

# An All-in-One Bioinspired Neural Network

Shiva Subbulakshmi Radhakrishnan, Akhil Dodda, and Saptarshi Das\*



Cite This: *ACS Nano* 2022, 16, 20100–20115



Read Online

ACCESS |



Metrics & More



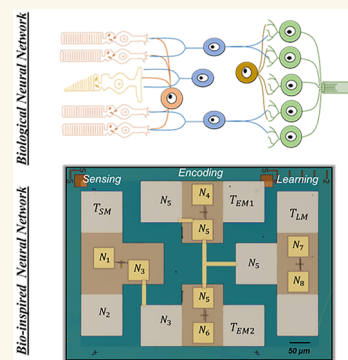
Article Recommendations



Supporting Information

**ABSTRACT:** In spite of recent advancements in artificial neural networks (ANNs), the energy efficiency, multifunctionality, adaptability, and integrated nature of biological neural networks remain largely unimitated by hardware neuromorphic computing systems. Here, we exploit optoelectronic, computing, and programmable memory devices based on emerging two-dimensional (2D) layered materials such as  $\text{MoS}_2$  to demonstrate a monolithically integrated, multipixel, and “all-in-one” bioinspired neural network (BNN) capable of sensing, encoding, learning, forgetting, and inferring at minuscule energy expenditure. We also demonstrate learning adaptability and simulate learning challenges under specific synaptic conditions to mimic biological learning. Our findings highlight the potential of in-memory computing and sensing based on emerging 2D materials, devices, and integrated circuits to not only overcome the bottleneck of von Neumann computing in conventional CMOS designs but also to aid in eliminating the peripheral components necessary for competing technologies such as memristors.

**KEYWORDS:** two-dimensional materials, monolayer  $\text{MoS}_2$ , field effect transistors, neural networks, photosensing, neuromorphic computing, gate-tunable persistent photoconductivity, charge trapping/detrapping



Biological neural networks comprising billions of neurons connected via trillions of synapses are incredibly diverse, integrated, and energy efficient in processing information that involves sensing, encoding, storage, and computation. For example, sensory neurons receive external/internal stimuli from various sensory organs and convert the information into spike trains following various encoding algorithms, which are then communicated via interneurons to the central nervous system (CNS) where spike-based computation leads to memory formation (learning) and/or decision making (inference). Spikes are electrical impulses or digital point events in time that enable ultralow-power neural computation as well as long-distance neural communication. Spiking activity between the presynaptic and postsynaptic neurons determines the potentiation or depression of their connection strengths or synaptic weights, which is ultimately responsible for learning and forgetting. Another key feature of the biological neural network is neuroplasticity, which allows adaptation to learning and decision-making under changing environmental conditions. For example, eyes can identify patterns under both low-light (scotopic vision) as well as bright-light (photopic vision) conditions. Finally, the balance between the relative strength of potentiation and depression of synaptic connections is critical, and any deviation can lead to neurological disorders including learning disabilities. Therefore, designing low-power neuromorphic hardware systems that resemble the functionality, organization, and plasticity of the biological neural network can not only

accelerate the development of hardware artificial intelligence (AI) but also benefit edge computing.

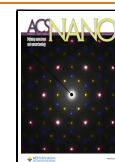
Artificial neural networks (ANNs) are a highly simplified yet common abstraction of biological neural networks that have already demonstrated breakthroughs in many applications, including image classification, speech recognition, and game playing.<sup>1,2</sup> However, hardware realization of ANNs using traditional complementary metal-oxide-semiconductor (CMOS) technology consumes orders of magnitude higher power compared to what the brain demands for similar tasks. One of the key differences is in the computing architecture; where CMOS-based computation embraces von Neumann architecture that physically separates the compute (logic) and storage (memory), biological neural networks dissolve such gaps by placing neurons, the computational primitives, and synapses, the storage units, right next to each other.

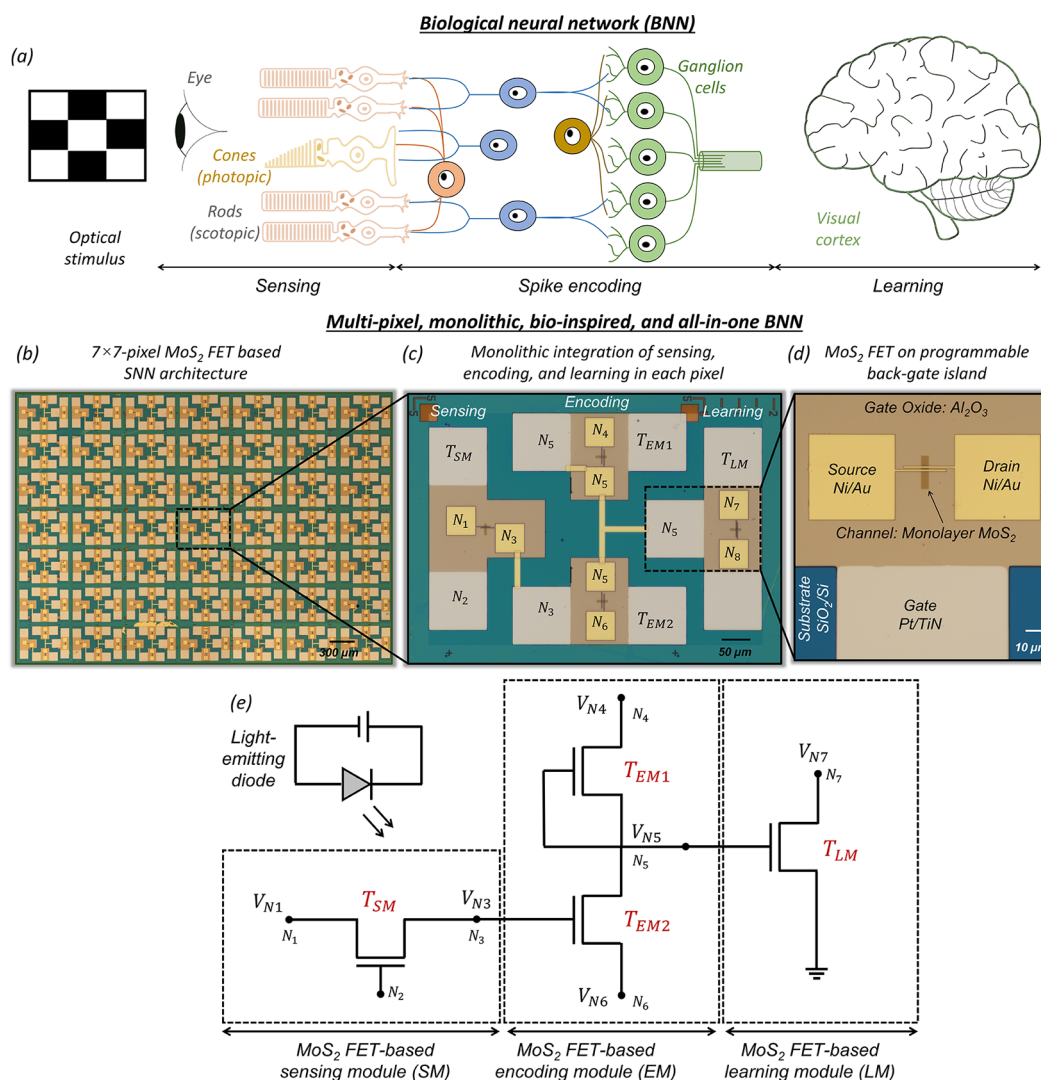
Acknowledging the energy gap, field-programmable gate arrays (FPGAs)<sup>3</sup> and crossbar architectures utilizing memristors,<sup>4,5</sup> resistive random-access memory (RRAM),<sup>6</sup> phase change memory (PCM),<sup>7–9</sup> etc., with tunable conductance states are accelerating the development of energy-efficient and non-von Neumann computing architectures. However, these

**Received:** March 3, 2022

**Accepted:** October 18, 2022

**Published:** November 15, 2022



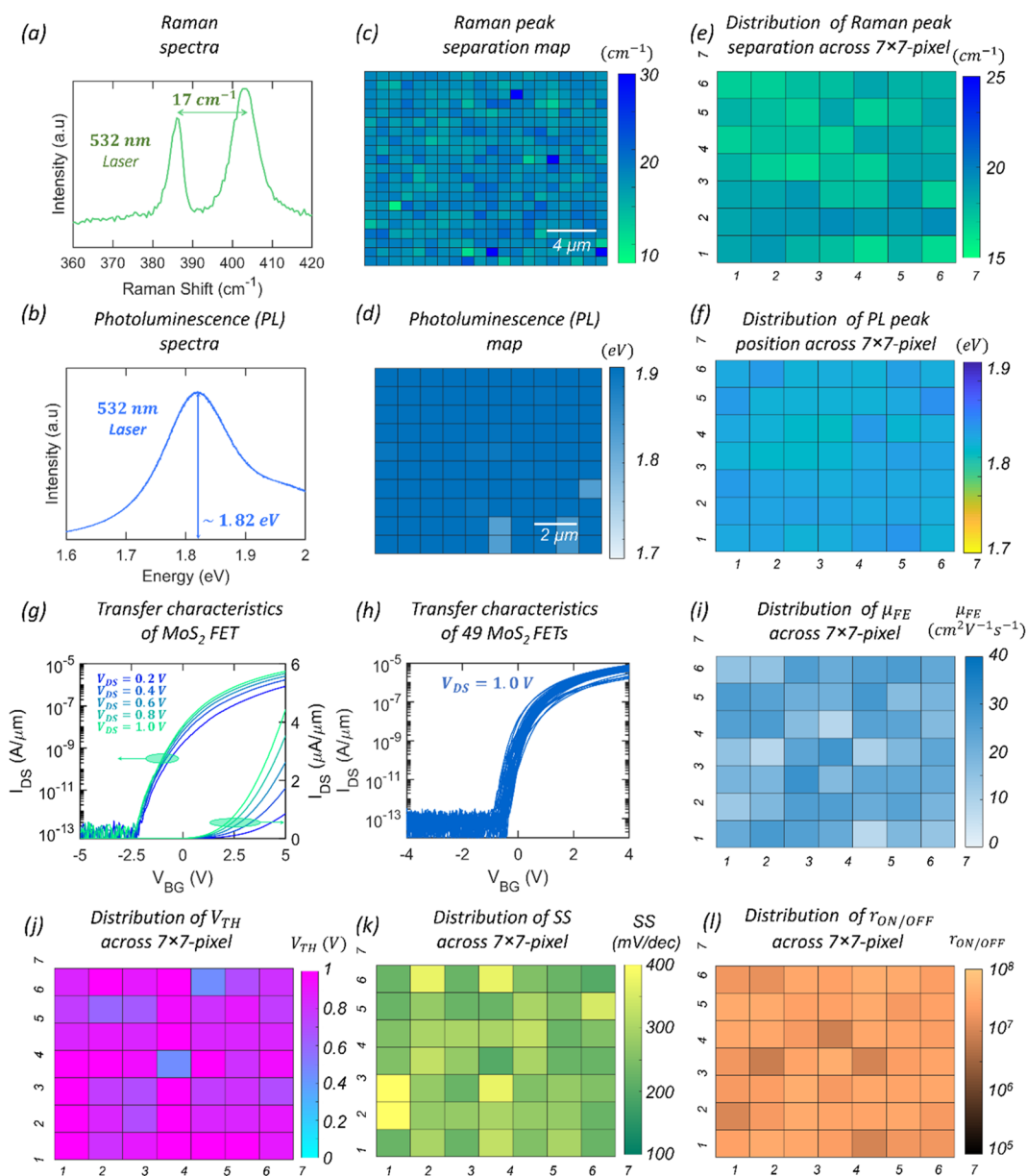


**Figure 1.** Monolithically integrated, multipixel, all-in-one biological neural network. a) Biological neural network (BNN) for processing visual information. Optical images of b)  $7 \times 7$ -pixel BNN architecture, c) a pixel comprising four monolayer MoS<sub>2</sub> FETs (4T cell) that monolithically integrates the sensing module (SM), encoding module (EM), and learning module (LM), and d) an individual monolayer MoS<sub>2</sub> field-effect transistor (FET), which is locally back-gated using a stack comprising atomic layer deposition (ALD) grown 50 nm Al<sub>2</sub>O<sub>3</sub> on sputter deposited 40/30 nm Pt/TiN. All back-gate islands were placed on a commercially purchased SiO<sub>2</sub>/p<sup>+</sup>-Si substrate. e) Circuit schematic for each pixel showing the connection between the SM, EM, and LM consisting of 1 ( $T_{SM}$ ), 2 ( $T_{EM1}$  and  $T_{EM2}$ ), and 1 ( $T_{LM}$ ) MoS<sub>2</sub> FETs, respectively. MoS<sub>2</sub> FETs used for the SMs and LMs have footprints ( $W \times L$ ) of  $5 \mu\text{m} \times 1 \mu\text{m}$ , MoS<sub>2</sub> FETs used for the EMs have a footprint of  $5 \mu\text{m} \times 3 \mu\text{m}$  excluding the contact pads, and each pixel has a footprint of  $400 \mu\text{m} \times 600 \mu\text{m}$ . Each pixel is designed to achieve functional and organizational resemblance with different neuronal cells found in the visual BNN. For example, the SM is functionally equivalent to photoreceptor cells (rods and cones) in the human eyes that convert external optical stimuli into corresponding graded potentials ( $V_{N3}$ ) at node  $N_3$ . Rods primarily enable low-light (scotopic) vision, whereas cones are responsible for bright-light (photopic) vision, both of which can be achieved using  $T_{SM}$  by exploiting gate-tunable persistent photoconductivity. Similarly, the EM mimics the functionality of retinal ganglion cells that encode the graded potentials into spike trains and transmit to the visual cortex for further processing. Finally, the LM imitates the visual cortex where learning, forgetting, and inference take place.

non-von Neumann platforms still require CMOS-based peripheral transducers for converting external stimuli into electrical impulses, unlike biological neural networks where specialized afferent neurons transduce sensed information into electrical signals, i.e., spike trains. Such preprocessing can ultimately limit the energy efficiency and scalability of emerging non-von Neumann architectures.<sup>10,11</sup> Finally, neuroplasticity of learning in changing environments and modeling of learning disabilities even at a high level of abstraction is yet to be demonstrated.

