SCIENCE CHINA
Information Sciences

• LETTER •

December 2026, Vol. 69, Iss. 12, 229406:1–229406:2
https://doi.org/10.1007/s11432-025-4975-9

# Robust forward-forward training of physical neural networks enabled by 2D materials

Xueyuan ZHANG[1†], Pengfei WANG[2†], Yichen ZHAO[2], Chen PAN[1*], Yinzhi HUANG[2], Yuekun YANG[3], Cong WANG[3], Bin CHENG[1,2], Wentao YU[1] & Shi-Jun LIANG[2,4*]

[1]*Institute of Interdisciplinary Physical Sciences, School of Physics, Nanjing University of Science and Technology, Nanjing 210094, China*
[2]*Institute of Brain-Inspired Intelligence, National Laboratory of Solid State Microstructures, Collaborative Innovation Center of Advanced Microstructures, Jiangsu Physical Science Research Center, School of Physics, Nanjing University, Nanjing 210093, China*
[3]*State Key Laboratory for Novel Software Technology, School of Intelligence Science and Technology, Nanjing University, Suzhou 215163, China*
[4]*Chemistry and Biomedicine Innovation Center (ChemBIC), Nanjing University, Nanjing 210093, China*

Received 18 December 2025/Revised 31 January 2026/Accepted 15 May 2026/Published online 27 October 2026

**Citation** Zhang X Y, Wang P F, Zhao Y C, et al. Robust forward-forward training of physical neural networks enabled by 2D materials. Sci China Inf Sci, 2026, 69(12): 229406, https://doi.org/10.1007/s11432-025-4975-9

Neuromorphic hardware provides a promising route toward low-power, parallel information processing by mapping neural network operations directly onto physical systems [1–3]. Two-dimensional (2D) material-based electronic devices, which exhibit tunable nonlinear responses, are particularly attractive for such implementations [4]. However, their intrinsic stochastic fluctuations present challenges for conventional backpropagation (BP) training algorithms, which depend on accurate and idealized device models [1]. In contrast, the forward-forward (FF) algorithm performs forward inference directly on hardware and updates weights locally based on the resulting outputs [3,5]. By leveraging the intrinsic nonlinear responses of the devices without relying on idealized models, FF offers inherent robustness to device nonidealities and is well suited for training physical neural networks. Here, we demonstrate a nonlinear physical neural network implemented on $WSe_2$ floating-gate arrays and trained using a hardware-compatible FF algorithm. Under hardware-induced nonuniformity, the FF-trained system maintains accuracy close to 80%, whereas backpropagation collapses to below 50%. These results establish FF as a robust, hardware-compatible training strategy for nonlinear device arrays, enabling efficient neuromorphic computing with emerging 2D devices.

*Results and discussion.* The physical deep neural network (PDNN) is implemented using floating-gate transistor arrays based on the ambipolar 2D semiconductor $WSe_2$, where synaptic weighting and neuronal nonlinear computation are directly realized through the device physics. As shown in Figure 1(a), the intrinsic nonlinear response of each neuron-neuron connection implements both activation and weighting, while subsequent network-level summation integrates these outputs to realize the complete PDNN computation. Leveraging this principle, a single-layer PDNN is implemented using a crossbar array of $WSe_2$ floating-gate transistors whose topology mirrors the network structure (Figure 1(b)), enabling physical vector-matrix multiplication. Each computational node consists of a pair of devices with tunable nonlinear responses, driven by differential voltage pulses ($+V_1/-V_1$). Nonlinear activation and linear weighting arise inherently from the device's current-voltage relationship, $I_{\text{out}} = V_{\text{in}} \cdot G(V_{\text{in}})$, where $G(V_{\text{in}})$ is the voltage-dependent conductance. The resulting currents are aggregated along shared signal lines according to Kirchhoff's law, completing the required weighted summation. Column-wise aggregation naturally produces network-level accumulation of weighted activations, allowing the array to simultaneously perform vector-matrix multiplication, nonlinear transformation, and summation entirely in hardware.

