# Effect of Stochastic Activation Function on Reconstruction Performance of Restricted Boltzmann Machines with Stochastic Magnetic Tunnel Junctions

Minggang Zeng,[1, a] Zitian Li,[2] Jing Wen Saw,[3] and Bingjin Chen[4, b]

[1] *Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore 138632*

[2] *Department of Mathematics, National University of Singapore, Singapore 119076*

[3] *Department of Computer Science, National University of Singapore, Singapore 117417*

[4] *Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore, 138632*

Stochastic Magnetic Tunnel Junctions (SMTJs) emerge as a promising candidate for neuromorphic computing. The inherent stochasticity of SMTJs makes them ideal for implementing stochastic synapses or neurons in neuromorphic computing. However, the stochasticity of SMTJs may impair the performance of neuromorphic systems. In this study, we conduct a systematic examination of the influence of three stochastic effects (shift, change of slope, and broadening) on the sigmoid activation function. We further explore the implications of these effects on the reconstruction performance of Restricted Boltzmann Machines (RBMs). We find that the trainability of RBMs is robust against the three stochastic effects. However, reconstruction error is strongly related to the three stochastic effects in SMTJs-based RBMs. Significant reconstruction error is found when the stochastic effect is strong. Last, we identify the correlation of the reconstruction error with each stochastic factor. Our results might help develop more robust neuromorphic systems based on SMTJs.

[a] Authors to whom correspondence should be addressed; zengmg@i2r.a-star.edu.sg

[b] Authors to whom correspondence should be addressed; chen_bingjin@ihpc.a-star.edu.sg

Neuromorphic computing that mimics the human brain's structure and function might be a potential revolution in computing paradigms.[1–5] Such systems promise excellent energy efficiency with real-time processing capabilities and advanced cognitive abilities, surpassing conventional digital computing.[6–12] Consequently, the development of specialized hardware and algorithms for neuromorphic computing has gained significant momentum, with Stochastic Magnetic Tunnel Junctions (SMTJs) emerging as a particularly promising candidate. SMTJs, composed of two ferromagnetic layers separated by an insulating layer, display stochastic resistance states owing to the rapid switching of the free layer,[13–16] which makes them ideal for implementing stochastic synapses or neurons in neuromorphic computing.[6–12] When compared to other forms of artificial synapses or neurons, SMTJs excel in terms of speed,[15] power consumption,[11,17] and scalability.[18] Their implementation in various neuromorphic computing systems, including spiking neural networks and Boltzmann machines,[19–21] has demonstrated efficacy in tasks such as invertible logic,[6–9] image classification,[12] accelerating Monte Carlo simulation,[11] and solving Ising models.[10]

A distinctive advantage of SMTJs lies in their suitability for probability programming, a powerful framework for reasoning about uncertainty and making predictions based on incomplete or noisy data.[6,9–11] Traditionally, Boltzmann machines, a genre of generative probabilistic models, exploit deterministic activation functions, such as sigmoid or softmax, to compute the activations of hidden units.[22,23] Substituting these deterministic activations with SMTJs imbues the neurons with inherent stochasticity, aligning with the probabilistic nature of Boltzmann machines. Such stochastic advantages include reduced circuit complexity and power consumption, making it apt for specific computational tasks like probabilistic neural networks and Monte Carlo simulations.[10,11] Stochastic behaviors in SMTJs stem from various sources. For instance, SMTJs inherently exhibit stochasticity in their switching dynamics. Moreover, manufacturing processes and material imperfections introduce device-to-device variations in their electrical properties. Thermal variation further influence SMTJ device's switching behavior, impacting its switching probability, critical current, and the energy barrier for magnetization reversal. Finally, the length of sampling window can modulate the observed stochastic results.

While the inherent variability and randomness in the switching behavior of SMTJs enable probabilistic computing, they may also impair the accuracy and reliability of SMTJ-based circuits. The variability in critical parameters can introduce performance fluctuations in neural network models, potentially compromising model accuracy and learning capability. This paper delves into three random effects of the sigmoid activation function of SMTJ neurons, particularly focusing on MNIST reconstruction using Restricted Boltzmann Machines (RBMs).[24] RBMs are easy to train, extensively explored in machine learning, and play a pivotal role in the resurgence of neural networks in the 2000s.[23,25] Several groups have explored the application of SMTJs for RBMs, demonstrating proof-of-concept analogous building blocks,[26] and suggesting SMTJs as intrinsic stochastic neurons in RBMs.[20,21] However, there is a conspicuous gap in the investigation on reconstruction performance of RBMs when randomness effects of SMTJ stochasticity are taken into account. Reconstruction captures the underlying structure of dataset. The reconstruction performance is key for tasks such as data generation, anomaly detection, and denoising. In this paper, we systematically investigate the impact of three stochastic effects (shift, change of slope and broadening) on the sigmoid activation function and its implications on the reconstruction performance of RBMs. We find that even with different stochastic sources and at different scales, RBMs with SMTJs can still be effectively trained using the Contrastive Divergence (CD) algorithm.[23,25] However, the significant reconstruction error on the test samples suggests that the stochasticity of SMTJs can impair the reconstruction performance of RBMs. We further identify the correlation between reconstruction error and the three stochastic effects. The results of this study provide valuable insights on developing more robust neuromorphic systems for reconstruction applications with SMTJs.