Here, we mitigate the aforementioned challenges by introducing a monolithically integrated, multipixel, and “all-

in-one” bioinspired neural network (BNN) which is capable of sensing, encoding, learning, forgetting, and inferring using monolayer MoS<sub>2</sub>-based multifunctional field effect transistors (FETs). First, we use gate-tunable persistent photoconductivity in a monolayer MoS<sub>2</sub> FET to convert optical information into graded potentials using a neuromorphic sensing module (SM). Next, we demonstrate the MoS<sub>2</sub>-based neuromorphic encoding module (EM) comprising two MoS<sub>2</sub> FETs to transform the graded potentials into spike-count- and spike-duration-based programming voltages. Finally, we exploit the electrical programmability of MoS<sub>2</sub>-FET-based nonvolatile synapses for realizing a neuromorphic learning module (LM)



**Figure 2.** Characterization and device-to-device variation of MoS<sub>2</sub> FETs. a) Raman and b) photoluminescence (PL) spectra of a representative MoS<sub>2</sub> channel region. The Raman peak separation between the characteristic A<sub>1g</sub> and E<sub>2g</sub> modes was consistent with monolayer MoS<sub>2</sub> at 17 cm<sup>-1</sup>; the PL peak location at 1.82 eV was also consistent with monolayer MoS<sub>2</sub>. c) Raman map of 21 μm × 20 μm and d) PL map of 10 μm × 10 μm taken across the as-grown MoS<sub>2</sub> film, with each grid corresponding to an area of 1 μm<sup>2</sup>. Colormaps of the distribution of e) the Raman peak separation and f) the PL peak position across 49 MoS<sub>2</sub> channels corresponding to each of the 7 × 7 pixels of our BNN architecture. The mean and standard deviation values were extracted to be 18 cm<sup>-1</sup> and 0.8 cm<sup>-1</sup>, respectively, for Raman peak separation and 1.82 and 0.01 eV, respectively, for PL peak location. g) Transfer characteristics, i.e., source-to-drain current (*I*<sub>DS</sub>) as a function of the local back-gate voltage (*V*<sub>BG</sub>), at different drain biases (*V*<sub>DS</sub>) for a representative MoS<sub>2</sub> FET with *L* = 1 μm. h) Device-to-device variation in the transfer characteristics across 49 MoS<sub>2</sub> FETs corresponding to each of the 7 × 7 pixels. Colormaps of the distribution of i) the electron field-effect mobility (*μ*<sub>FE</sub>) extracted from the peak transconductance, j) the current on/off ratio (*r*<sub>ON/OFF</sub>), k) the subthreshold slope (SS) over 3 orders of magnitude change in *I*<sub>DS</sub>, and l) the threshold voltage (*V*<sub>TH</sub>) extracted at an iso-current of 100 nA/μm for these 49 MoS<sub>2</sub> FETs. Extracted mean values for *μ*<sub>FE</sub>, *r*<sub>ON/OFF</sub>, SS, and *V*<sub>TH</sub> were found to be 21 cm<sup>2</sup> V<sup>-1</sup> s<sup>-1</sup>, 2.6 × 10<sup>7</sup>, 275 mV/decade, and 0.9 V, respectively, with corresponding standard deviation values of 5.5 cm<sup>2</sup> V<sup>-1</sup> s<sup>-1</sup>, 0.8 × 10<sup>7</sup>, 59 mV/decade, and 0.2 V, respectively.

for spike-based learning, forgetting, and inference. Furthermore, we demonstrate low-power operation and adaptability of our BNN to learning under different ambient conditions, mimicking the neuroplasticity of biological neural networks. Our BNN hardware also offers a platform to model learning disabilities and disorders at a high level of abstraction. Our work experimentally demonstrates an integrated BNN

exploiting in-memory computing and sensing based on emerging two-dimensional (2D) layered materials, devices, and circuits that can accelerate the development of energy-efficient neuromorphic systems.

The motivation behind using 2D layered MoS<sub>2</sub> as a hardware platform for neuromorphic computing is multifold. First, there are several demonstrations of photodetectors,<sup>12</sup>



chemical sensors,<sup>13</sup> biological sensors,<sup>13</sup> touch sensors,<sup>14</sup> and radiation sensors<sup>15</sup> using MoS<sub>2</sub>-based devices; these can naturally serve as artificial sensory afferent neurons, eliminating the need for peripheral sensors in MoS<sub>2</sub>-based intelligent systems. Next, MoS<sub>2</sub> being a semiconductor, almost all peripheral analog or digital signal processing units can be built using MoS<sub>2</sub> FETs<sup>16</sup>, thus eliminating the need for hybrid design involving CMOS circuitry. Additionally, the atomically-thin-body nature of MoS<sub>2</sub> allows aggressive channel length scaling without the loss of superior gate electrostatics, benefiting high integration density. In fact, recent studies show high performance monolayer MoS<sub>2</sub> FETs with the channel and contact lengths scaled to 29 and 13 nm, respectively.<sup>17</sup> Moreover, some of the early criticism of 2D FETs has also been successfully addressed in recent years through the realization of low contact resistance,<sup>18</sup> high ON currents,<sup>19</sup> integration of ultrathin and high-k gate dielectrics,<sup>20</sup> and wafer scale growth using chemical vapor deposition (CVD)<sup>21</sup> and metal–organic CVD (MOCVD).<sup>22,23</sup> Similarly, MoS<sub>2</sub>-based microprocessors,<sup>24</sup> analogue operational amplifiers,<sup>25</sup> RF electronics components,<sup>26</sup> and neuromorphic, security,<sup>27–29</sup> and biomimetic hardware platforms<sup>30–32</sup> have been reported. Finally, MoS<sub>2</sub> can enable flexible<sup>33</sup> and printable optoelectronics, adding value toward a MoS<sub>2</sub>-based hardware platform similar to ultrathin silicon on insulators.<sup>34,35</sup>

## RESULTS AND DISCUSSION

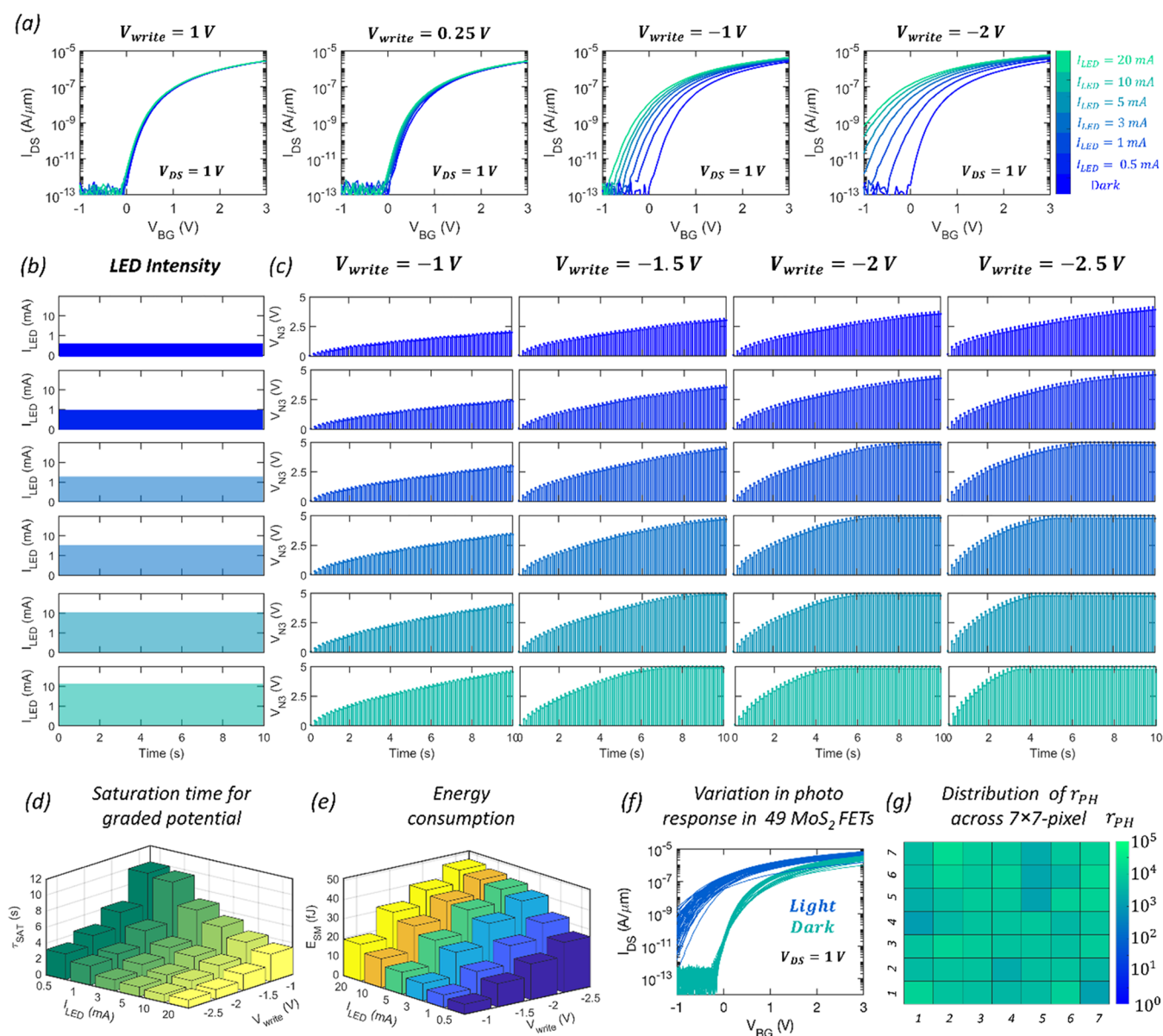
**Monolithically Integrated, Multipixel, All-in-One BNN Platform.** Figure 1a shows the neurobiological architecture for processing visual information, and Figure 1b–d, respectively, show optical images of our multipixel (7 × 7) BNN hardware platform, a pixel comprising four monolayer MoS<sub>2</sub> FETs (4T cell) that monolithically integrates the sensing module (SM), encoding module (EM), and learning module (LM), and an individual MoS<sub>2</sub> FET, which is locally back-gated using a stack comprising atomic layer deposition (ALD) grown 50 nm Al<sub>2</sub>O<sub>3</sub> on sputter deposited 40/30 nm Pt/TiN. All back-gate islands were placed on a commercially purchased SiO<sub>2</sub>/p<sup>++</sup>-Si substrate to isolate each MoS<sub>2</sub> FET (see Supporting Information 1–3 for enlarged optical images of the entire chip, 7 × 7 pixels, and each pixel, respectively). In fact, any other rigid or flexible substrate could potentially be used instead of the SiO<sub>2</sub>/p<sup>++</sup>-Si substrate to build our “all-in-one” bioinspired hardware platform. Within each pixel, the SM consists of 1 MoS<sub>2</sub> FET ( $T_{SM}$ ), the EM consists of 2 MoS<sub>2</sub> FETs ( $T_{EM1}$  and  $T_{EM2}$ ), and the LM consists of 1 MoS<sub>2</sub> FET ( $T_{LM}$ ), each of which are connected using the circuit diagram shown in Figure 1e. MoS<sub>2</sub> FETs used for the SMs and LMs have footprints ( $W \times L$ ) of 5  $\mu\text{m} \times 1 \mu\text{m}$ , MoS<sub>2</sub> FETs used for the EMs have a footprint of 5  $\mu\text{m} \times 3 \mu\text{m}$  excluding the contact pads, and each pixel has a footprint of 400  $\mu\text{m} \times 600 \mu\text{m}$ . For the purposes of these experiments, the scalability of our devices was limited by our measurement setup, which requires large contact pads for probing the devices. In fact, the use of atomically-thin monolayer MoS<sub>2</sub> as the channel material for the FETs makes this technology aggressively scalable.<sup>36–38</sup> Therefore, it is possible to accomplish denser arrays of sensors, memory devices, and compute elements using MoS<sub>2</sub> FETs, which can then be exploited for hardware implementation of advanced deep neural networks as well as bioinspired vision architectures such as ON versus OFF ganglion cells or subsequent layers of processing. Nevertheless, the circuit schematic in Figure 1e shows that each pixel is designed to

achieve functional and organizational resemblance with different neuronal cells found in the vision pathways in primates, which is depicted schematically in Figure 1a. For example, the monolayer MoS<sub>2</sub> FET-based SM is functionally equivalent to photoreceptor cells (rods and cones) in the human eyes that convert external optical stimuli into corresponding graded potentials. Rods primarily enable low-light (scotopic) vision, whereas cones are responsible for bright-light (photopic) vision, both of which can be achieved using our adaptive SM. Similarly, the MoS<sub>2</sub> FET-based EM mimics the functionality of retinal ganglion cells that encode the graded potentials into spike trains and transmit to the visual cortex, or midbrain, for higher order processing and computation. Finally, the MoS<sub>2</sub> FET-based LM imitates the visual cortex where learning, forgetting, and inference take place. As we will elucidate later, the Al<sub>2</sub>O<sub>3</sub>/Pt/TiN gate stack allows nonvolatile programming of our MoS<sub>2</sub> FETs owing to the trapping/detrapping of charge carriers at and near the MoS<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub> interface when subjected to large positive and negative gate biases. This, in turn, empowers our BNN architecture to overcome the von Neumann bottleneck and enable in-memory sensing and computing capabilities, which are presently lacking for the conventional silicon technology. The programming capability is also central toward the realization of a reconfigurable BNN platform that allows adaptation to different learning conditions (e.g., scotopic conditions) similar to biological neuroplasticity as well as offers a platform to model and study the origin of various learning disabilities found in humans (e.g., autism disorder).

The MoS<sub>2</sub> used in this study was obtained from the 2D Crystal Consortium (2DCC) and was grown epitaxially on a sapphire substrate using MOCVD at 1000 °C.<sup>22,39</sup> As we will elucidate, high-temperature growth ensures high film quality and low device-to-device variability, which are critical for the successful demonstration of our BNN platform. The monolayer MoS<sub>2</sub> film was transferred from the growth substrate to the target application substrate, i.e., the SiO<sub>2</sub>/p<sup>++</sup>-Si substrate with predefined islands of Al<sub>2</sub>O<sub>3</sub>/Pt/TiN, for subsequent FET fabrication and monolithic integration of the SM, EM, and LM. Details on the fabrication of the back-gate stack, monolayer MoS<sub>2</sub> synthesis, film transfer, fabrication of MoS<sub>2</sub> FETs, and monolithic integration can be found in the Methods section.

