PERSPECTIVE

NANO · MICRO small
www.small-journal.com

Opportunities for 2D-Material-Based Multifunctional Devices and Systems in Bioinspired Neural Networks

*Jin Feng Leong, Maheswari Sivan, Jieming Pan, Zihang Fang, Jianan Li, Zefeng Xu, Shi Zhao, Quanzhen Wan, Evgeny Zamburg, and Aaron Voon-Yew Thean**

The growing demand for intelligent, real-time systems pushes artificial intelligence beyond the confines of centralized data centers toward distributed, edge-based applications such as autonomous robotics, mobile platforms, and IoT sensors. However, the energy and space requirements of conventional artificial intelligence (AI) hardware such as graphic processing units and AI-specific application-specific integrated circuits, pose fundamental limitations for deployment at the edge. Bioinspired computing offers a compelling alternative, emulating the efficiency and adaptability of biological systems to achieve low-power, real-time intelligence. Among these approaches, spiking neural networks stand out for their sparse, event-driven computation and have demonstrated orders-of-magnitude energy efficiency gains on neuromorphic platforms such as SpiNNaker and Intel's Loihi. Yet, to realize the full potential of bioinspired intelligence in edge environments, a new class of customized hardware is imperative. Emerging innovations in material science, particularly the integration of 2D materials, can enable the design of compact, reconfigurable neuromorphic devices that mimic complex neuronal dynamics with minimal power consumption. These advances promise a new generation of scalable, multifunctional edge AI systems that are capable of perception, adaptation, and autonomous decision-making, heralding a transformative leap in energy-efficient computing for pervasive intelligent technologies.

1. Introduction

In the era of accelerated artificial intelligence (AI) computing and the proliferation of generative models, neural-network-based algorithms have become the dominant paradigm, driving breakthroughs across vision, language, and decision-making tasks.[1,2] Convolutional neural networks are foundational for spatial tasks such as image classification and object detection. Recurrent neural networks, including long short term memories (LSTMs) and gated recurrent units (GRUs), have excelled at modeling temporal dynamics in sequential data like speech and time series. Lately, transformers, driven by self-attention mechanisms, have achieved state-of-the-art results in natural language processing and are expanding into vision domains[3] due to their ability to model long-range dependencies in parallel. As these models scale in depth, width, and parameter count, their computational demands have increased exponentially. This has spurred significant interest and investment in high-performance, energy-intensive data processing architectures, particularly graphics processing units and AI-specific application-specific integrated circuits, such as Google's tensor processing unit (TPU)[4] and NVIDIA's H100,[5,6] which are designed for massive memory bandwidth and ultrafast data movement. These architectures have become the backbone of modern AI infrastructure, enabling the training and deployment of large-scale models across data centers and research labs. However, the insatiable thirst for intensive computation resources and energy of such approaches are increasingly seen as unsustainable and are fundamentally incompatible with processing at the edge, where energy supply and area are constrained.

Unlike conventional models, bioinspired computing designs computer algorithms and models based on biological mechanisms and living phenomena. This bioinspired approach offers a compelling alternative architecture for energy-constrained and real-time intelligence at the edge. For example, genetic algorithms[7] optimize desired objective by mimicking natural selection, artificial bee colony algorithm[8] explores the search space using the collective behaviors of bee colonies, and spiking neural

J. F. Leong, M. Sivan, J. Pan, Z. Fang, J. Li, Z. Xu, S. Zhao, Q. Wan, E. Zamburg, A. V.-Y. Thean
Department of Electrical and Computer Engineering
National University of Singapore
Singapore 117583, Singapore
E-mail: aaron.thean@nus.edu.sg
J. F. Leong, M. Sivan, J. Pan, Z. Fang, J. Li, Z. Xu, S. Zhao, Q. Wan, E. Zamburg, A. V.-Y. Thean
Singapore Next-Generation Hybrid μ-Electronics Research Center (SHINE)
National University of Singapore
Singapore 117608, Singapore

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smll.202506638

DOI: 10.1002/smll.202506638

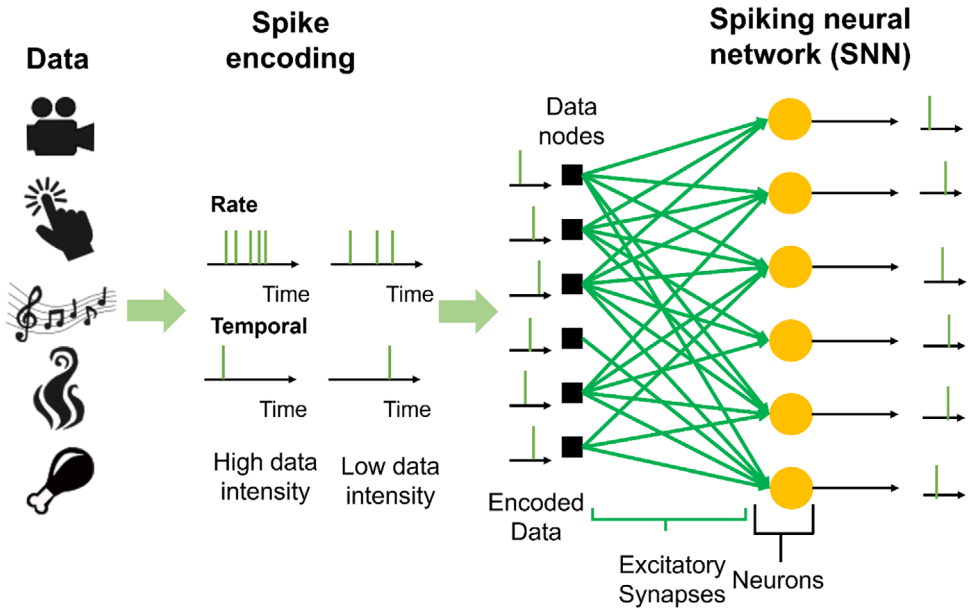

Figure 1. Spiking neural network architecture. In spiking neural networks, the processing of the data can be divided into three categories, encoding data into spikes, synaptic computation, and neurons. Data obtained from visual, sound, and chemical sensors need to be encoded into spike signals through various methods that include rate coding and temporal coding.[22] Subsequently, the encoded spikes signals will pass through the excitatory synaptic connections that strengthen connections between data nodes and the neuron for data nodes with higher data intensities. After the accumulation of sufficient spikes from the synapses such that the neuron membrane potential exceeds the threshold potential, the neuron fires spikes.

networks (SNNs)[9] learn through discrete input spikes similar to the functioning of the biological brain, enabling sparse computation that can significantly reduce energy consumption. With the rise of physical agentic AI, robotics, and distributed IoT sensors, a hierarchical integration of algorithms and modalities will become increasingly important with the growing need for data generation and preprocessing at the edge.[10]

The event-driven nature of SNNs that supports real-time, asynchronous processing with minimal overhead, makes them prime candidates of bioinspired computing for effective next generation of edge and agentic AI systems (e.g., autonomous drones, continuous monitoring systems, etc.). For instance, chips like BrainScaleS-2,[11] SpiNNaker,[12] and Intel's Loihi series chips[13,14] implement networks of spiking neurons in silicon. By leveraging spike-based computation, such systems have demonstrated energy usage orders of magnitude lower than conventional deep learning hardware. Intel's Loihi family[15,16] has demonstrated more than 50× speed up and at least 100× energy reductions on sparse, event-driven workloads such as constraint satisfaction, sparse coding, and other recurrent/temporal problems that align with spiking computation. For the future development of multiperception AI, innovations in customized hardware for advanced SNN implementation are required. This can benefit with exploration into novel materials such as 2D materials (2DMs) for the development of a new class of nonvolatile memories (NVMs) that can be used compute-in-memory which saves energy by minimizing data transfer. Early works of 2DM-based NVMs used in in-memory compute approaches have already shown $\approx 10^4$ gigaoperations (GOP) s^{-1} W^{-1}, which is 10× more energy efficient than traditional static random-access memories (SRAM) counterparts.[17,18] Furthermore, 2DMs can offer unique properties that cannot be found in bulk materials. These properties can open new opportunities for the creation of multifunctional devices. Low chalcogen vacancy migration energies and ambipolarity (n/p) of 2DMs enable the development of reconfigurable memtransistors[16] for bioinspired edge processing systems with tunable excitatory and inhibitory synapses. Second, leveraging on the enhanced charge confinement and low electronic screening in ultrathin channel bodies, single neuron implementations of 2DM can demonstrate greater energy savings and complex neuron models such as Hodgkin–Huxley with significantly smaller footprint than their silicon counterparts.[19,20] Therefore, the multifunctionality, device count reduction, power efficiency, accelerated learning, and even noise resilience offered by novel materials, devices, and codesigned systems will be key in supporting the next phase of multiperception AI.[21]

This article will discuss the opportunities and challenges of new material, device, and systems in bioinspired computing. Research and development in 2DMs and SNN algorithms will be used as a vehicle for in-depth evaluation and discussion. Through this discussion, we hope to show that the advantage of 2DM over other semiconducting bulk materials lies in the capability of creating novel multifunctional devices that enables reduction in implementation footprint and improvements of energy efficiency in SNN-based architectures.

2. Limitations of Digital Implementations of SNNs and Opportunities with 2D Materials

Bioinspired computation like SNN (**Figure 1**) processes information via discrete, time-dependent "spikes," closely mimicking the event-driven communication of biological neurons. Each neuron integrates incoming current until its membrane potential reaches a threshold, at which point it emits a brief voltage spike

and then resets. This temporal dynamics enables SNNs to encode information in the rate of spikes produced or time of spike production,[22] yielding rich representational power with high energy efficiency. During the training of these synapses, Hebbian and anti-Hebbian learning provide the local, spike-based synaptic update rules to ensure excitation–inhibition balance and unsupervised feature learning.[9] Hebbian plasticity strengthens the synaptic connection between two interlayer neurons when their activations are correlated. By contrast, anti-Hebbian learning weakens the synaptic connection when activations are uncorrelated. Finally, neurons integrate the signals received from the synapses, and fire forward once attaining a threshold to the next layer of neurons. The network of neurons will eventually achieve learning, logic, and even reasoning.

2.1. Existing Implementations of Spiking Neural Networks and Its Limitations

Due to the largely parallel computation across multiple synapses and neurons in an asynchronous fashion, traditional von Neumann serial computing systems are noncompatible with the computation of SNNs on the hardware level. This has led to the development of specialized digital neuromorphic chips like Loihi that parallelized computing through a mesh network of computational cores,[13,14] giving rise to a significant reduction in computational latency and improved energy efficiency.[21] However, such chips will eventually reach a limit in scaling down footprint and power due to the excessive number of devices required to implement synapses and neurons. Specifically, they use digital SRAMs for synaptic computation that are area-inefficient and drive up implementation area.[13,14] The volatile nature of SRAMs also leads to high off-state power to maintain the memory state which raises power consumption.[23] Neuronal implementations also take up at least 18 transistor devices per neuron, further exacerbating the problem with footprint and power.[24] In comparison, the coupling of emerging devices with novel computing architectures can lead to significant reductions in footprint and power for synaptic and neuron computation. Emerging NVMs[18] such as memristors[25,26] and memtransistors[27–29] offer a lower device count (from six devices per cell to one) for memory, which is significantly more efficient than SRAM. These advantages on the device level can be extended through the novel computing architecture such as in-memory computing (IMC). Emerging IMC-based neuromorphic chips have utilized NVMs to achieve at least 10 times improvement in performance density in GOP s^{-1} mm^{-2} compared to SRAM-based chips.[18] These IMC-based neuromorphic chips have also demonstrated a significant power efficiency on learning tasks compared to their digital counterparts that utilize SRAMs.[18] Particularly, novel NVM-based neuromorphic chips show on average 10^5 GOP s^{-1} W^{-1} which is 100 times higher than their SRAM counterparts.[18] For tackling the issue of a large neuron footprint, recent bioinspired implementations have also looked at potential single silicon neuron device implementations.[24,30] Hence, there is potential to further drive down the energy and area consumption of bioinspired computing through research into novel NVMs and potential single-device neuron candidates.

2.2. Opportunities with 2D Materials in Multifunctional Devices for Spiking Neural Networks

The exploration of novel device candidates for bioinspired computing can benefit from research into new materials that can offer additional functionalities that are lacking in state-of-the-art silicon complementary metal oxide semiconductors (CMOS). Emerging 2DMs hold promise in overcoming several limitations of silicon and other potential semiconducting materials candidates like semiconducting transition metal oxides (TMOs) in the development of multifunctional devices for SNNs, which will be discussed in this section. We limit our discussion of 2DMs to MoS_2 and WSe_2 as they are the materials in line with the International Roadmap for Devices and Systems roadmap and have a propensity for industry adoption in the future.[31]

While silicon has seen widespread use in transistor electronics, it is not an ideal material for multifunctional memtransistors that have potential for compact implementation of complex synaptic functions.[27,28,51] As shown in literature, the creation of memtransistors can be made through a few approaches, namely, resistive switching through electric-field-induced dopant migration, phase change, and ferroelectrics. Vacancy-modulated memtransistors are known to leverage on chalcogen (O, S, Se, Te) vacancy–charge carrier electrostatic interactions and electrostatic movement of chalcogen vacancies with n-doping properties for memory storage.[41] Silicon is not suitable for construction of memtransistors as silicon vacancies have high vacancy formation energies (≈4.0 eV in **Table 1**) and they also do not dope the material to elicit conductivity changes for memory storage. Furthermore, silicon dopants, such as boron (B) and phosphorus (P), have high dopant diffusion activation energies of 3.12 and 2.74 eV, respectively (in Table 1), which will require high electric field to move. This would lead to high operational voltages. On the other hand, 2DMs possess ambipolarity (n/p) tunability, relatively low chalcogen vacancy formation energy, and low dopant diffusion energies that can used for multifunctional vacancy-modulated memtransistors.[27,28] Specifically, the chalcogen vacancies of 2D-MoS_2 and WSe_2 are ≈2.8 eV (Table 1), which is significantly lower that of silicon, enabling vacancy formation without excessive damage to the device channel that would limit transistor operations. In addition, the vacancy migration energies of 2DMs are ≈3 times lower than that of silicon extrinsic dopants (Table 1), which can lower operational voltages of memtransistors. Although the dopant diffusion energy of O vacancies and vacancy formation energies in TMOs are around the same as 2DMs, they lack ambipolarity (n/p) due to their mostly large bandgap (Table 1) to leverage carrier–vacancy electrostatic interactions for additional memtransistor functionalities such as Hebbian/anti-Hebbian tunability.[27] Due to the high switching energies and low carrier mobility even in their crystalline state, phase change materials also face limitations in their potential as memtransistor devices. Furthermore, as the resistive switching mechanism of phase change is temperature-dependent, it is a challenge to couple charge carrier dynamics with phase change resistive switching mechanism to introduce multifunctionality in memtransistors. For organic semiconductors,[52] the disordered crystalline order in organic semiconductors limits their carrier mobilities.[32] This would limit the potential of transistor functionality in memtransistors. While traditional ferroelectrics are

Table 1. Comparison of common semiconducting materials candidates for electronics. For transition metal oxides, indium gallium zinc oxide (IGZO) is chosen for comparison as a common n-type semiconducting transition metal oxides (TMOs). Nickel oxide (NiO) is chosen as a common semiconducting candidate for p-type TMOs. For 2D materials, the MoS_2 and WSe_2 are chosen as the materials for comparison given their huge interest for semiconducting electronics from the industry. HZO is hafnium zirconium oxide and GST is germanium antimony telluride.

Material properties	Silicon	Transition metal oxides (NiO, IGZO)	Organic semiconductors	Traditional ferroelectrics (HZO)	Traditional phase change materials (GST)	2D materials (MoS_2 and WSe_2)
Ambipolar (n/p)	Yes	No	Yes[32]	No	Yes[33]	Yes
Dopant/vacancy diffusion energy/domain/phase change energy [eV]	Boron (B) (3.12)[34] Phosphorus (P) (2.74)[35]	O vacancy (≈0.85)[36]	–	≈0.45–0.65[37,38]	≈2[39,40]	S vacancy (≈0.66)[29] Se vacancy (≈0.7)[41]
Vacancy formation energy [eV]	4.0[42]	IGZO: 2–3[43] NiO: 1.17–4.03[44]	–	–	–	MoS_2: 2.8 WSe_2: 2.8[45]
Critical electric field for impact ionization [kV cm^{-1}]	300[46]	–	–	–	–	WSe_2: 63.75[20]
Carrier mobility [cm^2 V^{-1} s^{-1}]	$\approx 10^2$[47]	$\approx 10^2$[48]	≈ 10[32,49]	–	≈ 10[50]	$\approx 10^2$[47]

insulators and are used as gate dielectric in memtransistors. As a result, ferroelectric materials are highly susceptible to charge trapping at the gate dielectric interface that gives rise to depolarization field which limits memory retention to $<10^5$ s.[53,54] Comparatively, 2DM-based vacancy-modulated memtransistors have shown excellent retention >10 years.[28] Thus, among all the prevalent neuromorphic and traditional materials, 2DM vacancy modulation mechanisms have been showed to have the properties most suitable for the creation of novel reconfigurable memtransistors for SNNs.

Emerging 2DMs can also leverage on their ultrathin bodies that can enable low critical electric and weak electrostatic screening for effective implementations of compact neurons. One of the more promising silicon-based single neuron device is known to leverage on impact ionization.[24] However, such processes are not that efficient in silicon due to the high critical electric field of 300 kV cm^{-1}[46] required to create impact ionization that will lead to increased power consumption (Table 1). 2DMs can achieve impact ionization with ≈5 times lower critical electric field (Table 1) due to high charge confinement in their ultrathin bodies and absence of dangling bonds essential for impact ionization. Furthermore, the 2DMs also have significantly lower dielectric constants compared to silicon and TMOs below 5 nm thickness (Table 1), which results in weak electrostatic screening. This enables the ease of forming gate-tunable band alignment in heterojunctions that can be used to build gaussian transistors for simplified implementations of the Hodgkin–Huxley neuron model used in SNNs.[19] As such, 2DMs have showcased unique physical properties that are suitable for the development of compact neuron hardware.

