST dataset and observed similar trends

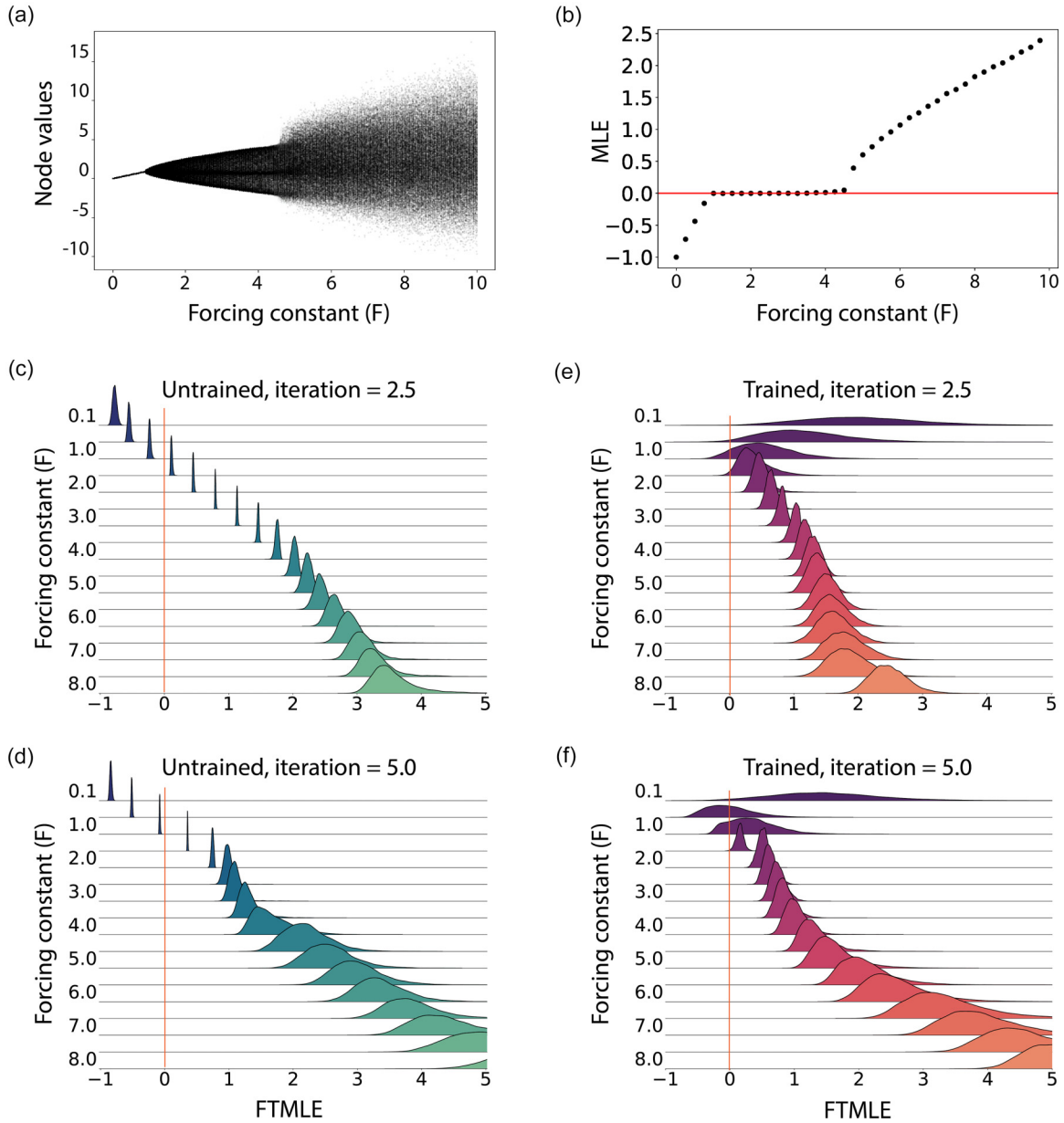


FIG. 5. Bifurcation analyses of the Lorenz 96 system on the MNIST task. (a) Global bifurcation diagram depicting the system behavior with random input over an iteration time  $T = 20$ . (b) The MLE identifies global dynamics of fixed-point convergence, quasiperiodic orbits, and chaos. (c), (d) Illustration of the FTMLE spectra for the untrained Lorenz 96 models. (e), (f) Distribution of the FTMLE for the trained Lorenz 96 models, particularly highlighting the divergent transient dynamics observed at smaller  $F$  values (e.g., less than two) compared to their untrained counterparts.

concerning the region of optimal accuracy and convergence speed. Furthermore, to explore the impact of chaos on convergence speed in a general case, we evaluated the convergence speed of MLPs initialized at weights with various spectral radius, as shown in Sec. V B in the Supplemental Material [21]. We discovered that chaos plays an active role in enhancing the speed of convergence.

## 2. Lorenz 96

We extended our framework to continuous-time internal layers to evaluate the effectiveness of real-world chaos. We employed the Lorenz 96 [37], a model of the atmosphere that shows high-dimensional spatiotemporal chaos with

continuous time, with trainable linear  $W_{\text{in}}$  and  $W_{\text{out}}$  added before and after the dynamical system, as formulated by the following equation:

$$\mathbf{x}(0) = W_{\text{in}} \mathbf{u}(n), \quad (7)$$

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t \dot{\mathbf{x}}(s) ds, \text{ and} \quad (8)$$

$$\mathbf{y}(n) = W_{\text{out}} [\mathbf{x}(T)^\top; 1]^\top, \quad (9)$$

where  $T \in \mathbb{R}$  represents iteration time, and  $x \in \mathbb{R}^{500}$  is the state of the system. As depicted in Figs. 5(a) and 5(b), the Lorenz 96 system's global dynamics reveal three salient regimes: fixed-point convergence, periodic oscillations, and