**Characterization and Device-to-Device Variation of MoS<sub>2</sub> FETs.** Before diving deeper into each functional unit of our hardware BNN platform, i.e., SM, EM, and LM, it is important to thoroughly characterize the basic building blocks, i.e., the MoS<sub>2</sub> FETs. Figure 2a–b, respectively, show the Raman and photoluminescence (PL) spectra of a representative MoS<sub>2</sub> channel region. The Raman peak separation between the characteristic  $A_{1g}$  and  $E_{2g}^1$  modes was consistent with monolayer MoS<sub>2</sub> at 17 cm<sup>−1</sup>; the PL peak location at 1.82 eV was also consistent with monolayer MoS<sub>2</sub>.<sup>40–42</sup> Figure 2c–d, respectively, show the Raman map of a 21  $\mu\text{m} \times 20 \mu\text{m}$  region and the PL map of a 10  $\mu\text{m} \times 10 \mu\text{m}$  region, respectively, of the as-grown MoS<sub>2</sub> film. Raman peak separation and PL peak position vary less than 4% over the entire map, confirming the high quality and uniformity of the monolayer film. In fact, a similar assessment of film uniformity for the entire chip can be made from Figure 2e–f, which shows Raman peak separation and PL peak position across 49 MoS<sub>2</sub> channels corresponding to each of the 7 × 7 pixels (see Supporting Information 4 and 5





**Figure 3.** MoS<sub>2</sub>-FET-based neuromorphic sensing module (SM): a) Transfer characteristics of a monolayer MoS<sub>2</sub> FET measured at  $V_{DS} = 1$  V before and after illumination from a blue LED with input currents ranging from  $I_{LED} = 0.5$  mA (low-brightness) to  $I_{LED} = 20$  mA (high-brightness) at different  $V_{BG} = V_{write}$  for  $t_{write} = 100$  ms. b) Analog valued and continuous time input optical stimuli from the blue LED. c) Corresponding temporal evolution of the graded potential,  $V_{N3}$ , at node N<sub>3</sub> for different  $V_{write}$ 's obtained by using the circuit layout for the SM shown in Figure 1e. A constant voltage,  $V_{N1} = 5$  V, is applied to node N<sub>1</sub>, which is the drain terminal of  $T_{SM}$ , and a clocking signal toggling between  $V_{read}$  and  $V_{write}$  is applied to node N<sub>2</sub>, which is the local back-gate of  $T_{SM}$ , with  $\tau_{CLK} = 100$  ms. The source terminal of  $T_{SM}$  is connected to the local back-gate of  $T_{EM1}$  at node N<sub>3</sub>. d) Time for  $V_{N3}$  to reach the same magnitude ( $\tau_{SAT}$ ) as a function of  $I_{LED}$  and  $V_{write}$ . e) Average energy consumption by the SM ( $E_{SM}$ ) during each  $\tau_{CLK}$  for different  $I_{LED}$ 's and  $V_{write}$ 's. f) Device-to-device variation in the photoresponse of 49 MoS<sub>2</sub> FETs corresponding to the SMs of each of the 7 × 7 pixels of our BNN hardware after  $t_{write} = 1$  s exposure to  $I_{LED} = 20$  mA at  $V_{write} = -2.5$  V. g) Colormap of the distribution of the ratio of postillumination photoconductance to dark conductance ( $r_{PH}$ ) measured at  $V_{BG} = 0$  V. The mean and standard deviation values were found to be  $6.7 \times 10^3$  and  $3.8 \times 10^3$ , respectively.

for the Raman and PL scans of each of these 49 MoS<sub>2</sub> channels, respectively). The mean and standard deviation values were extracted to be 18 cm<sup>-1</sup> and 0.8 cm<sup>-1</sup>, respectively, for Raman peak separation and 1.82 and 0.01 eV, respectively, for PL peak location.

Figure 2g shows the transfer characteristics, i.e., source-to-drain current ( $I_{DS}$ ) as a function of the local back-gate voltage ( $V_{BG}$ ), at different drain biases ( $V_{DS}$ ) for a representative MoS<sub>2</sub> FET with channel width and length of 5  $\mu$ m and 1  $\mu$ m, respectively. Figure 2h shows the device-to-device variation in

the transfer characteristics across 49 MoS<sub>2</sub> FETs of dimension 5  $\mu$ m × 1  $\mu$ m (width × length) corresponding to each of the 7 × 7 pixels (see Supporting Information 6 for the transfer characteristics for each of these 49 MoS<sub>2</sub> FETs). Note that, as expected, MoS<sub>2</sub> FETs show unipolar, n-type characteristics owing to the pinning of the metal Fermi level close to the conduction band allowing only electron transport through the channel. Figure 2i shows the map of electron field-effect mobility values ( $\mu_{FE}$ ) extracted from the peak transconductance for these 49 MoS<sub>2</sub> FETs with a mean of  $\sim 21$  cm<sup>2</sup> V<sup>-1</sup> s<sup>-1</sup>

and a standard deviation of  $5.5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ . Figure 2j–l, respectively, show similar colormaps on device-to-device variation in current on/off ratio ( $r_{\text{ON/OFF}}$ ), subthreshold slope (SS) over 3 orders of magnitude change in  $I_{\text{DS}}$ , and threshold voltage ( $V_{\text{TH}}$ ) extracted at an iso-current of  $100 \text{ nA}/\mu\text{m}$ , with mean values of  $2.6 \times 10^7$ ,  $275 \text{ mV/decade}$ , and  $0.9 \text{ V}$ , respectively, and standard deviation values of  $0.8 \times 10^7$ ,  $59 \text{ mV/decade}$ , and  $0.2 \text{ V}$ , respectively. Our  $\mu_{\text{FE}}$ ,  $r_{\text{ON/OFF}}$ , SS, and  $V_{\text{TH}}$  values and the corresponding device-to-device variations are on par with the state-of-the-art literature on large-area-grown  $\text{MoS}_2$ , which can be attributed to high-quality growth, relative damage-free transfer, and clean device fabrication. However, we also believe that it is possible to further reduce the device-to-device variation by improving the growth and the transfer process. Supporting Information 7 shows the output characteristics, i.e.,  $I_{\text{DS}}$  versus  $V_{\text{DS}}$ , at different  $V_{\text{BG}}$ 's for a representative  $\text{MoS}_2$  FET with  $L = 1 \mu\text{m}$ . While we mostly exploit the off-state and subthreshold regimes of FET operation in our SM, EM, and LM, the on-state current reaches as high as  $\sim 100 \mu\text{A}/\mu\text{m}$  at  $V_{\text{DS}} = 5 \text{ V}$  for an inversion charge carrier density of  $\sim 1.5 \times 10^{13}/\text{cm}^2$ ; this is yet another piece of evidence indicating high film quality.

**$\text{MoS}_2$  FET-Based Neuromorphic Sensing Module (SM).** Monolayer  $\text{MoS}_2$ -based phototransistors have been studied extensively in recent years, including in our own work.<sup>12,43–47</sup> The phototransduction mechanism in  $\text{MoS}_2$  FETs is typically attributed to two mechanisms: photocarrier generation in the  $\text{MoS}_2$  channel and photogating effect arising due to charge trapping/detrapping at the  $\text{MoS}_2$ /gate-dielectric interface. Figure 3a shows the transfer characteristics taken at  $V_{\text{DS}} = 1 \text{ V}$  for a representative monolayer  $\text{MoS}_2$  FET before and after illumination from a blue light emitting diode (LED) with input currents ranging from  $I_{\text{LED}} = 0.5 \text{ mA}$  (low-brightness) to  $I_{\text{LED}} = 20 \text{ mA}$  (high-brightness) at different  $V_{\text{BG}} = V_{\text{write}}$  for  $t_{\text{write}} = 100 \text{ ms}$ . The corresponding incident optical power is in the range of  $0.1$ – $10 \text{ Wm}^{-2}$ , obtained by calibrating using a commercially-purchased silicon PIN photodiode as described in Supporting Information 8. Given that the channel area of each  $\text{MoS}_2$  FET used in the SM is  $5 \mu\text{m} \times 1 \mu\text{m}$ , the estimated incident power on each pixel is  $0.5$ – $50 \text{ pW}$ . See Supporting Information 9 for the optical images showing corresponding LED brightness levels. Note that, instead of the LASER illumination conventionally used to study photoresponse in monolayer  $\text{MoS}_2$ ,<sup>47</sup> we have used an LED to provide optical stimuli since it represents a more realistic lighting ambience akin to where most neuromorphic sensors will be deployed.

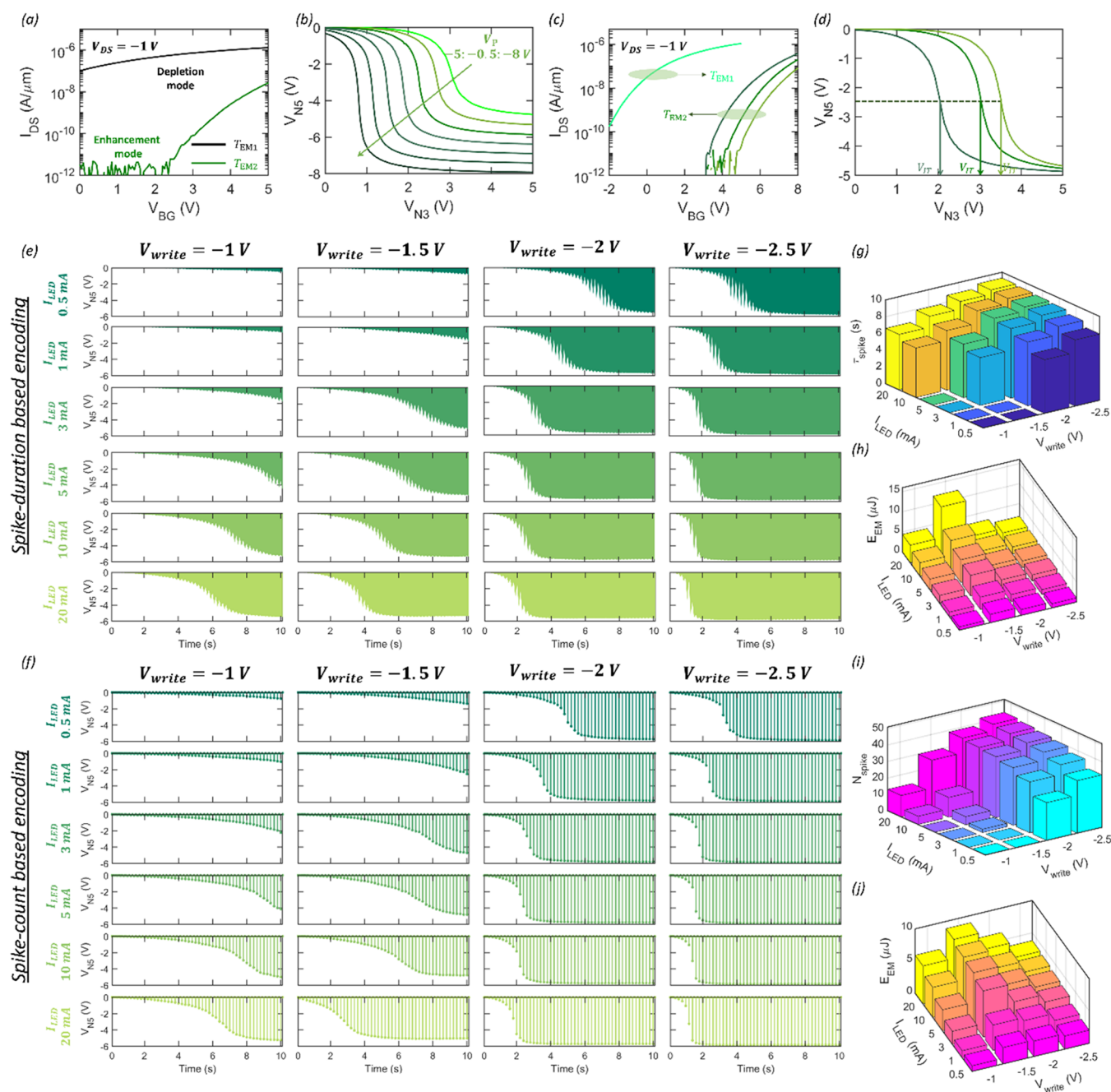
Two distinct types of photoresponse are observed in Figure 3a. For  $V_{\text{write}} > 0 \text{ V}$ , i.e., illuminations in the on-state ( $V_{\text{write}} = 2.0 \text{ V}$ ) and in the subthreshold regime ( $V_{\text{write}} = 0.5 \text{ V}$ ) of the  $\text{MoS}_2$  FET, there is no visible shift in the device characteristics postillumination irrespective of the brightness level of the LED ( $I_{\text{LED}}$ ). This can be ascribed to photocarrier generation in the  $\text{MoS}_2$  channel, which are swept across by the applied  $V_{\text{DS}}$ , and hence, there is no persistent photocurrent beyond the optical exposure. However, for  $V_{\text{write}} < 0 \text{ V}$ , i.e., illuminations in the off-state ( $V_{\text{write}} = -1.5 \text{ V}$  and  $V_{\text{write}} = -2.5 \text{ V}$ ) of the  $\text{MoS}_2$  FET, there are significant shifts in the device characteristics postillumination. This is a feature of the photogating effect, where photocarrier trapping at the  $\text{MoS}_2$ /dielectric interface leads to the shift in the device threshold voltage ( $V_{\text{TH}}$ ). The detrapping mechanism is rather slow and can take hours to several days, which is why the  $V_{\text{TH}}$  shift is visible

postillumination. Higher  $I_{\text{LED}}$ , more negative  $V_{\text{write}}$ , and longer  $t_{\text{write}}$  naturally result in more photocarrier trapping ( $Q_{\text{trap}}$ ) and hence larger  $V_{\text{TH}}$  shifts ( $\Delta V_{\text{TH}}$ ). Supporting Information 10 shows  $\Delta V_{\text{TH}}$  and the corresponding  $Q_{\text{trap}} (= C_{\text{OX}}\Delta V_{\text{TH}})$  as a function of  $V_{\text{write}}$  and  $t_{\text{write}}$  for  $I_{\text{LED}} = 20 \text{ mA}$ , where  $C_{\text{OX}} \approx 2 \times 10^{-3} \text{ Fm}^{-2}$  is the back-gate oxide capacitance per unit area.

We exploit the gate-tunable photogating effect in  $\text{MoS}_2$  FETs ( $T_{\text{SM}}$ ) for the conversion of analog optical stimuli into a graded potential,  $V_{\text{N3}}$ , at node  $N_3$  using the circuit layout shown in Figure 1e. A constant voltage,  $V_{\text{N1}} = 5 \text{ V}$ , is applied to the node  $N_1$ , which is the drain terminal of  $T_{\text{SM}}$ , and a clocking signal toggling between  $V_{\text{read}}$  and  $V_{\text{write}}$  is applied to the node  $N_2$ , which is the local back-gate of  $T_{\text{SM}}$  with  $\tau_{\text{CLK}} = 100 \text{ ms}$ . The source terminal of  $T_{\text{SM}}$  is connected to the local back-gate of  $T_{\text{EM2}}$  at node  $N_3$ . Note that  $T_{\text{SM}}$  and  $T_{\text{EM1}}$  form an RC circuit, with  $T_{\text{SM}}$  serving as the resistor and the local back-gate of  $T_{\text{EM1}}$  serving as the capacitor. Prior to illumination, the time constant for charging node  $N_3$  remains very high,  $>100 \text{ s}$ , since  $T_{\text{SM}}$  is biased in the off-state. Figure 3b shows analog-valued and continuous-time input optical stimuli from the LED, and Figure 3c shows the corresponding temporal evolution of  $V_{\text{N3}}$  for different  $V_{\text{write}}$ 's with  $V_{\text{read}} = 0 \text{ V}$ . Some key observations can be made from the results: 1)  $V_{\text{N3}}$  increases monotonically for any given  $I_{\text{LED}}$  and  $V_{\text{write}}$  owing to the photogating effect, which results in a gradual negative shift in the  $V_{\text{TH}}$  of  $T_{\text{SM}}$ , switching it from the off-state through the subthreshold to the on-state and thereby reducing the charging time constant for node  $N_3$ ; 2) for any given  $I_{\text{LED}}$ ,  $V_{\text{N3}}$  increases faster for more negative  $V_{\text{write}}$  since more trap states are available at the  $\text{MoS}_2$ /dielectric interface, resulting in a greater  $V_{\text{TH}}$  shift and hence higher photoconductance; 3) the time required by  $V_{\text{N3}}$  to reach  $V_{\text{N1}} = 5 \text{ V}$  scales inversely with  $I_{\text{LED}}$  for any given  $V_{\text{write}}$ , i.e., higher  $I_{\text{LED}}$  allows the graded potential to reach its maximum strength earlier and *vice versa*, and finally, 4) a lower  $I_{\text{LED}}$  (e.g.,  $5 \text{ mA}$ ) can invoke a similar response in  $V_{\text{N3}}$  like a higher  $I_{\text{LED}}$  (e.g.,  $20 \text{ mA}$ ), when the former is measured using a more negative  $V_{\text{write}} = -2 \text{ V}$  compared to the latter measured using  $V_{\text{write}} = -1.5 \text{ V}$ , allowing adaptation to different illumination levels. These observations are summarized in Figure 3d, which shows the time for  $V_{\text{N3}}$  to reach the same magnitude as a function of  $I_{\text{LED}}$  and  $V_{\text{write}}$ . Also, see Supplementary Video 1 for time evolution of the graded potential for various  $I_{\text{LED}}$ 's using different  $V_{\text{write}}$ 's.