3. Multifunctional 2D-Material-Based Devices: Opportunities and Limitations

In the previous section, 2DMs have demonstrated salient advantages in their physical properties for the development of multifunctional memtransistors and devices for compact neuron implementations. In this section, we will first describe the operational principles of each device and discuss how these devices fare compared to existing state-of-the-art devices and their limitations.

3.1. Reconfigurable Vacancy-Modulated Memtransistors

Vacancy-modulated memtransistors (**Figure 2a**) are three-terminal memory devices that utilize gate-controlled electrostatic modulation to alter the distribution of chalcogen vacancies. This modulation enables nonvolatile changes in conductance, allowing for memory storage (Figure 2b). The additional gate terminal has enabled multifunctional synapses for learning rate modulation[51] and tunable Hebbian/anti-Hebbian capabilities[27] without having to change the properties of the spikes carrying information about the input data. For example, a two-terminal vacancy-modulated TiO_x memristor requires changes to the spike pulse amplitude of the input data pulse train to adjust the learning rate of the synaptic device.[56] On the other hand, vacancy-modulated memtransistors made from MoS_2 can tune the learning rates from the gate terminal of the device[51] (Figure 2c). This type of device has been demonstrated to simplify circuitry used for implementation of continuous learning algorithms, which is a vital part of SNNs to accommodate incremental data continuously in dynamic environments. The naïve solution of complete retraining the entire model inevitably raises energy costs significantly. A more efficient approach is to regulate the learning of certain synapses to retain existing knowledge while the rest of synapses are free to incorporate new information to existing SNN model, achieving a satisfactory recognition accuracy for all tasks. The key mechanism of learning rate regularization can be efficiently implemented with the additional gate terminal in vacancy-modulated memtransistors, eliminating the need for complete retraining and peripheral input encoder circuit. This would be critical for enlarging the learning capability of the edge systems that could encounter new scenarios continuously during the actual deployment despite limited energy and compute resources.

For existing silicon implementations, the demonstration of Hebbian/anti-Hebbian capabilities would require at least five

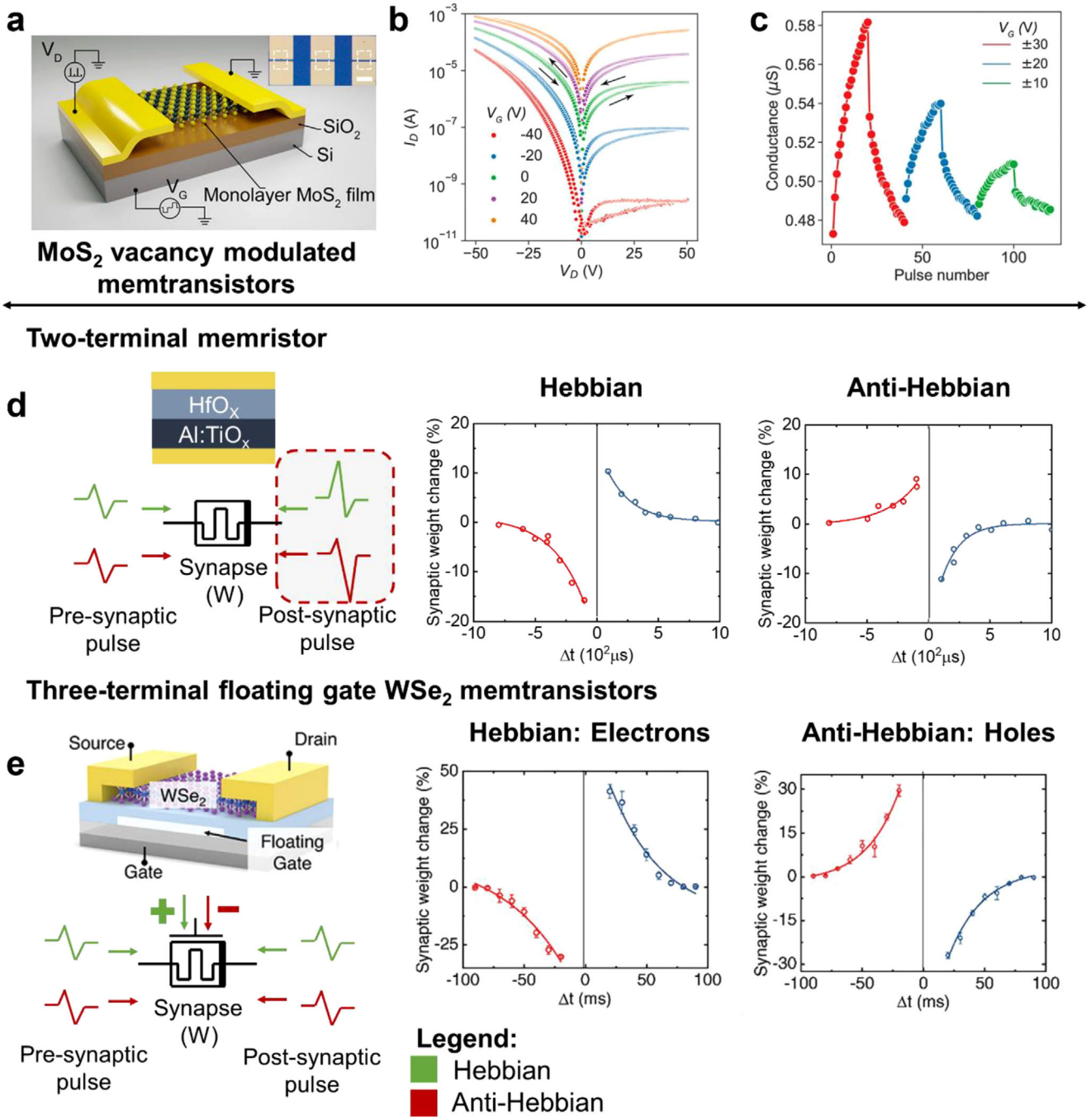

Figure 2. Tunable Hebbian/anti-Hebbian learning in synaptic devices. a) Vacancy-modulated memtransistors made using MoS_2. b) Resistive switching I_d–V_d curves of memtransistors. The varying degrees of hysteresis when changing the gate voltage applied indicates gate tunable resistive switching. c) Gate tunable varying potentiation and depression curves that enables tunable learning rates in the device via the gate. d) An example of a two-terminal resistive switching synaptic memristor device that can tune between Hebbian and anti-Hebbian learning modes in spike-timing-dependent plasticity (STDP) by changing the postsynaptic pulse profile (highlighted by red dotted line box). Data are taken from ref. [55] e) Floating-gate reconfigurable memtransistor. It uses the floating gate to program the majority carriers within the device. The electrostatic carrier–vacancy interaction can determine whether the floating gate memtransistor behaves either in Hebbian or anti-Hebbian mode. (a–c) Reproduced with permission.[51] Copyright 2021, American Chemical Society.(e) Reproduced with permission.[27] Copyright 2023, John Wiley & Sons.

Table 2. Comparison of various Hebbian/anti-Hebbian tunable synaptic devices. Operational voltage is the voltage required to operate the device when $\Delta t > 0$ for the STDP operation. Δt is the time difference between the postsynaptic and presynaptic pulse. The on/off ratio refers to the ratio of the largest conductance state and lowest conductance state of the device. Here, (H/A-H) refers to Hebbian/anti-Hebbian. Linearity is a measure of how linear the conductance change is with the number of pulses applied in the potentiation and depression curve.[18] Input pulse change refers to the change in either pre- and postsynaptic pulse profile to tune between Hebbian/anti-Hebbian behavior. The endurance metric measures the highest number of switching cycles recorded. D.C. and A.C. indicate the direct current and pulse switching measuring conditions, respectively.

Material	Resistive switching mechanism	Operational voltage [V] (H/A-H)	Endurance (cycles)	Retention [h]	On/off ratio (H/A-H)	Linearity	Is an input pulse change required?
Two-terminal $HfO_x/Al/TiO_2$[55]	Charge-trapping	4/−4	>200 (D.C)	≈2	$10^3/10^3$	4.7/−7.7 (A-H) 5.2/−7.3 (H)	Yes
Two-terminal $La_{0.67}Sr_{0.33}MnO_3/SrTiO_3$[58]	Oxygen vacancy migration	3/−3	–	≈3	≈100	17.6/−7.5 (H)	Yes
Two-terminal $Ge_2Sb_2Te_5$[59]	Phase change	1.1/−1.1	>500 (A.C.)	–	20	2.8/2.5 (H)	Yes
Two-terminal a-hBN[60]	Conductive bridge (Ag)	0.6/−1.9	$>10^4$ (A.C.)	300	10^4	−0.8/−5.0 (H)	Yes
Three-terminal WSe_2[61]	Ferroelectric	−8/−8	>750 (A.C.)	–	10^3	1.5/−2.8 (A-H) 0.6/−1.2 (H)	No
Three-terminal WSe_2[27]	Floating-gate and selenium vacancy migration	11/11	$>10^4$ (A.C.)	≈3	10	0.1/−2.3 (A-H) 2.8/−3.1 (H)	No

floating gate transistors coupled with capacitors and operational amplifiers that require a large circuit footprint.[57] This onerous design can be replaced with a multifunctional emerging resistive memory that is capable of tunable Hebbian/anti-Hebbian modes, effectively reducing hardware footprint by at least 5 times. However, most Hebbian/anti-Hebbian tunable two-terminal devices require a change of the pulse profile of input pulses to the pre- and postsynaptic terminals of the device to tune between Hebbian and anti-Hebbian modes (Figure 2d). Three-terminal vacancy-modulated memtransistors can leverage carrier–vacancy electrostatic interactions to tune between Hebbian and anti-Hebbian by adjusting the gate terminal that alters majority carriers (electrons/holes) within the channel (Figure 2e). This can lead to significant savings in device count which would be elaborated in the next section. For example, a dual memory floating-gate vacancy-modulated memtransistor[27] demonstrated tunable Hebbian and anti-Hebbian behavior by programming the device in n-/p-modes through the floating-gate and using the second memory element to store synaptic information with electrostatic migration of Se vacancies under high electric field. By applying a strong electrostatic field through a source drain bias of $V_d = 10$ V, the positively charged Se vacancies can be moved to generate a change in the vacancy distribution across the channel and under the contacts. As Se vacancies are positively charged n-dopants, an increase in Se vacancies within the channel relative to the contacts would lead to a decrease (increase) in source-to-channel barrier for majority electron (hole) operation, leading to a negative (positive) change in threshold voltage of the memtransistor. This is analogous to a Hebbian (anti-Hebbian) learning synapse, the connection between two neurons should get stronger (weaker) with positive data inputs through the connection. Similarly, as more positively charged selenium vacancies are electrostatically driven into the channel, the electrons (holes) in n-(p-)dominant conduction regime becomes more (less) conductive. This device can reduce spike pulse generator components used to tune between Hebbian and anti-Hebbian learning modes in the device.

To compare the performance of Hebbian and anti-Hebbian tunable devices, we use standard metrics[18] that are used to compare NVMs such as operational voltage, retention, on/off ratio, linearity, and also whether they require pulse modulation at the input to configure between Hebbian and anti-Hebbian modes. Here, the on/off ratio refers to the ratio between the lowest conductance state and highest conductance state reported for the memory device. Linearity is an important measure of how linear the device changes in conductance against the number of pulses applied, which is measured from the device's potentiation and depression curves. The closer the value is to the 0, the more linear it is. This is critical for circuit and system designers for binning of memory states to determine the synaptic weights and would merit a fairer comparison and assessment for memory devices used in this type of application. Other important metrics such as switching speed and temperature stability are required to enable better benchmarking with existing memory devices. Switching speed directly affects the memory write time, which can significantly affect the training time of the neural network. Similarly, temperature stability at the industrial standard of 85 °C is also an important factor in considering the robustness and usability of such devices as reliable memory chips.[62]

We show a comparison of the three-terminal memtransistors to some of the reported works of Hebbian/anti-Hebbian tunable devices in **Table 2**. An observation is that two-terminal vertical memristors typically have lower operational voltages than three-terminal memtransistors. It is possible to significantly reduce the operational voltages of vacancy-modulated memtransistors as shown by Feng et al. to ≈0.35 V[63] with proper optimization of the grain boundaries for chemical-vapor-deposited grown 2DM. The use of high k dielectric is also known to have reduced the MoS_2 switching voltages to 1 V.[64] Thus, with proper material engineering of the channel of the vacancy-modulated memtransistors, it does present huge opportunities for low-power, reconfigurable synaptic memory device for SNNs.

However, one critical issue with memtransistors would be the linearity of the resistive switching.[27,28,63,65] Feng et al.[63] have tried to improve the linearity by using increasing voltage pulses in the pulse train but this is not feasible in actual operations as the device can be at various points of memory states during training and it is a challenge to keep track of all synaptic weights in a memory array as to which pulse height to use. Drawing lessons from the study of vacancy-modulated memristors, we understand that the linearity of the memory device can be improved by thermal enhancement layer[66] and tuning of chalcogen concentration distribution within the material medium.[67] Thus, future works can be explored in these areas to make memtransistors more viable options as memory devices for future IMCs.

3.2. Explorations of Vacancy-Modulated Memtransistor Device Operations with Modeling

As discussed in the previous section, the unique physical mechanisms governing vacancy-modulated memtransistor behavior in 2D materials necessitate a thorough understanding of their physical properties through physics-based modeling. Such insight is critical for the development of accurate compact models and, ultimately, product design kits for circuit design. Furthermore, it can offer valuable design guidelines for ensuring reliable operation, selecting appropriate channel materials, improving scalability, and optimizing device architecture.

One of the major bottlenecks in advancing memtransistor-based technologies is the limited understanding of the complex electrostatic interactions among the three terminals and the memristive channel, including the roles of defects and carriers across different operating regimes. Particularly critical is the understanding of the dramatic changes in space-charge distribution as the device dynamically transitions between cutoff, linear, and saturation modes in response to varying gate and drain potentials. This gap often leads to unreliable memory-erase operations and operational uncertainties across different applications. Therefore, the development of a robust, physics-based model is imperative. Such a model can address key challenges, including high power consumption, poor control over defect redistribution, and inefficient memory-erase mechanisms.

The fundamental understanding of the resistive switching mechanism in memtransistor proposed by Sangwan et al.[28] is based on conductance modulation which arises from field-induced vacancy migration, leading to modulation of the Schottky barrier (SB) width (**Figure 3a**). However, memtransistor being controlled by an additional gate terminal is likely to have contributions not solely from SB modulation, but also from the channel lateral electrostatics. Hence solving for interactions between gate potential, ions, and carriers under different operational bias conditions including the drift-conduction-dominated transport is essential to overcome the limitation of the Schottky barrier model, whose validity ceases beyond the subthreshold conduction regime.

Another important aspect to consider is the influence of lattice temperature on vacancy migration and, consequently, on conductance modulation (Figure 3b). This is particularly significant because lattice self-heating is a common occurrence in devices with sub-micrometer channel lengths and during operations in the pinch-off regime. Since vacancy hopping rates and mobility increase with temperature, it becomes essential to include vacancy dynamics under thermal effects as well. To address this, Sivan et al.[41] developed a self-consistent, physics-based model that integrates lattice temperature, vacancy behavior, and channel electrostatics to accurately capture the interactions between gate potential, ionic species, and charge carriers during memristive switching. This model is calibrated using an ambipolar WSe_2 memtransistor, allowing for detailed understanding of how carrier type (electrons vs holes) influences vacancy transport. Expanding on the concept of Schottky barrier height modulation, the study highlights how vacancies significantly impact threshold voltage modulation in transistors. This insight can be strategically used to reduce the write bias. Specifically, the study demonstrates that threshold voltage shifts, driven by substantial changes in vacancy distribution near the source–channel interface, can lead to a 25% improvement in power efficiency.

An equally important consideration in modeling 2D-material-based memtransistors is the influence of grain boundaries, which are frequently observed in chemical-vapor-deposition (CVD)-grown polycrystalline 2D materials. However, most existing compact models for 2D devices overlook grain boundary (GB)-induced transport mechanisms, thereby limiting the identification of an appropriate state variable for resistive switching. Wang et al.[68,69] addressed that gap by introducing a compact model based on an exponential distribution of localized states, where transport is dominated by the GB energy barrier (Figure 3c). The model leverages theories of GB energy barriers and space-charge-limited current (SCLC) to explain the scaling behavior with respect to channel length and grain size. Under high lateral electric fields, SCLC effects emerge, leading to nonuniform charge distribution with higher density near the injecting contact. These SCLC effects can result in a sudden increase in current at intermediate voltages. Additionally, electric-field-induced generation and recombination of vacancies cause redistribution of trap centers at GBs and contacts. By treating the trap center density as a state variable, RS behavior is captured using state-variable transition functions and appropriate boundary conditions, where GB barrier modulation and asymmetric trap redistribution due to SCLC are identified as key contributors to RS behavior.