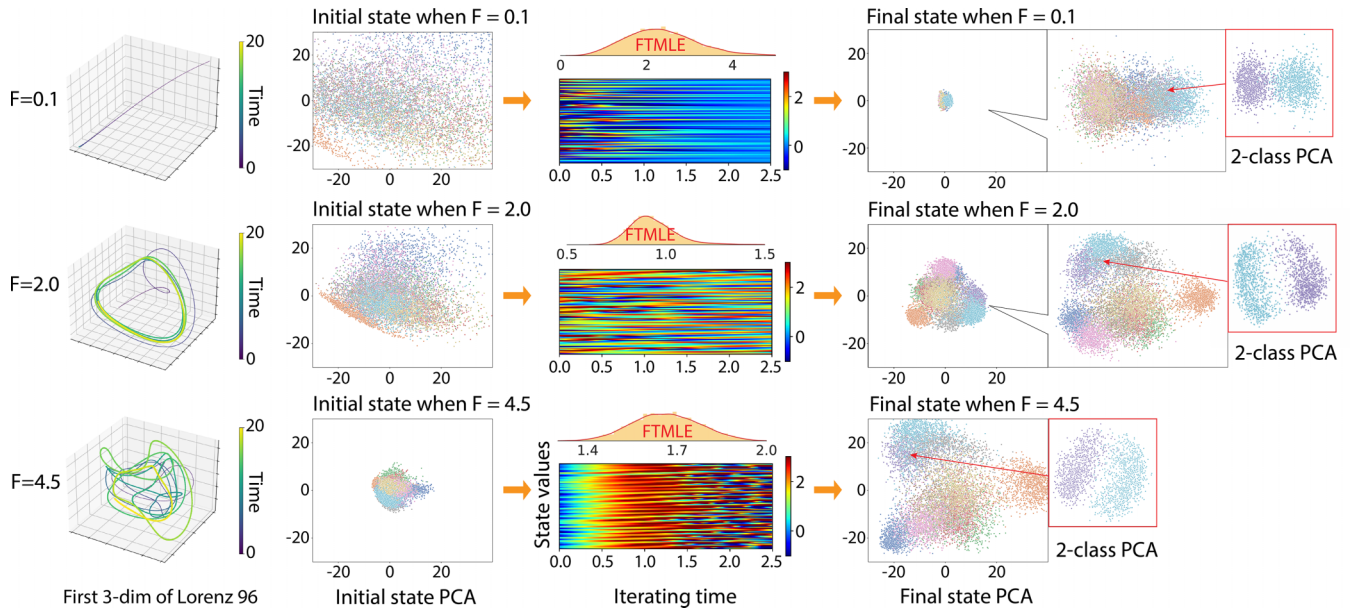


FIG. 6. PCA projections of initial and final states with FTMLE indicators of Lorenz 96 systems on the MNIST dataset. 3D trajectory diagrams show state evolution in Lorenz 96 system's first three dimensions with  $T$  spanning from 0 to 20. PCA results use  $T = 2.5$  particularly. The horizontal axis of internal state dynamics represents iteration time, and the vertical axis shows state values. The color gradient indicates neuron value.

chaotic behavior, each dictated by varying values of the forcing term ( $F$ ). The MLE diagram elucidates these dynamics in relation to  $F$ . Notably, our experimental focus was on harnessing the transient dynamics of the Lorenz 96 system for small  $T$ , which was assessed through FTMLE. Instead of applying conventional backpropagation to discrete internal layers, we employed the adjoint sensitivity method for error propagation within continuous-time internal systems [38,39].

In Figs. 5(c)–5(f), a comparative analysis between the FTMLE spectra of both trained and untrained systems across  $T = 2.5$  and 5.0 is presented. The untrained system demonstrates a monotonic trend: with a rise in  $F$ , the FTMLE values tend to increase. In contrast, the trained system exhibits a deviating pattern. Specifically, at lower  $F$  values, there is a positive FTMLE distribution. This behavior can be attributed to the training process, which allows the system to identify an optimal initial state by adjusting  $W_{in}$ , thereby suppressing the system's natural tendency to contract. Additionally, a consistent observation across all systems was the predominance of positive FTMLE distributions, hinting at the utilization of an expansion map during propagation.

Further insights are provided in Fig. 6, which highlights three representative systems corresponding to  $F = 0.1$ , 2.0, and 4.5. The depicted 3D trajectory diagrams illuminate the longer trajectory ( $T$  up to 20) for the initial three dimensions of the Lorenz 96 system, each visualizing a unique global dynamic. The subsequent state analysis was conducted at  $T = 2.5$  to assess transient dynamics. From the PCA visualizations, the contrasts between the initial and final states of the iterative internal system were evident. Specifically, at  $F = 0.1$ , the initial state appeared markedly expanded, corroborating our hypothesis regarding the expansive resistance to contraction. This initially high-variance state underwent significant contraction throughout the iteration. State

dynamics diagrams further illustrate the evolution of state values throughout the iterating period. A salient observation across all state dynamics was the presence of positive FTMLE values, again reaffirming the system's inclination toward expansion mapping. For small  $F$  values, such as  $F = 0.1$ , even when the feature space generally contracted, the distinct category clusters retained discernible boundaries, as illustrated in the two-class PCA visualization (PCA applied to two selected categories only).

Figure 7(a) shows the averaged accuracy heat map, with higher values (lighter colors) indicating higher accuracy. Notably, the regions of optimal accuracy, as represented by the luminous yellow hues, predominantly lie in regions that include (i)  $F < 2$  and  $T < 1$ , and (ii)  $F \in (1, 2.5)$  and  $T > 2.5$ , where higher  $F$  with relative small  $T$  still yields good performance. Figure 7(b) shows the results of recording the average number of epochs required to reach optimal accuracy. Disregarding the divergence region and the early overfitting region (low numbers of epochs and low accuracy), we observed that the optimal region for fast convergence overlapped with the optimal accuracy region. Notably, as shown in Fig. 7(c), all areas of the averaged FTMLE heat map display positive FTMLE values. The optimal region, distinguished by higher accuracy and faster convergence speed, exhibits relatively lower FTMLE values, yet still demonstrates expansion behavior. Furthermore, in contrast to FFESN in our main content, the Lyapunov time does not serve as an indicator for the region of optimal performance.