Figure 3e shows the average energy consumption by the SM ( $E_{\text{SM}}$ ), given by  $E_{\text{SM}} = \frac{1}{2} C_{\text{OX}} V_{\text{write}}^2$ , during each  $\tau_{\text{CLK}}$  for different  $I_{\text{LED}}$ 's and  $V_{\text{write}}$ 's. Even for the brightest LED illumination at the most negative  $V_{\text{write}}$ ,  $E_{\text{SM}}$  is  $\sim 50 \text{ fJ}$ , which suggests energy-efficient phototransduction by our  $\text{MoS}_2$  FET-based SM. Finally, Figure 3f shows the pre- and postillumination transfer characteristics of 49  $\text{MoS}_2$  FETs corresponding to the SMs of each of the  $7 \times 7$  pixels of our BNN hardware after  $t_{\text{write}} = 1 \text{ s}$  exposure to  $I_{\text{LED}} = 20 \text{ mA}$  at  $V_{\text{write}} = -2.5 \text{ V}$ , and Figure 3g shows the colormap of the ratio of postillumination photoconductance to dark conductance ( $r_{\text{PH}}$ ) measured at  $V_{\text{BG}} = 0 \text{ V}$  (see Supporting Information 11 for the pre- and postillumination transfer characteristics for each of these 49  $\text{MoS}_2$  FETs). The mean and standard deviation values were found to be  $6.7 \times 10^3$  and  $3.8 \times 10^3$ , respectively.

**$\text{MoS}_2$  FET-Based Neuromorphic Encoding Module (EM).** The EM converts the graded potentials received from the SM into corresponding programming waveforms using

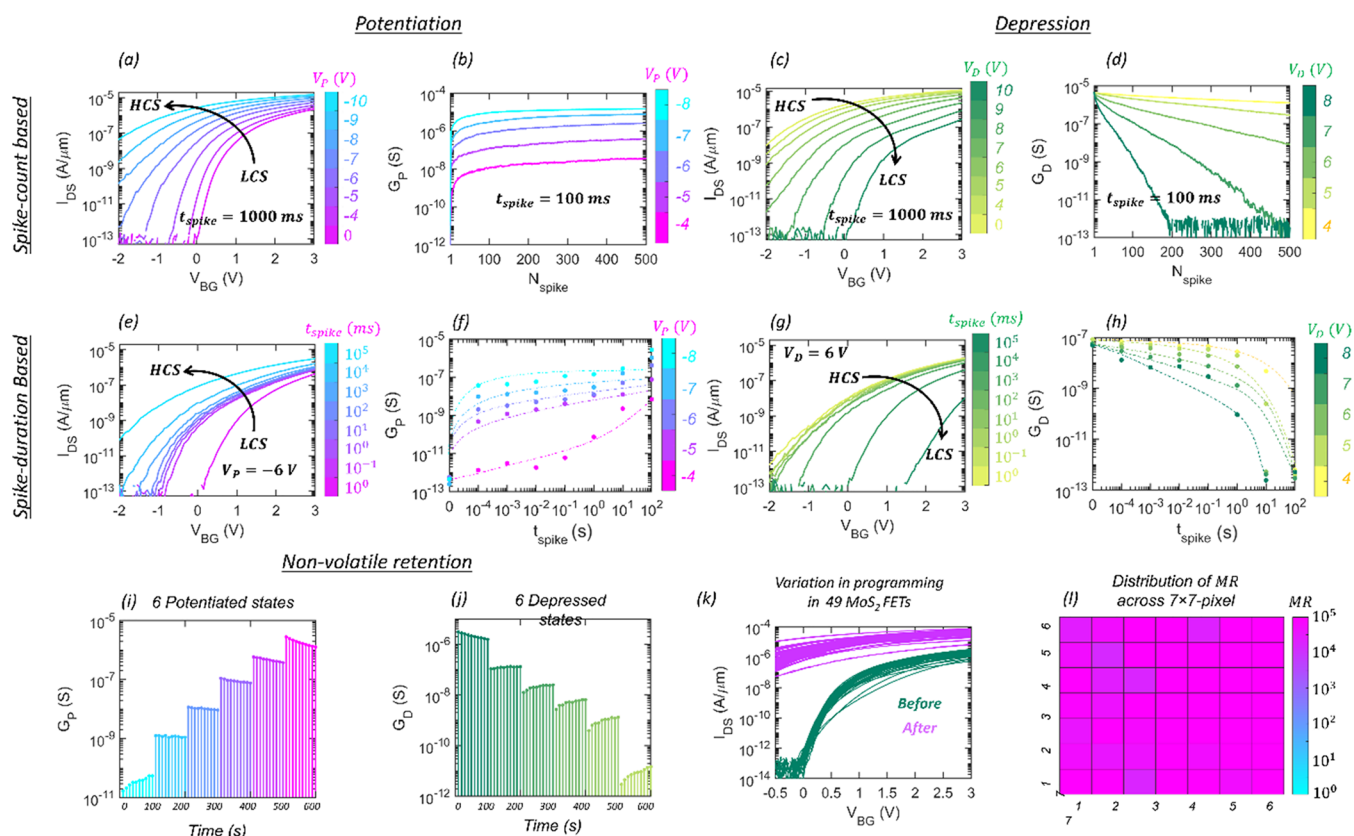


**Figure 4.** MoS<sub>2</sub>FET-based neuromorphic encoding module (EM). a) Transfer characteristics of MoS<sub>2</sub> FETs used as  $T_{EM1}$  and  $T_{EM2}$  in the EM.  $T_{EM1}$  is programmed to operate as a depletion mode (normally on) n-channel FET, whereas  $T_{EM2}$  operates as an enhancement mode (normally off) n-channel FET. Based on the circuit layout shown in Figure 1e, the EM serves as an NMOS inverter with a depletion load. b) Input ( $V_{N3}$ ) versus output ( $V_{N5}$ ) characteristics of the EM for different  $V_P$  values applied to the source terminal of  $T_{EM2}$ , i.e.,  $N_6$ . The drain terminal of  $T_{EM1}$ , i.e.,  $N_4$ , is kept grounded. c) Various programming states of  $T_{EM2}$  and d) corresponding EM characteristics for  $V_P = -5$  V. The inverting threshold ( $V_{IT}$ ), i.e., the magnitude of  $V_{N3}$  at which  $V_{N5}$  reaches  $V_P/2$ , can be adjusted by reconfiguring  $T_{EM1}$  and  $T_{EM2}$ . e) Spike-duration- and f) spike-count-based encoding of the graded potential ( $V_{N3}$ ) received from the SM, corresponding to different  $V_{write}$ 's and  $I_{LED}$ 's, into a programming voltage ( $V_{N5}$ ) for transmission to the learning module (LM). g) Total spiking time ( $\tau_{spike}$ ) and h) the corresponding average encoding energy expenditure ( $E_{EM}$ ) per clock cycle for spike-duration-based encoding and (i) the total number of spikes ( $N_{spike}$ ) and j) the corresponding  $E_{EM}$  for spike-count-based encoding as a function of  $V_{write}$  and  $I_{LED}$ . The input stimulus is presented for a duration of 10 s and  $V_P = -6$  V. Spike-duration and spike numbers are counted once  $V_{N5}$  reaches 75% of  $V_P$ , i.e.,  $-4.5$  V.

spike-count- and spike-duration-based algorithms and transmits them to the LM as summarized in Figure 4a–j. Each EM comprises two MoS<sub>2</sub> FETs,  $T_{EM1}$  and  $T_{EM2}$ , connected in series as shown in Figure 1e. Note that the local back-gate of  $T_{EM1}$  is shorted to its source at node  $N_5$ , which is also the drain of  $T_{EM2}$  and the output node of the EM. The drain terminal of

$T_{EM1}$ , i.e.,  $N_4$ , is kept grounded, and the programming voltage,  $V_P$ , is applied to the source terminal of  $T_{EM2}$ , i.e.,  $N_6$ . Furthermore, to realize an NMOS inverter,  $T_{EM1}$  and  $T_{EM2}$  should operate as depletion mode (normally on) and enhancement mode (normally off) n-channel FETs, respectively. Since all of our MoS<sub>2</sub> FETs are programmable, we can



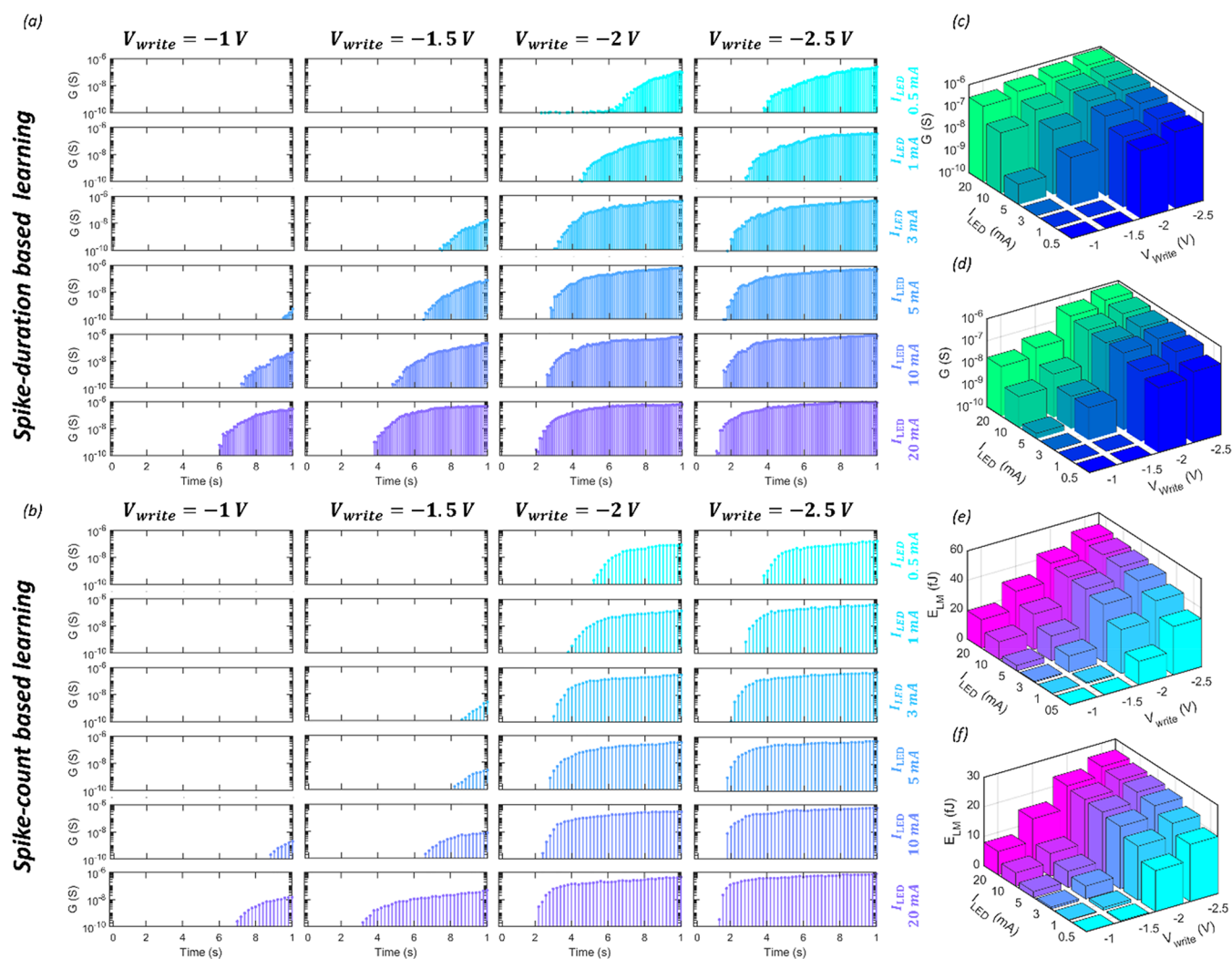


**Figure 5.** Analog and nonvolatile programming of MoS<sub>2</sub>-FET-based synapses. a) Potentiation of an MoS<sub>2</sub> synapse from a low conductance state (LCS) after the application of a fixed number of programming spikes ( $N_{\text{spike}} = 10$ ) of different amplitudes and negative polarity ( $V_p$ ); each spike was applied for  $t_{\text{spike}} = 100$  ms. b) Postpotentiated conductance states ( $G_p$ ) measured at  $V_{\text{BG}} = 0$  V as a function of  $N_{\text{spike}}$  for different  $V_p$ 's. c) Depression of an MoS<sub>2</sub> synapse from a high conductance state (HCS) after the application of a fixed number of programming spikes ( $N_{\text{spike}} = 10$ ) of different amplitudes and positive polarity ( $V_D$ ); each spike was applied for  $t_{\text{spike}} = 100$  ms. d) Postdepressed conductance states ( $G_D$ ) measured at  $V_{\text{BG}} = 0$  V as a function of  $N_{\text{spike}}$  for different  $V_D$ 's. e) Potentiation of the MoS<sub>2</sub> synapse from LCS after the application of a single spike of constant magnitude  $V_p = -6$  V for different  $t_{\text{spike}}$ 's. f) Postpotentiated  $G_p$  measured at  $V_{\text{BG}} = 0$  V as a function of  $t_{\text{spike}}$  for different  $V_p$ 's. g) Depression of the MoS<sub>2</sub> synapse from HCS after the application of a single spike of constant magnitude  $V_D = 6$  V for different  $t_{\text{spike}}$ 's. h) Postdepressed  $G_D$  measured at  $V_{\text{BG}} = 0$  V as a function of  $t_{\text{spike}}$  for different  $V_D$ 's. Retention characteristics of (i) 6 potentiated and (j) 6 depressed conductance states for 100 s. k) Device-to-device variation in the pre- and postprogrammed transfer characteristics and l) corresponding colormap of the distribution of the memory ratio (MR) measured at  $V_{\text{BG}} = 0$  V for 49 monolayer MoS<sub>2</sub> FETs from each LM of our  $7 \times 7$  BNN platform when programmed using  $N_{\text{spike}} = 10$  with spike magnitude,  $V_p = -8$  V, and spike width,  $t_{\text{spike}} = 100$  ms. The mean and standard deviation values for MR were found to be  $6 \times 10^5$  and  $0.5 \times 10^5$ , respectively.