The $WSe_2$ floating-gate transistor (Figure 1(c)) enables gate-field-controlled modulation of the dominant carrier type in the ambipolar $WSe_2$ channel, thereby tuning the Schottky barriers at the $WSe_2$-electrode interfaces. As a result, the device functions as a nonvolatile memory element and exhibits highly tunable nonlinear rectification. As shown in Figures 1(d) and (e), programming pulses with opposite polarities continuously tune both the type and density of charges stored in the floating gate, thereby reversibly modulating the polarity and effective height of the Schottky barrier. This enables stable bidirectional rectification with more than 32 programmable current levels in each polarity (detailed in Appendix A). The measured current-voltage characteristics follow an exponential transfer function, $f(x) = W(e^{\beta x} - 1)$, where $W$ reflects the stored charge state and $\beta$ is an intrinsic device constant ($\approx 6$). This programmable nonlinear response unifies synaptic weighting and nonlinear transformation, providing a tunable hardware foundation for implementing the PDNN.

To assess forward-only training on the $WSe_2$ floating-gate transistor-based PDNN, we applied the FF algorithm, which up-

* Corresponding author (email: chenpan@njust.edu.cn, sjliang@nju.edu.cn)
† These authors contributed equally to this work.

© Science China Press 2026 

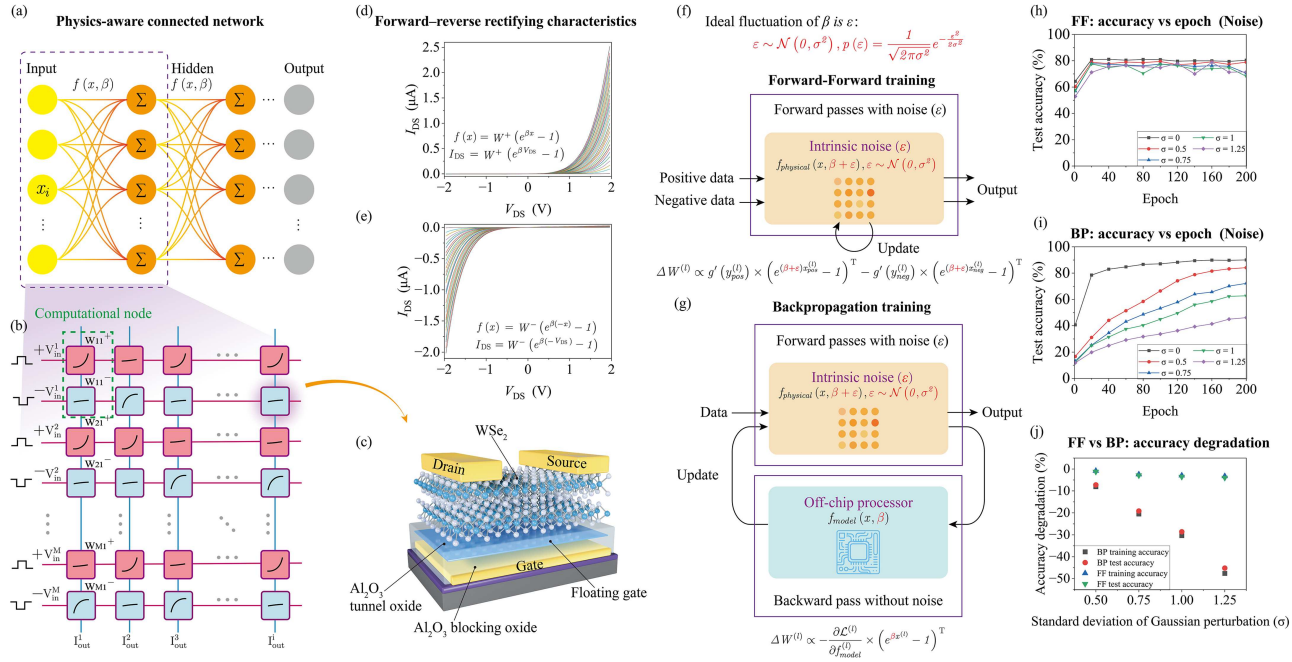



**Figure 1** (Color online) (a) Schematic of the PDNN. (b) Schematic of the $WSe_2$ floating-gate transistor array. (c) The structure of the $WSe_2$ floating-gate transistor. (d), (e) Electrically programmable nonlinear $I$-$V$ characteristics of the $WSe_2$ floating-gate transistor. $W^+$ and $W^-$ are programmable weights extracted from $I_{DS}$-$V_{DS}$ fits under forward and reverse rectification, corresponding to different programmed states shown in (d) and (e). (f) Schematic of weight updates under the FF training algorithm. (g) Schematic of weight updates under the BP algorithm. MNIST classification accuracy of the FF-trained (h) and BP-trained (i) PDNN under different levels of hardware nonuniformity (noise). (j) Comparison of FF and BP accuracy degradation as a function of hardware nonuniformity level.