It is also important to develop a numerical modeling framework that enables efficient exploration of the large parameter space associated with 2D memristive synaptic devices. Recently, Spetzler et al.[70] reported a high-throughput numerical model to study the tunable synaptic behavior of 2D MoS_2 memristive

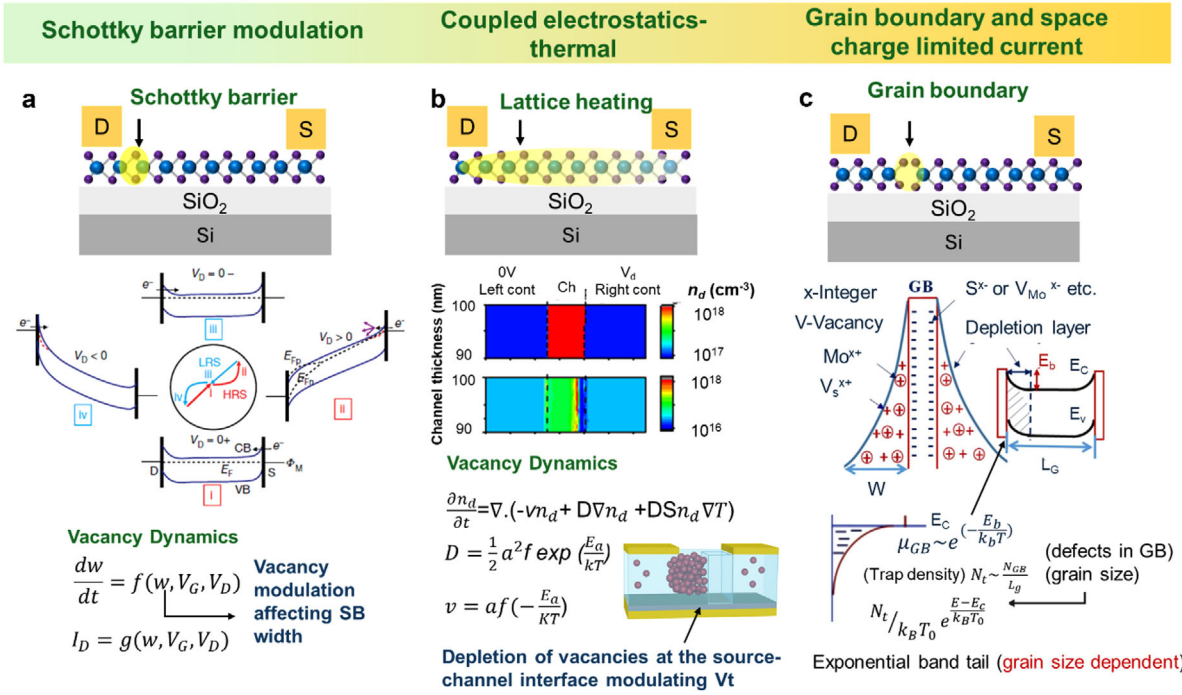

Figure 3. Modeling considerations for vacancy-modulated memtransistors. a) Vacancy accumulation/depletion driven Schottky barrier modulation as the dominant switching mechanism chemical-vapor-deposition (CVD)-grown polycrystalline monolayer MoS_2.[28] b) Effect of lattice temperature and threshold voltage on the resistive switching mechanism in ambipolar WSe_2 memtransistor.[41] c) Impact of grain boundaries and space charge limited current on the resistive switching characteristics of CVD-grown MoS_2 memtransistor.[68] Band diagram from Figure 1. (a) Reproduced with permission.[28] Copyright 2018, Springer Nature Limited. b) Adapted with permission.[41] Copyright 2022, American Chemistry Society. c) Adapted with permission.[68] Copyright 2019, IEEE.

devices, focusing on the voltage pulse characteristics in lateral 2D memristors and evaluating synaptic device metrics across different weight-update schemes.

While the prospects for implementing memtransistors in in-memory computing architectures are promising, several challenges remain to be addressed. The lateral configuration of these devices often requires high write bias voltages, resulting in increased energy consumption that may offset the energy savings gained from reduced data movement. Although exploiting threshold voltage shifts can yield some energy benefits, memtransistors still typically consume more power than their vertical memristor counterparts. Lowering the write voltage presents one possible solution, but it may weaken the driving force necessary for effective vacancy migration. In this context, solution-processed 2D materials present a valuable opportunity for further optimization. Their tunable properties—such as flake size—can influence vacancy concentrations, particularly at flake edges, offering an additional degree of control. This tunability could enable new strategies for performance enhancement. These possibilities will be explored further in the section on future developments.

3.3. Multifunctional Devices for Compact Implementation of Neurons

An emerging device for compact spiking neuron implementation of integrate and fire functions is impact ionization field-effect transistors (I^2FETs). In fact, a recent study published in Nature earlier this year[24] also demonstrated an artificial-neural-network (ANN)-based silicon neuron using this phenomenon. The opportunity for 2DM here is that I^2FETs can be implemented with a lower energy consumption can compared to silicon-based I^2FETs due to the lower critical electric field required to elicit this phenomenon of impact ionization, as shown in the earlier section. Here, we will look at an example of 2DM implemented spiking neuron from Choi et al.[20] using a WSe_2-based I^2FET (**Figure 4a**). The device follows the leaky integrate-and-fire (LIF) model: excitatory voltage pulses raise the membrane potential, and when it exceeds a threshold, a spike event occurs. The structure consists of gated and ungated regions. The operational principles of the I^2FETs are described step by step (Figure 4b) from impact ionization to integrate and fire and finally reset with the derived drain current profile against time in Figure 4c. Initially, (step i) the device is in equilibrium. Upon applying a voltage above the threshold (step ii), impact ionization induces an exponential rise in current as generated carriers outnumber leakage carriers. This corresponds to the integrate and fire process of the neuron. Over time (step iii), carrier scattering slows the current increase. When the current reaches around 0.5 mA μm^{-1}, a reset voltage (step iv) flushes out carriers, restoring the initial state for the next spike event. This corresponds to the reset of the neuron after firing. A comparison among the state-of-the-art spiking neurons was made in the work[20] and it was demonstrated that the 11.3 pJ per spike μm^{-1} is one of the lowest among all other possible integrate and fire neuron devices which include volatile memristors, phase

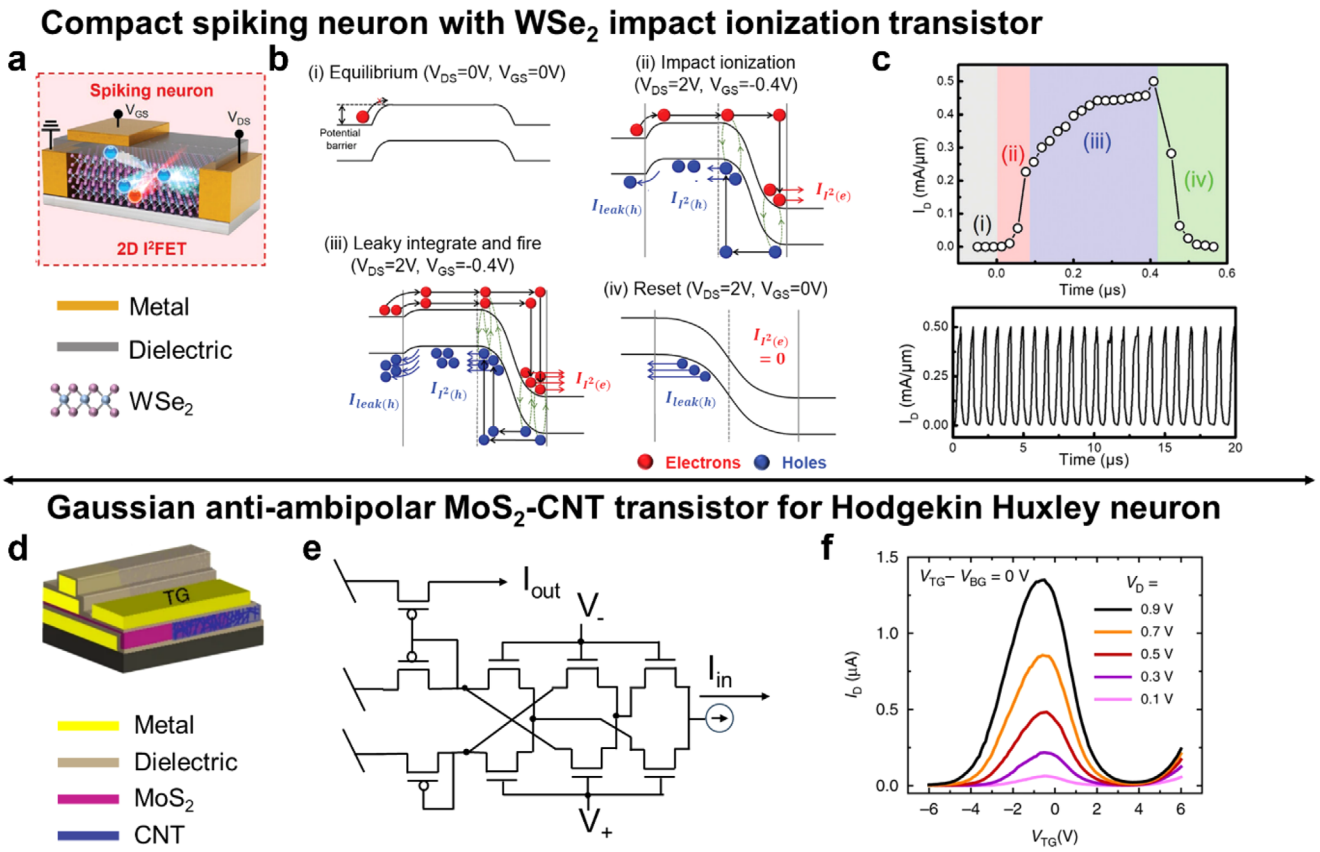

Figure 4. Multifunctional devices for neuron implementation. a) Spiking neuron using impact ionization field-effect transistors (I^2FETs). b) Operational principles of I^2FETs. The figure depicts at each step of the operation in the behavior of holes and electrons through impact ionization, leaky integrate and fire, and reset process necessary for neuron implementation. c) The corresponding current against time curve during the impact ionization, integrate and fire, as well as reset process. The spiking process is repeatable for multiple charging and discharging cycles. d) Heterojunction device of MoS_2 and carbon nanotube formed Gaussian transistor. e) Gibert Gaussian circuit that requires at least nine transistors to form the Gaussian profile.[71] f) Profile of the Gaussian transistor in(d) with varying drain voltages. (a–c) Reproduced with permission.[20] Copyright 2024, John Wiley & Sons. (d and f) Reproduced under the terms of the CC-BY license.[19] Copyright 2020, Megan et al., Published by Springer Nature Limited.

change memories, partially depleted silicon on insulator. Here, the energy consumption is also normalized to the device channel width. Comparing it with recent work on silicon I^2FETs for ANN[24] which has an operational voltage of 75 pJ per spike μm^{-1}, the 2DM variant is at least 6 times lower in power consumption. The potential of I^2FETs can be extended with the incorporation of high-*k* dielectric that enables more efficient gate control over the device, thus improving the electric field concentration and enabling further reductions in power consumption of the device. The conditions used to characterize the energy consumption of the I^2FETs is shown in **Table 3**.

Table 3. Comparison of energy per spike of impact ionization field effect transistors built from silicon and WSe_2 and its corresponding measurement conditions.

Channel material	Spiking frequency [Hz]	Pulse width [μs]	Energy per spike [pJ per spike μm^{-1}]
Si	10^4–10^5	1–50 000	75–18 × 10^6@25 °C
WSe_2	1.77 × 10^6	0.5	11.3@25 °C

However, akin to other works on single device neuron implementations,[30] they neglect to mention that such single device neurons are insufficient to implement the entire neuron. Additional pulse generator device components are required to generate bipolar pulses[72] for training of synapses and multimodal pulse waveforms for other forms of neuron models like Hodgkin–Huxley.[73] The components required for implementing pulse generators is nontrivial, as seen in a recent work by Zhu et al.[72] The demand of pulses adjusting synaptic weights comes from synaptic devices used, and it is true proper synapse–neuron codesign,[27,24,74] that all aspects of neuron implementation are considered.

For more complex neuron models such as Hodgkin–Huxley neuron model, Beck et al.[19] introduced a novel approach to neuromorphic circuit design by leveraging dual-gated Gaussian heterojunction transistors (Figure 4d). These transistors, built from wafer-scale, mixed-dimensional van der Waals heterostructures, incorporate monolayer MoS_2 (grown via CVD) and solution-processed semiconducting single-walled carbon nanotubes. A key advantage of these devices is their ability to form gate-tunable p–n heterojunctions, leading to an antiambipolar response with a distinct Gaussian transfer characteristic. The working

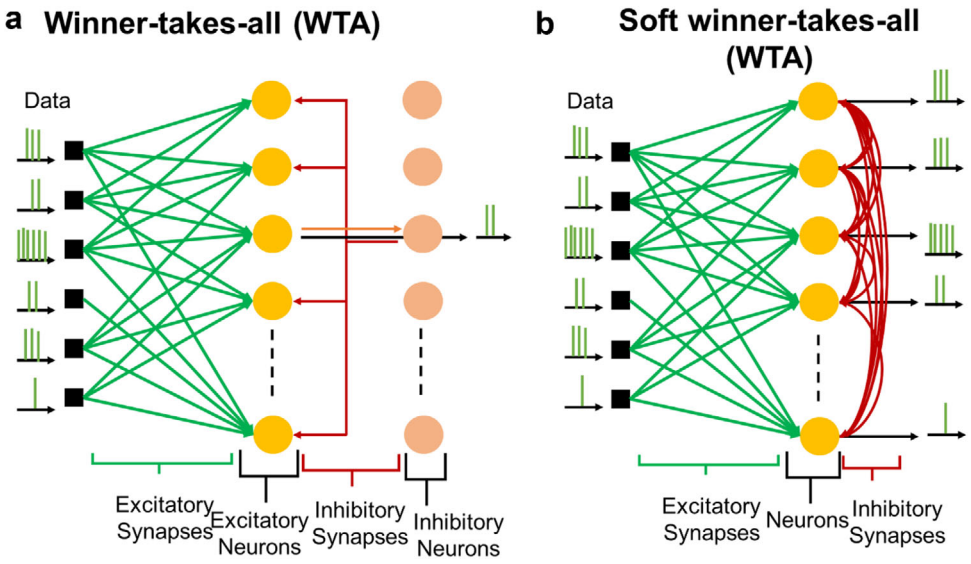

Figure 5. Comparison between winner-takes-all and soft winner-takes-all in SNN. a) A winner-takes-all implementation of SNN. The data shown here are encoded by rate encoding. After a winning neuron fires, it sends the signal to the inhibitory neuron that fires inhibitory pulses to neighboring nonfiring neurons to totally inhibit their firing activity. b) A soft winner-takes-all implementation of SNN. The data shown here are encoded by rate encoding. After a winning neuron fires, it sends inhibitory pulses to neighboring neuron that do not totally inhibit the firing activity of neighboring neurons. The degree of inhibition is dependent on the strength of the inhibitory synapses between the winning neuron and the neighboring neurons.

principle of such Gaussian transistors stems from the basis of using the gate to tune between dominant majority carrier operational regimes in a forward biased p–n junction diode channel. On the left sides of the Gaussian, it is typical that the n-type conduction is more dominant while the right side of the Gaussian, the p-type conduction is more dominant.[19,75] Compared to silicon-based implementations of Gaussian functions, this device can reduce the transistor count from nine to one[71] (Figure 4e), thus showing its superiority in implementation footprint. Next, we compare other metrics of the device such as peak to valley ratio (PVR), peak position, and the full-width half maximum (FWHM) of the Gaussian curve formed. For the silicon implementation, the PVR is 189, the peak position is 0, and FWHM is 0.19. On the other hand, the average PVR of Gaussian transistor used in Megan et al. is for $V_{tg} - V_{bg} = 0$ V and $V_d = 1$ V is ≈130 (Figure 4f), with the average peak position of −0.42 V and FWHM of 2.92.

It is easier to tune the profile of the Gaussian and attain the PVR, peak position, and FWHM desired as it involves the inclusion and exclusion of transistors in circuit design.[71] However, the Gaussian device design to tune the parameters to attain the desired Gaussian parameters has not been discussed in the work and we believe it is nontrivial.[19] Here, we deduced the device parameters that influence the Gaussian formation in the device. We believe that FWHM of the device can be lowered with gate oxide thickness and the proper selection of gate metals and doping to tune the V_{th} of the device. Currently, the gate oxide thickness for the top and bottom gate of the Gaussian transistor is 35 nm, which can result in relatively high subthreshold swing (SS) to turn on the device. With the improvement of SS, the device will be able to tune on faster, leading to more steep profiles when climbing to the peak of the Gaussian. The V_{th} can be tuned such that the device n- and p-sides can turn on closer to the center of the Gaussian with proper work function metal and doping profile of the semiconductors. Consequently, the V_{th} will determine the peak position as well. For PVR, it is more complicated as it is determined by the overlap of n and p on current as well as the off-state leakage current, which is influenced by the doping profile, contact resistance, SS, as well as the V_{th} of the device. The determination of the PVR can only be made with the proper simulation of the device, which can be potential avenues for exploration in future works.

4. System Level Savings Derived from Multifunctional 2D Devices

By incorporating the multifunctional 2DM-based devices into the SNN hardware, it can potentially derive some savings in energy consumption and implementation footprint. In this section, we will look at how these multifunctional devices can address some of the challenges faced in existing implementations of SNNs.