### C. Coupled spin-torque oscillators

We implemented our approach on spin-torque oscillators (STOs), which are a type of spintronics device, as depicted in Fig. 8(a). STOs are considered a promising option for

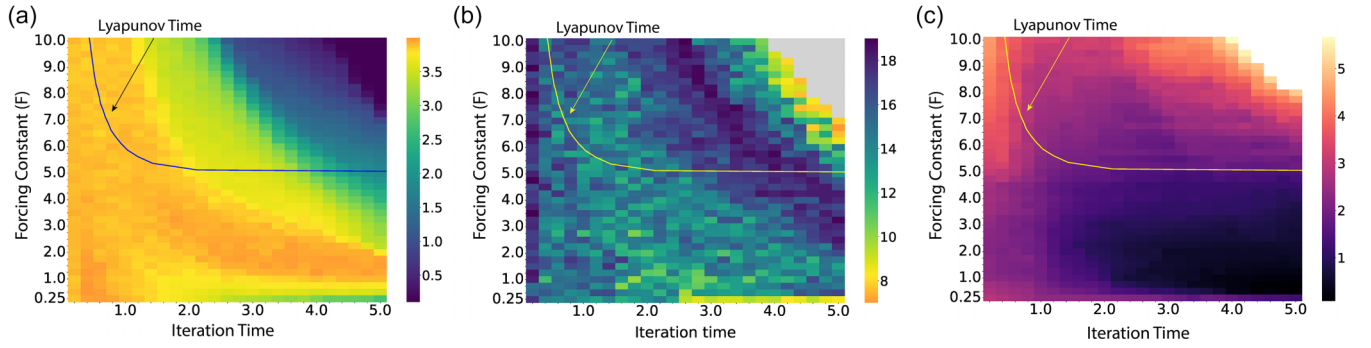


FIG. 7. Heat maps illustrating classification accuracy, training epoch, and averaged FTMLE across varied  $F$  and  $T$  of Lorenz 96 systems. (a) The accuracy map has a color spectrum indicating  $|\log(\epsilon)|$ , where  $\epsilon$  is the prediction error. (b) The convergence metric illustrates the number of training epochs necessary to achieve optimal accuracy, considering an error tolerance of  $5 \times 10^{-5}$ . (c) The FTMLE map shows the averaged FTMLE over the testing set.

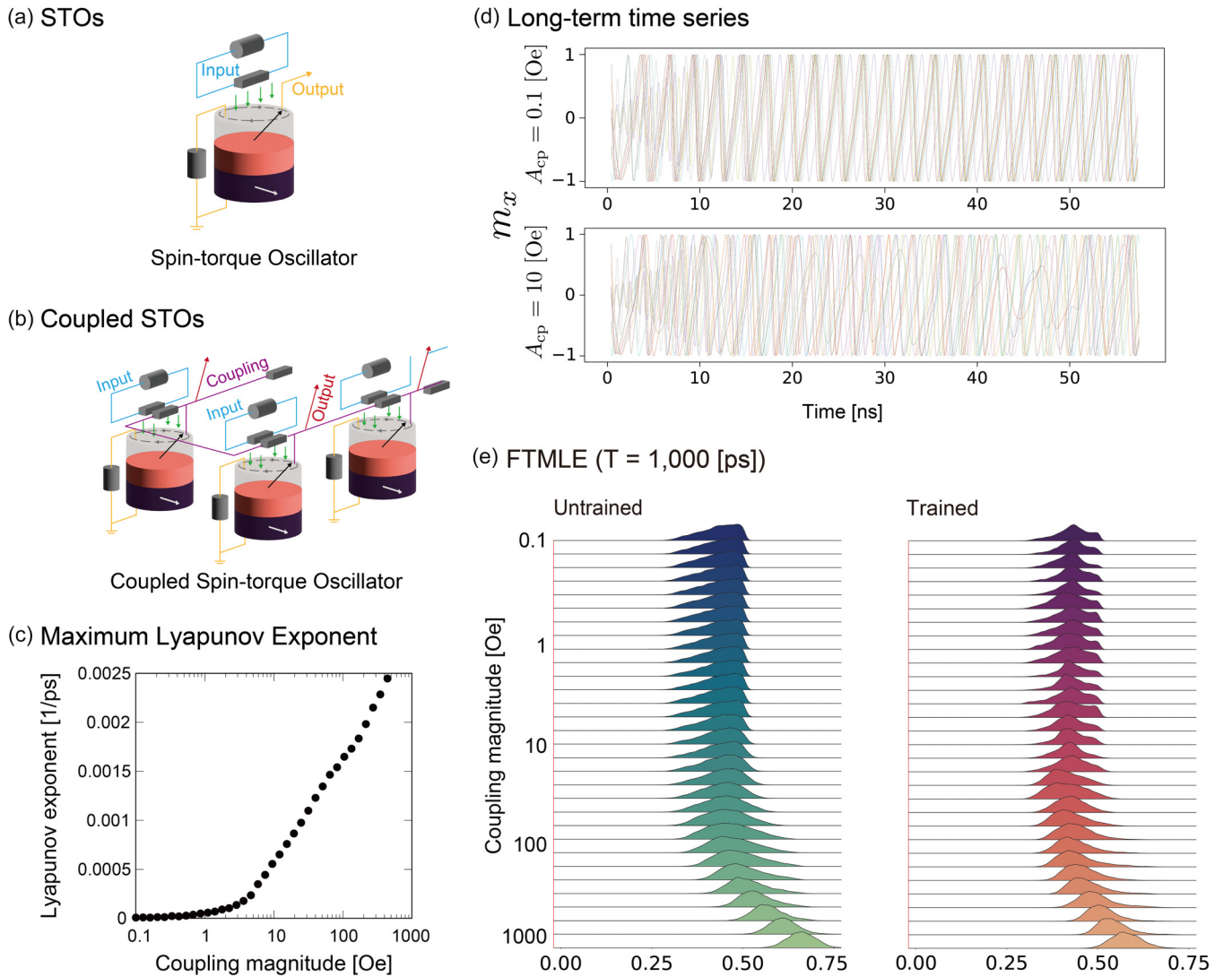


FIG. 8. Dynamical analyses for coupled STOs and solving the MNIST task using coupled STOs. (a), (b) The schematic diagrams of STOs and coupled STOs. (c) MLEs of coupled STOs through coupling magnitude  $A_{cp}$ . (d) Typical long-term dynamics of periodic and chaotic coupled STOs with random initial states. (e) FTMLE distributions for coupled STOs for MNIST input data through coupling magnitude  $A_{cp}$ . The left and right graphs show the FTMLE distributions of untrained and trained coupled STOs, respectively.

