An All-in-One Biomimetic 2D Spiking Neural Network

Shiva Subbulakshmi Radhakrishnan¹, Akhil Dodda¹, and Saptarshi Das¹,²,³,*

¹Department of Engineering Science and Mechanics, Pennsylvania State University, University Park, PA 16802, USA
²Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA
³Materials Research Institute, Pennsylvania State University, University Park, PA 16802, USA

Abstract: In spite of recent advancements in bio-realistic artificial neural networks such as spiking neural networks (SNNs), the energy efficiency, multifunctionality, adaptability, and integrated nature of biological neural networks (BNNs) largely remain unimitated in hardware neuromorphic computing systems. Here we exploit optoelectronic and programmable memory devices based on emerging two-dimensional (2D) layered materials such as MoS₂ to demonstrate an “all-in-one” hardware SNN system which is capable of sensing, encoding, unsupervised learning, and inference at miniscule energy expenditure. In short, we have utilized photogating effect in MoS₂ based neuromorphic phototransistor for sensing and direct encoding of analog optical information into graded spike trains, we have designed MoS₂ based neuromorphic encoding module for conversion of spike trains into spike-count and spike-timing based programming voltages, and finally we have used arrays of programmable MoS₂ non-volatile synapses for spike-based unsupervised learning and inference. We also demonstrate adaptability of our SNN for learning under scotopic (low-light) and photopic (bright-light) conditions mimicking neuroplasticity of BNNs. Furthermore, we use our hardware SNN platform to show learning challenges under specific
synaptic conditions, which can aid in understanding learning disabilities in BNNs. Our findings highlight the potential of in-memory computing and sensing based on emerging 2D materials, devices, and circuits not only to 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. as well as bridge the understanding between neuroscience of learning and machine learning.

Biological neural networks (BNNs) comprising of 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/external 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 stereotypical electrical impulses or all-or-none (digital) point events in time that enable energy efficient neural computation and long-distance neural communication. Spiking activity between the pre-synaptic and post-synaptic neurons determines the potentiation or depression of their connection strengths or synaptic weights, which are ultimately responsible for learning. Another key feature of BNN is neuroplasticity, which allows adaptation to learning and decision making under changing environment. For example, eyes can identify patterns in starlight (scotopic vision) as well as in bright sunlight (photopic vision) in spite of illumination levels differing by ~ 9 orders of magnitude. Finally, spiking activity manifesting in loss or excessive potentiation or depression of synaptic connections can lead to various neurological conditions resulting in learning disability.
Therefore, designing spike-based and low-power neuromorphic hardware systems that resemble the functionality, organization, and plasticity of BNN can not only accelerate the development of hardware artificial intelligence (AI) and benefit edge computing and smart sensing for Internet of Things (IoT), but also offer a platform to model plasticity-related learning disorders of the CNS.

Artificial neural networks (ANNs) are highly simplified, but most prevalent abstraction of BNNs that have already demonstrated breakthroughs in many applications including image classification, speech recognition, and game playing [1]. While in early days, performance of ANNs primarily relied on supervised learning, more recently, reinforcement and unsupervised learning using deep neural networks (DNNs) have shown remarkable improvements in challenging domains such as mastering the game of Go without input from human experts [2]. However, these computer-science-oriented learning algorithms require tremendous computational resources when implemented in neuromorphic hardware using traditional complementary metal-oxide-semiconductor (CMOS) technology leading to orders of magnitude higher power consumption compared to brain. One of the key differences is in the computing architecture, whereas CMOS-based computation embrace von Neumann architecture that physically separates the compute (logic) and storage (memory), BNNs dissolve such gap by placing neurons, the computational primitives, and synapses, the storage units, right next to each other.

Acknowledging the aforementioned limitations, non–von Neumann architectures leveraging silicon CMOS technology have been developed, such as the True North from IBM [3]. While the chip shows remarkably low power consumption of 63 mW for multiobject detection and classification in real-time using 1 million artificial spiking neurons and 256 million artificial
synapses, at the implementation level, only marginal similarities with brain-like spike-based computing can be recognized. The benefits of spike-based computation was recently demonstrated through the hardware realization of spiking neural network (SNN) in a chip named Loihi from Intel [4]. Note that SNNs require stronger interaction between memory and compute in mimicking the rich spatiotemporal dynamics of spike-based encoding and learning rules. While SNNs promise to bridge the energy gap between ANNs and BNNs [5-7] hybrid ANN/SNN artificial general intelligence systems such as the Tianjic chip [8] are equally attractive accommodating both computer-science-based and neuroscience-based learning algorithms. However, these chips are based on CMOS technology, which is experiencing a steady decline in scaling and may not compete well in the emerging IoT market necessitating material discovery and device level innovations to closely imitate the functionalities of biological neurons.

In this context, field programmable gate arrays (FPGAs) [9] and crossbar architectures utilizing memristors [10, 11], resistive random-access memory (RRAM) [12], phase change memory (PCM) [13-15], etc. with tunable conductance states are accelerating the development of energy efficient and non von Neumann computing architectures. These devices naturally lend themselves towards unsupervised learning using spike-time dependent plasticity and spike-rate dependent plasticity found in neurobiology. In fact, hardware SNNs have been constructed using memristor-based artificial synapses [16] and spiking neurons [17]. However, unlike BNNs, where specialized afferent neurons transduce the continuous time and analog valued information obtained from the environment into spike trains, memristive SNNs involve extensive CMOS-based peripheral circuits for spike encoding. Such pre-processing can ultimately limit the energy efficiency and scalability of memristive SNN architectures [18, 19]. Furthermore, sensing is an integral part of
BNN, which is unfounded in memristive networks necessitating integration of peripheral sensors. Finally, neuroplasticity of learning in changing environment, and modeling of learning disabilities even at a high level of abstraction is yet to be demonstrated using memristive SNNs.