As shown in Fig. 1a, SMTJs have two magnetic layers, a free layer and a fixed layer, separated by a thin insulating layer. The switching of free layer results in the variation of tunnel magnetic resistance. In SMTJs, random thermal fluctuations can switch the free layer due to their extremely low thermal barrier. The stochastic behavior of SMTJs can be described mathematically by a sigmoid function, which is a type of S-shaped curve describing the probability of the free layer in its parallel (+1) or antiparallel (-1) states with respect to the fixed layer. With an external current biases ($I_i$), the magnetization of SMTJ ($m_i$), as a function of $I_i$, can be described mathematically by

$$m_i = sign[tanh(I_i) - r]$$

, where r is a random number uniformly distributed between -1 and +1. In order to cascade the SMTJ with other devices to form a neural network, we need to convert the current to voltage. Zanh et al. proposed a building block (Fig. 1b) by putting the SMTJ in series with a CMOS transistor, and the output is converted by an additional inverter.[17] Interestingly, the output voltage can be approximately described by an expression similar to SMTJs,

$$\frac{\overbrace{V_{out,i}}^{m_i}}{V_{DD}/2} \approx sign\{tanh \frac{\overbrace{V_{in,i}}^{I_i}}{V_0} - r\}. \tag{1}$$

Here, $V_{in,i}$ is the input voltage of the SMTJ, $V_0$ is the scaling constant,[6] $V_{DD}$ is the supply voltage, and $V_{out,i}$ is the output voltage of the SMTJ. The sigmoid behavior of SMTJs is a key feature that makes them attractive for neuromorphic computing, where the sigmoid function is commonly used as an activation function and represents the time-averaged resistance state of the SMTJ. Variations in the manufacturing process and device characteristics can affect the shape and position of this sigmoid curve, which in turn can affect the behavior of the neuromorphic system, such as RBMs.

The architecture of RBMs (Fig. 1c) includes a visible layer and a hidden layer, both of which are made up of binary units connected by weighted edges ($W_{ij}$). The visible and hidden units can be represented by binary vectors $\mathbf{v} \in \{\mathbf{0}, \mathbf{1}\}^{\mathbf{m}}$ and $\mathbf{h} \in \{\mathbf{0}, \mathbf{1}\}^{\mathbf{n}}$, respectively. Physically, $\mathbf{v}$ and $\mathbf{h}$ correspond to the magnetization state of the SMTJ, or the binary voltage output, as discussed later. To compute the probability of a configuration $(\mathbf{v}, \mathbf{h})$, RBMs use the Boltzmann distribution, with an energy function given by:

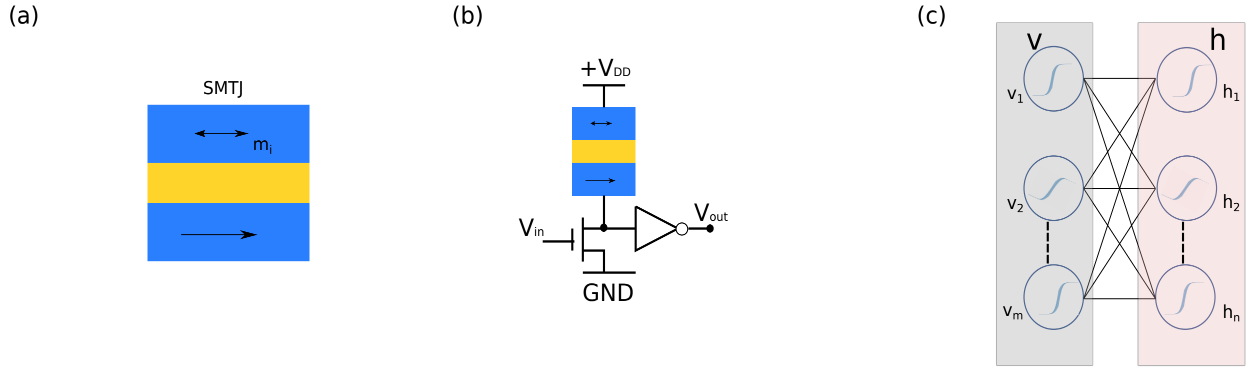
$$E(\mathbf{v}, \mathbf{h}) = -\sum_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{a_i v_i} - \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{b_j h_j} - \sum_{\mathbf{i=1}}^{\mathbf{m}} \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{v_i h_j W_{ij}}$$

Here, $\mathbf{a}$ and $\mathbf{b}$ represent the biases of visible and hidden units, respectively. The probability of a configuration $(\mathbf{v}, \mathbf{h})$ is given by $P(\mathbf{v}, \mathbf{h}) = \frac{\mathbf{e^{-\beta E(v,h)}}}{\sum_{\mathbf{(v,h)}} \mathbf{e^{-\beta E(v,h)}}}$, where the denominator sums over all possible configurations involving the visible and hidden units, and $\beta$ is the inverse temperature. The probabilities of these units are updated using the Gibbs sampling:

$$P(\mathbf{h_j} = \mathbf{1}|\mathbf{v}) = \sigma(\sum_{\mathbf{i=1}}^{\mathbf{m}} (\mathbf{W_{ij} v_i} + \mathbf{b_j}))$$

$$P(\mathbf{v_i} = \mathbf{1}|\mathbf{h}) = \sigma(\sum_{\mathbf{j=1}}^{\mathbf{n}} (\mathbf{W_{ij} h_j} + \mathbf{a_i}))$$

Gibbs sampling involves iteratively sampling the states of hidden and visible units based on their conditional probabilities, allowing the RBM to learn the underlying patterns in




$$m_i = \mathrm{sgn}[\tanh(I_i) - r]$$

$$\overbrace{\frac{V_{out,i}}{V_{DD/2}}}^{m_i} \approx \mathrm{sgn}[\tanh \overbrace{\frac{V_{in,i}}{V_0}}^{I_i} - r]$$

$$E(v,h) = -\sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m} \sum_{j=1}^{n} v_i h_j W_{ij}$$

FIG. 1. (a) Schematic of SMTJs. The relationship between the magnetization of SMTJs ($m_i$) and the input current $I_i$ exhibits a sigmoid behavior. (b) Schematic of a SMTJ neuron circuit block. The output voltage $V_{out,i}$ as a function of $V_{in,i}$ also displays a sigmoid behavior. (c) RBM architecture with SMTJ neurons in the visible layer and the hidden layer. The shift, change of slope and broadening effect of the Sigmoid curve in each neuron are highlighted. We assume that all the SMTJs in the network have different shift, slope, and broadening

. $E_{v,h}$ is the energy model of RBM.

data. The sigmoid curve, commonly represented by the logistic function, is often used as the activation function in RBMs. The sigmoid function maps the input values to a range between 0 and 1, representing probabilities. The relationship between the Gibbs sampling of RBMs and the sigmoid curve of SMTJs is significant. In traditional RBMs, deterministic activation functions like the sigmoid function are used to compute the activations of the hidden layer neurons. However, by replacing these deterministic activations with SMTJs, the neurons can exhibit inherent stochasticity, making them suitable for modeling probabilistic computations in RBMs.