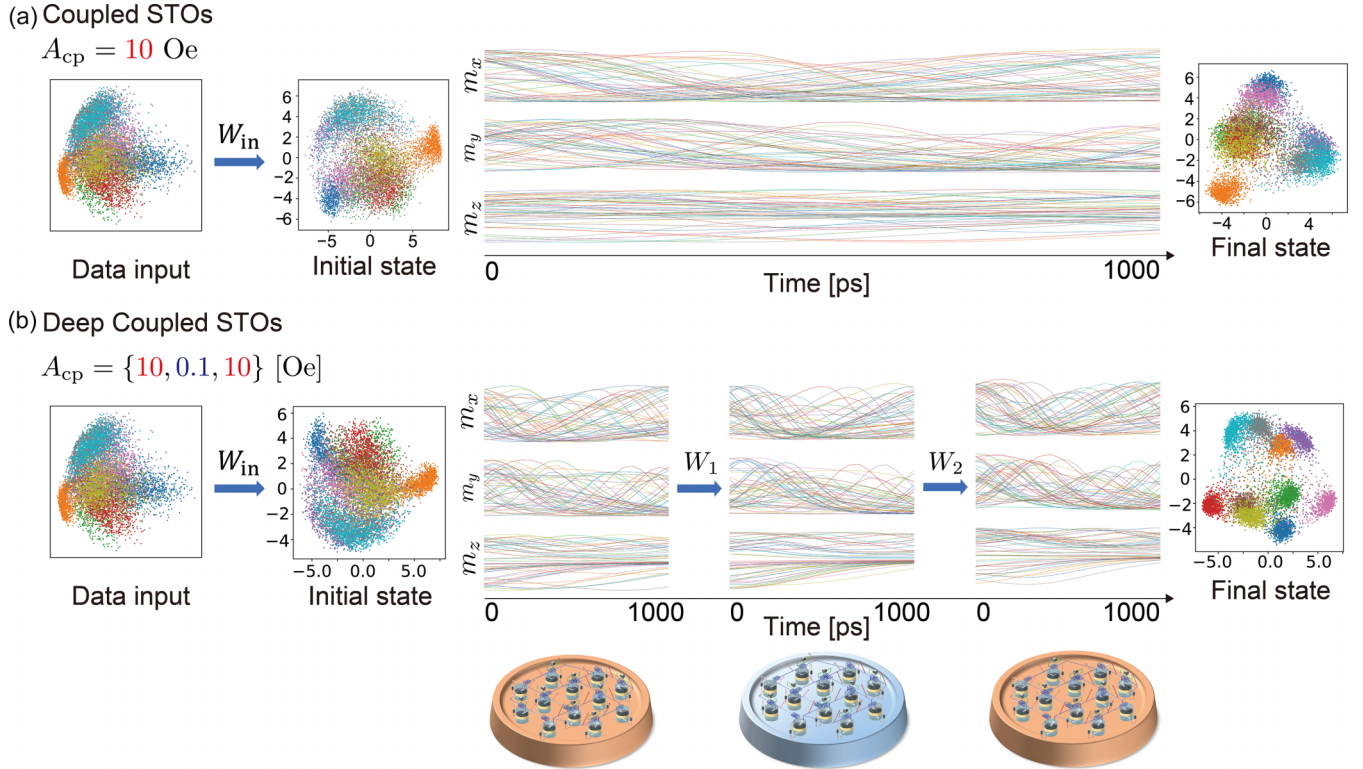


FIG. 9. (a) The architecture using coupled STOs as an internal layer. (b) The deep architecture uses three coupled STOs as internal layers.

neuromorphic computing devices due to their small size, high-speed dynamics, and high energy efficiency [8,40–43]. Typically, magnetization in STOs exhibits limit cycle oscillation dynamics, although there are reports of chaotic dynamics influenced by input [2,44,45], feedback [46,47], and coupling [48]. However, previous research has not effectively utilized chaotic dynamics for information processing. Thus, we concentrated on leveraging high-dimensional chaotic dynamics in coupled STOs, simulating these dynamics by solving the Landau-Lifshitz-Gilbert equation [48]. Details on device configurations and simulation settings are provided in Sec. III C in the Supplemental Material [21].

The behavior of magnetization in coupled STOs exhibits a bifurcation between periodic and chaotic dynamics, influenced by the coupling magnitude  $A_{cp}$ . Figure 8(c) presents the MLE of the coupled STOs across different values of  $A_{cp}$ . We observed that the MLE shifted from zero to positive, signifying a transition from periodic to chaotic dynamics, as  $A_{cp}$  increased. Figure 8(d) illustrates a typical extended time series of the magnetizations in these coupled STOs. Specifically, the magnetization in coupled STOs with  $A_{cp} = 0.1$  and 10 Oe showed periodic and chaotic dynamics, respectively.

In our proposed framework, we utilized the dynamics of coupled STOs as an internal backbone model with the mathematical formulations provided in Sec. III C in the Supplemental Material [21]. For the MNIST dataset, the input data were converted into the states of coupled STOs. These states represent the magnetization of each STO in a three-dimensional form  $\mathbf{m}_k = (m_{k,x}, m_{k,y}, m_{k,z})^T$ , based on the input weight. Our network comprised 200 coupled STOs, resulting in a total of 600 computational nodes, corresponding to the three dimensions of each STO's magnetization.