shift the threshold voltage of the device by pulsing appropriate voltage spikes to the back-gate. Therefore, we applied a depression pulse to  $T_{\text{EM1}}$  and a potentiation pulse to  $T_{\text{EM2}}$  to operate them as enhancement and depletion mode FETs, respectively, as shown in Figure 4a. The EM, therefore, serves as an NMOS inverter with a depletion load. Figure 4b shows the input ( $V_{\text{N3}}$ ) versus output ( $V_{\text{N5}}$ ) characteristics of the EM for different  $V_p$  values. Also, note that the inverting threshold ( $V_{\text{IT}}$ ), i.e., the magnitude of  $V_{\text{N3}}$  at which  $V_{\text{N5}}$  reaches  $V_p/2$ , can be adjusted by reconfiguring  $T_{\text{EM1}}$  and  $T_{\text{EM2}}$ . Figure 4c-d, respectively, show the various programming states of  $T_{\text{EM2}}$  and the corresponding EM characteristics for  $V_p = -5$  V. Tunability in the EM characteristic is an additional benefit of our BNN platform, allowing adaptation to various learning conditions as we will elucidate later. Finally, a constant  $V_p$  applied to node  $N_6$  transforms the graded potential into a spike-duration-based programming voltage, as shown in Figure 4e, whereas a clocking signal toggling between  $V_p$  and 0 V with  $\tau_{\text{CLK}} = 100$  ms applied to node  $N_6$  transforms the graded potential into a spike-count-based programming voltage, as shown in Figure 4f (see Supporting Information 12 for the

biasing configuration of the EM for spike-duration- and spike-count-based encoding). Figure 4g and Figure 4i, respectively, show the total spiking time ( $\tau_{\text{spike}}$ ) and the total number of spikes ( $N_{\text{spike}}$ ) as a function of  $V_{\text{write}}$  and  $I_{\text{LED}}$  when the input stimulus is presented for a duration of 10 s and  $V_p = -6$  V. Note that we start to count the spike time and spike number once  $V_{\text{N5}}$  reaches 75% of  $V_p$ , i.e.,  $-4.5$  V in Figure 4e-f. As expected, for any given  $V_{\text{write}}$  graded potentials received from the SM module that correspond to higher values of input stimuli ( $I_{\text{LED}}$ ) invoke longer  $\tau_{\text{spike}}$  and more  $N_{\text{spike}}$  at the output of the EM for spike-duration- and spike-count-based encodings, respectively. Similarly, for any given  $I_{\text{LED}}$ , more negative  $V_{\text{write}}$  invokes longer  $\tau_{\text{spike}}$  and more  $N_{\text{spike}}$ . Note that  $\tau_{\text{spike}}$  and  $N_{\text{spike}}$  can also be controlled by configuring  $V_{\text{IT}}$  (Figure 4d). For example, the scotopic condition will benefit from lower  $V_{\text{IT}}$  since spikes will reach the programming voltage,  $V_p$ , earlier for any given  $I_{\text{LED}}$  and  $V_{\text{write}}$ .

Alternatively, a higher  $V_p$  value can be used to encode a shorter duration or lower number of programming spikes. See Supporting Information 13 for encoding of the same graded potential using different  $V_p$ 's for both spike-duration- and



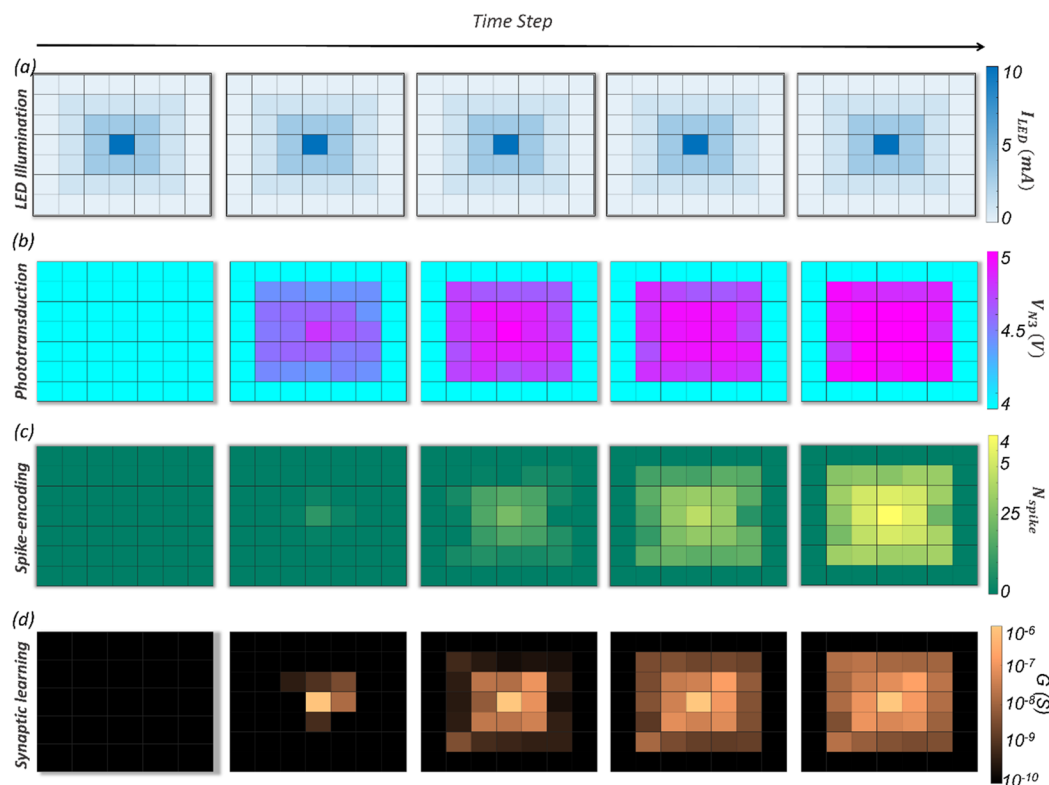
**Figure 6.** MoS<sub>2</sub>-FET-based neuromorphic learning module (LM). a) Spike-duration- and b) spike-count-based conductance evolution in the MoS<sub>2</sub>-FET-based LM when input programming waveforms ( $V_{NS}$ ) are received from the EM corresponding to different  $I_{LED}$ 's and  $V_{write}$ 's. Final conductance state achieved by the LM for c) spike-duration- and d) spike-count-based input spiking patterns received from the EM corresponding to different  $I_{LED}$ 's and  $V_{write}$ 's. The average learning energy expenditure ( $E_{LM}$ ) per clock cycle by the LM for e) spike-duration- and f) spike-count-based learning at different  $I_{LED}$ 's and  $V_{write}$ 's.

spike-count-based adaptive encoding. The reconfigurability of the EM can also be exploited for modeling learning disabilities. For example, if bright light is encoded into low-magnitude  $V_p$  spikes, potentiation of synapses can be severely limited, invoking learning difficulty. Finally, the average encoding energy expenditure ( $E_{EM}$ ) per clock cycle by the EM, given by  $E_{EM} = \frac{1}{2}C_{ox}V_{N3}^2 + V_pI_p\tau_{CLK}$ , is shown in Figure 4h and Figure 4j for both spike-count and spike-duration, respectively. The relatively higher energy expenditure for the EM is a direct consequence of using a depletion mode NMOS inverter and can be reduced significantly by using a CMOS inverter. This will require the development of p-channel MoS<sub>2</sub> or the use of another 2D material such as WSe<sub>2</sub>.

**MoS<sub>2</sub> FET-Based Neuromorphic Learning Module (LM).** Optical information encoded in spikes is delivered from the EM to the LM for pattern learning using memory augmented reinforcement. Each learning module comprises one MoS<sub>2</sub> FET ( $T_{LM}$ ), as shown in Figure 1e, which serves as a nonvolatile synapse with analog conductance states programmable by applying electrical voltage spikes to the local back-

gate terminal,  $N_s$ , which also serves as the output terminal of the EM. Figure 5a–l show that MoS<sub>2</sub>-FET-based synapses allow both spike-count- and spike-duration-based nonvolatile programming and can achieve both potentiation and depression analogous to chemical synapses in BNNs with low programming energy expenditure.

Figure 5a shows the potentiation of a representative MoS<sub>2</sub> electrical synapse from a low conductance state (LCS) after the application of a fixed number of programming spikes ( $N_{spike} = 10$ ) of different amplitudes and negative polarity ( $V_p$ ), with each spike being applied for  $t_{spike} = 100$  ms. Figure 5b shows the postpotentiated conductance states ( $G_p$ ) measured at  $V_{BG} = 0$  V as a function of  $N_{spike}$  for different  $V_p$ 's. As expected, a lower  $N_{spike}$  value invokes lower potentiation, i.e., smaller change in  $G_p$ , and *vice versa*, which can be exploited for spike-count-based learning. Similarly, Figure 5c shows the depression of a potentiated MoS<sub>2</sub> synapse, i.e., from a high conductance state (HCS) to an LCS, after the application of a fixed number of programming spikes ( $N_{spike} = 10$ ) of different amplitudes and positive polarity ( $V_D$ ), with each spike again being applied for  $t_{spike} = 100$  ms. Figure 5d shows the



**Figure 7.** Multipixel demonstration of analog image sensing, encoding, and learning. a) Analog  $7 \times 7$  input pattern obtained by illuminating the blue LED. Temporal evolution of corresponding b) graded potential ( $V_{N3}$ ) at the output of the SMs, c) programming spike-count ( $N_{\text{spike}}$ ) at the output of the EMs, and d) programmed conductance values at the output of LMs. Clearly, the input LED pattern is learned by our  $7 \times 7$  BNN hardware. For this demonstration, all  $\text{MoS}_2$ -FET-based synapses belonging to the LMs were initially programmed in their LCS (100 pS), and different LED illuminations were presented one by one to the corresponding pixels of our  $7 \times 7$  BNN hardware.

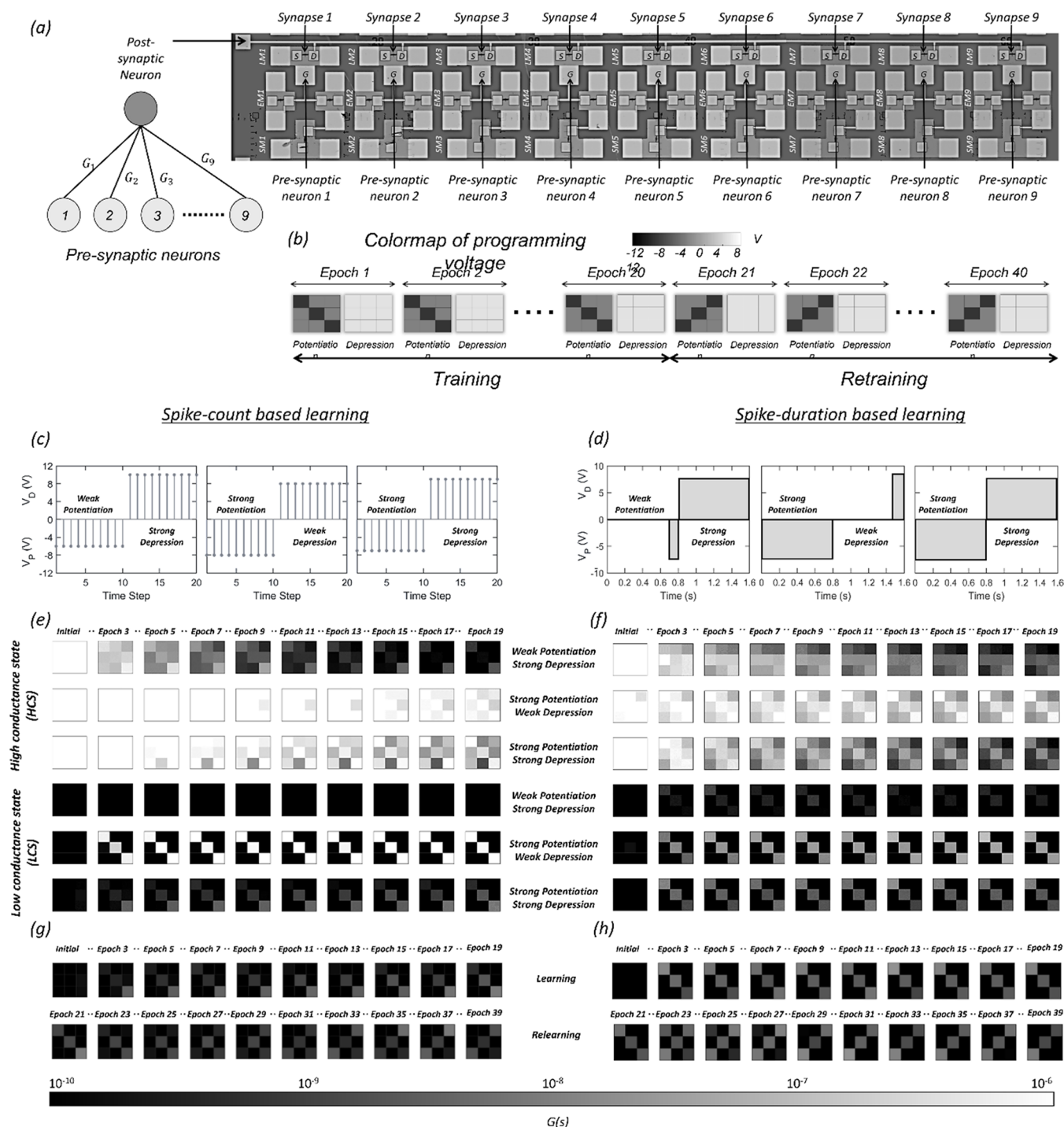
postdepressed conductance states ( $G_D$ ) measured at  $V_{BG} = 0$  V as a function of  $N_{\text{spike}}$  for different  $V_D$ 's. As expected, a lower  $N_{\text{spike}}$  value invokes lower depression and *vice versa*, which can be exploited for spike-count-based forgetting. As we will elucidate later, forgetting capabilities enable relearning of new patterns using the same synapses that have learned in a previous pattern. Also, note that a smaller  $N_{\text{spike}}$  value can achieve higher potentiation/depression if encoded using a higher  $V_{P/D}$ . As mentioned earlier, this aspect can be exploited to achieve learning plasticity.

Figure 5e–h show the spike-duration-based potentiation and depression of  $\text{MoS}_2$  synapses. Figure 5e shows the potentiation of an  $\text{MoS}_2$  synapse from an LCS after the application of a single spike of constant magnitude  $V_P = -6$  V for different  $t_{\text{spike}}$ 's, and Figure 5f shows the postpotentiated  $G_P$  measured at  $V_{BG} = 0$  V as a function of  $t_{\text{spike}}$  for different  $V_P$ 's. Similarly, Figure 5g shows the depression of an  $\text{MoS}_2$  synapse from an HCS after the application of a single spike of constant magnitude  $V_D = 6$  V for different  $t_{\text{spike}}$ 's, and Figure 5h shows the postdepressed  $G_D$  measured at  $V_{BG} = 0$  V as a function of  $t_{\text{spike}}$  for different  $V_D$ 's. Here, shorter  $t_{\text{spike}}$  invokes lower potentiation/depression and *vice versa*, which can be used for spike-duration-based learning/forgetting. Note that similar to spike-count-based learning/forgetting, higher potentiation/depression can be achieved for shorter spike durations when encoded using a higher magnitude of  $V_{P/D}$ , thus enabling spike-duration-based learning plasticity under scotopic conditions.