4.1. Competitive Learning in Spiking Neural Networks with Adaptive Lateral Inhibition

Competitive learning algorithms have been developed for SNNs to help enhance their recognition accuracy.[76,77] Competitive learning introduces lateral inhibitory synapses between neurons to prevent two neurons from learning similar features, thus ensuring that the distinctive features between different images are enhanced and recognition accuracy is improved. This is typically applied with a winner-takes-all (WTA) approach which totally inhibits all other neurons with every winning neuron firing event (**Figure 5a**). However, this is an oversimplification of biological

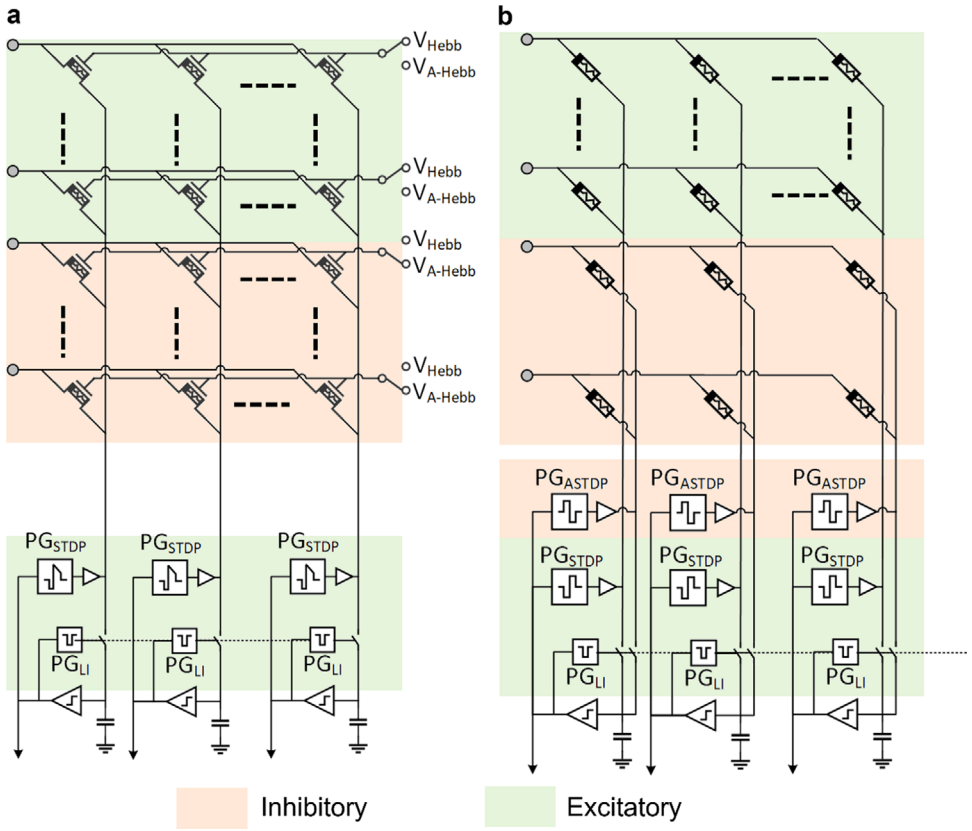

Figure 6. Implementation schemes for adaptive lateral inhibition with reconfigurable floating gate WSe_2 memtransistors. a) Circuit implementation with floating gate dual mode memtransistors for adaptive lateral inhibition. It does not require additional pulse generator components per neuron for tuning between Hebbian and anti-Hebbian learning like typical two-terminal memristors with similar Hebbian and anti-Hebbian tunable properties. b) Circuit implementation with typical Hebbian/anti-Hebbian reconfigurable memristors. Reproduced with permission.[27] Copyright 2023, John Wiley & Sons.

approaches that possess variable inhibition for different stimuli. Based on previous works, the lack of variable inhibition would result in delayed overall learning time.[9,76] With longer training time corresponding to higher power consumption, there should be a tuning parameter to the amount of inhibition received in SNN. This can be achieved in soft-WTA networks,[76,78,79] where adaptive lateral inhibition performing anti-Hebbian learning can tune the amount of inhibition received by the nonfiring neurons during a winning neuron firing event (Figure 5b). Here, a faster convergence can be achieved with more adaptive inhibition, but distinction in feature learning of neurons would be reduced. This tunability enables customizable inhibition for different environments and user needs.[9,76,80]

However, it is challenging to implement adaptive lateral inhibitions for soft-WTA due to the additional learnable parameters introduced. Current implementation typically handles anti-Hebbian connections with either an additional layer of inhibitory neurons[81,82] or additional pulse generators that generate different pulse profiles to achieve anti-Hebbian dynamics with two-terminal memristor devices[55] (Figure 2d). Leong et al.[27] proposed to use the configurable Hebbian/anti-Hebbian property in floating gate WSe_2 memtransistors to implement similar competitive spiking neural networks without the inhibitory neurons or need for additional pulse generators to tune between Hebbian/anti-Hebbian learning in each neuron. This novel implementation scheme of the lateral inhibition process is described with reference to **Figure 6a**. The same feedback pulses generated by the firing neuron can thus be multiplexed for lateral anti-Hebbian connections. This is performed by first totally discharging the neighboring nonfiring neurons, followed by charging the membrane potential of each neighboring neuron through a lateral pulse fired toward the neighboring neurons through the inhibitory synapses. The amount of charging the membrane potential of nonfiring neurons is weighted by the synaptic weight of inhibitory synapses that undergo anti-Hebbian learning. Since the same pulses can be utilized for both Hebbian and anti-Hebbian connections, the additional layer of inhibitory neurons or inverted polarity pulse generators is no longer necessary. This approach uses 27% lesser devices for anti-Hebbian pulses that would be used in previously reported architectures,[55] as shown in Figure 6b. This architecture can be applied to a feature extractor in a spiking convolutional neural network,[83] which has demonstrated at least 93.5% in recognition of MNIST handwritten digits. Furthermore, it was demonstrated that the use of soft-WTA algorithm can improve convergence rate of up to 5.7× in MNIST recognition tasks compared to conventional WTA. Thus, the use of a reconfigurable floating gate WSe_2 memtransistor as synapses can simplify and enable SNN algorithms to converge faster in visual recognition tasks.

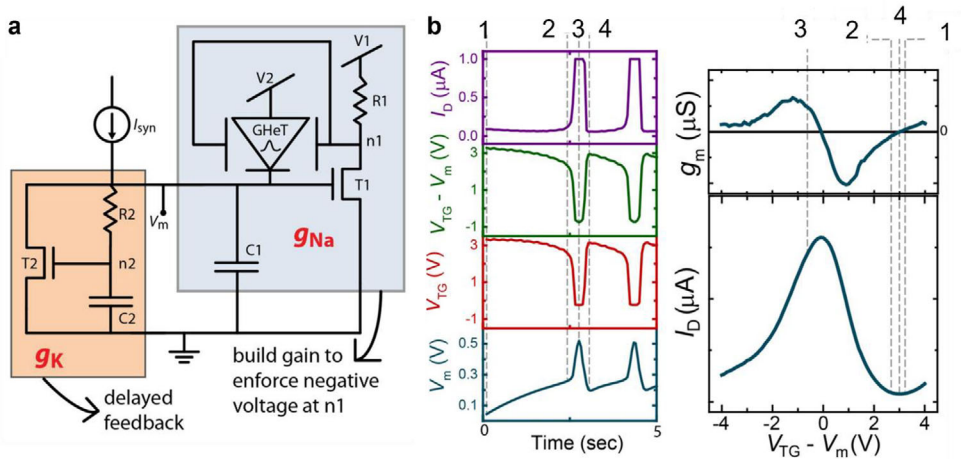

Figure 7. a) Gaussian-heterojunction-transistor (GHeT)-based spiking neuron circuit. The GHeT device along with T1–R1–C1 emulates the Na+ ion channel conductance (g_{Na}) while T2–R2–C2 emulates the K+ ion channel conductance (g_K). b) Temporal evolution of GHeT source voltage, V_m, applied top gate voltage, V_{TG}, relative gate bias, $V_{TG} - V_m$, and drain current, I_D. Transconductance and current evolution as a function of gate bias for the GHeT device.(a and b) Reproduced under the terms of the CC-BY license.[19] Copyright 2020, Megan et al., Published by Springer Nature Limited.

4.2. Multimodal Spike Encoding Using Gaussian Transistors for Circuit Component Reduction

Research has been made into exploring complex neuron dynamics to mimic high-level neuron response that potentially enables advanced neuromorphic algorithm. The Hodgkin–Huxley neuron model describes characteristics of biological ion (K+/Na+) channels and thus can capture a wide range of neuron spike dynamics. The potassium channels contribute to membrane repolarization and their efficacy are proportional to the membrane potential. The sodium channels contribute to both the rapid rise phase of the membrane potential and membrane depolarization.

However, the complexity of ion channel behaviors in the Hodgkin–Huxley (H–H) model make compact implementations with conventional mixed signal circuits a considerable challenge. Beck et al.[19] proposed to utilize the Gaussian transfer characteristics of a dual-gated Gaussian heterojunction transistor to mimic sodium channels. This can then be integrated with the potassium channel circuit and membrane charging capacitor to achieve a compact circuit implementation of the H–H model. The sodium channel dynamic is achieved by connecting the Gaussian transistor with a common-source amplifier (**Figure** 7a). The integration of the synaptic current increases the membrane potential V_m. As V_m exceeds the threshold voltage of T1, the voltage applied to the gates of the Gaussian transistor is pulled down due to the negative gain of the amplifier. This results in a negative gate-to-source voltage across the Gaussian transistor, which falls into the negative transconductance region, thus increasing the slope of V_m. As V_m continues to increase, the gate-to-source voltage continues to decrease, accessing the positive transconductance region, thus allowing the circuit to stabilize (Figure 7b). The positive transconductance, combined with the activation of the potassium channel circuit, resets the membrane potential. Hence, by traversing the Gaussian transfer curve of the Gaussian transistor, the neuron circuit is capable of the spontaneous generation of the action potential. Further, the dual gate of the Gaussian transistor modulates the peak position and amplitude of the Gaussian transfer curve. For example, the voltage difference between the top and the bottom gates modulates the peak position. This effectively mimics the sodium channel behavior. By introducing additional feedback controls from the membrane potential to the gates of the Gaussian transistor, such a neuron can demonstrate multiple biological spiking neuron responses, including phasic spiking, tonic bursting, and dampened tonic bursting. This significantly simplifies the spiking neuron implementation of the H–H model. By contrast, digital CMOS design requires complex circuits and a look-up table to mimic Gaussian responses, while analog CMOS circuits suffer from limited programmability and high bias current.

4.3. Comparison of Multifunctional Device Integrated Systems with Existing CMOS-Based Architectures

The 2DM-based devices have enabled functionalities that are typically challenging or inefficient for conventional silicon CMOS. As discussed, the reconfigurability of the floating gate memtransistors discussed in Sections 3.1 and 4.1[27] can significantly reduce memory interface logic complexity while offering flexibility between Hebbian and anti-Hebbian modes. Such novel device functionalities motivate algorithm innovation, accelerating learning and noise resilience. Similarly, complex neuron circuits, such as the Hodgkin–Huxley neuron circuit, can be demonstrated in a more compact form with Gaussian heterojunction transistor[19] discussed in Sections 3.2 and 4.2. **Table** 4 presents a benchmark of 2DM-based neuron implementations against silicon CMOS to showcase potential energy and area savings with the use of multifunction 2DM devices. The use of Gaussian transistor and I^2FETs can reduce the amount of devices required to implement the desired circuit functionality compared to their CMOS counterparts. Comparing the LIF implementations, the 2DM I^2FETs do not require a charging capacitor which provides significant savings to the circuit layout as capacitors consume a huge area that is significantly bigger than transistors.[30] For soft-WTA implementation,

Table 4. Comparison of different neuron implementations. H–H represents Hodgkin–Huxley. LIF represents leaky integrate and fire. * Values are estimated based on the reported data. "T" represents the number of transistors used for the implementation of the neuron functionality in the circuit. †Estimated from designed chips using 0.35 µm technology node.

	Neuron functions	Technology	Energy per spike per neuron	Number of device components	Neuron area [µm^2]	Charging capacitor?
Beck et al.[19]	H–H	2DM GHeT	250 nJ	1 GHeT + 2T	–	Yes
Saïghi et al.[84]	H–H	CMOS	≈850 µJ*	≈1000T*	335 435	Yes
Choi et al.[20]	LIF	2DM I^{2}FET	2 pJ	1 I^{2}FET + 6T*	–	No
Joubert et al.[85]	LIF	CMOS	2 pJ	33T	120	Yes
Leong et al.[27]	LIF/soft-WTA	2DM memtransistor	141.3 µJ	45T	177†	Yes
Diehl and Cook[82]	LIF/soft-WTA	Memristor-like	180.6 µJ	90T	324†	Yes

the use of a Hebbian/anti-Hebbian memtransistor can reduce the number of devices used by half compared to traditional CMOS implementation.

The efficient implementation of neuron and synapse dynamics can enable the realization of large-scale neuronal circuits and lead to breakthroughs in current computing systems. Here, we highlight that bioinspired computing could be the key enabler for sensor fusion,[86] as it naturally supports spatiotemporal sensor data. The spike encoding scheme provides a universal framework for data transmission and processing. Information generated by various biological sensors can thus be encoded and combined to create a more comprehensive and accurate understanding of the environment and the system itself. Sparse sensor data can be compressed with minimal data loss by such computing systems, such as bioinspired adaptive feature extractor, to be efficiently transmitted for compute-intensive processing. The event-driven processing further reduces energy overhead, which is critical for sensor data edge processing. The significant efficiency improvements provide an outlook on where bioinspired computing systems can be integrated with large generative AIs to enable embodied intelligence.[87]

5. Future Developments and Potential for Scalability of 2D Materials

With the system codesign of 2D materials, SNNs can be implemented more effectively with improved energy efficiency and reduced implementation footprint. One of the major limitations of 2D material is manufacturability and performance stability which has been an active domain of research. Current state-of-the-art fabrication method of 2DM films would be through CVD approaches, which has demonstrated polycrystalline films of high quality on a wafer level.[88] However, as CVD grown materials require high temperatures (>750 °C) to form such high-quality films, it would be a challenge to integrate multimaterial and devices into a single platform. Recently developed transfer methods from growth substrates to desired chips can introduce mechanical stress during transfer, leading to the formation of wrinkles and cracks, which disrupt charge transport and induce local strain variations that alter the material's band structure.[89] Thus, there is growing interest in other scalable approaches of fabrication of 2DM films. One of this approaches would be solution-processed methodologies, which has shown to be a transfer free, low cost approach[90] that has demonstrated potential in fabricating wafer-scale films.[91] Furthermore, such films have demonstrated the capability to vary defect concentrations through 2DM flake size and film morphology, which has influence on memristor operational voltages. As such, this method can provide a potential scalable approach of defect engineering of reconfigurable memtransistors, which will be discussed in this section.

5.1. Defect Engineering Using Solution-Processed Method for Vacancy-Modulated Memtransistors

This section demonstrates that resistive switching voltages (set/reset) in solution-processed 2D films can be effectively tuned by engineering the flake size, which modulates the edge defect density. In particular, memristors fabricated from aerosol-jet-printed MoS_2 flakes—typically <100 nm (**Figure 8a**), exhibit ultralow switching voltages (as low as 0.18 V) and operate at femtojoule energy levels (Figure 8b). These values are among the lowest reported, even when compared to conventional oxide-based memristors (Figure 8c). The key mechanism is the high density of sulfur vacancies induced by the ultrasonic atomizer, which fragments the flakes and increases edge-site exposure. The random stacking of small flakes further introduces multiple possible filament paths, facilitating easier switching. However, aerosol jet printing inherently couples flake size reduction and defect formation in a way that is difficult to control. To decouple these processes and improve reproducibility, we adopt a spin-coating approach. By optimizing the MoS_2 ink through sonication and liquid cascade centrifugation, we can gain precise control over flake size and defect density prior to deposition. Spin-coated memristors, with more ordered morphology and fewer sulfur vacancies, exhibit higher set voltages (up to 1.5 V for large flakes shown in Figure 8d–f) compared to the aerosol-printed counterparts.

We further observe a clear flake-size-dependent switching behavior (Figure 8g): smaller flakes (with higher edge-to-area ratios) yield lower switching voltages due to increased edge defect density. This is supported by atomic force microscopy and Kelvin probe force microscopy (KPFM) data (Figure 8h,i), which show lower surface potential near flake edges—indicating the presence of positively charged sulfur vacancies. Thus, by reducing flake size, one can systematically increase the defect density and lower the switching threshold.