dates weights using paired positive and negative forward passes rather than backpropagation (Figure B1). By relying on a local optimality criterion, FF avoids cross-layer gradient transport and remains resilient to device stochastic fluctuations. Numerical simulations on the MNIST dataset demonstrate that the FF-trained PDNN achieves high recognition accuracy under both supervised and unsupervised learning tasks. These results establish FF as a practical training strategy for in-memory computing hardware (Figure D1).

To demonstrate the robustness of the FF algorithm, we performed MNIST simulations in which device-intrinsic nonuniformity was modeled by introducing Gaussian perturbations into the transfer function. Specifically, in $f(x) = W(e^{\beta x} - 1)$, the intrinsic parameter $\beta$ was replaced with $\beta + \mathcal{N}(0, \sigma^2)$ to represent stochastic variations in the device characteristics, yielding $f(x) = W(e^{(\beta+\mathcal{N}(0,\sigma^2))x} - 1)$. During FF training, forward inference is performed directly on the physical $WSe_2$ floating-gate array based on its intrinsic nonlinear transport characteristics, and weight updates are subsequently computed from the resulting outputs using a local optimality criterion (Figure 1(f)). This approach allows the learning process to inherently incorporate device-level nonlinearities and stochastic fluctuations. In contrast, BP computes gradients off-chip (Figure 1(g)) using idealized device models, causing its updates to diverge from actual hardware behavior and reducing its tolerance to noise (see Appendix C). Performance comparisons across multiple noise levels highlight this difference (Figures 1(h)–(j)). The FF-trained network sustains high test accuracy (80.98%–78.2%) with only modest degradation (1.2%–3.98%), whereas BP accuracy collapses from 84.13% to 46.2%, with reductions up to 45.15%. The network architecture used for this evaluation is summarized in Figure E1. These results show that FF achieves stable convergence even under substantial device variability, while BP remains highly sensitive to nonidealities. Overall, FF provides a hardware-compatible, noise-tolerant training strategy that aligns naturally with nonlinear device physics.

*Conclusion.* We demonstrate that the FF algorithm can be directly implemented on in-memory computing hardware based on $WSe_2$ floating-gate transistors, where learning naturally leverages the devices' intrinsic nonlinear responses for PDNN computation. The resulting system exhibits stable learning performance under device variability and stochastic perturbations, while conventional backpropagation degrades markedly in noisy conditions. These results establish FF as a hardware-compatible and noise-tolerant training paradigm, offering a viable route toward scalable neuromorphic computing.

**Acknowledgements** This work was supported in part by National Natural Science Foundation of China (Grant Nos. 62204119, 62404099, 62304104, 62305155), Natural Science Foundation of Jiangsu Province (Grant No. BK20233001), and Fundamental Research Funds for the Central Universities (Grant No. KG202508).

**Supporting information** Appendixes A–E. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

**References**

1 Wright L G, Onodera T, Stein M M, et al. Deep physical neural networks trained with backpropagation. Nature, 2022, 601: 549–555
2 Momeni A, Rahmani B, Malléjac M, et al. Backpropagation-free training of deep physical neural networks. Science, 2023, 382: 1297–1303
3 Xue Z, Zhou T, Xu Z, et al. Fully forward mode training for optical neural networks. Nature, 2024, 632: 280–286
4 Huang X, Liu C, Tang Z, et al. An ultrafast bipolar flash memory for self-activated in-memory computing. Nat Nanotechnol, 2023, 18: 486–492
5 Hinton G. The forward-forward algorithm: some preliminary investigations. 2022. ArXiv:2212.13345