Our analysis encompassed two scenarios involving input data from both untrained (randomly initialized) and trained input weights. Despite the occurrence of bifurcation in the long-term dynamics, the FTMLE consistently remained positive in both scenarios, as shown in Fig. 8(e). This consistency indicates that the transient dynamics of the systems after training were chaotic.

Figure 9 illustrates the flow of MNIST data transformed by a specific example of chaotic coupled STOs and a chaotic deep architecture. Additionally, we present heat maps depicting the performance and convergent speed of coupled STOs in Fig. 10. Systems with an iteration time around 300 ps

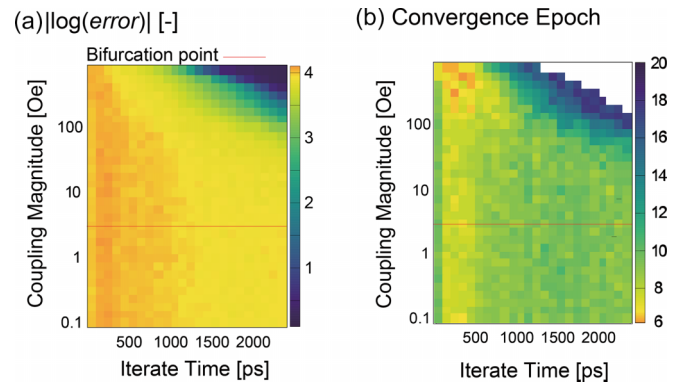


FIG. 10. Accuracy and training convergence speed of coupled STOs systems with different coupling magnitudes. (a) The MNIST performance heat map in terms of iteration time and coupling magnitude. The color represents the  $|\log(\epsilon)|$ . (b) The MNIST training speed heat map through the iteration time and coupling magnitude. The color represents the convergence epoch.

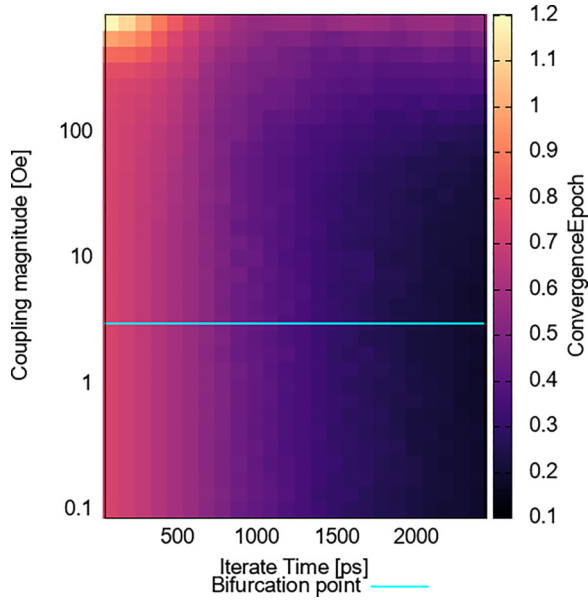


FIG. 11. The color map of average FTMLEs in the coupled STOs. The inputs of coupled STOs are MNIST data that were transformed by the trained input weight.

exhibited optimal performance. However, it is noteworthy that the peak training speed occurred in the intensely chaotic region ( $A_{cp} > 100$  Oe). This performance is on a par with state-of-the-art neuromorphic spin-torque devices for the MNIST task [49]. Such a pattern aligns with observations made in both the FFESN and Lorenz 96 systems. For the Lyapunov time, we found no correlation with either performance or convergence speed, and thus it is not presented here.

To conduct a thorough examination of the transient dynamics within the coupled STOs systems, we calculated the FTMLE values for the coupled STOs after training, as depicted in Fig. 11. We observed positive average FTMLEs across all the regions analyzed, even including areas where the global dynamics were nonchaotic, as denoted by  $A_{cp} < 2.0$  in Fig. 8(c). This divergence between global dynamics and transient dynamics indicates that the expansion behavior was generally utilized in our coupled STOs systems in finite iteration timesteps to facilitate the classification process.

For a direct comparison between discrete-time and continuous-time dynamical systems, as well as conventional

DNNs, we conducted experiments using FFESN, Lorenz 96, and coupled STOs under specific settings. These were compared with a linear regression model, an MLP with one hidden layer, and a CNN with one convolutional layer and one hidden fully connected layer. To ensure a fair comparison, the number of hidden neurons in the MLP, FFESN, and Lorenz 96 models was identical. Furthermore, we applied a convolutional layer to the coupled STOs, as shown in Fig. 12, with a read-in layer equivalent to the baseline convolutional layer. The highest accuracy and the average accuracy across five trials for each model were recorded, as presented in Table I. The specific settings for each model are detailed in Sec. IV in the Supplemental Material [21]. We observed that FFESN and Lorenz 96 systems showed superior performance over MLP, and coupled STOs with a convolutional input layer outperformed the conventional CNN as well, presenting state-of-the-art performance in neuromorphic computing [49].

For a more comprehensive analysis of the transient dynamics of the system, we computed the FTMLE values for the coupled STOs after training, as shown in Fig. 11. We found that the average FTMLEs were positive in all analyzed regions, even in regions where the global dynamics of  $A_{cp} < 2.0$  were nonchaotic indicated in Fig. 8(d). Moreover, a smaller iteration time coupled with greater coupling magnitude led to higher average FTMLEs. Notably, regions with high average FTMLEs corresponded to those with shorter convergence epochs, as shown in Fig. 10(f). These results imply that the expansion properties enhance the efficiency of the search process in stochastic gradient descent.