This scalable strategy opens a pathway to engineer switching behavior in solution-processed memristors and can potentially extend to WSe_2-based memtransistors. Although spin-coated

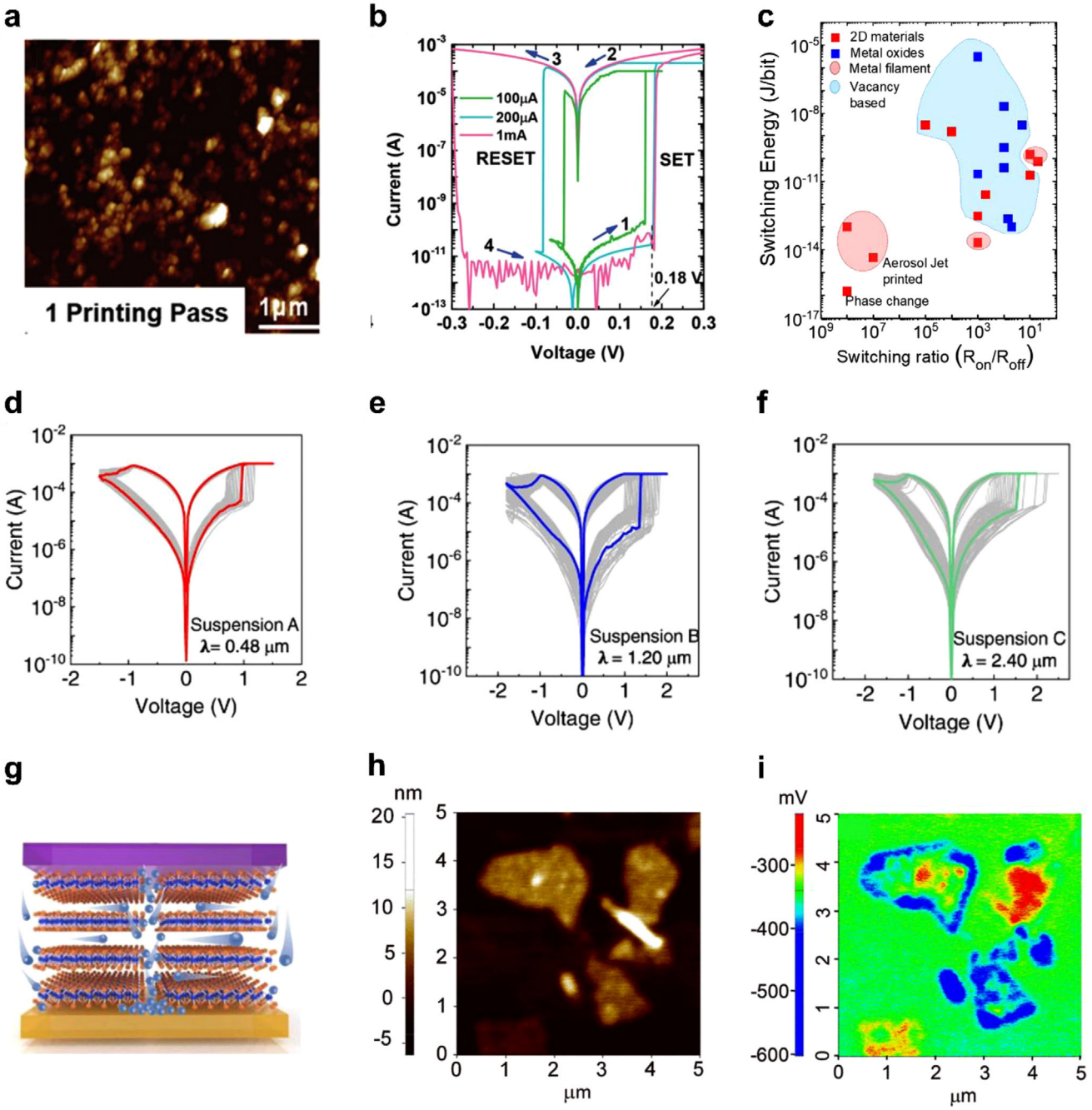

Figure 8. Resistive switching with solution-processed 2D materials formed aerosol jet printing and spin-coated films. a) Topography image of the MoS_2 flakes after printed for 1 printing pass. The flakes are small in order of <100 nm. b) I_d–V_d of the resistive random-access memory (ReRAM) made from aerosol-jet-printed MoS_2. V_{set} is extremely low at 0.18 V. c) Benchmark of resistive switching ratio against set voltage. Data points taken from refs.[26,91–106] d–f) I_d–V_d sweep showcasing the flake-size-dependent resistive switching set voltage of memristor made from spin-coated MoS_2 films. g) Schematic diagram showcasing the unique flake edge directed formation of filament in spin-coated MoS_2 memristors. h) Topographic image of the MoS_2 flakes used for spin-coated memristors. i) Kelvin probe force microscopy image of the MoS_2 flakes shown in (h). The blue regions indicate regions with lower surface potential and more n-doping from sulfur vacancies. (a and b) Reproduced under the terms of the CC-BY license.[26] Copyright 2019, Feng et al., Published by John Wiley and Sons. (d–i) Reproduced under the terms of the CC-BY license.[91] Copyright 2022, Tang et al., Published by Springer Nature Limited.

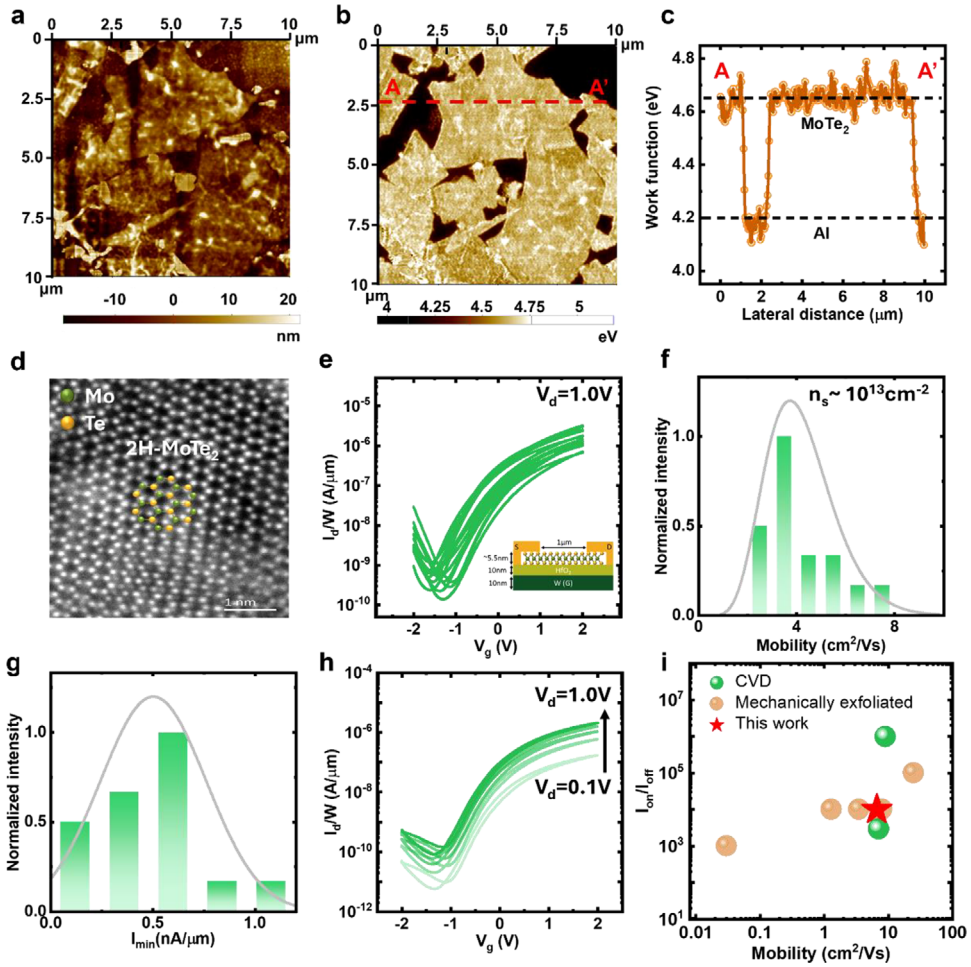

Figure 9. Assessment of quality of Langmuir–Schaeffer deposited films with high quality 2H-$MoTe_2$ flakes. a) Topographic image of the EALPE 2H-$MoTe_2$ flakes. b) Work function image of the 2H-$MoTe_2$. c) Work function profile along red line A–A' indicated in(b). The profile shows that the work function of $MoTe_2$ is ≈4.65 eV. d) Scanning transmission electron microscope image of the 2H-$MoTe_2$ flake. The image shows that the atoms are arranged in a hexagonal manner without vacancies. e) I_d–V_g curves of 2H-$MoTe_2$ devices made using Langmuir–Schaefer method. (Inset) Schematic diagram of device measured. f) Mobility extracted from 2H-$MoTe_2$ devices. g) I_{min} of the devices. h) I_d–V_g of oxygen plasma thinned device. i) Benchmark of 2H-$MoTe_2$ with other $MoTe_2$ production methods. References for benchmark are in refs.[107–110].

films may suffer from random stacking compared to Langmuir–Schaeffer assembled ones, their process simplicity and flake-size tunability make them attractive for large-area, low-voltage neuromorphic devices. In the following section, we explore how further defect control across the basal plane can complement this strategy through extrinsic engineering methods.

Preliminary data from 2H-$MoTe_2$ suggest that high quality solution-processed flakes, coupled with Langmuir–Schaeffer deposition method, can achieve comparable quality to their mechanically exfoliated counterparts. Given that 2H-$MoTe_2$ has one of the lowest vacancy formation energies among transition metal dichalcogenides (TMDCs), it is selected here as a representative 2DM. This approach may be generalized to other 2DMs, suggesting that with proper defect control, similar results could be achieved across a broader range of TMDCs. Using KPFM, we examined the 2H-$MoTe_2$ flakes that have been only sonicated for 15 min during the ink preparation process. **Figure 9a** shows the topographic image of the 2H-$MoTe_2$ flakes deposited on aluminum used for KPFM. The work function potential image (Figure 9b) shows that the flakes are highly uniform in surface potential, which means that the basal planes of the flakes are uniform across multiple flakes. By extracting along the line labeled A–A' in Figure 9b, we can determine that the work function of the 2H-$MoTe_2$ flakes is ≈4.65 eV (Figure 9c). The measured work function corresponds to the mid-gap Fermi level ≈4.6 eV of 2H-$MoTe_2$ shown in previous works,[111,112] thus indicating that exfoliated flakes are pristine with the absence of doping caused by impurities or vacancy defects in the sample. This is further supported by scanning transmission electron microscopy performed on the flakes (Figure 9d) that shows that basal plane of the flake is defect free. As mentioned earlier, Langmuir–Schaeffer approach can maximize the basal plane to edge ratio, which has one of the highest control over defect densities of the film. To evaluate the mobility of the 2H-$MoTe_2$ films, we built transistors using Langmuir–Schaeffer films. We show the set of I_d–V_g curves of the 15 devices we have measured in Figure 9e. It shows that there are device-to-device variations, which are likely to stem from the thickness variations in the flake distribution which is

Table 5. Performance metrics of memristors made from MoS_2 memristive devices. The first three processes are solution-processed methods. ECE indicates electrochemical-assisted exfoliation discussed in ref. [91] LPE stands for liquid phase exfoliation[117] which consists of only sonication exfoliation from MoS_2. CVD is chemical vapor deposition. D.C. represents direct current measurements of endurance while A.C. is pulse-based switching endurance measurements. σ is the standard deviation of the set/reset voltages of the memristive devices. V_{mean} is mean voltage of the set and reset voltages of the memristive devices.

Method	Endurance (cycles)	Retention [s]	σ/V_{mean} (Set)	σ/V_{mean} (Reset)
ECE and aerosol jet printed[26]	>100 (D.C.)	$>10^4$ @25 °C	–	–
ECE and spin-coating[91]	$>10^7$ (A.C.)	$>10^8$ @85 °C	0.197	0.185
LPE and Langmuir–Schaeffer[118]	$>10^3$ (D.C.)	$>10^4$ @25 °C	0.363	0.105
CVD[119]	>100 (D.C.)	$>10^4$ @25 °C	–	–
CVD[120]	>500 (D.C.)	$>10^7$ @50 °C	0.4	0.375
Sputtering[121]	>400 (D.C.)	$>10^4$ @25 °C	0.109	–

fundamental issue with solution-processing methodologies. The variations are reflected in the mobility distribution plot in Figure 9f which shows that the devices have an average electron field-effect mobility of 4.19 $cm^2\ V^{-1}\ s^{-1}$ with a standard deviation of 1.5 $cm^2\ V^{-1}\ s^{-1}$. However, the average off-current I_{min} of ≈0.5 nA μm^{-1} in Figure 9g is still relatively high as it is likely limited by the difficulty of controlling the channel due to the relatively thick channel.

In this aspect, a postdeposition approach was developed to control the thickness of the film with minimal additional defects introduced to the film. We perform remote oxygen plasma onto the 2H-$MoTe_2$ flakes followed by removal of the oxide layer using KOH. The thickness of 2H-$MoTe_2$ removed was ≈2.2 nm. Comparatively, there is an improvement in subthreshold swing of the device and decrease in off current, which are both strong indication of a thinning effect (Figure 9h). The oxygen-plasma-thinned device is comparable in mobility with mechanically exfoliated flaked devices and chemical-vapor-deposited 2H-$MoTe_2$ devices (Figure 9i). Notably, there are reported results[113,114] on 2H-$MoTe_2$ that have better mobilities and on/off ratios than the ones showed in the benchmark which have used encapsulation methods like hBN to reduce charge scattering and improve device mobility. However, our goal here is to show that our Langmuir–Schaeffer-flakes-formed devices are comparable in mobility and I_{on} to other mechanical-exfoliated flake devices, indicating that they are of high quality. Comparing to other solution-processed deposition methodologies, this approach enables creation of films that are determined by the intrinsic thickness of the flakes. This is difficult to achieve with other solution-processed methods. Furthermore, the liquid–liquid assembly approach in the Langmuir–Schaeffer film formation is highly compatible with industrial roll-to-roll printing technologies[115,116] that can achieve exceptionally high throughput compared to methods like inkjet or aerosol jet printing. This could be a potential avenue for exploration toward scalable industrial manufacturing of such films.

We show that solution-processed methodologies can produce memristive devices that are competitive to other known scalable approaches like CVD. To demonstrate the competitiveness of solution-processed methods in the production of memristive devices, we compared the various performance metrics of MoS_2-based memristive devices in **Table 5**. MoS_2 was chosen for comparison. From Table 5, solution-processed methods of spin-coating with electrochemically exfoliated flakes showcase the best reported cycling endurance compared to all other fabrication approaches. In addition, it has a retention of >10 years at 85 °C, showcasing environmental stability demanded by state-of-the-art memristive devices.[62]

However, the device-to-device variabilities for all reported methods are high relative to state-of-the-art reported memristive devices.[122] Other than common issues such as large unit cell size and low device count in accounting for statistics that give rise to variability observed in most memristive devices, there are potential fundamental reasons that give rise to these device variabilities pertaining to solution-processed methods. One of these reasons is the critical limitation of spin-coated films is the lack of ability to control over the morphology of the films formed, which tend to overlap each other in a random orientation.[91] Langmuir–Schaeffer methods used can propose to overcome this issue with the layer-by-layer self-assembly of solution-processed flakes. Another reason for large variability in solution-processed films would be exfoliation methods. Liquid phase exfoliation requires long processing time under aggressive ultrasonic conditions that would introduce random defects in the flakes used in the films.[123] In Figure 9, we have shown that electrochemical-assisted liquid exfoliation shows promise in minimalizing the defects on the basal plane with less aggressive exfoliation conditions. As demonstrated, Langmuir–Schaeffer films still face issues with intrinsic variances in thicknesses and flake size which give rise to poor film uniformity. The challenges of poor film uniformity can be addressed at two different stages of the film preparation process: fabrication of the 2DM ink dispersion and methods to control thickness and defects at the post-film-deposition stage.

In the ink dispersion stage, the conventional approach to control the flake size and thickness is to perform flake sorting using centrifuge approaches like liquid cascade centrifuge[124] and density gradient ultracentrifuge.[125] However, this approach does not address the fundamental issue of improving the yield of high aspect ratio flakes (i.e., thin flakes with large planar area). The yield of high aspect ratio flakes is determined by the effectiveness of the exfoliation process, which can be improved through preintercalation of the bulk 2DM crystal as shown in previous works.[126,127] The most notable approach is electrochemical intercalation,[126] which has achieved great yield of high aspect ratio flakes for high conductivity 2DMs, but fails to replicate this in low conductivity or insulating materials. Thus, exploration into other methods of intercalation, could be of interest in further investigation to achieving great yield of high aspect ratio flakes.

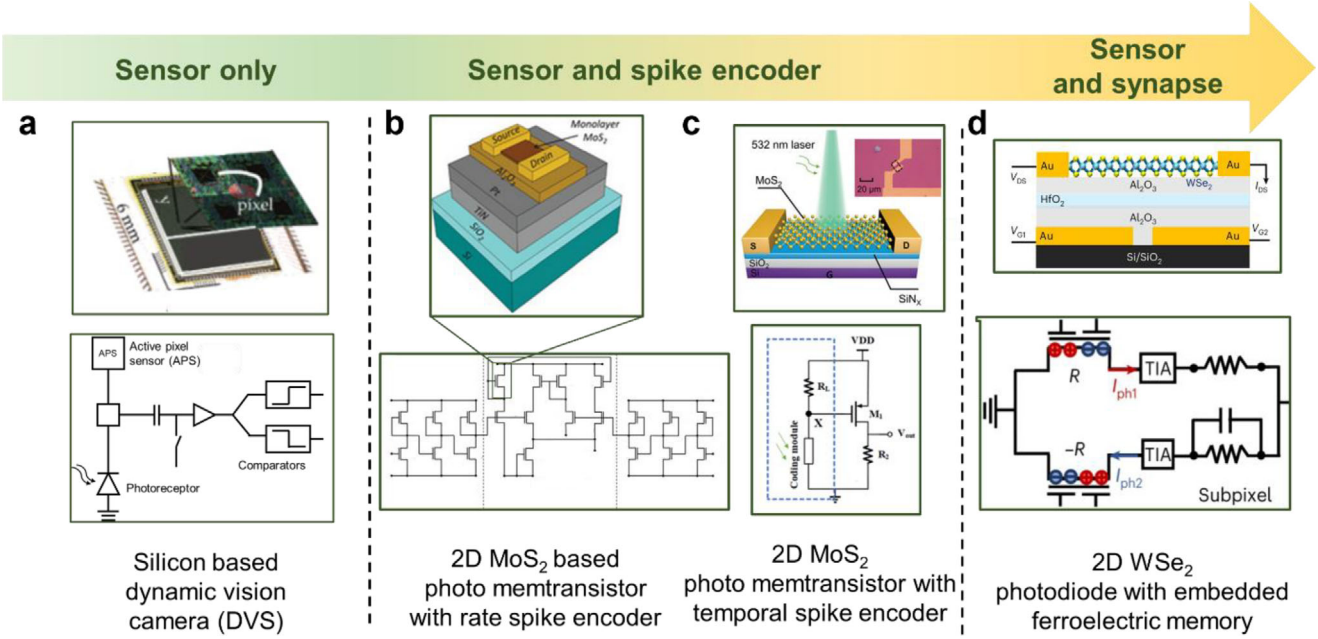

Figure 10. Multifunctional 2D-material-based sensing systems. a) Silicon-based dynamic vision camera (DVS) is an event-based sensing device that can detect motion,[132] but requires interfacing with additional analog to digital convertors and digital processors to convert data from DVS to spike for SNNs. b,c) 2D MoS_2-based photomemtransistors can leverage photogating phenomenon and be integrated into sensing-spike encoding cointegrated systems. (b) is a representation of rate encoding. (c) is a representation of temporal encoding. Image taken from refs.[133,134] d) 2D WSe_2 photodiode can be incorporated with ferroelectric memory to form a sensing device with memory. This is used in sensor-synapse cointegration system. (a) Adapted under the terms of the CC-BY license.[132] Copyright 2020, Gallego et al., Published by IEEE. (b) Reproduced with permission.[134] Copyright 2022, John Wiley & Sons. (c) Reproduced with permission.[133] Copyright 2024, John Wiley & Sons. (d) Reproduced with permission.[135] Copyright 2023, Springer.