To study the effect of stochasticity of SMTJs, we vary three random factors in the sigmoid curve, including shift ($\Delta$), change of slope ($\alpha$), and broadening ($\sigma$). First, shift ($\Delta$ at Fig. 2a), a horizontal movement of the sigmoid curve, could be caused by variations in the device's threshold voltage, which might occur due to factors like variations in the thickness of the

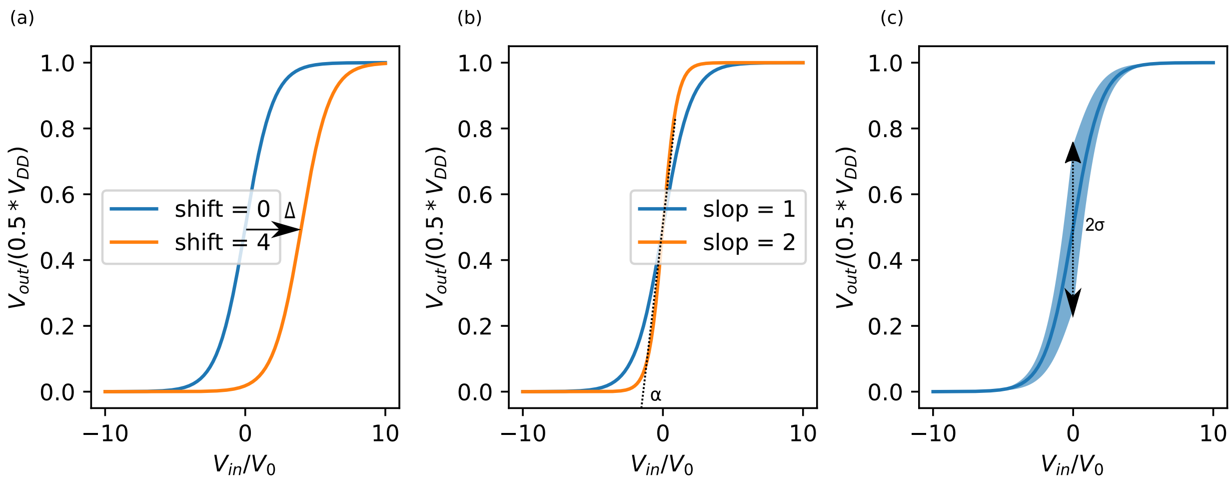



FIG. 2. (a) Shift ($\Delta$) of the sigmoid curve. (b) Change of slope ($\alpha$) on the sigmoid curve. (c) Broadening effect ($\sigma$) of the sigmoid curve.

insulating layer in the MTJ or the offset field from the anti-ferromagnetic pinning layer. A shift in the sigmoid curve means that different voltages would be required to achieve the same probabilities. Second, change of slope ($\alpha$ at Fig. 2b) is reflected by the steepness of the sigmoid curve. This could be caused by variations in the thermal stability factor of the device. A change in slope affects the sensitivity of the SMTJ to changes in voltage.[27] Last, broadening effect ($\sigma$ at Fig. 2c) refers to a widening of the sigmoid curve. This could be caused by variations in the thermal temperature in the device, or the length of sampling windows. Broadening of the sigmoid curve means the transition between the two states becomes less sharp, which could lead to increased uncertainty and noise in the neuromorphic system. In practice, broadening is an ensemble average (mean and standard deviation) of different activation curves caused by variations in devices, circuits and environment.

In a neuromorphic computing system, these effects, arising from the variations in the behavior of SMTJs, could in turn affect the overall system. For example, if different SMTJs in the system have different sigmoid curves due to process variations, this might affect the accuracy of computations and the ability of the system to learn and adapt. Therefore, understanding and managing these effects is key for neuromorphic systems based on SMTJs.

In this study, we explore the effects of different sigmoid characteristics by varying the broadness of the sigmoid function, the shift along the voltage axis, and the slope at the midpoint. Each of these parameters is given at three different scales, yielding a total of 27 scale levels. We set the broadness scale of the sigmoid function as 0.2, 0.4, and 0.6, and a smaller (larger) sigmoid value means the device switches more deterministically (stochastically). By varying the shift scale as 2, 4, and 6, we explore the effects of different threshold voltages for the SMTJs. A larger shift means a higher voltage is required for the device to switch states. By varying the slope scale as 0.5, 1.0, and 2.0, we examine the effects of different sensitivities to changes in voltage. A steeper slope (larger value) means the device is more sensitive to voltage changes. We assume that all the SMTJs in the network have different shift, slope, and broadening. The actual probability of the sigmoid function at each SMTJ neuron is sampled from the broadened sigmoid curve, $[-\sigma * (0.5 - |p - 0.5|), \sigma * (0.5 - |p - 0.5|)]$, where p is the probability of the sigmoid function $\frac{1}{1+e^{-\alpha*(x-\Delta)}}$. Since the sigmoid curve is affected by the scaling voltage, $V_0$ and $V_{DD}$, our sigmoid curves may look different with those in other literature.[17] Moreover, the development of SMTJs is still in its early stage, without much information on the characteristics of experimentally observed sigmoid curves, our choice of simulation range for slope, shift, and broadening may not be realistic, but it is useful to demonstrate the trend or possible worst case study.