The underlying mechanism behind the spike-count- and spike-duration-based potentiation and depression of  $\text{MoS}_2$  synapses can be explained using the shift in  $V_{TH}$  observed in

the transfer characteristics of  $\text{MoS}_2$  FETs. The  $V_{TH}$  shift is attributed to charge trapping/detrapping at and near the  $\text{MoS}_2/\text{Al}_2\text{O}_3$  interface, which is also responsible for the photogating effect described earlier. In our previous work,<sup>48</sup> we performed a bias-temperature instability (BTI) test to confirm the charge trapping in the dielectric. A high magnitude of  $V_P$  and  $V_E$  is required for the charge trapping and detrapping process since a minimal hysteresis loop is observed in as-fabricated, postdepressed, and postpotentiated devices for a narrow gate voltage range, as shown in Supporting Information 14. Interestingly, the trapping and detrapping processes were found to be nonvolatile as evident from the retention measurements displayed in Figure 5i–j for 6 representative potentiated ( $G_P$ ) and depressed ( $G_D$ ) conductance states, respectively, measured over 100 s. We also examined long-term memory retention characteristics of two representative postprogrammed analog conductance states for  $\sim 10^4$  seconds, as shown in Supporting Information 15. The memory ratio (MR) between these two states was found to change from  $\sim 1.1 \times 10^2$  to  $0.6 \times 10^2$  following an exponential decay with a time constant of  $1.6 \times 10^4$  seconds. The projected time before the two states become indistinguishable, or MR reaches 1, was found to be  $\sim 1$  day. Note that, while conventional memory devices require nonvolatile retention for years, many neuromorphic applications including those used by edge devices and sensors relax the requirement for long-term retention and can be well served with short-term memory retention of several hours to days. The retention window demonstrated by the  $\text{MoS}_2$  FETs was adequate for the successful realization of our proof-of-concept “all-in-one” BNN. Certainly, it is desirable and





**Figure 8.** Importance of forgetting (synaptic depression) in learning and inference. a) Schematic and optical image of a 2-layer BNN with 9 presynaptic neurons and 1 postsynaptic neuron for learning and inferring patterns from  $3 \times 3$  pixelated images. b) Training and retraining schedule with  $M = 40$  epochs, with each epoch having potentiation and depression cycles. During the potentiation, the pattern to be learned is presented to the BNN, whereas during the depression, all synapses are uniformly depressed. Spiking profiles used for c) spike-count- and d) spike-duration-based learning. For each type of learning, three BNN configurations are used: 1) weak potentiation and strong depression, 2) strong potentiation and weak depression, and 3) strong potentiation and strong depression. The strength of potentiation ( $V_P$ ) and depression ( $V_D$ ) is adjusted using the spike magnitude and spike duration for spike-count- and spike-duration-based learning, respectively. The time evolution of the colormap of synaptic weights, i.e., the conductance states of the 9 synapses during e) spike-count- and f) spike-duration-based learning. For each type of learning, all synapses are initialized either in an HCS where  $G_{HCS} = 100$  nS or in an LCS where  $G_{LCS} = 100$  pS (also see the [Supplementary Videos 3 and 4](#)). Learning of the left diagonal is followed by relearning of the right diagonal when potentiation and depression are both strong for g) spike-count- and h) spike-duration-based learning (also see the [Supplementary Videos 5 and 6](#)).

possible to improve the memory retention window by optimizing the design of the local back-gate stack, e.g., by

mimicking the floating-gate architecture used by conventional FLASH memory devices.

The device-to-device variation in the pre- and postprogrammed transfer characteristics and the corresponding color-map of MR measured at  $V_{BG} = 0$  V for 49 monolayer MoS<sub>2</sub> FETs from each LM of our  $7 \times 7$  BNN platform, when programmed using  $N_{\text{spike}} = 10$  with spike magnitude,  $V_p = -8$  V, and spike width,  $t_{\text{spike}} = 100$  ms, are shown in Figure S5-k-l, respectively (see Supporting Information 16 for the pre- and postprogrammed transfer characteristics for each of these 49 MoS<sub>2</sub> FETs). The mean and standard deviation values for MR were found to be  $6 \times 10^5$  and  $0.5 \times 10^5$ , respectively. In our work, we have characterized the device-to-device variation in the electrical memory and in the persistent photocurrent (PPC) observed in MoS<sub>2</sub> FETs and correlated these effects with the material characterization.

Finally, Figure 6a-b, respectively, show the spike-duration- and spike-count-based conductance evolution in the MoS<sub>2</sub> FET-based LM when input programming waveforms ( $V_{NS}$ ) are received from the EM corresponding to different  $I_{LED}$ 's and  $V_{\text{write}}$ 's. As shown in Figure 6c-d, for any given  $V_{\text{write}}$ , the spiking patterns received from the EM that correspond to higher values of input stimuli ( $I_{LED}$ ) result in higher values of final conductance, and *vice versa*, for both spike-duration- and spike-count-based learning, respectively. Similarly, for any given  $I_{LED}$ , more negative  $V_{\text{write}}$  results in better learning, i.e., higher final conductance. The average learning energy expenditure ( $E_{LM}$ ) per clock cycle by the LM, given by  $E_{LM} = \frac{1}{2} C_{ox} V_{NS}^2$ , for both spike-duration- and spike-count-based learning is shown in Figure 6e-f, respectively. The energy expenditure for the LM was found to be miniscule at  $\sim 50$  fJ per clock cycle even for the brightest illumination and most negative  $V_{\text{write}}$ .

**Multipixel Demonstration of Analog Image Sensing, Encoding, and Learning.** Figure 7a-d and Supplementary Video 2 show a complete demonstration of our BNN hardware involving the multipixel and monolithically integrated SM, EM, and LM. A  $7 \times 7$  analog input pattern obtained by illuminating the LED (Figure 7a) is transduced into corresponding graded potentials using the SMs (Figure 7b) and encoded into corresponding programming spikes following the spike-count-based encoding algorithm by the EMs (Figure 7c), which are subsequently used by the LMs to potentiate the MoS<sub>2</sub>-FET-based nonvolatile synapses (Figure 7d). Clearly, the input LED pattern is learned by our  $7 \times 7$  BNN hardware. For this demonstration, all synapses were initially programmed in their LCS, and different LED illuminations were presented one by one to the corresponding pixels of our  $7 \times 7$  BNN hardware. For simultaneous illumination, a lensing system will be needed to focus the image pixels onto the corresponding SMs of our BNN platform. In our future endeavors, we will attempt to integrate the lensing system with the BNN hardware.

Note that MoS<sub>2</sub> FETs used in the SMs are biased in the deep off-state by applying negative  $V_{\text{write}}$  to harness the photogating effect, resulting in the transduction of optical illuminations into corresponding graded potentials. However, the MoS<sub>2</sub> FETs used in the EM and LM are biased either in the subthreshold or in the on-state, where these devices remain insensitive to illumination and hence their operation is not impacted by illumination. Also note that, while the input pattern is learned by our BNN architecture, device-to-device variation in the photogating effect, transfer characteristics, and programming of MoS<sub>2</sub> FETs are translated into variation in the graded potential, spike-count, and learned conductance values

corresponding to the same input LED signal, as seen in Figure 7. There is no doubt that further reduction in device-to-device variation is desirable. As MoS<sub>2</sub> technology matures further through the optimization of growth conditions to reduce point defects, cleaner and damage-free techniques are developed for large area transfer, and polymer residues are eliminated from device fabrication processes, it will be possible to mitigate device-to-device variation to a larger extent and achieve near-ideal learning. Nevertheless, our proof-of-concept demonstration highlights the fully-integrated nature of our MoS<sub>2</sub>-FET-based hardware BNN that combines sensing, computing (encoding), and storage and thereby distinguishes it from other hardware BNN architectures based on CMOS or emerging technologies such as RRAM, PCM, memristor, and all-optic, as well as hybrid, approaches.

### Importance of Forgetting in Learning and Inference.

Forgetting has traditionally been considered to be a passive brain process, which ensures unused information fades over time so that neural resources can be reallocated for storing newer and more important information. When machines learn with unrestricted storage resources (e.g., cloud servers), forgetting is irrelevant. However, when storage capacity is either limited or not accessible, for example, in the Internet of things (IoT) edge devices deployed in remote locations, forgetting can play an active role in smart learning. Here, we demonstrate the role of forgetting in relearning without any external supervision and by directly interacting with the changing environment.

Figure 8a shows the schematic and optical image of a fully connected 2-layer BNN with 9 presynaptic input neurons and 1 postsynaptic output neuron for learning and inferring patterns from  $3 \times 3$  pixelated images (see Supporting Information 17 for an enlarged version of the optical image). Figure 8b shows the training and retraining schedule consisting of  $M = 40$  epochs, with each epoch having two cycles: potentiation and depression. See Supporting Information 18 for the biasing configuring of the EM for introducing the depression cycle. To introduce depression in the EM,  $V_D$  is applied to the drain terminal of  $T_{EM1}$ , i.e.,  $N_4$ , instead of keeping it grounded with clocking profiles as shown in Figure S17a-b for spike-duration- and spike-count-based encoding, respectively. Figure S17c-d, respectively, show the output ( $V_{NS}$ ) of the EM at a constant graded potential,  $V_{N3} = 5$  V, and under various combinations of  $V_D$  and  $V_p$  for spike-duration- and spike-count-based encoding, respectively. During the potentiation cycle, the image pattern to be learned is presented to the corresponding synapses of the  $9 \times 1$  BNN, whereas during the depression cycle, all 9 synapses are uniformly depressed. The first pattern (left diagonal) is presented for 20 epochs followed by the second pattern (right diagonal) for another 20 epochs to test whether our BNN can forget previously learned patterns and relearn new patterns. Figure 8c-d, respectively, show the spiking profiles used for spike-count- and spike-duration-based learning. For each type of learning, we consider three configurations of the BNN: 1) weak potentiation and strong depression, 2) strong potentiation and weak depression, and 3) strong potentiation and strong depression. For spike-count-based learning, the strength of potentiation ( $V_p$ ) and depression ( $V_D$ ) is adjusted using the spike magnitude, for example,  $V_p = -10$  V for strong and  $V_p = -8$  V for weak potentiation and  $V_D = 12$  V for strong and  $V_D = 10$  V for weak depression. Similarly, for the pattern to be learned, each pixel in the  $3 \times 3$  images is encoded with  $N_{\text{spike}} =$

10 if it is bright and  $N_{\text{spike}} = 0$  if it is dark. For spike-duration-based learning, the strength of potentiation and depression is adjusted using the spike duration, i.e.,  $t_{\text{spike}} = 800$  ms for strong and  $t_{\text{spike}} = 100$  ms for weak potentiation/depression; for the pattern to be learned, each pixel in the  $3 \times 3$  images is encoded with the appropriate  $t_{\text{spike}}$  (weak/strong) if it is bright and  $t_{\text{spike}} = 10$  ms if it is dark.

Figure 8e–f, respectively, show the time evolution of the colormap of synaptic weights, i.e., the conductance states of the 9 synaptic devices, during the spike-count- and spike-duration-based learning cycles. For each type of learning, all synapses are initialized either in an HCS, with  $G_{\text{HCS}} = 100$  nS, or in an LCS, with  $G_{\text{LCS}} = 100$  pS (also see [Supplementary Videos 3 and 4](#)). Following are the key observations. When potentiation is weak but depression is strong, it is difficult to learn irrespective of the initial state of the synapses; however, when potentiation is strong but depression is weak, learning from the LCS is fast, but forgetting, and thus relearning, from the HCS is slow. This is expected since synapses that are potentiated get stuck in their HCS owing to weak depression making it difficult for them to forget their respective states. Finally, if both potentiation and depression are strong, learning and forgetting become faster irrespective of the initial synaptic state. This is demonstrated in Figure 8g–h, which show learning of the left diagonal followed by relearning of the right diagonal when potentiation and depression are both strong for spike-count- and spike-duration-based learning, respectively (also see [Supplementary Videos 5 and 6](#)). Our findings indicate that the relative strengths of potentiation and depression play a critical role in learning using our BNN. This is similar to natural BNNs; for example, autism spectrum disorder (ASD), which includes a broad range of conditions such as challenges with learning social skills, repetitive behaviors, etc., has been related to dysregulation or a deficit in long-term depression in several mouse models.<sup>49,50</sup> Therefore, our hardware BNN platform offers an opportunity to bridge the gap between the neuroscience of learning and machine learning. [Supporting Information 19](#) shows inference using our BNN architecture. We have used a  $9 \times 2$  fully-connected neural network implemented using two sets of  $9 \times 1$  synapses, as shown in [Figure S18a](#). The synapses between the 9 presynaptic neurons and the “Yes” postsynaptic neuron are trained with the actual pattern, whereas the synapses between the 9 presynaptic neurons and the “No” postsynaptic neuron are trained with the inverse of the pattern to obtain the respective conductance maps ( $G_{i-\text{Yes/No}}$ ,  $i = 1, 2, 3, \dots, 8, 9$ ), as shown in [Figure S18b](#). Any input pattern from the LED is converted to corresponding graded potentials by the SMs and transduced into spike trains by the EMs. For spike-count-based inference, the output voltage spikes ( $V_{ij}$ ,  $i = 1, 2, 3, \dots, 8, 9$ ;  $j = 1, 2, 3, \dots, N_{\text{spike}}$ ) obtained at the output of the encoding module and corresponding to each pixel of the  $3 \times 3$  image are applied to the drain terminals of the 9 presynaptic neurons. The output currents from the common source terminal, i.e., postsynaptic “Yes” and “No” neurons, are integrated using capacitors ( $C_{\text{Yes/No}}$ ) to obtain  $V_{\text{Yes}}$  and  $V_{\text{No}}$ , as shown in [Figure S18c](#) and in [eq 1](#).

$$V_{\text{Yes}} = \frac{t_{\text{spike}}}{C_{\text{Yes}}} \sum_{i=1}^9 \sum_{j=1}^{N_{\text{spike}}} G_{i-\text{Yes}} V_{ij} \quad V_{\text{No}} = \frac{t_{\text{spike}}}{C_{\text{No}}} \sum_{i=1}^9 \sum_{j=1}^{N_{\text{spike}}} G_{i-\text{No}} V_{ij} \quad (1)$$

For spike-duration-based inference, a similar approach is adopted, except for the fact that only one voltage spike ( $V_i$ ,  $i =$

1, 2, 3, ..., 8, 9) is obtained at the output of the encoding module, corresponding to each pixel of the  $3 \times 3$  image with different spiking durations. In this case,  $V_{\text{Yes}}$  and  $V_{\text{No}}$  are given by [eq 2](#).

$$V_{\text{Yes}} = \frac{1}{C_{\text{Yes}}} \sum_{i=1}^9 \int_0^{t_{\text{spike}}} G_{i-\text{Yes}} V_i \quad V_{\text{No}} = \frac{1}{C_{\text{No}}} \sum_{i=1}^9 \int_0^{t_{\text{spike}}} G_{i-\text{No}} V_i \quad (2)$$

For the “Yes” neuron to be a winner,  $V_{\text{Yes}} > V_{\text{No}}$  and  $V_{\text{Yes}} \geq V_{\text{Win}}$ , where  $V_{\text{Win}}$  is the winning threshold determined by the learned pattern. Clearly, the “Yes” neuron should be the winner only when the pattern similar to the learned one is inferred, whereas the “No” neuron should win for all other patterns. However, the experimental inference accuracy was found to be  $\sim 96\%$ . This is because the patterns which contain one or two off-diagonal pixels in addition to the diagonal pixels also make the “Yes” neuron the winner. There are a total of  ${}^6C_1 + {}^6C_2 = 21$  such patterns, which accounts for  $\sim 4\%$  of all  $2^9 = 512$  patterns that are wrongly inferred. Note that if 3 or more pixels in addition to the diagonal pixels are bright, the “No” neuron wins. The inference accuracy was improved to 100% by making  $V_{\text{No}} \geq V_{\text{Win}}$  even when only one off-diagonal pixel is present in the input pattern. This was accomplished through greater potentiation of the synaptic connections between the input neurons and the “No” neuron during training, with the inverse pattern resulting in an order of magnitude higher learned conductance value.