At the post-film-deposition stage, the thickness can be controlled through development of planarization techniques for 2DMs. Currently, one known approach with such capabilities would be through self-limiting etching with focused laser thinning,[128] which has been demonstrated for a wide range of 2DMs previously. The flaw with this approach is the low throughput, which would limit its scalability to large scale manufacturing. On the other hand, high throughput industrial compatible chemical mechanical polishing technique is yet to be developed and validated for 2DMs.[129,130] The key parameter to examine for such self-limiting etch techniques would the concentration of additional defects introduced to the 2DMs during the etching process as well as the throughput of the etching process, both of which are essential for practical implementation.

Given that defects like chalcogen vacancies can naturally exist or be introduced during the self-limiting etching process, further defect passivation methods with the introduction of organic molecular dopants like 1,4-benzenedithiol (BDT)[131] can be used to passivate the chalcogen vacancies on the edges of solution-processed films. The concentration of BDT used can influence the degree of vacancies passivated as well as improve the overall mobility of the film by the connecting between flakes. Further exploration on the interaction of BDT with chalcogen vacancies in resistive switching processes of memristors and memtransistors would demand further investigation for assessment on potential scalable applications.

5.2. Cointegrated Sensors and Spike Encoders with Photomemtransistors

An important aspect to bioinspired computational systems would be data collection. For SNNs, this is particularly critical as they work asynchronously, requiring complementary event-based sensing to collect data. The collected data need to be encoded into spikes, which demand additional hardware to interface between the sensor and SNN processing chips.[136] Recent works on 2DMs have aimed to cointegrate sensors and encoding together, to minimize footprint for sensor and encoding applications.[137] This will be discussed in this section.

Event-based sensors such as DVS cameras require the use of excessively bulky capacitors and comparators to implement (**Figure 10a**). They also lack the capability to encode spikes that is necessary for processing in SNNs. Thus, to overcome these issues, recent works in 2DMs have explored the use of the photogating mechanism for in-sensor spike encoding in visual sensors (Figure 10b). The photogating mechanism requires the Fermi level of the channel material to be close to the energy trap levels in the gate dielectric which can be illustrated with the MoS_2/Al_2O_3 system shown by Subbulakshmi Radhakrishnan et al.[134] In the MoS_2/Al_2O_3 system, the hole traps can be revealed within Al_2O_3 by applying a negative gate bias, which is capable of trapping photoactivated hole carriers within the channel. The trapping of positively charged hole carriers lowers the threshold voltage of the MoS_2 photomemtransistor, enabling it to use this change in

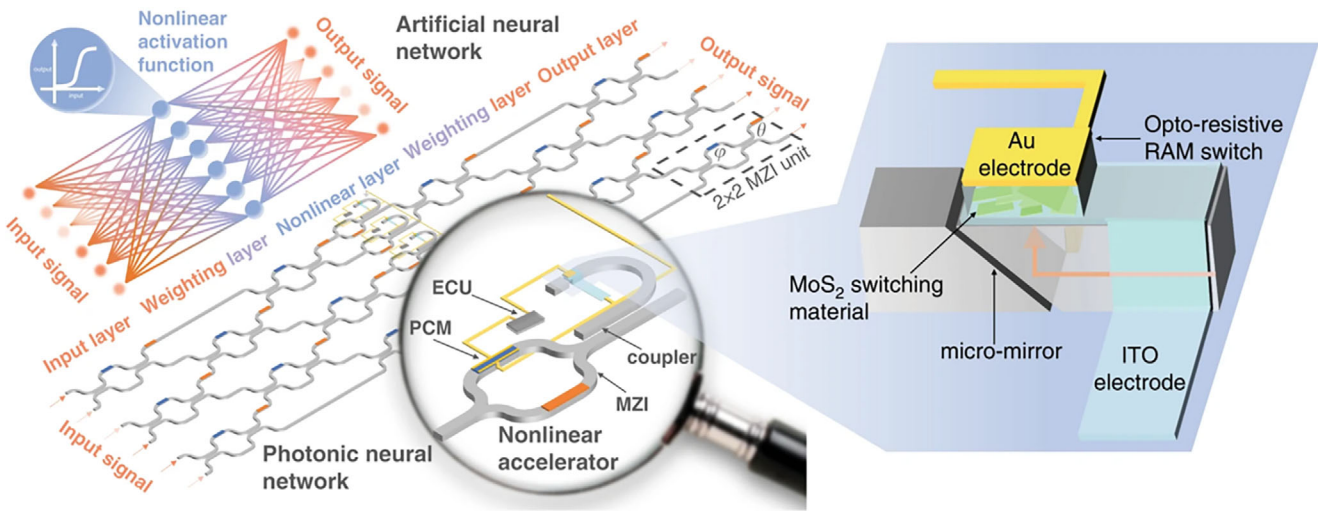

Figure 11. Photonic computation systems that leverage on novel optical resistive switching memory made using solution-processed MoS_2 flakes. The device can be used as a nonlinear activation function that is required in neuronal models in artificial neural networks and transformers. Reproduced under the terms of the CC-BY license.[145] Copyright 2022, Copyright holder, Published by Springer Nature.

conductance to reduce the time to spike in a cascaded invertor and XOR circuit. The time to spike can be tuned by pretrapping carriers with the application of a 100 ms pulse to the back-gate. This is similar to the property of the biological eye, in which the spiking signals are modulated in the event of high illumination to prevent oversaturation of signals emitted from the photoreceptors. The varying time to spike and number of spikes in different light conditions can be used either in rate coding or TTFS spike encoding used in SNNs. However, the cascaded invertors and XOR circuit takes a total of 21 memtransistors to implement, which significantly raised footprint costs. A recent work by Su et al.[133] (Figure 10c) have employed a simpler circuit consisting of the photogating device, transistor, and two resistors. Using similar photoinduced charge trapping mechanism at the channel–gate dielectric interface, the conductivity of the photogating device increases with the intensity of the light. This turns the transistor M1 (Figure 10c) from off-state to on-state, which triggers a spike generation at V_{out}. The higher the intensity of the light, the faster the electron charge trapping occurs, the faster the spike is generated. Thus, this system can be used for the TTFS coding with minimal devices. Photogating mechanisms demonstrated here are known to be extremely low power (≈nJ) compared to existing state-of-the-art systems such as LiDaR and image sensors, which use power in order of 10–10^3 J.[138] In terms of footprint savings, they have also significantly reduced the number of devices required compared to state-of-the-art DVS cameras that requires comparators and capacitors to implement.

Although the photogating effect has been demonstrated to be nonvolatile, it has limited retention time with maximum demonstration of only up to 100 s. Environmental effects have a heavy influence on photogating performance. Existing literature[139] has shown that water and oxygen molecules in the environment can undergo desorption with photogenerated holes which also introduce threshold voltage shifts. This can be easily mitigated with the passivation through deposition of dielectrics such as HfO_2 and Al_2O_3[140] which have demonstrated capabilities of isolating the channel from environmental effects. In addition, the prolonged voltage bias stress on the device under illumination can introduce additional interface trap states that will cause the device to drift from its intended operational range.[141] A potential solution to improve the reliability of the overall system is to build in redundancy in the sensing array.[142]

To reduce data movement between the sensor and SNN, there are works that built in the SNN within the sensor. In the recent work of Zhou et al.[135] (Figure 10d), a split floating-gate photomemtransistor device is demonstrated to serve both as a sensor and synapse in an in-sensor SNN. In this work, a split floating-gate for the WSe_2 is first programmed such that a p–n junction is formed between the two gates. The synaptic weights of the SNN are stored by adjusting the charge trapped in the floating gate, which tunes the degree of n-/p-doping and responsivity of the device toward incident light. By laying out the photomemtransistors in an array, the system can be trained to detect motion. Since the synaptic weights are built in to the responsivity of the device, the photogenerated signals can be directly fed to the neuron without the need for spike encoding. In this manner, in-sensor SNN can achieve 98% data reduction compared to conventional frame-based sensors, which significantly reduce computation, thus avoiding the need to spike encoders like previous works. However, translation of this approach to other forms of sensors would be a challenge as the existing mechanism leverages on the varying response of light to the p–n junctions connected in a reverse fashion that might not be replicable to other sensing elements such sound, smell, etc. Thus, there is opportunity to look into other multifunctional devices that can incorporate sensing into the synapse, attaining the ultimate footprint savings. Alternatively, other sensory systems can still consider sensor-spike encoding system that is incorporated with reconfigurable memtransistors for adaptive lateral inhibition, which does not depend on novel stimuli–synaptic interactions to perform preprocessing of data like with the case of sensor–synaptic integrated device.

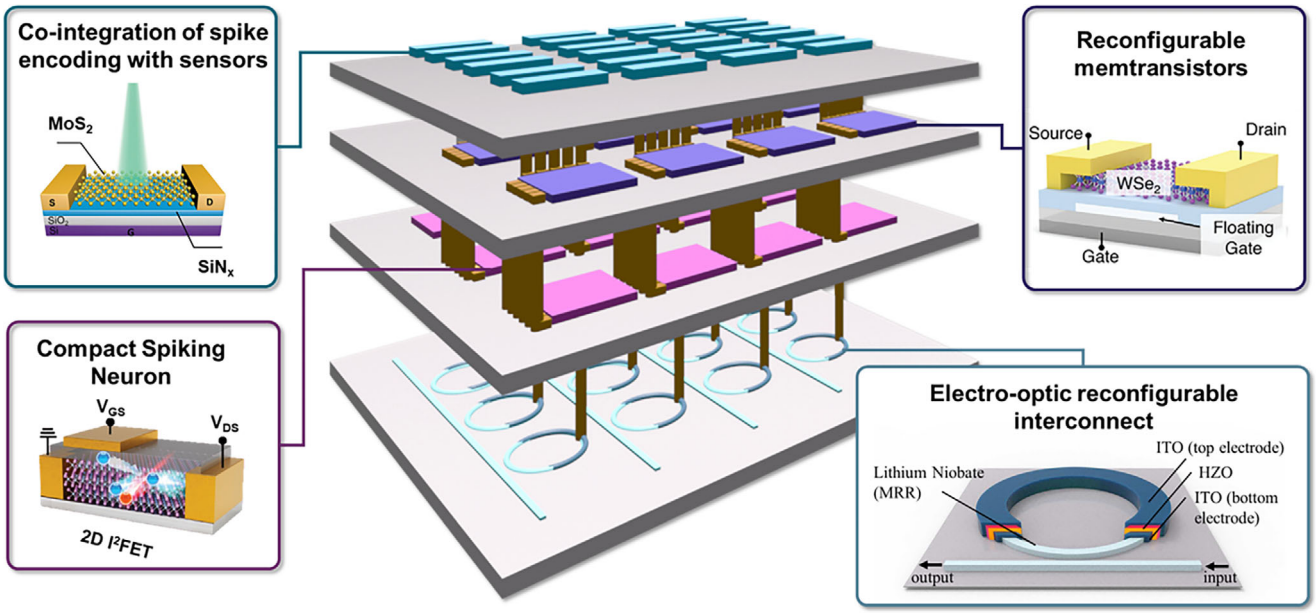

Figure 12. Monolithic 3D integration of spiking neural networks on the edge. In addition to the competitive spiking neural network (SNN), edge computation can be realized through investigation of other critical components. Sensors should generate spike-encoded signals to minimize additional components required for interfacing with SNN feature extractor. Compact neuron implementations can leverage on 2DM approaches discussed in Section 3. For multisensor communication, analysis, and recognition to be performed at a higher level without data loss through transmission, electro-optic reconfigurable interconnects can be used to convert electrical signals to optical signals and alter connections across different SNN networks. The signals will then be sent to a neural network system at higher hierarchy (generative AI or ANN) for photonic neural network computation. Figure of spike encoding and sensor adapted with permission.[133] Copyright 2024, John Wiley & Sons. Figure of reconfigurable memtransistor adapted with permission.[27] Copyright 2023, John Wiley & Sons. Figure of cointegration of spike encoding with sensors adapted from figure of compact neuron adapted with permission.[151] Copyright 2024, John Wiley & Sons.

5.3. Photonics Interconnects for Integration across Multisensor Systems

In line of the recent trend of photonic development,[143] bioinspired computing on the edge can benefit from the near lossless connections through photonic interconnects for implementation of hierarchical computing systems. In a hierarchical computing system, SNNs serve to provide data preprocessing on edge devices for advanced training of logic and inference models such as generative AI. The transmission lines to generative AI computing systems can be implemented with photonic interconnects so that multiple sensors can be connected for low loss, high speed communication. These networks can then operate across various scales, from short-range applications within a house to large-scale implementations across a city.[144]

However, the use of photonic interconnects would demand the use of electrical-to-optical and optical-to-electrical conversion devices. Electrical-to-optical devices are required to convert spiking encoded signals from SNN neurons to optical signals for lossless data transmission. Energy efficient optical to electrical signal conversion is also necessary to enable potential hybrid electrical–optical computation paradigms for generative AI. While mature technologies such as high-speed photodetectors[146] (for optical-to-electrical conversion) and modulators[147] (for electrical-to-optical conversion) exist, heterogeneous integration of novel materials with silicon photonics leveraging on unique physics can enable potential energy savings. Emerging lithium niobate and lithium tantalate can be integrated with heating elements or ferroelectrics for effective electro-optic modulation due to the Pockel's effect.[148,149] Compared to electrical interconnects interposers, it has been demonstrated that photonic lithium-niobate-based interconnect interposers leveraging on Pockel's effect can enable an improvement of 70% in energy efficiency.[149] For optical-to-electrical conversion, solution-processed 2DMs can serve as potential game changers with their strong light–matter interaction that can be coupled with memory elements to enable energy-efficient hybrid electrical optical computation systems. A spin-coated MoS_2-based optoelectronic random-access memory was demonstrated as a photon-dependent, low power resistance-switching device that enables a nonlinear activation function[145] that is necessary for computing thresholding elements in neurons of both ANNs and generative AI models such as transformers[150] (**Figure 11**). As such, design of edge systems could leverage on the strengths of photonics to deliver a heterogeneous system that is low power and can serve multiple sensory nodes over long distances.

5.4. Outlook and Perspective

A bioinspired system on the edge consists of sensor systems with spike encoding capabilities, multifunctional synaptic memtransistors, compact spiking neurons, and photonic components can be monolithic 3D integrated onto a chip (**Figure 12**). This will result in the ultimate reduction of implementation footprint as the different layers are now stacked instead of being laid

out in a planar manner. The use of 2DM-based multifunctional devices will enable area and energy savings across various components of the chip. First, cointegrated sensors with spike encoding capabilities will avoid the need of bulky analog to digital convertors and digital processors to convert data into spikes.[133,134,136] Second, reconfigurable floating gate memtransistors can be used to enable excitatory and inhibitory synapses for near edge feature extraction using adaptive lateral inhibition.[27] The use of reconfigurable memtransistors can enable reduction in device count used for pulse generation in the neurons compared to implementations that use two-terminal memristors. Although the existing work demonstrated high power operations, proposed solution-processed methods of fabrication of the 2DMs can enable better control over defect distributions to lower operational voltages and power. Explorations into other reconfigurable memtransistors can also be made to perform other synaptic functions. Third, the use of impact ionization-based single device neuron that replaces bulky capacitors and comparators in neurons can also lead to further savings in terms of area.[24,30] The 2DM-implemented version of the impact-ionization-based single device neuron is 6 times more energy efficient per spike compared to their silicon counterparts. Considering that other neuronal models exist, such as the Hodgkin–Huxley model, compact implementations that integrate multifunctional 2DM devices like Gaussian transistor[19] can also be considered. Thus, it is demonstrated that the use of 2DM-based multifunctional devices can present significant savings in terms of area and energy. Rather than introducing 2DM for all elements of neuromorphic computing, 2DMs should only be used if it does bring significant advantages to existing state-of-the-art as shown in this article.

As bioinspired computing like SNNs are considered for on edge computing, logic and reasoning is performed at a higher computational hierarchy by more complex models like generative AI. For lossless transmission across multiple SNN systems, the use of photonic interconnects provides an improvement in energy by 70% compared to pure electrical interposers.[149] Future works should examine proper integration techniques for these devices to truly realize a low power, multisensory system with bioinspired computing. As such, we briefly introduce some challenges toward monolithic 3D integration (M3D) as well as 2DM-based systems that can serve as case studies in building M3D-based bioinspired computing systems. The key limitations of current M3D include chip overheating, intertier signal delay, and electrical interference between tiers. First, the thermal challenge in 3D ICs occur across various scales from device to chip level. At the device level, localized hot spots are generated by high drain electric fields in FETs that create energetic electrons that generate optical phonons with low velocity. By stacking these devices, the localized hot spots spread across tiers, which results in both performance and reliability degradation. Additionally, the reduced dimensions of interconnects in M3D increases self-heating which impedes the ability to further scaling. Second, Jiang et al. highlighted detailed simulation on the effect of intertier signal delay caused by lengthy monolithic intertier vias (MIVs). In their study, the simulated FO4 driver–interconnect–load circuits in 7 nm technology highlights greater intertier than intratier signal delays due to the high parasitics of MIVs.[152] Finally, the thin dielectric layers in the monolithic stack fails to shield signal and interference from different layers, resulting in noise and significant cross-talk.