TABLE I. Three random factors in the sigmoid curve with three scales and the corresponding sampling range for each factor.

| Effect | scale | sampling range |
|---|---|---|
| shift | 2, 4, 6 | $\Delta \sim [-shift, +shift]$ |
| slope | 0.5, 1, 2 | $\alpha \sim [0, slope]$ |
| Broadening ($\sigma$) | 0.2, 0.4, 0.6 | $p \sim [-\sigma * (0.5 - \|p - 0.5\|), \sigma * (0.5 - \|p - 0.5\|)]$ |

We first investigate how these variations in the sigmoid characteristics of SMTJ neurons affect the trainability (Fig. 3). Each SMTJ's behavior is modeled with a sigmoid function characterized by varying broadness, shift, and slope parameters. The RBM is trained using the Contrastive Divergence (CD) method with a Gibbs sampling step of k=1 (CD-1), a technique known for its efficiency in training RBMs. We find that, across all configurations of the sigmoid parameters, the loss curve of mean squared error (MSE) at Figs. 3(a∼c) during training consistently converges to a stable value. This result indicates robust trainability of the RBM under a wide range of SMTJ characteristics, demonstrating the effectiveness of the CD-1 method. Importantly, this stability in training occurs despite the variations in the sigmoid characteristics of the SMTJs, which may introduce significant variability in

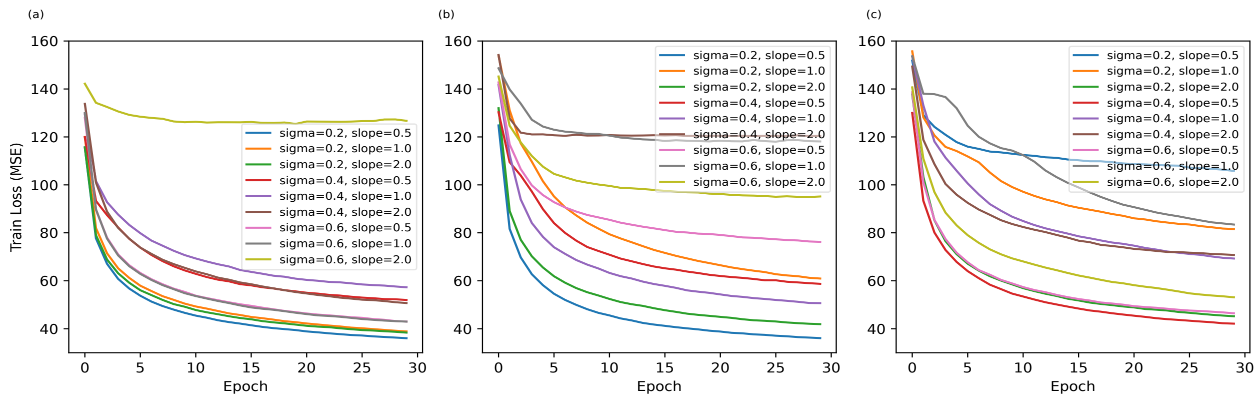



FIG. 3. Training loss (MSE) for the SMTJ-based RBM with different scales of ($slope$, $\sigma$) and (a) $shift = 2$ (b)$shift = 4$ and (c)$shift = 6$. Stable trainability is found in all cases, which indicates the effectiveness of CD-k method.

the behavior of the neuromorphic system. This evidence of stable training across different device and process variations is a promising finding. It suggests that RBMs can be effectively trained in neuromorphic systems based on SMTJs, even in the presence of device and process variations that are likely to be encountered in practical manufacturing scenarios.

Following the training of the RBM using SMTJs neuromorphic system, we evaluate its performance on the reconstruction of the MNIST test set. This step is critical to understanding how the trained model generalizes to unseen data, and particularly, how the intrinsic stochasticity of the SMTJs impacts the model's predictive ability. In this evaluation, we keep the model parameters fixed and observe that the stochastic properties of the SMTJs lead to a degree of uncertainty in the model's predictions. We provide the detailed reconstruction MSE of RBM with different scales of ($shift$, $slope$, $\sigma$) in Table II. The mean and standard variation of each case are obtained by the statistics in an ensemble of 20 calculations. Interestingly, a correlation between reconstruction mean squared error (MSE) and prediction uncertainty is found. The smaller the reconstruction MSE, the smaller the model's prediction uncertainty. This result provides a valuable insight into the relationship between prediction accuracy and uncertainty in neuromorphic systems based on SMTJs. For instance, targeting lower reconstruction MSE can reduce prediction uncertainty.

Next, we perform a detailed analysis to understand the individual effects of the three factors, shift, change of slope, and broadening ($\sigma$), on the reconstruction of the MNIST test set. Figure 4 shows the violin plot of reconstruction MSE for different scales of (a) $\sigma$, (b)

TABLE II. Reconstruction MSE of RBM with different scales of ($shift$, $slope$, $\sigma$). The mean and standard variation of each case are obtained by the statistics in an ensemble of 20 calculations.