In addition to the superiority of accuracy and convergence speed, our coupled STOs system also demonstrated robustness to noise. Physical dynamics are inevitably subject to various types of noise, and noises in nonlinear systems can manifest as noise-induced phenomena, potentially exerting critical impacts on a system. Therefore, analyzing robustness against noises is crucial when contemplating neuromorphic computations based on physical dynamics. Specifically, we analyzed the performance of the MNIST task using the coupled STOs when injecting two types of noise described by the following equations:

$$x(0) = f_{\text{norm}}(W_{\text{in}}u + \mathbf{p}_{\text{obs}}), \quad (10)$$

$$x(i+1) = f(W'x(i)), \text{ and} \quad (11)$$

$$y = W_{\text{out}}(x(n) + \mathbf{p}_{\text{dyn}}), \quad (12)$$

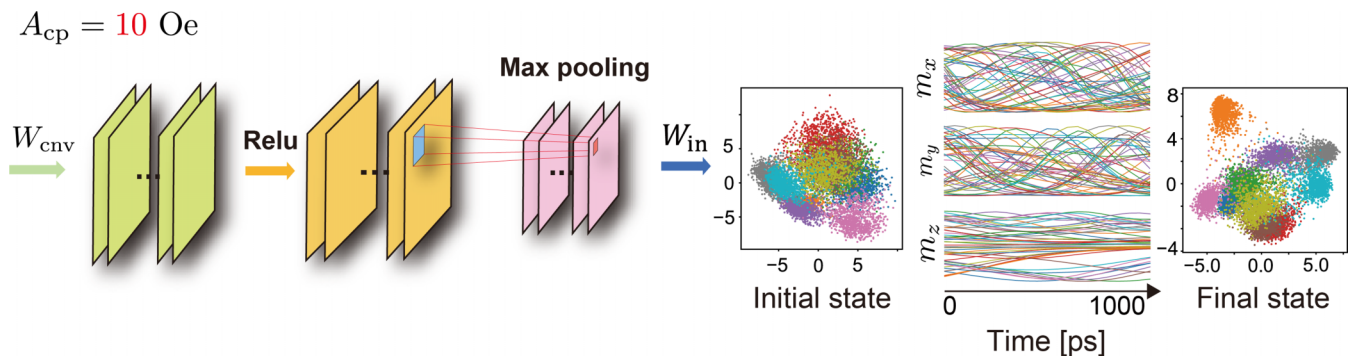


FIG. 12. The architecture of coupled STOs with a convolutional input layer. The scatter plots illustrate MNIST test data in the first and second principal component spaces. The time series plot shows the state transformations over time.

TABLE I. Accuracy comparison of the proposed framework for the MNIST dataset, showing the highest and average accuracy across ten trials for all models. For a fair comparison, the MLP and CNN models employed here all have only one hidden layer with  $N$  neurons.

| System                                                            | Highest (%)  | Average (%)  |
|-------------------------------------------------------------------|--------------|--------------|
| Linear Regression                                                 | 92.54        | 92.50        |
| MLP ( $N = 500$ )                                                 | 97.54        | 97.46        |
| FFESN ( $\rho = 0.9, N = 500, T = 1$ )                            | 98.18        | 98.08        |
| FFESN ( $\rho = 1.0, N = 500, T = 1$ )                            | 98.29        | 98.13        |
| FFESN ( $\rho = 1.8, N = 500, T = 2$ )                            | 98.34        | 98.24        |
| Lorenz 96 ( $F = 0.5, N = 500, T = 0.4$ )                         | 98.31        | 98.18        |
| Lorenz 96 ( $F = 4.75, N = 500, T = 0.2$ )                        | 98.28        | 98.12        |
| CNN ( $N = 600$ )                                                 | 98.79        | 98.71        |
| CSTOs ( $N = 600, A_{cp} = 17.8$ Oe, $T = 300$ ps)                | 98.44        | 98.34        |
| Deep CSTOs ( $N = 600, A_{cp} = \{10, 0.1, 10$ Oe}, $T = 100$ ps) | 98.56        | 98.34        |
| CNN + CSTOs ( $N = 600, A_{cp} = 750$ Oe, $T = 200$ ps)           | <b>99.05</b> | <b>99.00</b> |

where  $\mathbf{p}_{\text{obs}}$  and  $\mathbf{p}_{\text{obs}}$  represent observation and dynamical noises, respectively, generated from Gaussian distributions. We conducted five trials by varying the signal-to-noise (SN) ratio ( $A_s/A_n$ ) under two conditions: coupling strength  $A_{cp} = 0.1$  and 500 Oe, where  $A_s$  and  $A_n$  are variances of the state and noise, respectively. The results are depicted in Fig. 13. In both conditions, the performance change with respect to the SN ratio was continuous, confirming that noise did not have a fatal impact on information processing under observational and dynamic noises.

### III. DISCUSSION

In this study, we investigated the transient dynamics of state-of-the-art DNNs across diverse domains through the lens of dynamical systems, employing layerwise FTMLE analyses. Our findings underscore the pervasive use of expansion maps within DNNs to facilitate information processing, such as the encoding of input features. Furthermore, the assessment of FTMLE presents an approach to interpreting DNNs within the framework of dynamical systems, offering an intriguing perspective on the interpretability of DNNs.

Motivated by the prevalent emergence of expansion behavior in DNNs, we have introduced a straightforward framework that exploits chaotic dynamics as a computational medium. We focus on three distinct systems: (i) the discrete-time FFESN, (ii) the continuous-time Lorenz 96, and (iii) a neuro-

morphic system known as the STOs. Our experimental design encompasses a range of settings, representing various global dynamics. A subsequent deep dive into the transient dynamics of these systems revealed certain deviations from global dynamics that are particularly pronounced in continuous-time dynamical systems.

Through the experiments conducted on these varied systems, we discovered that chaotic dynamics can indeed be utilized for information processing, with our empirical tasks centered on classification. A salient observation was the role of expansion maps in these systems, as evidenced by dominant positive FTMLE distributions, which consistently led to optimal performance. This concurrence of expansion behavior with optimal classification accuracy underscores the potential of chaotic dynamics for practical applications. For example, our successful integration of STOs highlights the feasibility of leveraging real-world chaotic systems. Similarly, recent advancements in utilizing spatiotemporal chaos in photonic systems [50] have demonstrated both optimal performance and energy efficiency in image classification. These physical systems naturally progress over time and obviate the need for state computation, thereby opening new paths for building energy-efficient neuromorphic systems.