A review of previous literature suggests that M3D with 2D transistors offers potential in overcoming these challenges with its ultrathin profile which leads to lower intrinsic thermal resistance, excellent thermal conductivity, reduced intertier MIV length. With lower thermal resistance in 2DM transistor, the effect of local hotspots is reduced.[153] This is evident that the interfacial thermal conductance of 2DM MoS_2/SiO_2 of 15 MW m^{-2} K^{-1} is 10 times lower than that of silicon.[154,155] Additionally, the excellent thermal conductivity of 2DM enables development of small dimension interconnect for M3D with higher integration density without significant self-heating. For instance, intercalation-doped multilayer graphene nanoribbons (≈12 nm) exhibit superior in-plane thermal conductivity compared to Cu of the same thickness, which can alleviate thermal gradients and enhance thermal integrity in global interconnects.[156] The challenge of signal delay can also be addressed by the reduction in MIV length as the ultrathin bodies of 2DM can form tiers with smaller heights. In summary, 2DM-based devices and interconnects show promise in mitigating the challenges faced in M3D. With further development, we envision that 2DMs can deliver an electronic–photonic M3D system that can realize a truly energy efficient, bioinspired computing system for edge computation.

Acknowledgements

This work was supported in part by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 (Grant No. MOE-T2EP50123-0033), the Semiconductor Research Corporation (Contract No. 2023-LM-3145), as well as NRF's Medium Sized Centre: Singapore Hybrid-Integrated Next-Generation μ-Electronics (SHINE) Centre funding program and the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Science Program.

Conflict of Interest

The authors declare no conflict of interests.

Keywords

competitive learning, in-sensor computing, memristors, memtransistors, monolithic 3D integration, neuromorphic computing, photonics

Received: June 3, 2025
Revised: September 16, 2025
Published online:

[1] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[2] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, et al., in NIPS '20: Proc. 34th Int. Conf. Neural Information Processing Systems, Curran Associates, Inc, New York, United States of America, **2020**, pp. 1877–1901.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, arXiv preprint 2021, arXiv:2010.11929.
[4] N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, R. Boyle, P. Cantin, C. Chao, C. Clark, J. Coriell, M. Daley, M. Dau, J. Dean, B. Gelb, T. V. Ghaemmaghami, R. Gottipati, W. Gulland, R. Hagmann, C. R. Ho, D. Hogberg, J. Hu, R. Hundt, D. Hurt, J. Ibarz, A. Jaffey, et al., in *44th Int. Symp. Computer Architecture (ISCA)*, ACM, Toronto, Canada **2017**, pp. 1–17.
[5] R. Raina, A. Madhavan, A. Y. Ng, in *Proc. 26th Annual Int. Conf. Machine Learning*, ACM, Montreal Quebec Canada **2009**, pp. 873–880.
[6] J. Choquette, *IEEE Micro* **2023**, *43*, 9.
[7] S. Katoch, S. S. Chauhan, V. Kumar, *Multimedia Tools Appl.* **2021**, *80*, 8091.
[8] F. S. Abu-Mouti, M. E. El-Hawary, in *2012 IEEE Int. Systems Conf. SysCon 2012*, IEEE, Piscataway, NJ **2012**, pp. 1–6.
[9] J. D. Nunes, M. Carvalho, D. Carneiro, J. S. Cardoso, *IEEE Access* **2022**, *10*, 60738.
[10] M. Mayberry, in *2020 IEEE Symp. VLSI Technology*, IEEE, Piscataway, NJ **2020**, pp. 1–4.
[11] C. Pehle, S. Billaudelle, B. Cramer, J. Kaiser, K. Schreiber, Y. Stradmann, J. Weis, A. Leibfried, E. Müller, J. Schemmel, *Front. Neurosci.* **2022**, *16*, 795876.
[12] E. Painkras, L. A. Plana, J. Garside, S. Temple, F. Galluppi, C. Patterson, D. R. Lester, A. D. Brown, S. B. Furber, *IEEE J. Solid-State Circuits* **2013**, *48*, 1943.
[13] G. Orchard, E. P. Frady, D. B. D. Rubin, S. Sanborn, S. B. Shrestha, F. T. Sommer, M. Davies, in *2021 IEEE Workshop Signal Processing System (SiPS)*, IEEE, Piscataway, NJ **2021**, pp. 254–259.
[14] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y.-H. Weng, A. Wild, Y. Yang, H. Wang, *IEEE Micro* **2018**, *38*, 82.
[15] M. Davies, A. Wild, G. Orchard, Y. Sandamirskaya, G. A. F. Guerra, P. Joshi, P. Plank, S. R. Risbud, *Proc. IEEE* **2021**, *109*, 911.
[16] M. Isik, K. Tiwari, M. B. Eryilmaz, I. C. Dikmen, in *2024 IEEE High Performance Extreme Computing Conf. (HPEC)*, IEEE, Piscataway, NJ **2024**, pp. 1–7.
[17] S. Li, M.-E. Pam, Y. Li, L. Chen, Y.-C. Chien, X. Fong, D. Chi, K.-W. Ang, *Adv. Mater.* **2022**, *34*, 2103376.
[18] W. Zhang, B. Gao, J. Tang, P. Yao, S. Yu, M.-F. Chang, H.-J. Yoo, H. Qian, H. Wu, *Nat. Electron.* **2020**, *3*, 371.
[19] M. E. Beck, A. Shylendra, V. K. Sangwan, S. Guo, W. A. Gaviria Rojas, H. Yoo, H. Bergeron, K. Su, A. R. Trivedi, M. C. Hersam, *Nat. Commun.* **2020**, *11*, 1565.
[20] H. Choi, S. Baek, H. Jung, T. Kang, S. Lee, J. Jeon, B. C. Jang, S. Lee, *Adv. Mater.* **2024**, *37*, 2406970.
[21] Y. Choi, S. Jeong, H. Jeong, S. Han, J. Ko, S. E. Yu, Z. Xu, M. S. Chae, M. Son, Y. Meng, S. Xu, J.-H. Kang, S. Mun, S.-H. Bae, *npj Unconv. Comput.* **2025**, *2*, 8.
[22] W. Guo, M. E. Fouda, A. M. Eltawil, K. N. Salama, *Front. Neurosci.* **2021**, *15*, 638474.
[23] K. Khare, R. Kar, D. Mandal, S. P. Ghoshal, in *2014 Int. Conf. Communication Signal Processing*, IEEE, Piscataway, NJ **2014**, pp. 523–527.
[24] S. Pazos, K. Zhu, M. A. Villena, O. Alharbi, W. Zheng, Y. Shen, Y. Yuan, Y. Ping, M. Lanza, *Nature* **2025**, *640*, 69.
[25] D. Kuzum, R. G. D. Jeyasingh, B. Lee, H.-S. P. Wong, *Nano Lett.* **2012**, *12*, 2179.
[26] X. Feng, Y. Li, L. Wang, S. Chen, Z. G. Yu, W. C. Tan, N. Macadam, G. Hu, L. Huang, L. Chen, X. Gong, D. Chi, T. Hasan, A. V.-Y. Thean, Y.-W. Zhang, K.-W. Ang, *Adv. Electron. Mater.* **2019**, *5*, 1900740.
[27] J. F. Leong, Z. Fang, M. Sivan, J. Pan, B. Tang, E. Zamburg, A. V.-Y. Thean, *Adv. Funct. Mater.* **2023**, *33*, 2302949.
[28] V. K. Sangwan, H.-S. Lee, H. Bergeron, I. Balla, M. E. Beck, K.-S. Chen, M. C. Hersam, *Nature* **2018**, *554*, 500.
[29] L. Wang, W. Liao, S. L. Wong, Z. G. Yu, S. Li, Y.-F. Lim, X. Feng, W. C. Tan, X. Huang, L. Chen, L. Liu, J. Chen, X. Gong, C. Zhu, X. Liu, Y.-W. Zhang, D. Chi, K.-W. Ang, *Adv. Funct. Mater.* **2019**, *29*, 1901106.
[30] J.-K. Han, S.-Y. Yun, S.-W. Lee, J.-M. Yu, Y.-K. Choi, *Adv. Funct. Mater.* **2022**, *32*, 2204102.
[31] L. Shi, *Nat. Nanotechnol.* **2024**, *19*, 145.
[32] Y. Zhang, Y. Wang, C. Gao, Z. Ni, X. Zhang, W. Hu, H. Dong, *Chem. Soc. Rev.* **2023**, *52*, 1331.
[33] S. G. Sarwat, B. Kersting, T. Moraitis, V. P. Jonnalagadda, A. Sebastian, *Nat. Nanotechnol.* **2022**, *17*, 507.
[34] S. Mirabella, D. De Salvador, E. Napolitani, E. Bruno, F. Priolo, *J. Appl. Phys.* **2013**, *113*, 031101.
[35] J. S. Christensen, H. H. Radamson, A. Y. Kuznetsov, B. G. Svensson, *Appl. Phys. Lett.* **2003**, *82*, 2254.
[36] S. Kim, S. Choi, W. Lu, *ACS Nano* **2014**, *8*, 2369.
[37] M. Lederer, R. Olivo, D. Lehninger, S. Abdulazhanov, T. Kämpfe, S. Kirbach, C. Mart, K. Seidel, L. M. Eng, *Phys. Status Solidi RRL* **2021**, *15*, 2100086.
[38] A. Sünbül, D. Lehninger, M. Lederer, H. Mähne, R. Hoffmann, K. Bernert, S. Thiem, F. Schöne, M. Döllgast, N. Haufe, L. Roy, T. Kämpfe, K. Seidel, L. M. Eng, *Phys. Status Solidi A* **2023**, *220*, 2300067.
[39] K. Kohary, C. D. Wright, *Phys. Status Solidi B* **2013**, *250*, 944.
[40] S. Senkader, C. D. Wright, *J. Appl. Phys.* **2004**, *95*, 504.
[41] M. Sivan, J. F. Leong, J. Ghosh, B. Tang, J. Pan, E. Zamburg, A. V.-Y. Thean, *ACS Nano* **2022**, *16*, 14308.
[42] N. Fukata, A. Kasuya, M. Suezawa, *Jpn. J. Appl. Phys.* **2001**, *40*, L854.
[43] H. Song, G. Kang, Y. Kang, S. Han, *Phys. Status Solidi B* **2019**, *256*, 1800486.
[44] W.-B. Zhang, N. Yu, W.-Y. Yu, B.-Y. Tang, *Eur. Phys. J. B* **2008**, *64*, 153.
[45] A. Dabral, A. K. A. Lu, D. Chiappe, M. Houssa, G. Pourtois, *Phys. Chem. Chem. Phys.* **2019**, *21*, 1089.
[46] J. L. Hudgins, G. S. Simin, E. Santi, M. A. Khan, *IEEE Trans. Power Electron.* **2003**, *18*, 907.
[47] Y. Liu, X. Duan, H.-J. Shin, S. Park, Y. Huang, X. Duan, *Nature* **2021**, *591*, 43.
[48] Y. Yu, N. Lv, D. Zhang, Y. Wei, M. Wang, *IEEE Electron Device Lett.* **2021**, *42*, 1480.
[49] J. Chen, W. Zhang, L. Wang, G. Yu, *Adv. Mater.* **2023**, *35*, 2210772.
[50] K. S. Siegert, F. R. L. Lange, E. R. Sittner, H. Volker, C. Schlockermann, T. Siegrist, M. Wuttig, *Rep. Prog. Phys.* **2014**, *78*, 013001.
[51] J. Yuan, S. E. Liu, A. Shylendra, W. A. Gaviria Rojas, S. Guo, H. Bergeron, S. Li, H.-S. Lee, S. Nasrin, V. K. Sangwan, A. R. Trivedi, M. C. Hersam, *Nano Lett.* **2021**, *21*, 6432.
[52] L. Müller, S.-Y. Rhim, V. Sivanesan, D. Wang, S. Hietzschold, P. Reiser, E. Mankel, S. Beck, S. Barlow, S. R. Marder, A. Pucci, W. Kowalsky, R. Lovrincic, *Adv. Mater.* **2017**, *29*, 1701466.
[53] C. Leblanc, S. Song, D. Jariwala, *Curr. Opin. Solid State Mater. Sci.* **2024**, *32*, 101178.
[54] X. Pan, T. P. Ma, *Appl. Phys. Lett.* **2011**, *99*, 013505.
[55] C.-C. Chang, P.-C. Chen, B. Hudec, P.-T. Liu, T.-H. Hou, in *2018 IEEE Int. Electron Devices Meeting, IEDM 2018*, IEEE, Piscataway, NJ **2018**, pp. 15.5.1–15.5.4.
[56] J. Jang, S. Gi, I. Yeo, S. Choi, S. Jang, S. Ham, B. Lee, G. Wang, *Adv. Sci.* **2022**, *9*, 2201117.
[57] M. Figueroa, E. Matamala, G. Carvajal, S. Bridges, in *IEEE Computer Society Annual Symp. Emerging VLSI Technologies Architectures (ISVLSI06)*, IEEE, Piscataway, NJ **2006**, p. 6.