| sigma | shift | slope | reconstruction error (mse) | reconstruction error (std) |
|---|---|---|---|---|
| 0.20 | 2 | 0.50 | 76.78 | 2.51 |
| 0.20 | 2 | 1.00 | 91.68 | 4.07 |
| 0.20 | 2 | 2.00 | 89.27 | 4.02 |
| 0.20 | 4 | 0.50 | 142.01 | 5.83 |
| 0.20 | 4 | 1.00 | 172.17 | 6.38 |
| 0.20 | 4 | 2.00 | 147.94 | 7.31 |
| 0.20 | 6 | 0.50 | 217.74 | 7.32 |
| 0.20 | 6 | 1.00 | 220.00 | 6.95 |
| 0.20 | 6 | 2.00 | 189.14 | 6.82 |
| 0.40 | 2 | 0.50 | 89.85 | 2.40 |
| 0.40 | 2 | 1.00 | 112.46 | 2.77 |
| 0.40 | 2 | 2.00 | 104.09 | 4.07 |
| 0.40 | 4 | 0.50 | 151.00 | 5.17 |
| 0.40 | 4 | 1.00 | 156.74 | 6.92 |
| 0.40 | 4 | 2.00 | 202.87 | 8.11 |
| 0.40 | 6 | 0.50 | 188.21 | 5.50 |
| 0.40 | 6 | 1.00 | 216.56 | 6.98 |
| 0.40 | 6 | 2.00 | 216.30 | 7.66 |
| 0.60 | 2 | 0.50 | 82.94 | 3.50 |
| 0.60 | 2 | 1.00 | 96.79 | 4.38 |
| 0.60 | 2 | 2.00 | 187.07 | 5.24 |
| 0.60 | 4 | 0.50 | 173.60 | 5.03 |
| 0.60 | 4 | 1.00 | 193.90 | 5.38 |
| 0.60 | 4 | 2.00 | 193.23 | 6.41 |
| 0.60 | 6 | 0.50 | 194.53 | 7.52 |
| 0.60 | 6 | 1.00 | 223.45 | 8.03 |
| 0.60 | 6 | 2.00 | 202.05 | 8.07 |

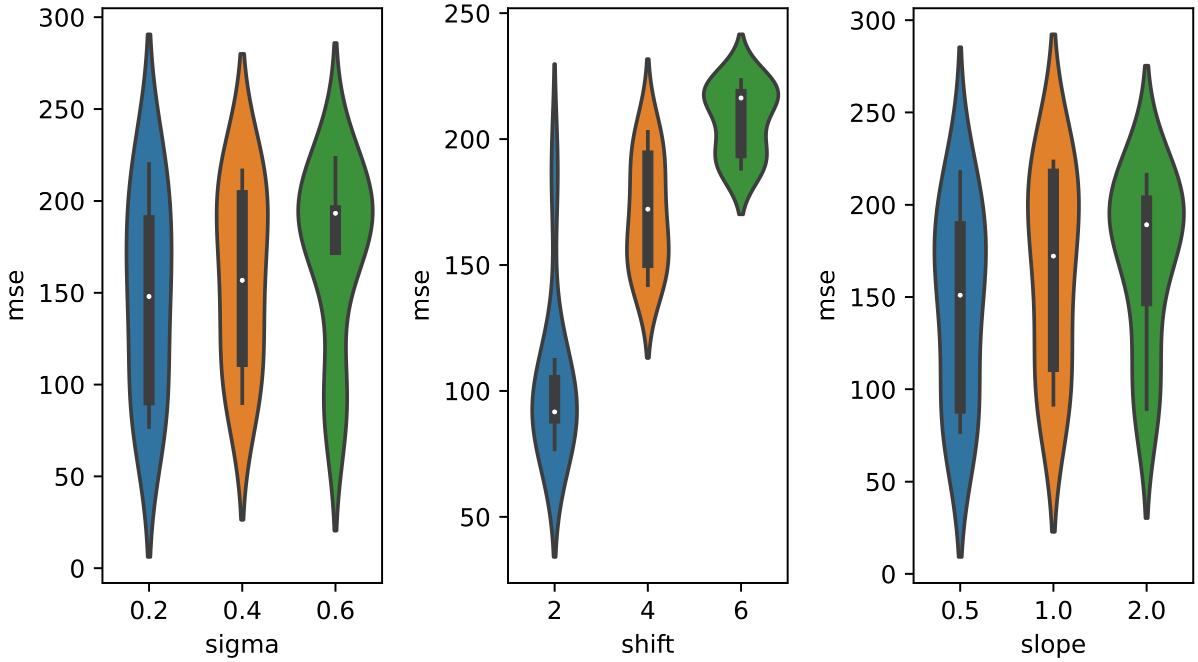



FIG. 4. Violin plot of reconstruction MSE for different scales of (a) $\sigma$, (b) $shift$, (c) *slope*.

$shift$, and (c) *slope*. The violin plot is a combination of a box plot and a kernel density plot, offering information on the distribution of data, including median (the white dot), interquartile range (the thick bar) and 95% confidence interval (the thin bar), as well as the probability density of data (the outskirt of violin plot). Our results in Fig. 4 show that among the three factors, the shift along voltage axis has the most significant positive correlation with reconstruction MSE, indicating that variations in the threshold voltage of the SMTJs has the greatest impact on the model's performance. With $shift = 2$, almost all cases have a reconstruction MSE below 150. However, with $shift = 4$ and $shift = 6$, the reconstruction MSE increases significantly. The majority cases of $shift = 4$ have a reconstruction MSE above 150, and all cases in $shift = 6$ above 150. This result suggests that the threshold voltage of the SMTJs is a critical factor that affects the reconstruction performance of the RBM. Meanwhile, broadening ($\sigma$) and change of slope have comparatively less influence on the reconstruction MSE. Furthermore, we find that smaller values of sigma, shift, and slope result in better reconstruction performance, as indicated by a smaller MSE. This suggests that more deterministic switching behavior (smaller $\sigma$), lower threshold voltages (smaller $shift$), and less sensitivity to voltage changes (smaller *slope*) are beneficial for