However, our study has certain limitations. First among these is the fact that our framework predominantly focuses on basic image classification tasks in which DNNs demonstrate marked proficiency. For the more difficult tasks, it remains a considerable challenge to attain performances that are superior to those achieved by state-of-the-art DNNs. A barrier to this challenge is the inherent difficulty in increasing the depth of integrated systems. Nonetheless, our research unveils the usefulness of chaotic dynamics in which expansion maps are employed. This insight presents an avenue for future researchers to incorporate chaotic components into the design of DNN architectures. A secondary concern pertains to the practical integration of STOs in real-world applications. Specifically, managing the initial state, represented by the electron's spin angle, is difficult to control with precision in practice. However, it is plausible to explore methods for controlling the initial state via input signals, drawing on the principles of input synchronization. Our empirical approach to employing chaotic dynamics in neuromorphic computing highlights the untapped potential of

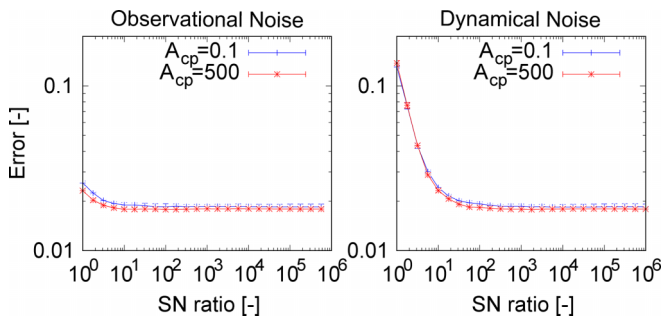


FIG. 13. The performances of the MNIST task using the coupled STOs with noise. The left- and right-hand sides show results with observational and dynamical noises, respectively. The vertical axes represent the errors as  $(1 - \text{accuracy})$ .

chaotic systems—systems that were once considered inconsequential. This revelation not only presents the utility of chaotic systems but also suggests that other physical reservoirs, with STOs being just one example, might offer even greater ease of implementation for this framework. These insights pave the way for exploring more complex integrations in future neuromorphic architectures, expanding beyond our current examples.

### ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Sumito Tsunegi and Dr. Tomohiro Taniguchi for their helpful suggestions on

the physical settings of the spin-torque oscillators used for neuromorphic simulation. This work was supported by Cross-ministerial Strategic Innovation Promotion Program (SIP) on “Integrated Health Care System” Grant No. JPJ012425, and was partially supported by a JST CREST Grant, Number JPMJCR2014, Japan.

### DATA AVAILABILITY

We have open source our code in Ref. [51] for the ease of accessibility.

- 
- [1] E. N. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.* **20**, 130 (1963).
  - [2] Z. Yang, S. Zhang, and Y. C. Li, Chaotic dynamics of spin-valve oscillators, *Phys. Rev. Lett.* **99**, 134101 (2007).
  - [3] S. J. Schiff, K. Jerger, D. H. Duong, T. Chang, M. L. Spano, and W. L. Ditto, Controlling chaos in the brain, *Nature (London)* **370**, 615 (1994).
  - [4] H. Korn and P. Faure, Is there chaos in the brain? II. Experimental evidence and related models, *C.R. Biol.* **326**, 787 (2003).
  - [5] W. S. Pritchard and D. W. Duke, Measuring chaos in the brain—a tutorial review of EEG dimension estimation, *Brain Cognit.* **27**, 353 (1995).
  - [6] G. Deco, V. K. Jirsa, and A. R. McIntosh, Emerging concepts for the dynamical organization of resting-state activity in the brain, *Nat. Rev. Neurosci.* **12**, 43 (2011).
  - [7] J. Backus, Can programming be liberated from the von Neumann style? A functional style and its algebra of programs, *Commun. ACM* **21**, 613 (1978).
  - [8] D. Marković, A. Mizrahi, D. Querlioz, and J. Grollier, Physics for neuromorphic computing, *Nat. Rev. Phys.* **2**, 499 (2020).
  - [9] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**, 78 (2004).
  - [10] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, *Neural Comput.* **14**, 2531 (2002).
  - [11] K. Nakajima and I. Fischer, *Reservoir Computing: Theory, Physical Implementations, and Applications* (Springer, Singapore, 2021).
  - [12] S. Ishihara and K. Kaneko, Magic number  $7 \pm 2$  in networks of threshold dynamics, *Phys. Rev. Lett.* **94**, 058102 (2005).
  - [13] A. M. Saxe, J. L. McClelland, and S. Ganguli, Exact solutions to the nonlinear dynamics of learning in deep linear neural networks, in *Proceedings of the International Conference on Learning Representations (ICLR)* (Banff, AB, Canada, 2014).
  - [14] Y. Mototake and T. Ikegami, The dynamics of deep neural networks, in *Proceedings of the Twentieth International Symposium on Artificial Life and Robotics (AROB20)* (Beppu, Oita, Japan, 2015), Vol. 20.
  - [15] S. S. Schoenholz, J. Gilmer, S. Ganguli, and J. Sohl-Dickstein, Deep information propagation, in *Proceedings of the International Conference on Learning Representations (ICLR)* (Toulon, France, 2017).
  - [16] B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli, Exponential expressivity in deep neural networks through transient chaos, *Advances in Neural Information Processing Systems* **29**, 3360 (2016).
  - [17] H. Lin, C. Wang, W. Yao, and Y. Tan, Chaotic dynamics in a neural network with different types of external stimuli, *Commun. Nonlinear Sci. Numer. Simul.* **90**, 105390 (2020).
  - [18] C. Keup, T. Kühn, D. Dahmen, and M. Helias, Transient chaotic dimensionality expansion by recurrent networks, *Phys. Rev. X* **11**, 021064 (2021).
  - [19] K. Inoue, S. Ohara, Y. Kuniyoshi, and K. Nakajima, Transient chaos in bidirectional encoder representations from transformers, *Phys. Rev. Res.* **4**, 013204 (2022).
  - [20] R. Engelken, F. Wolf, and L. F. Abbott, Lyapunov spectra of chaotic recurrent neural networks, *Phys. Rev. Res.* **5**, 043044 (2023).
  - [21] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/l8dq-mphg> for additional literature reviews, analysis of transient dynamics in neural networks, implementation specifics, and supplementary experimental results.
  - [22] M. Kondo, S. Sunada, and T. Niiyama, Lyapunov exponent analysis for multilayer neural networks, *Nonlinear Theory Appl. IEICE* **12**, 674 (2021).
  - [23] L. Storm, H. Linander, J. Bec, K. Gustavsson, and B. Mehlig, Finite-time Lyapunov exponents of deep neural networks, *Phys. Rev. Lett.* **132**, 057301 (2024).
  - [24] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, USA, 2016), pp. 770–778.
  - [25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Miami, FL, USA, 2009), pp. 248–255.
  - [26] A. Krizhevsky and G. Hinton, Learning multiple layers of features from tiny images, Master’s thesis, University of Toronto, Toronto, ON, Canada, 2009.
  - [27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proceedings of the Conference of the North American Chapter of the Association for Computational*