[58] H. Tan, S. Majumdar, Q. Qin, J. Lahtinen, S. van Dijken, *Adv. Intell. Syst.* **2019**, *1*, 1900036.
[59] Y. Li, Y. Zhong, L. Xu, J. Zhang, X. Xu, H. Sun, X. Miao, *Sci. Rep.* **2013**, *3*, 1619.
[60] A. C. Khot, T. D. Dongale, K. A. Nirmal, J. H. Sung, H. J. Lee, R. D. Nikam, T. G. Kim, *ACS Appl. Mater. Interfaces* **2022**, *14*, 10546.
[61] Y. Zhou, Y. Wang, F. Zhuge, J. Guo, S. Ma, J. Wang, Z. Tang, Y. Li, X. Miao, Y. He, Y. Chai, *Adv. Mater.* **2022**, *34*, 2107754.
[62] M. Lanza, H.-S. P. Wong, E. Pop, D. Ielmini, D. Strukov, B. C. Regan, L. Larcher, M. A. Villena, J. J. Yang, L. Goux, A. Belmonte, Y. Yang, F. M. Puglisi, J. Kang, B. Magyari-Köpe, E. Yalon, A. Kenyon, M. Buckwell, A. Mehonic, A. Shluger, H. Li, T.-H. Hou, B. Hudec, D. Akinwande, R. Ge, S. Ambrogio, J. B. Roldan, E. Miranda, J. Suñe, K. L. Pey, et al., *Adv. Electron. Mater.* **2019**, *5*, 1800143.
[63] X. Feng, S. Li, S. L. Wong, S. Tong, L. Chen, P. Zhang, L. Wang, X. Fong, D. Chi, K.-W. Ang, *ACS Nano* **2021**, *15*, 1764.
[64] S. E. Liu, T. T. Zeng, R. Wu, V. K. Sangwan, M. C. Hersam, *J. Mater. Res.* **2024**, *39*, 1463.
[65] H.-S. Lee, V. K. Sangwan, W. A. G. Rojas, H. Bergeron, H. Y. Jeong, J. Yuan, K. Su, M. C. Hersam, *Adv. Funct. Mater.* **2020**, *30*, 2003683.
[66] W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, H. Qian, *IEEE Electron Device Lett.* **2017**, *38*, 1019.
[67] Y. Jiang, K. Zhang, K. Hu, Y. Zhang, A. Liang, Z. Song, S. Song, F. Wang, *Mater. Sci. Semicond. Process.* **2021**, *136*, 106131.
[68] L. Wang, L. Wang, K.-W. Ang, A. V.-Y. Thean, G. Liang, *IEEE Trans. Electron Devices* **2019**, *66*, 4092.
[69] L. Wang, L. Wang, K.-W. Ang, A. V.-Y. Thean, G. Liang, in *2018 IEEE Int. Electron Devices Meeting, IEDM 2018*, IEEE, Piscataway, NJ **2018**, pp. 24.5.1–24.5.4.
[70] B. Spetzler, V. K. Sangwan, M. C. Hersam, M. Ziegler, *npj 2D Mater. Appl.* **2025**, *9*, 17.
[71] K. Kang, T. Shibata, *IEEE Trans. Circuits Syst.* **2010**, *57*, 1513.
[72] K. Zhu, S. Pazos, F. Aguirre, Y. Shen, Y. Yuan, W. Zheng, O. Alharbi, M. A. Villena, B. Fang, X. Li, A. Milozzi, M. Farronato, M. Muñoz-Rojo, T. Wang, R. Li, H. Fariborzi, J. B. Roldan, G. Benstetter, X. Zhang, H. N. Alshareef, T. Grasser, H. Wu, D. Ielmini, M. Lanza, *Nature* **2023**, *618*, 57.
[73] G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. Saïghi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, K. Boahen, *Front. Neurosci.* **2011**, *5*, 00073.
[74] J.-K. Han, J. Oh, G.-J. Yun, D. Yoo, M.-S. Kim, J.-M. Yu, S.-Y. Choi, Y.-K. Choi, *Sci. Adv.* **2021**, *7*, abg8836.
[75] Y. Wakayama, R. Hayakawa, *Adv. Funct. Mater.* **2020**, *30*, 1903724.
[76] S. Ambrogio, N. Ciocchini, M. Laudato, V. Milo, A. Pirovano, P. Fantini, D. Ielmini, *Front. Neurosci.* **2016**, *10*, 00056.
[77] G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, D. Ielmini, *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2018**, *8*, 77.
[78] Y. Chen, *Front. Comput. Neurosci.* **2017**, *11*, 00020.
[79] J. Binas, U. Rutishauser, G. Indiveri, M. Pfeiffer, *Front. Comput. Neurosci.* **2014**, *8*, 00068.
[80] D. Zhao, Y. Zeng, Y. Li, *Neural Networks* **2022**, *154*, 68.
[81] X. Zhang, J. Lu, Z. Wang, R. Wang, J. Wei, T. Shi, C. Dou, Z. Wu, J. Zhu, D. Shang, G. Xing, M. Chan, Q. Liu, M. Liu, *Sci. Bull.* **2021**, *66*, 1624.
[82] P. U. Diehl, M. Cook, *Front. Comput. Neurosci.* **2015**, *9*, 00099.
[83] C. Lee, P. Panda, G. Srinivasan, K. Roy, *Front. Neurosci.* **2018**, *12*, 00435.
[84] S. Saïghi, Y. Bornat, J. Tomas, G. Le Masson, S. Renaud, *IEEE Trans. Biomed. Circuits Syst.* **2011**, *5*, 3.
[85] A. Joubert, B. Belhadj, O. Temam, R. Héliot, in *2012 Int. Joint Conf. Neural Networks (IJCNN)*, IEEE, Piscataway, NJ **2012**, pp. 1–5.
[86] E. Ceolini, C. Frenkel, S. B. Shrestha, G. Taverni, L. Khacef, M. Payvand, E. Donati, *Front. Neurosci.* **2020**, *14*, 00637.
[87] C. Bartolozzi, G. Indiveri, E. Donati, *Nat. Commun.* **2022**, *13*, 1024.
[88] Y. Wan, E. Li, Z. Yu, J.-K. Huang, M.-Y. Li, A.-S. Chou, Y.-T. Lee, C.-J. Lee, H.-C. Hsu, Q. Zhan, A. Aljarb, J.-H. Fu, S.-P. Chiu, X. Wang, J.-J. Lin, Y.-P. Chiu, W.-H. Chang, H. Wang, Y. Shi, N. Lin, Y. Cheng, V. Tung, L.-J. Li, *Nat. Commun.* **2022**, *13*, 4149.
[89] M. Nakatani, S. Fukamachi, P. Solís-Fernández, S. Honda, K. Kawahara, Y. Tsuji, Y. Sumiya, M. Kuroki, K. Li, Q. Liu, Y.-C. Lin, A. Uchida, S. Oyama, H. G. Ji, K. Okada, K. Suenaga, Y. Kawano, K. Yoshizawa, A. Yasui, H. Ago, *Nat. Electron.* **2024**, *7*, 119.
[90] G. Hu, J. Kang, L. W. T. Ng, X. Zhu, R. C. T. Howe, C. G. Jones, M. C. Hersam, T. Hasan, *Chem. Soc. Rev.* **2018**, *47*, 3265.
[91] B. Tang, H. Veluri, Y. Li, Z. G. Yu, M. Waqar, J. F. Leong, M. Sivan, E. Zamburg, Y.-W. Zhang, J. Wang, A. V.-Y. Thean, *Nat. Commun.* **2022**, *13*, 3037.
[92] M. Sivan, Y. Li, H. Veluri, Y. Zhao, B. Tang, X. Wang, E. Zamburg, J. F. Leong, J. X. Niu, U. Chand, A. V.-Y. Thean, *Nat. Commun.* **2019**, *10*, 5201.
[93] X. Yan, Q. Zhao, A. P. Chen, J. Zhao, Z. Zhou, J. Wang, H. Wang, L. Zhang, X. Li, Z. Xiao, K. Wang, C. Qin, G. Wang, Y. Pei, H. Li, D. Ren, J. Chen, Q. Liu, *Small* **2019**, *15*, 1901423.
[94] T. Yu, Z. Zhao, H. Jiang, Z. Weng, Y. Fang, Y. Yu, W. Lei, S. B. Shafe, M. N. Mohtar, *Mater. Today Chem.* **2023**, *27*, 101268.
[95] X. F. Lu, Y. Zhang, N. Wang, S. Luo, K. Peng, L. Wang, H. Chen, W. Gao, X. H. Chen, Y. Bao, G. Liang, K. P. Loh, *Nano Lett.* **2021**, *21*, 8800.
[96] W. Hou, A. Azizimanesh, A. Dey, Y. Yang, W. Wang, C. Shao, H. Wu, H. Askari, S. Singh, S. M. Wu, *Nat. Electron.* **2024**, *7*, 8.
[97] F. Zhang, H. Zhang, P. R. Shrestha, Y. Zhu, K. Maize, S. Krylyuk, A. Shakouri, J. P. Campbell, K. P. Cheung, L. A. Bendersky, A. V. Davydov, J. Appenzeller, in *2018 IEEE Int. Electron Devices Meeting IEDM 2018*, IEEE, Piscataway, NJ **2018**, pp. 22.7.1–22.7.4.
[98] S. Lee, J. Sohn, Z. Jiang, H.-Y. Chen, H.-S. P. Wong, *Nat. Commun.* **2015**, *6*, 8407.
[99] X. Yan, Z. Zhou, B. Ding, J. Zhao, Y. Zhang, *J. Mater. Chem. C* **2017**, *5*, 2259.
[100] L. Cheng, M.-Y. Zhang, Y. Li, Y.-X. Zhou, Z.-R. Wang, S.-Y. Hu, S.-B. Long, M. Liu, X.-S. Miao, *J. Phys. Appl. Phys.* **2017**, *50*, 505102.
[101] B. Govoreanu, G. S. Kar, Y.-Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D. J. Wouters, J. A. Kittl, M. Jurczak, in *2011 Int. Electron Devices Meeting*, IEEE, Piscataway, NJ **2011**, pp. 31.6.1–31.6.4.
[102] X. Zhao, S. Liu, J. Niu, L. Liao, Q. Liu, X. Xiao, H. Lv, S. Long, W. Banerjee, W. Li, S. Si, M. Liu, *Small* **2017**, *13*, 1603948.
[103] Y. L. Song, Y. Liu, Y. L. Wang, M. Wang, X. P. Tian, L. M. Yang, Y. Y. Lin, *IEEE Electron Device Lett.* **2011**, *32*, 1439.
[104] A. A. Bessonov, M. N. Kirikova, D. I. Petukhov, M. Allen, T. Ryhänen, M. J. A. Bailey, *Nat. Mater.* **2015**, *14*, 199.
[105] T. Breuer, L. Nielen, B. Roesgen, R. Waser, V. Rana, E. Linn, *Sci. Rep.* **2016**, *6*, 23967.
[106] P. Huang, J. Kang, Y. Zhao, S. Chen, R. Han, Z. Zhou, Z. Chen, W. Ma, M. Li, L. Liu, X. Liu, *Adv. Mater.* **2016**, *28*, 9758.
[107] Y.-F. Lin, Y. Xu, S.-T. Wang, S.-L. Li, M. Yamamoto, A. Aparecido-Ferreira, W. Li, H. Sun, S. Nakaharai, W.-B. Jian, K. Ueno, K. Tsukagoshi, *Adv. Mater.* **2014**, *26*, 3263.
[108] J. Chen, Z. Feng, S. Fan, S. Shi, Y. Yue, W. Shen, Y. Xie, E. Wu, C. Sun, J. Liu, H. Zhang, W. Pang, D. Sun, W. Feng, Y. Feng, S. Wu, D. Zhang, *ACS Appl. Mater. Interfaces* **2017**, *9*, 30107.
[109] C. Kim, I. Moon, D. Lee, M. S. Choi, F. Ahmed, S. Nam, Y. Cho, H.-J. Shin, S. Park, W. J. Yoo, *ACS Nano* **2017**, *11*, 1588.

[110] Y.-M. Chang, S.-H. Yang, C.-Y. Lin, C.-H. Chen, C.-H. Lien, W.-B. Jian, K. Ueno, Y.-W. Suen, K. Tsukagoshi, Y.-F. Lin, *Adv. Mater.* **2018**, *30*, 1706995.
[111] R. Schlaf, O. Lang, C. Pettenkofer, W. Jaegermann, *J. Appl. Phys.* **1999**, *85*, 2732.
[112] Y. Balaji, Q. Smets, Á. Szabo, M. Mascaro, D. Lin, I. Asselberghs, I. Radu, M. Luisier, G. Groeseneken, *Adv. Funct. Mater.* **2020**, *30*, 1905970.
[113] M. J. Mleczko, A. C. Yu, C. M. Smyth, V. Chen, Y. C. Shin, S. Chatterjee, Y.-C. Tsai, Y. Nishi, R. M. Wallace, E. Pop, *Nano Lett.* **2019**, *19*, 6352.
[114] Y. J. Park, A. K. Katiyar, A. T. Hoang, J.-H. Ahn, *Small* **2019**, *15*, 1901772.
[115] R. A. Wells, H. Johnson, C. R. Lhermitte, S. Kinge, K. Sivula, *ACS Appl. Nano Mater.* **2019**, *2*, 7705.
[116] I.-T. Chen, E. Schappell, X. Zhang, C.-H. Chang, *Microsyst. Nanoeng.* **2020**, *6*, 22.
[117] Z. Li, R. J. Young, C. Backes, W. Zhao, X. Zhang, A. A. Zhukov, E. Tillotson, A. P. Conlan, F. Ding, S. J. Haigh, K. S. Novoselov, J. N. Coleman, *ACS Nano* **2020**, *14*, 10976.
[118] P. Saha, M. Sahad E, S. Sathyanarayana, B. C. Das, *ACS Nano* **2024**, *18*, 1137.
[119] M. Kim, R. Ge, X. Wu, X. Lan, J. Tice, J. C. Lee, D. Akinwande, *Nat. Commun.* **2018**, *9*, 2524.
[120] M. Naqi, M. S. Kang, N. Liu, T. Kim, S. Baek, A. Bala, C. Moon, J. Park, S. Kim, *npj 2D Mater. Appl.* **2022**, *6*, 53.
[121] E. Lee, J. Kim, J. Park, J. Hwang, H. Jang, K. Cho, W. Choi, *ACS Appl. Mater. Interfaces* **2023**, *15*, 15839.
[122] J. Song, Y. Wang, L. Wang, Z. Liu, Y. Lei, M. Cheng, Y. Zhang, W. Zhou, Z. Xu, X. Li, M. S. Saleem, L. Chen, B. Huang, W. Wang, C. Li, *Appl. Phys. Rev.* **2025**, *12*, 031403.
[123] W. Qiao, S. Yan, X. He, X. Song, Z. Li, X. Zhang, W. Zhong, Y. Du, *RSC Adv.* **2014**, *4*, 50981.
[124] C. Backes, B. M. Szydłowska, A. Harvey, S. Yuan, V. Vega-Mayoral, B. R. Davies, P. Zhao, D. Hanlon, E. J. G. Santos, M. I. Katsnelson, W. J. Blau, C. Gadermaier, J. N. Coleman, *ACS Nano* **2016**, *10*, 1589.
[125] J. Kang, J.-W. T. Seo, D. Alducin, A. Ponce, M. J. Yacaman, M. C. Hersam, *Nat. Commun.* **2014**, *5*, 5478.
[126] Z. Lin, Y. Liu, U. Halim, M. Ding, Y. Liu, Y. Wang, C. Jia, P. Chen, X. Duan, C. Wang, F. Song, M. Li, C. Wan, Y. Huang, X. Duan, *Nature* **2018**, *562*, 254.
[127] J. Li, P. Song, J. Zhao, K. Vaklinova, X. Zhao, Z. Li, Z. Qiu, Z. Wang, L. Lin, M. Zhao, T. S. Herng, Y. Zuo, W. Jonhson, W. Yu, X. Hai, P. Lyu, H. Xu, H. Yang, C. Chen, S. J. Pennycook, J. Ding, J. Teng, A. H. Castro Neto, K. S. Novoselov, J. Lu, *Nat. Mater.* **2021**, *20*, 181.
[128] A. Castellanos-Gomez, M. Barkelid, A. M. Goossens, V. E. Calado, H. S. J. van der Zant, G. A. Steele, *Nano Lett.* **2012**, *12*, 3187.
[129] G. R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M. S. Strano, V. R. Cooper, L. Liang, S. G. Louie, E. Ringe, W. Zhou, S. S. Kim, R. R. Naik, B. G. Sumpter, H. Terrones, F. Xia, Y. Wang, J. Zhu, D. Akinwande, N. Alem, J. A. Schuller, R. E. Schaak, M. Terrones, J. A. Robinson, *ACS Nano* **2015**, *9*, 11509.
[130] A. Sebastian, F. Zhang, A. Dodda, D. May-Rawding, H. Liu, T. Zhang, M. Terrones, S. Das, *ACS Nano* **2019**, *13*, 78.
[131] S. Ippolito, F. Urban, W. Zheng, O. Mazzarisi, C. Valentini, A. G. Kelly, S. M. Gali, M. Bonn, D. Beljonne, F. Corberi, J. N. Coleman, H. I. Wang, P. Samorì, *Adv. Mater.* **2023**, *35*, 2211157.
[132] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, D. Scaramuzza, *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 154.
[133] X. Su, B. Zhang, C. Liang, M. Tian, T. Zhang, Z. Bian, J. Miao, Q. Yang, Y. Xu, B. Yu, Y. Chai, P. Lin, Y. Zhao, *Adv. Funct. Mater.* **2024**, *34*, 2315323.
[134] S. Subbulakshmi Radhakrishnan, S. Chakrabarti, D. Sen, M. Das, T. F. Schranghamer, A. Sebastian, S. Das, *Adv. Mater.* **2022**, *34*, 2202535.
[135] Y. Zhou, J. Fu, Z. Chen, F. Zhuge, Y. Wang, J. Yan, S. Ma, L. Xu, H. Yuan, M. Chan, X. Miao, Y. He, Y. Chai, *Nat. Electron.* **2023**, *6*, 870.
[136] D. Kudithipudi, C. Schuman, C. M. Vineyard, T. Pandit, C. Merkel, R. Kubendran, J. B. Aimone, G. Orchard, C. Mayr, R. Benosman, J. Hays, C. Young, C. Bartolozzi, A. Majumdar, S. G. Cardwell, M. Payvand, S. Buckley, S. Kulkarni, H. A. Gonzalez, G. Cauwenberghs, C. S. Thakur, A. Subramoney, S. Furber, *Nature* **2025**, *637*, 801.
[137] J. Chen, W. Wang, X. Yan, *npj Unconv. Comput.* **2025**, *2*, 19.
[138] D. Jayachandran, A. Pannone, M. Das, T. F. Schranghamer, D. Sen, S. Das, *ACS Nano* **2023**, *17*, 1068.
[139] P. Han, E. R. Adler, Y. Liu, L. St Marie, A. El Fatimy, S. Melis, E. Van Keuren, P. Barbara, *Nanotechnology* **2019**, *30*, 284004.
[140] K. M. Price, K. E. Schauble, F. A. McGuire, D. B. Farmer, A. D. Franklin, *ACS Appl. Mater. Interfaces* **2017**, *9*, 23072.
[141] S. G. Seo, J. H. Ryu, W. Y. Lee, S. H. Jin, *Phys. Status Solidi A* **2022**, *219*, 2200052.
[142] C. Di Natale, A. D'Amico, F. A. M. Davide, *Sens. Actuators, A* **1993**, *37–38*, 612.
[143] S. Shekhar, W. Bogaerts, L. Chrostowski, J. E. Bowers, M. Hochberg, R. Soref, B. J. Shastri, *Nat. Commun.* **2024**, *15*, 751.
[144] A. Leal-Junior, L. Avellar, W. Blanc, A. Frizera, C. Marques, *IEEE Internet Things J.* **2024**, *11*, 9587.
[145] Z. Xu, B. Tang, X. Zhang, J. F. Leong, J. Pan, S. Hooda, E. Zamburg, A. V.-Y. Thean, *Light: Sci. Appl.* **2022**, *11*, 288.
[146] G. Chen, Y. Yu, Y. Shi, N. Li, W. Luo, L. Cao, A. J. Danner, A.-Q. Liu, X. Zhang, *Laser Photonics Rev.* **2022**, *16*, 2200117.
[147] J. Witzens, *Proc. IEEE* **2018**, *106*, 2158.
[148] M. Li, J. Ling, Y. He, U. A. Javid, S. Xue, Q. Lin, *Nat. Commun.* **2020**, *11*, 4123.
[149] Z. Xu, C.-K. Chen, H.-L. Lin, Y. Gao, W. Ke, B. Xu, P. Dmitriev, C. Arbiz, E. Zamburg, S. Touzard, X. Cai, J. Lee, S. Venkatesan, A. Danner, A. V.-Y. Thean, in *2023 Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ **2023**, pp. 1–4.
[150] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, in *31st Annual Conf. Neural Information Processing Systems (NIPS 2017)*, Curran Associates, Inc, New York, United States of America **2017**, pp. 1–15.
[151] H. Choi, S. Baek, H. Jung, T. Kang, S. Lee, J. Jeon, B. C. Jang, S. Lee, *Adv. Mater.* **2024**, *37*, 2406970.
[152] J. Jiang, K. Parto, W. Cao, K. Banerjee, *IEEE J. Electron Devices Soc.* **2019**, *7*, 878.
[153] J. Jiang, J. Kang, K. Banerjee, in *2017 IEEE Int. Reliability Physics Symp. (IRPS)*, IEEE, Piscataway, NJ, **2017**, pp. 6B-1.1-6B-16.
[154] K. Parto, A. Pal, X. Xie, W. Cao, K. Banerjee, in *2018 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ **2018**, p. 24.
[155] E. Yalon, B. Aslan, K. K. H. Smithe, C. J. McClellan, S. V. Suryavanshi, F. Xiong, A. Sood, C. M. Neumann, X. Xu, K. E. Goodson, T. F. Heinz, E. Pop, *ACS Appl. Mater. Interfaces* **2017**, *9*, 43013.
[156] J. Jiang, J. Kang, W. Cao, X. Xie, H. Zhang, J. H. Chu, W. Liu, K. Banerjee, *Nano Lett.* **2017**, *17*, 1482.