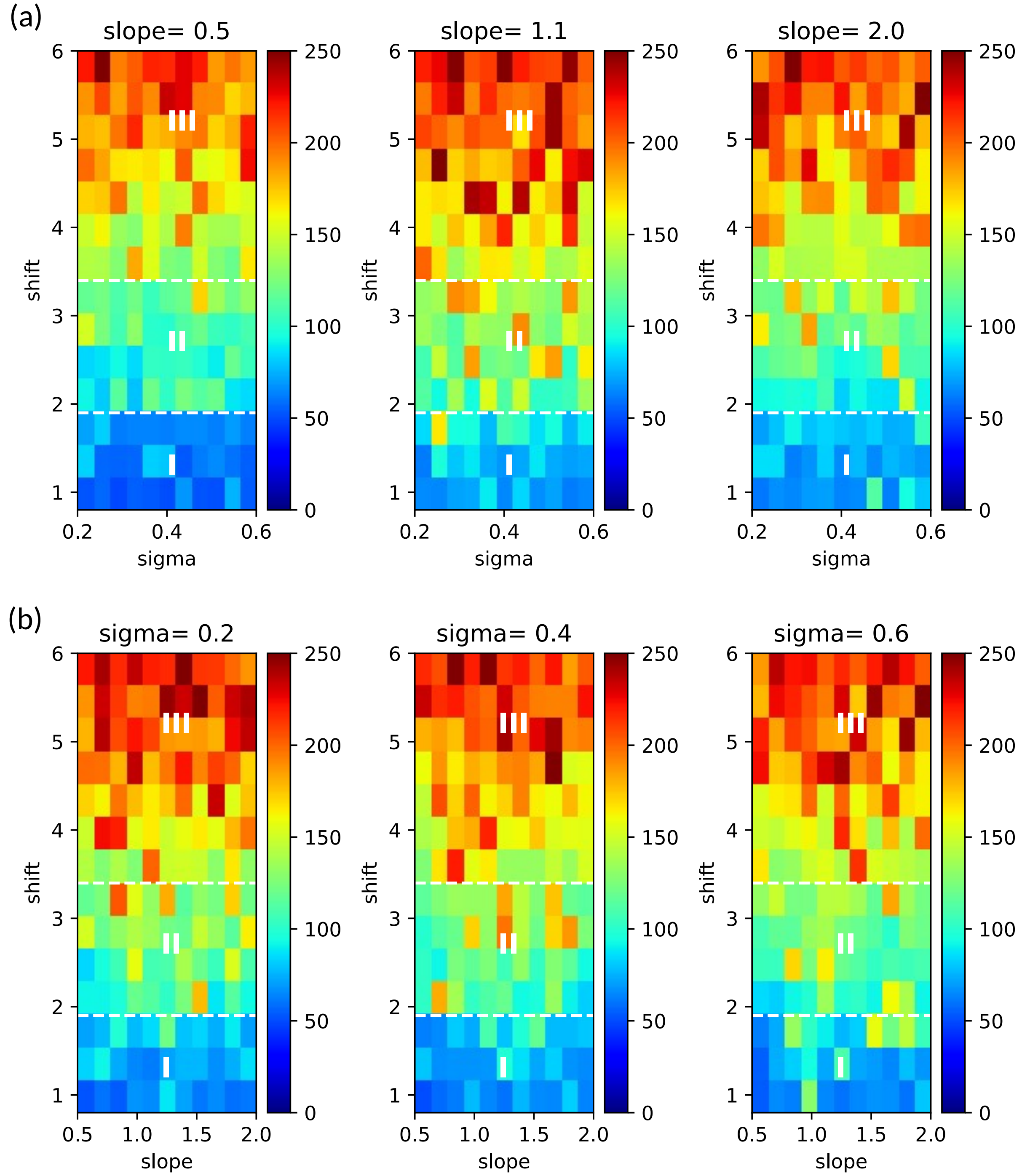



FIG. 5. (a) Reconstruction of RBM at different (sigma, shift) configurations and three different slope values. (b) Reconstruction of RBM at different (slope, shift) configurations and three different sigma values.