## CONCLUSION

In conclusion, we have experimentally demonstrated a fully integrated, multipixel, and biomimetic BNN hardware platform based on monolayer  $\text{MoS}_2$  that combines sensing, encoding, learning, and inference. We have employed both spike-count- and spike-duration-based encoding, learning, and inference inspired by the energy efficiency of spike-based computing in the brain. Similarly, we were able to show adaptive learning in photopic and scotopic conditions and the impact of the relative strengths of synaptic potentiation and depression on learning and forgetting. Our accomplishments can be attributed to the photoresponse of monolayer  $\text{MoS}_2$ -based phototransistors for sensing,  $\text{MoS}_2$ -based neuromorphic circuit modules for encoding, and programmable and nonvolatile  $\text{MoS}_2$  synapses enabled by our local back-gate memory stack for unsupervised and adaptive learning. Our findings highlight the potential of in-memory computing and sensing based on emerging 2D materials, devices, and circuits that not only overcome the bottleneck of von Neumann computing in conventional CMOS designs but also aid in eliminating peripheral components necessary for competing technologies such as memristors, RRAM, PCM, etc. We believe that our  $\text{MoS}_2$ -based low-power and fully integrated hardware BNN system is more biorealistic in terms of functionality, organization, and plasticity of BNN and, therefore, can not only accelerate the development of hardware artificial intelligence (AI) and benefit edge computing and smart sensing for the Internet of Things (IoT) but also offer a platform for adaptive learning and for modeling plasticity-related learning disorders in natural BNNs.

## METHODS

**Fabrication of Local Back-Gate Islands.** To define the back-gate island regions, the substrate (285 nm  $\text{SiO}_2$  on  $\text{p}^{++}\text{-Si}$ ) was spin-coated with a bilayer photoresist consisting of Lift-Off-Resist (LOR 5A) and Series Photoresist (SPR 3012) baked at 185 and 95 °C, respectively. The bilayer photoresist was then exposed using a



Heidelberg Maskless Aligner (MLA 150) to define each island and subsequently developed using MF CD26 microposit, followed by a deionized (DI) water rinse. The back-gate electrode of 20/50 nm TiN/Pt was deposited using reactive sputtering. The photoresist was removed using acetone and Photo Resist Stripper (PRS 3000) and cleaned using 2-propanol (IPA) and DI water. An atomic layer deposition (ALD) process was then implemented to grow 50 nm  $\text{Al}_2\text{O}_3$  across the entire substrate, including the island regions. To access the individual Pt back-gate electrodes, etch patterns were defined using the same bilayer photoresist consisting of LOR 5A and SPR 3012. The bilayer photoresist was then exposed using the MLA 150 and developed using MF CD26 microposit. 50 nm  $\text{Al}_2\text{O}_3$  was subsequently dry-etched using a  $\text{BCl}_3$  reactive ion etch (RIE) chemistry at 5 °C for 20 s, which was repeated four times to minimize heating in the substrate. Next, the photoresist was removed to give access to the individual Pt electrodes.<sup>29,48</sup>

**Large Area Monolayer  $\text{MoS}_2$  Film Growth.** Monolayer  $\text{MoS}_2$  was deposited on an epi-ready 2'' c-sapphire substrate by metal-organic chemical vapor deposition (MOCVD). An inductively heated graphite susceptor equipped with wafer rotation in a cold-wall horizontal reactor was used to achieve uniform monolayer deposition as previously described.<sup>51</sup> Molybdenum hexacarbonyl ( $\text{Mo}(\text{CO})_6$ ) and hydrogen sulfide ( $\text{H}_2\text{S}$ ) were used as precursors.  $\text{Mo}(\text{CO})_6$  maintained at 10 °C and 950 Torr in a stainless-steel bubbler were used to deliver 0.036 sccm of the metal precursor for the growth, while 400 sccm of  $\text{H}_2\text{S}$  was used for the process.  $\text{MoS}_2$  deposition was carried out at 1000 °C and 50 Torr in  $\text{H}_2$  ambient, where monolayer growth was achieved in 18 min. The substrate was first heated to 1000 °C in  $\text{H}_2$  and maintained for 10 min before the growth was initiated. After growth, the substrate was cooled in  $\text{H}_2\text{S}$  to 300 °C to inhibit the decomposition of the  $\text{MoS}_2$  films. More details can be found in our earlier work.<sup>39,44,52</sup>

**$\text{MoS}_2$  Film Transfer to Local Back-Gate Islands.** To fabricate the  $\text{MoS}_2$  FETs, the MOCVD-grown monolayer  $\text{MoS}_2$  film was transferred from the sapphire growth substrate to the  $\text{SiO}_2/\text{p}^{++}\text{-Si}$  application substrate with local back-gate islands using a PMMA (polymethyl-methacrylate) assisted wet transfer process. First, the  $\text{MoS}_2$  on the growth substrate was spin-coated with PMMA and left over night to achieve good adhesion. The corners of the spin-coated film were then scratched using a razor blade and immersed inside a 2 M NaOH solution kept at 90 °C. Capillary action caused the NaOH to be drawn into the substrate/film interface, separating the PMMA/ $\text{MoS}_2$  film from the sapphire substrate. The separated film was rinsed multiple times inside a water bath and transferred onto the  $\text{SiO}_2/\text{p}^{++}\text{-Si}$  substrate with local back-gate islands. The substrate was then baked at 50 and 70 °C for 10 min each to remove moisture and promote adhesion. Finally, the PMMA supporting layer was removed by immersing the substrate in acetone and the substrate was cleaned using an IPA bath.<sup>29,48</sup>

**Fabrication of Monolayer  $\text{MoS}_2$  FET.** To define the channel regions of the  $\text{MoS}_2$  FETs, the substrate was spin-coated with PMMA and baked at 180 °C for 90 s. The resist was then patterned using electron beam (e-beam) lithography and developed using a 1:1 mixture of 4-methyl-2-pentanone (MIBK) and 2-propanol (IPA) for 60 s and an IPA rinse for 45 s. The monolayer  $\text{MoS}_2$  film was subsequently etched using a sulfur hexafluoride ( $\text{SF}_6$ ) RIE chemistry at 5 °C for 30 s. Next, the sample was rinsed in acetone and IPA to remove the e-beam resist. To define the source and drain contacts, the sample was then spin-coated with methyl methacrylate (MMA) followed by A3 PMMA. Then, e-beam lithography was again used to pattern the source/drain contacts and development was again conducted using a 1:1 mixture of MIBK and IPA for 60 s and IPA for 45 s. 40 nm nickel (Ni) and 30 nm gold (Au) were deposited using e-beam evaporation to act as the contact metals. Finally, a lift-off process was performed to remove the evaporated Ni/Au except from the contact regions by immersing the sample in acetone for 30 min followed by IPA for another 30 min. In the final design, each island contains one  $\text{MoS}_2$  FET to allow for individual gate control.<sup>29,48</sup>

**Monolithic Integration.** Each pixel of our multipixel ( $7 \times 7$ ) BNN hardware consists of 4  $\text{MoS}_2$  FETs, as shown using the circuit

schematic in Figure 1e. Within each pixel, the SM consists of 1  $\text{MoS}_2$  FET ( $T_{\text{SM}}$ ), the EM consists of 2  $\text{MoS}_2$  FETs ( $T_{\text{EM1}}$  and  $T_{\text{EM2}}$ ), and the LM consists of 1  $\text{MoS}_2$  FET ( $T_{\text{LM}}$ ). To fabricate the connections between the respective nodes of  $T_{\text{SM}}$ ,  $T_{\text{EM1}}$ ,  $T_{\text{EM2}}$ , and  $T_{\text{LM}}$ , the substrate was spin-coated with MMA and PMMA, the e-beam lithography and development processes previously described were used to define the connections, and e-beam evaporation was used to deposit 60 nm of Au. Finally, the e-beam resist was rinsed away by the same lift-off process mentioned previously.<sup>29,48</sup>

**Electrical Characterization.** Electrical characterization of the fabricated devices was performed using a Lake Shore CRX-VF probe station under atmospheric conditions with a Keysight B1500A parameter analyzer.

**Code Availability.** The codes used for plotting the data are available from the corresponding authors on reasonable request.

## ASSOCIATED CONTENT

### Data Availability Statement

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsnano.2c02172>.

Video 1 (MOV)

Video 2 (MOV)

Video 3 (MP4)

Video 4 (MP4)

Video 5 (MP4)

Video 6 (MP4)

Optical image of all-in-one multipixel chip (1 cm  $\times$  1 cm), optical image of  $7 \times 7$  pixel SNN platform, optical image of one pixel, device-to-device variation in Raman spectra, PL spectra, transfer characteristics of monolayer  $\text{MoS}_2$  FET across  $7 \times 7$  pixel, output characteristics of  $\text{MoS}_2$  FET, calibration of input optical power, optical images of different LED brightness, postillumination transfer characteristics of  $\text{MoS}_2$  FET for different exposure time and back-gate bias, device-to-device variation in photoresponse, biasing configuration of encoding module for both spike-timing and spike count, graded potential from sensing module corresponding to different light intensity, hysteresis loop for as fabricated, postpotentiated and postdepressed, long-term retention of two conductance states, device-to-device variation in programmability of monolayer  $\text{MoS}_2$  FET, optical image of fully connected 2-layer network, biasing configuration of encoding module for introducing both potentiation and depression cycle, and inference using our network architecture (PDF)

## AUTHOR INFORMATION

### Corresponding Author

Saptarshi Das — Department of Engineering Science and Mechanics, Department of Materials Science and Engineering, Materials Research Institute, and Department of Electrical Engineering and Computer Science, Pennsylvania State University, University Park, Pennsylvania 16802, United States; [orcid.org/0000-0002-0188-945X](https://orcid.org/0000-0002-0188-945X); Email: [sud70@psu.edu](mailto:sud70@psu.edu), [das.sapt@gmail.com](mailto:das.sapt@gmail.com)

### Authors

Shiva Subbulakshmi Radhakrishnan — Department of Engineering Science and Mechanics, Pennsylvania State

University, University Park, Pennsylvania 16802, United States

Akhil Dodda – Department of Engineering Science and Mechanics, Pennsylvania State University, University Park, Pennsylvania 16802, United States

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acsnano.2c02172>

## Author Contributions

S.D. and S.S.R. conceived the idea and designed the experiments. S.D., S.S.R., and A.D. performed the experiments, analyzed the data, discussed the results, and agreed on their implications. All authors contributed to the preparation of the manuscript.

## Notes

Preprint: Shiva Subbulakshmi Radhakrishnan; Akhil Dodda; Saptarshi Das. An All-in-one Biomimetic 2D Spiking Neural Network. **2021**, *Research Square*.[10.21203/rs.3.rs-249741/v1](https://doi.org/10.21203/rs.3.rs-249741/v1) (02 19, 2021). Shiva Subbulakshmi Radhakrishnan; Akhil Dodda; Saptarshi Das. An All-in-one Bioinspired Neural Network. **2021**, *Research Square*.[10.21203/rs.3.rs-969671/v1](https://doi.org/10.21203/rs.3.rs-969671/v1) (10 25, 2021).

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The work was supported by the Army Research Office (ARO) through Contract Number W911NF1920338 and the National Science Foundation (NSF) through CAREER Award under Grant Number ECCS-2042154. The authors also acknowledge Mr. Amritanand Sebastian and Mr. Thomas F. Schranghamer for help with device fabrication. The authors also acknowledge the materials support from the National Science Foundation (NSF) through the Pennsylvania State University 2D Crystal Consortium—Materials Innovation Platform (2DCCMIP) under NSF cooperative agreement DMR-1539916.