- Linguistics: Human Language Technologies (NAACL-HLT)* (Association for Computational Linguistics, Minneapolis, MN, USA, 2019), pp. 4171–4186.
- [28] I. Turc, M.-W. Chang, K. Lee, and K. Toutanova, Well-read students learn better: On the importance of pre-training compact models, [arXiv:1908.08962v2](https://arxiv.org/abs/1908.08962).
- [29] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, Learning word vectors for sentiment analysis, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)* (Association for Computational Linguistics, Portland, OR, USA, 2011), pp. 142–150.
- [30] M. Cernansky and M. Makula, Feed-forward echo state networks, in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)* (IEEE, Montreal, QC, Canada, 2005), Vol. 3, pp. 1479–1482.
- [31] B. Igel'nik and Y.-H. Pao, Stochastic choice of basis functions in adaptive function approximation and the functional-link net, *IEEE Trans. Neural Networks* **6**, 1320 (1995).
- [32] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks, in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)* (IEEE, Budapest, Hungary, 2004), Vol. 2, pp. 985–990.
- [33] E. Galaris, G. Fabiani, I. Gallos, I. Kevrekidis, and C. Siettos, Numerical bifurcation analysis of PDEs from lattice boltzmann model simulations: A parsimonious machine learning approach, *J. Sci. Comput.* **92**, 34 (2022).
- [34] G. Fabiani, E. Galaris, L. Russo, and C. Siettos, Parsimonious physics-informed random projection neural networks for initial value problems of ODEs and index-1 DAEs, *Chaos* **33**, 043128 (2023).
- [35] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Chaos in random neural networks, *Phys. Rev. Lett.* **61**, 259 (1988).
- [36] D. Verstraeten, B. Schrauwen, M. d'Haene, and D. Stroobandt, An experimental unification of reservoir computing methods, *Neural Networks* **20**, 391 (2007).
- [37] E. N. Lorenz, Predictability: A problem partly solved, in *Predictability of Weather and Climate*, edited by T. Palmer and R. Hagedorn (Cambridge University Press, Cambridge, UK, 2006), pp. 40–58.
- [38] L. S. Pontryagin, *Mathematical Theory of Optimal Processes*, 1st ed. (CRC press, Routledge, London, UK, 1987).
- [39] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, Neural ordinary differential equations, in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., Montreal, QC, Canada, 2018), Vol. 31, pp. 6571–6583.
- [40] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima *et al.*, Neuromorphic computing with nanoscale spintronic oscillators, *Nature (London)* **547**, 428 (2017).
- [41] S. Tsunegi, T. Taniguchi, K. Nakajima, S. Miwa, K. Yakushiji, A. Fukushima, S. Yuasa, and H. Kubota, Physical reservoir computing based on spin torque oscillator with forced synchronization, *Appl. Phys. Lett.* **114**, 164101 (2019).
- [42] J. Grollier, D. Querlioz, K. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, Neuromorphic spintronics, *Nat. Electron.* **3**, 360 (2020).
- [43] Y. Imai, S. Liu, N. Akashi, and K. Nakajima, Gradient-based optimization of spintronic devices, *Appl. Phys. Lett.* **126**, 082402 (2025).
- [44] T. Yamaguchi, N. Akashi, K. Nakajima, S. Tsunegi, H. Kubota, and T. Taniguchi, Synchronization and chaos in a spin-torque oscillator with a perpendicularly magnetized free layer, *Phys. Rev. B* **100**, 224422 (2019).
- [45] N. Akashi, T. Yamaguchi, S. Tsunegi, T. Taniguchi, M. Nishida, R. Sakurai, Y. Wakao, and K. Nakajima, Input-driven bifurcations and information processing capacity in spintronics reservoirs, *Phys. Rev. Res.* **2**, 043303 (2020).
- [46] T. Taniguchi, N. Akashi, H. Notsu, M. Kimura, H. Tsukahara, and K. Nakajima, Chaos in nanomagnet via feedback current, *Phys. Rev. B* **100**, 174425 (2019).
- [47] A. Kamimaki, T. Kubota, S. Tsunegi, K. Nakajima, T. Taniguchi, J. Grollier, V. Cros, K. Yakushiji, A. Fukushima, S. Yuasa *et al.*, Chaos in spin-torque oscillator with feedback circuit, *Phys. Rev. Res.* **3**, 043216 (2021).
- [48] N. Akashi, Y. Kuniyoshi, S. Tsunegi, T. Taniguchi, M. Nishida, R. Sakurai, Y. Wakao, K. Kawashima, and K. Nakajima, A coupled spintronics neuromorphic approach for high-performance reservoir computing, *Adv. Intell. Syst.* **4**, 2200123 (2022).
- [49] N. Leroux, A. D. Riz, D. Sanz-Hernández, D. Marković, A. Mizrahi, and J. Grollier, Convolutional neural networks with radio-frequency spintronic nano-devices, *Neuromorph. Comput. Eng.* **2**, 034002 (2022).
- [50] B. U. Kesgin and U. Teğin, *Photonic Neural Networks with Spatiotemporal Chaos in Multimode Fibers*, Nanophotonics (Walter de Gruyter, Berlin, 2025).
- [51] <https://github.com/ShuhongLL/Exploiting-Chaotic-Dynamics-as-DNNs>.