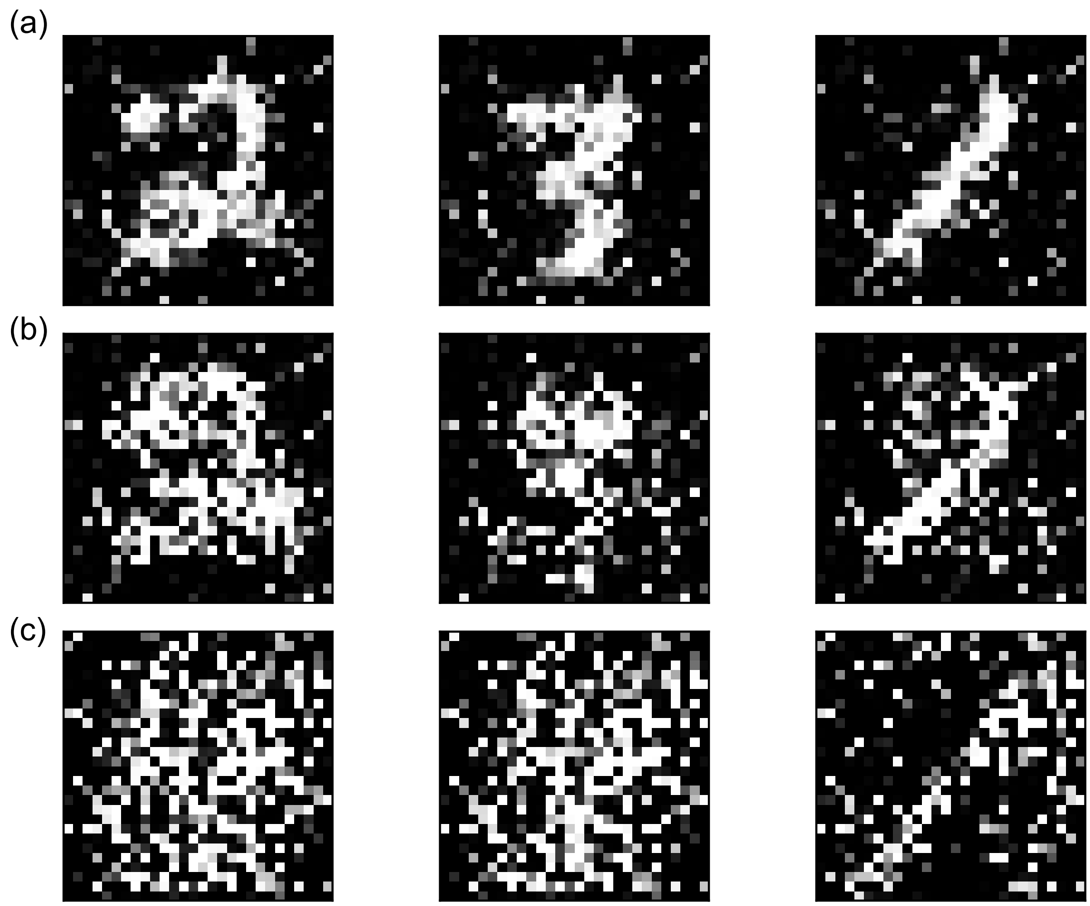



FIG. 6. Reconstruction of RBM on three test samples. (a) Low reconstruction MSE (MSE = 76, $\sigma = 0.2, shift = 2, slope = 0.5$), (b) medium reconstruction MSE (MSE = 142, $\sigma = 0.2, shift = 4, slope = 0.5$), (c)high reconstruction MSE (MSE = 712, $\sigma = 0.2, shift = 6, slope = 0.5$).

the performance of the SMTJ-based neuromorphic system. It should be noted that the three stochastic effects might be correlated with each other. For example, an increase in threshold voltage may affect thermal stability, resulting in a shift and an increase in the slope of the activation function. The correlation between the three schochastic effects complicates the accurate control of reconstruction MSE. For an example, use the case of $(\sigma, shift, slope) = (0.6, 6, 2)$ as a baseline, reducing $\sigma$ to 0.4 actually raises the MSE from 202 to 216, reducing the *slope* from 2 to 1 increases the MSE from 202 to 223, and reducing the $shift$ from 6 to 4 only lowers the MSE from 202 to 193. As a result, the $(\sigma, shift, slope)$ values may not monotonically tune the reconstruction MSE. Two effects could play a role in this. The first one is the intrinsic stochasticity in SMTJ-based RBMs. The second one is the correlation between $shift$, $\sigma$, and $slope$. Therefore, it is difficult to draw a clear point comparison on the reconstruction MSE between different $(\sigma, shift, slope)$ values. However, a general trend still can be provided for the reconstruction MSE. To clarify this, we investigated the reconstruction MSE for different values of sigma, shift, and slope, with (11, 14, 11) levels that evenly span over (0.2, 0.6), (0.8, 6.0), (0.5, 2.0), respectively. Since Tab. II indicates that the standard variation of reconstruction MSE is much smaller compared to their mean values, so we adopt a single-shot calculation in these total 1694 case studies. The results support our previous conclusion that the shift along the voltage axis has the most significant positive correlation with reconstruction MSE. As shown in Fig. 5(a), we plot the reconstruction results of RBM with different scales of ($shift$, $\sigma$) on three slopes. We divide the parameter space into three regions, from I, II to III, with low, medium and high reconstruction MSE, respectively. The pattern of reconstruction MSE is not affected by the slope value. A large shift ($\geq 3.4$) lead to a high reconstruction MSE, at region III. comparatively, shift in the range of 1.9 and 3.4 results in a medium reconstruction MSE, at region II, while a smaller shift ($\leq 1.9$) is necessary to achieve a low reconstruction MSE, as indicated by the region I. A similar pattern is found in Fig. 5(b), where we plot the reconstruction results of RBM with different scales of ($shift$, $slope$) on three $\sigma$ values. These results suggest that the shift along voltage axis has the most significant positive correlation with reconstruction MSE. The stochastic effect in each region reflects the feature of brain-like neuromorphic computing. Although it may pose challenges in accurate control of the reconstruction performance, it can benefit better generalization and improving fault tolerance. Our findings can give experimentalists direction on the main factors and trends in the design of neuromorphic systems based on SMTJs. It is possible to mitigate the shift effect in an electronic circuit by altering bias voltages or tuning resistors. Therefore, we anticipate that limiting the shift effect will help SMTJ-based neuromorphic systems perform well during reconstruction in experiments.

Last, as we identify the $shift$ of sigmoid curve plays the most essential role on the reconstruction error, we plot the reconstruction results with different $shift$ scales at Fig. 6. These three cases have low (MSE = 76), medium (MSE = 142) and large (MSE = 712) reconstruction error respectively. The result shows that RBM with stochastic sigmoid functions can only reconstruct the input image correctly with the low MSE (Fig. 6a) and fails to make a correct reconstruction with the medium and large MSE (Figs. 6(b, c)). Our results suggest that the stochasticity of the SMTJs play an essential role in RBM reconstruction performance, which should be considered in the design and optimization of SMTJ-based neuromorphic systems.

To summarize, our results provide valuable insights for the design and optimization of neuromorphic systems based on SMTJs. Minimizing the threshold voltage shift and controlling the broadness and slope of the sigmoid function can potentially lead to improvements in their reconstruction performance. This understanding could guide the development of manufacturing processes and device designs that aim to minimize these variations in SMTJ characteristics, leading to more efficient and reliable neuromorphic systems.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors upon request.

## ACKNOWLEDGMENTS

We thank Hlaing Min for his help in code review and refactoring. This work is supported by Agency for Science, Technology and Research (A*STAR) under Career Development Fund (Project No. C210812054).

## REFERENCES


[1]G. Bourianoff, D. Pinna, M. Sitte, and K. Everschor-Sitte, "Potential implementation of reservoir computing models based on magnetic skyrmions," AIP Advances **8**, 055602 (2018).

[2]J. Grollier, D. Querlioz, K. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, "Neuromorphic spintronics," Nature electronics **3**, 360 (2020).