## REFERENCES

- (1) Hassoun, M. H. *Fundamentals of artificial neural networks*; MIT Press: Cambridge, 1995.
- (2) Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; Chen, Y.; Lillicrap, T.; Hui, F.; Sifre, L.; van den Driessche, G.; Graepel, T.; Hassabis, D. Mastering the game of Go without human knowledge. *Nature* **2017**, *550* (7676), 354–359.
- (3) Glackin, B.; McGinnity, T. M.; Maguire, L. P.; Wu, Q.; Belatreche, A. In A novel approach for the implementation of large scale spiking neural networks on FPGA hardware. *International Work-Conference on Artificial Neural Networks*, Springer, Barcelona, Spain; 2005; pp 552–563, DOI: [10.1007/11494669\\_68](https://doi.org/10.1007/11494669_68).
- (4) Wang, Z.; Joshi, S.; Savel'ev, S.; Song, W.; Midya, R.; Li, Y.; Rao, M.; Yan, P.; Asapu, S.; Zhuo, Y. Fully memristive neural networks for pattern classification with unsupervised learning. *Nature. Electronics* **2018**, *1* (2), 137–145.
- (5) Al-Shedivat, M.; Naous, R.; Cauwenberghs, G.; Salama, K. N. Memristors empower spiking neurons with stochasticity. *IEEE journal on Emerging and selected topics in circuits and systems* **2015**, *5* (2), 242–253.
- (6) Milo, V.; Pedretti, G.; Carboni, R.; Calderoni, A.; Ramaswamy, N.; Ambrogio, S.; Ielmini, D. In Demonstration of hybrid CMOS/RRAM neural networks with spike time/rate-dependent plasticity. *2016 IEEE International Electron Devices Meeting (IEDM)*; IEEE: 2016; pp 16.8.1–16.8.4, DOI: [10.1109/IEDM.2016.7838435](https://doi.org/10.1109/IEDM.2016.7838435).
- (7) Kim, S.; Ishii, M.; Lewis, S.; Perri, T.; BrightSky, M.; Kim, W.; Jordan, R.; Burr, G.; Sosa, N.; Ray, A. In NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning. *2015 IEEE international electron devices meeting (IEDM)*; IEEE: 2015; pp 17.1.1–17.1.4, DOI: [10.1109/IEDM.2015.7409716](https://doi.org/10.1109/IEDM.2015.7409716).
- (8) Boybat, L.; Le Gallo, M.; Nandakumar, S.; Moraitis, T.; Parnell, T.; Tuma, T.; Rajendran, B.; Leblebici, Y.; Sebastian, A.; Eleftheriou, E. Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **2018**, *9* (1), 2514.
- (9) Ambrogio, S.; Ciocchini, N.; Laudato, M.; Milo, V.; Pirovano, A.; Fantini, P.; Ielmini, D. Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses. *Frontiers in neuroscience* **2016**, *10*, 56.
- (10) Chu, M.; Kim, B.; Park, S.; Hwang, H.; Jeon, M.; Lee, B. H.; Lee, B.-G. Neuromorphic hardware system for visual pattern recognition with memristor array and CMOS neuron. *IEEE Transactions on Industrial Electronics* **2015**, *62* (4), 2410–2419.
- (11) Li, C.; Hu, M.; Li, Y.; Jiang, H.; Ge, N.; Montgomery, E.; Zhang, J.; Song, W.; Dávila, N.; Graves, C. E. Analogue signal and image processing with large memristor crossbars. *Nature Electronics* **2018**, *1* (1), 52.
- (12) Yin, Z.; Li, H.; Li, H.; Jiang, L.; Shi, Y.; Sun, Y.; Lu, G.; Zhang, Q.; Chen, X.; Zhang, H. Single-layer MoS<sub>2</sub> phototransistors. *ACS Nano* **2012**, *6* (1), 74–80.
- (13) Wang, L.; Wang, Y.; Wong, J. I.; Palacios, T.; Kong, J.; Yang, H. Y. Functionalized MoS<sub>2</sub> nanosheet-based field-effect biosensor for label-free sensitive detection of cancer marker proteins in solution. *Small* **2014**, *10* (6), 1101–1105.
- (14) Park, M.; Park, Y. J.; Chen, X.; Park, Y. K.; Kim, M. S.; Ahn, J. H. MoS<sub>2</sub>-based tactile sensor for electronic skin applications. *Adv. Mater.* **2016**, *28* (13), 2556–2562.
- (15) Arnold, A. J.; Shi, T.; Jovanovic, I.; Das, S. Extraordinary Radiation Hardness of Atomically Thin MoS<sub>2</sub>. *ACS Appl. Mater. Interfaces* **2019**, *11* (8), 8391–8399.
- (16) Shen, P.-C.; Su, C.; Lin, Y.; Chou, A.-S.; Cheng, C.-C.; Park, J.-H.; Chiu, M.-H.; Lu, A.-Y.; Tang, H.-L.; Tavakoli, M. M.; Pitner, G.; Ji, X.; Cai, Z.; Mao, N.; Wang, J.; Tung, V.; Li, J.; Bokor, J.; Zettl, A.; Wu, C.-I.; Palacios, T.; Li, L.-J.; Kong, J. Ultralow contact resistance between semimetal and monolayer semiconductors. *Nature* **2021**, *593* (7858), 211–217.
- (17) Smets, Q.; Arutchelvan, G.; Jussot, J.; Verreck, D.; Asselberghs, I.; Mehta, A. N.; Gaur, A.; Lin, D.; El Kazzi, S.; Groven, B. In Ultra-scaled MOCVD MoS<sub>2</sub> MOSFETs with 42nm contact pitch and 250μA/μm drain current. *2019 IEEE International Electron Devices Meeting (IEDM)*; IEEE: 2019; pp 23.2.1–23.2.4, DOI: [10.1109/IEDM19573.2019.8993650](https://doi.org/10.1109/IEDM19573.2019.8993650).
- (18) Rai, A.; Valsaraj, A.; Movva, H. C.; Roy, A.; Ghosh, R.; Sonde, S.; Kang, S.; Chang, J.; Trivedi, T.; Dey, R.; Guchhait, S.; Larentis, S.; Register, L. F.; Tutuc, E.; Banerjee, S. K. Air Stable Doping and Intrinsic Mobility Enhancement in Monolayer Molybdenum Disulfide by Amorphous Titanium Suboxide Encapsulation. *Nano Lett.* **2015**, *15* (7), 4329–36.
- (19) English, C. D.; Smithe, K. K. H.; Xu, R. L.; Pop, E. In Approaching ballistic transport in monolayer MoS<sub>2</sub> <inf>2</inf> transistors with self-aligned 10 nm top gates. *2016 IEEE International Electron Devices Meeting (IEDM)*, 3–7 Dec. 2016; 2016; pp 5.6.1–5.6.4, DOI: [10.1109/IEDM.2016.7838355](https://doi.org/10.1109/IEDM.2016.7838355).
- (20) Price, K. M.; Schauble, K. E.; McGuire, F. A.; Farmer, D. B.; Franklin, A. D. Uniform Growth of Sub-5-Nanometer High-κ Dielectrics on MoS<sub>2</sub> Using Plasma-Enhanced Atomic Layer Deposition. *ACS Appl. Mater. Interfaces* **2017**, *9* (27), 23072–23080.
- (21) Liu, H.; Grasseschi, D.; Dodda, A.; Fujisawa, K.; Olson, D.; Kahn, E.; Zhang, F.; Zhang, T.; Lei, Y.; Branco, R. B. N. Spontaneous chemical functionalization via coordination of Au single atoms on monolayer MoS<sub>2</sub>. *Science advances* **2020**, *6* (49), eabc9308.
- (22) 2DCC 2d-crystal-consortium. <https://www.mri.psu.edu/2d-crystal-consortium/user-facilities/thin-films/list-thin-film-samples-available> (accessed 2022-10).
- (23) Kang, K.; Xie, S.; Huang, L.; Han, Y.; Huang, P. Y.; Mak, K. F.; Kim, C.-J.; Muller, D.; Park, J. High-mobility three-atom-thick

semiconducting films with wafer-scale homogeneity. *Nature* **2015**, 520 (7549), 656.

(24) Wachter, S.; Polyushkin, D. K.; Bethge, O.; Mueller, T. A microprocessor based on a two-dimensional semiconductor. *Nat. Commun.* **2017**, 8, 14948.

(25) Polyushkin, D. K.; Wachter, S.; Mennel, L.; Paur, M.; Paliy, M.; Iannaccone, G.; Fiori, G.; Neumaier, D.; Canto, B.; Mueller, T. Analogue two-dimensional semiconductor electronics. *Nature Electronics* **2020**, 3 (8), 486–491.

(26) Gao, Q.; Zhang, Z.; Xu, X.; Song, J.; Li, X.; Wu, Y. Scalable high performance radio frequency electronics based on large domain bilayer MoS<sub>2</sub>. *Nat. Commun.* **2018**, 9 (1), 4778.

(27) Oberoi, A.; Dodda, A.; Liu, H.; Terrones, M.; Das, S. Secure Electronics Enabled by Atomically Thin and Photosensitive Two-Dimensional Memtransistors. *ACS Nano* **2021**, 15, 19815.

(28) Dodda, A.; Radhakrishnan, S. S.; Schranghamer, T. F.; Buzzell, D.; Sengupta, P.; Das, S. Graphene-based physically unclonable functions that are reconfigurable and resilient to machine learning attacks. *Nat. Electron* **2021**, 4 (5), 364–374.

(29) Dodda, A.; Trainor, N.; Redwing, J.; Das, S. All-in-one, bio-inspired, and low-power crypto engines for near-sensor security based on two-dimensional memtransistors. *Nat. Commun.* **2022**, 13 (1), 3587.

(30) Das, S.; Dodda, A.; Das, S. A biomimetic 2D transistor for audiomorphic computing. *Nat. Commun.* **2019**, 10 (1), 3450.

(31) Sebastian, A.; Pannone, A.; Radhakrishnan, S. S.; Das, S. Gaussian synapses for probabilistic neural networks. *Nat. Commun.* **2019**, 10 (1), 4199.

(32) Arnold, A. J.; Razavieh, A.; Nasr, J. R.; Schulman, D. S.; Eichfeld, C. M.; Das, S. Mimicking neurotransmitter release in chemical synapses via hysteresis engineering in MoS<sub>2</sub> transistors. *ACS Nano* **2017**, 11 (3), 3110–3118.

(33) Kim, S. J.; Choi, K.; Lee, B.; Kim, Y.; Hong, B. H. Materials for Flexible, Stretchable Electronics: Graphene and 2D Materials. *Annu. Rev. Mater. Res.* **2015**, 45 (1), 63–84.

(34) Kim, D.-H.; Ahn, J.-H.; Choi, W. M.; Kim, H.-S.; Kim, T.-H.; Song, J.; Huang, Y. Y.; Liu, Z.; Lu, C.; Rogers, J. A. Stretchable and foldable silicon integrated circuits. *Science* **2008**, 320 (5875), 507–511.

(35) Song, Y. M.; Xie, Y.; Malyarchuk, V.; Xiao, J.; Jung, I.; Choi, K.-J.; Liu, Z.; Park, H.; Lu, C.; Kim, R.-H.; Li, R.; Crozier, K. B.; Huang, Y.; Rogers, J. A. Digital cameras with designs inspired by the arthropod eye. *Nature* **2013**, 497 (7447), 95–99.

(36) Sylvia, S. S.; Alam, K.; Lake, R. K. Uniform benchmarking of low-voltage van der Waals FETs. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits* **2016**, 2, 28–35.

(37) Lee, C.-S.; Cline, B.; Sinha, S.; Yeric, G.; Wong, H. S. P. 32-bit Processor core at 5-nm technology: Analysis of transistor and interconnect impact on VLSI system performance. 2016 *IEEE international electron devices meeting (IEDM)*; IEEE: 2016; pp 28.3.1–28.3.4, DOI: 10.1109/IEDM.2016.7838498.

(38) Agarwal, T.; Szabo, A.; Bardon, M. G.; Soree, B.; Radu, I.; Raghavan, P.; Luisier, M.; Dehaene, W.; Heyns, M.; Benchmarking of monolithic 3D integrated MX<sub>2</sub> FETs with Si FinFETs. 2017 *IEEE international electron devices meeting (IEDM)*, IEEE: 2017; pp 5.7.1–5.7.4, DOI: 10.1109/IEDM.2017.8268336.

(39) Sebastian, A.; Pendurthi, R.; Choudhury, T. H.; Redwing, J. M.; Das, S. Benchmarking monolayer MoS<sub>2</sub> and WS<sub>2</sub> field-effect transistors. *Nat. Commun.* **2021**, 12 (1), 693.

(40) Li, H.; Zhang, Q.; Yap, C. C. R.; Tay, B. K.; Edwin, T. H. T.; Olivier, A.; Baillargeat, D. From Bulk to Monolayer MoS<sub>2</sub>: Evolution of Raman Scattering. *Adv. Funct. Mater.* **2012**, 22 (7), 1385–1390.

(41) Das, S.; Robinson, J. A.; Dubey, M.; Terrones, H.; Terrones, M. Beyond Graphene: Progress in Novel Two-Dimensional Materials and van der Waals Solids. *Annu. Rev. Mater. Res.* **2015**, 45 (1), 1–27.

(42) Tongay, S.; Suh, J.; Ataca, C.; Fan, W.; Luce, A.; Kang, J. S.; Liu, J.; Ko, C.; Raghunathan, R.; Zhou, J. Defects activated photoluminescence in two-dimensional semiconductors: interplay between bound, charged, and free excitons. *Sci. Rep.-Uk* **2013**, 3, 2657.

(43) Nasr, J. R.; Simonson, N.; Oberoi, A.; Horn, M. W.; Robinson, J. A.; Das, S. Low-Power and Ultra-Thin MoS<sub>2</sub> Photodetectors on Glass. *ACS Nano* **2020**, 14, 15440.

(44) Dodda, A.; Oberoi, A.; Sebastian, A.; Choudhury, T. H.; Redwing, J. M.; Das, S. Stochastic resonance in MoS<sub>2</sub> photodetector. *Nat. Commun.* **2020**, 11 (1), 4406.

(45) Jayachandran, D.; Oberoi, A.; Sebastian, A.; Choudhury, T. H.; Shankar, B.; Redwing, J. M.; Das, S. A low-power biomimetic collision detector based on an in-memory molybdenum disulfide photodetector. *Nature Electronics* **2020**, 3, 646.

(46) Han, P.; Marie, L. S.; Wang, Q. X.; Quirk, N.; El Fatimy, A.; Ishigami, M.; Barbara, P. Highly sensitive MoS<sub>2</sub> photodetectors with graphene contacts. *Nanotechnology* **2018**, 29 (20), 20LT01.

(47) Mennel, L.; Symonowicz, J.; Wachter, S.; Polyushkin, D. K.; Molina-Mendoza, A. J.; Mueller, T. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **2020**, 579 (7797), 62–66.

(48) Subbulakshmi Radhakrishnan, S.; Chakrabarti, S.; Sen, D.; Das, M.; Schranghamer, T. F.; Sebastian, A.; Das, S. A Sparse and Spike-timing-based Adaptive Photo Encoder for Augmenting Machine Vision for Spiking Neural Networks. *Adv. Mater.* **2022**, 2202535.

(49) Piochon, C.; Kano, M.; Hansel, C. LTD-like molecular pathways in developmental synaptic pruning. *Nature neuroscience* **2016**, 19 (10), 1299.

(50) Hansel, C. Deregulation of synaptic plasticity in autism. *Neuroscience letters* **2019**, 688, 58–61.

(51) Xuan, Y.; Jain, A.; Zafar, S.; Lotfi, R.; Nayir, N.; Wang, Y.; Choudhury, T. H.; Wright, S.; Feraca, J.; Rosenbaum, L.; Redwing, J. M.; Crespi, V.; van Duin, A. C. T. Multi-scale modeling of gas-phase reactions in metal-organic chemical vapor deposition growth of WSe<sub>2</sub>. *J. Cryst. Growth* **2019**, 527, 125247.

(52) Dodda, A.; Das, S. Demonstration of Stochastic Resonance, Population Coding, and Population Voting Using Artificial MoS<sub>2</sub> Based Synapses. *ACS Nano* **2021**, 15 (10), 16172–16182.

## Recommended by ACS

### On-Chip Integrated Atomically Thin 2D Material Heater as a Training Accelerator for an Electrochemical Random-Access Memory Synapse for Neuromorphic Computing A...

Revannath Dnyandeo Nikam, Hyunsang Hwang, *et al.*

JULY 19, 2022

ACS NANO

READ 

### Multimode Synaptic Operation of a HfAlO<sub>x</sub>-Based Memristor as a Metaplastic Device for Neuromorphic Applications

Hyung Seok Shin, Kyung Min Kim, *et al.*

MAY 25, 2022

ACS APPLIED ELECTRONIC MATERIALS

READ 

### Multi-Stimuli-Responsive Synapse Based on Vertical van der Waals Heterostructures

Jiachao Zhou, Bin Yu, *et al.*

JULY 26, 2022

ACS APPLIED MATERIALS & INTERFACES

READ 

### Memristor-Based Neuromodulation Device for Real-Time Monitoring and Adaptive Control of Neuronal Populations

Catarina Dias, Paulo Aguiar, *et al.*

MAY 02, 2022

ACS APPLIED ELECTRONIC MATERIALS

READ 

Get More Suggestions >