[3]M. Musisi-Nkambwe, S. Afshari, H. Barnaby, M. Kozicki, and I. S. Esqueda, "The viability of analog-based accelerators for neuromorphic computing: a survey," Neuromorphic Computing and Engineering **1**, 012001 (2021).

[4]Y. Zhu, Y. Zhu, H. Mao, Y. He, S. Jiang, L. Zhu, C. Chen, C. Wan, and Q. Wan, "Recent advances in emerging neuromorphic computing and perception devices," Journal of Physics D: Applied Physics **55**, 053002 (2021).

[5]D. V. Christensen, R. Dittmann, B. Linares-Barranco, A. Sebastian, M. L. Gallo, A. Redaelli, S. Slesazeck, T. Mikolajick, S. Spiga, S. Menzel, I. Valov, G. Milano, C. Ricciardi, S.-J. Liang, F. Miao, M. Lanza, T. J. Quill, S. T. Keene, A. Salleo, J. Grollier, D. Markovic, A. Mizrahi, P. Yao, J. J. Yang, G. Indiveri, J. P. Strachan, S. Datta, E. Vianello, A. Valentian, J. Feldmann, X. Li, W. H. Pernice, H. Bhaskaran, S. Furber, E. Neftci, F. Scherr, W. Maass, S. Ramaswamy, J. Tapson, P. Panda, Y. Kim, G. Tanaka, S. Thorpe, C. Bartolozzi, T. A. Cleland, C. Posch, S.-C. Liu, G. Panuccio, M. Mahmud, A. N. Mazumder, M. Hosseini, T. Mohsenin, E. Donati, S. Tolu, R. Galeazzi, M. E. Christensen, S. Holm, D. Ielmini, and N. Pryds, "2022 roadmap on neuromorphic computing and engineering," Neuromorphic Computing and Engineering , 022501 (2022).

[6]K. Y. Camsari, S. Salahuddin, and S. Datta, "Implementing p-bits With Embedded MTJ," IEEE Electron Device Letters **38**, 1767 (2017).

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

[8]R. Faria, K. Y. Camsari, and S. Datta, "Low-Barrier Nanomagnets as p-Bits for Spin Logic," IEEE Magnetics Letters **8**, 1 (2017).

[9]K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, "Stochastic p-Bits for Invertible Logic," Physical Review X **7**, 031014 (2017).

[10] J. Kaiser and S. Datta, "Probabilistic computing with p-bits," Applied Physics Letters **119**, 150503 (2021).

[11] K. Kobayashi, N. Singh, Q. Cao, K. Selcuk, T. Hu, S. Niazi, N. A. Aadit, S. Kanai, H. Ohno, S. Fukami, and K. Y. Camsari, "Cmos + stochastic nanomagnets: heterogeneous computers for probabilistic inference and learning," (2023), arXiv:2304.05949 [cond-mat.mes-hall].

[12] S. Niazi, N. A. Aadit, M. Mohseni, S. Chowdhury, Y. Qin, and K. Y. Camsari, "Training deep boltzmann networks with sparse ising machines," (2023), arXiv:2303.10728 [cs.ET].

[13] K. Y. Camsari, M. M. Torunbalci, W. A. Borders, H. Ohno, and S. Fukami, "Double-Free-Layer Magnetic Tunnel Junctions for Probabilistic Bits," Physical Review Applied **15**, 044049 (2021).

[14] K. Kobayashi, W. A. Borders, S. Kanai, K. Hayakawa, H. Ohno, and S. Fukami, "Sigmoidal curves of stochastic magnetic tunnel junctions with perpendicular easy axis," Applied Physics Letters **119**, 132406 (2021).

[15] 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," Physical Review Letters **126**, 117202 (2021).

[16] S. Kanai, K. Hayakawa, H. Ohno, and S. Fukami, "Theory of relaxation time of stochastic nanomagnets," Physical Review B **103**, 094423 (2021).

[17] R. Zand, K. Y. Camsari, S. Datta, and R. F. Demara, "Composable Probabilistic Inference Networks Using MRAM-based Stochastic Neurons," ACM Journal on Emerging Technologies in Computing Systems **15** (2019).

[18] J. Kaiser, W. A. Borders, K. Y. Camsari, S. Fukami, H. Ohno, and S. Datta, "Hardware-aware-in-situ learning based on stochastic magnetic tunnel junctions," Physical Review Applied **17**, 014016 (2022).

[19] G. Srinivasan, A. Sengupta, and K. Roy, "Magnetic tunnel junction based long-term short-term stochastic synapse for a spiking neural network with on-chip STDP learning," Scientific Reports **6**, 29545 (2016).

[20] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, "Intrinsic optimization using stochastic nanomagnets," Scientific Reports **7**, 44370 (2017).

[21] J. Deng, V. P. K. Miriyala, Z. Zhu, X. Fong, and G. Liang, "Voltage-controlled spintronic stochastic neuron for restricted boltzmann machine with weight sparsity," IEEE Electron

Device Letters **41**, 1102 (2020).

[22] M. Roder, G. H. de Rosa, and J. P. Papa, "Learnergy: Energy-based machine learners," arXiv , 07443 (2020), arXiv:2003.07443 [cs.LG].

[23] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation **18**, 1527 (2006).

[24] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE **86**, 2278–2324 (1998).

[25] G. E. Hinton, "A practical guide to training restricted boltzmann machines," in *Neural networks: Tricks of the trade* (Springer, 2012) p. 599.

[26] B. Behin-Aein, V. Diep, and S. Datta, "A building block for hardware belief networks," Scientific Reports **6**, 29893 (2016).

[27] K. Kobayashi, W. A. Borders, S. Kanai, K. Hayakawa, H. Ohno, and S. Fukami, "Sigmoidal curves of stochastic magnetic tunnel junctions with perpendicular easy axis," Applied Physics Letters **119** (2021).