Here we demonstrate an “all-in-one” biomimetic hardware SNN which is capable of sensing, encoding, and spike-based unsupervised learning and inference using monolayer MoS$_2$ based multifunctional optoelectronic and programmable memory devices. First, we use photogating effect in MoS$_2$ phototransistor to directly sense and encode optical information into graded spike trains. Next, we develop MoS$_2$ based encoding cells to implement spike-count and spike-timing based encoding algorithms. And finally, we use MoS$_2$-based electrically programmable non-volatile synapses for spike-based unsupervised learning and inference. Furthermore, we demonstrate low-power operation and adaptability of our SNN to learning under different ambient conditions mimicking neuroplasticity of BNN. Our SNN hardware also offers a platform to model learning disabilities and disorders of BNN at a high level of abstraction. To the best of our knowledge, this is the first experimental demonstration of an integrated SNN exploiting in-memory computing and sensing based on emerging two-dimensional (2D) layered materials and devices that can accelerate the development of energy efficient neuromorphic hardware.

The motivation behind using two-dimensional (2D) layered MoS$_2$ as a hardware platform for neuromorphic computing is multifold. First, there are several demonstration of photodetectors [20], chemical sensors [21], biological sensors [21], touch sensors [22], and radiation sensors [23] using MoS$_2$ based devices, which can naturally serve as artificial sensory afferent neurons eliminating the need for peripheral sensors for MoS$_2$ based intelligent systems. Next, MoS$_2$ being
a semiconductor, almost all peripheral analog or digital signal processing units can be build using MoS$_2$ field effect transistors (FETs) largely eliminating the need for hybrid design involving CMOS circuitry. Additionally, the atomically thin body nature of MoS$_2$ allows aggressive channel length scaling without the loss of superior gate electrostatic benefiting high integration density. In fact, recent studies show high performance monolayer MoS$_2$ FETs with the channel and contact lengths scaled to 29 nm and 13 nm, respectively [24]. Moreover, some of the early criticism of 2D FETs have also been successfully addressed in recent years through the realization of low contact resistance [25], high ON current [26], integration of ultra-thin and high-k gate dielectric [27], and wafer scale growth using chemical vapor deposition (CVD) and metal organic CVD (MOCVD) [28, 29]. Similarly, MoS$_2$ based microprocessors [30], analogue operational amplifier [31], and RF electronics components [32] have been reported. Finally, unlike silicon CMOS, MoS$_2$ can enable flexible [33] and printable [34] electronic circuits adding value towards a MoS$_2$ based biomimetic and neuromorphic hardware platforms [35-37].

Fig. 1a-c show neurobiological architecture for processing visual information and Fig. 1d-g show our proposed MoS$_2$ based all-in-one biomimetic SNN architecture with remarkable resemblance between the two. For example, monolayer MoS$_2$ based neuromorphic phototransistors (PT) (Fig. 1e) are equivalent to photoreceptor cells (rods and cones) in the human eyes (Fig. 1b) that convert external optical stimuli into corresponding graded potentials. Rods primarily enable low-light or scotopic vision whereas cones are responsible for bright-light or photopic vision, both of which can be achieved using our MoS$_2$ PT. Similarly, MoS$_2$ based neuromorphic encoding cells (Fig. 1f) mimic the functionality of retinal ganglion cell (Fig. 1b) that encode the graded potentials into spike trains and transmit to visual cortex or midbrain for higher order processing and computation.
Figure 1. All-in-one biomimetic hardware SNN based on MoS2 FETs. a) Example optical stimulus. b) Phototransduction pathways in eyes for spike encoding. c) Visual cortex for information processing. d) Blue LED used as external optical stimulus for MoS2-based SNN. e) MoS2-based neuromorphic phototransistor, which is equivalent to photoreceptor cells (rods and cones) in the eyes that convert external optical stimuli into corresponding graded potentials. Rods primarily enable scotopic vision whereas cones are responsible for photopic vision, both of which can be achieved using our MoS2 PT. f) MoS2-based neuromorphic encoding cells mimicking the functionality of retinal ganglion cell that encode the graded potentials into spike trains. g) Optical image of MoS2-based non-volatile and electrically programmable synaptic array. An example experimental demonstration of h) pattern illumination using LED and corresponding i) sensory transduction using MoS2 phototransistor, j) spike-count based encoding using MoS2 encoding module, and k) conductance states of MoS2 synapses following unsupervised learning. l) Transfer characteristics, i.e., source to drain current ($I_{DS}$) as a function of the back-gate voltage ($V_{BG}$) at different drain biases ($V_{DS}$), m) output characteristics i.e., $I_{DS}$ versus $V_{DS}$ for different $V_{BG}$, n) phototransduction under different LED illumination, and o) spike-based programmed states for representative MoS2 FETs. Overall, the platform offers all capabilities including sensing, computing, and non-volatile storage based on monolayer MoS2 FETs integrated with programmable analog memory gate-stack ($Al_2O_3/Pt/TiN/p^{++}-Si$) allowing in-memory computing and sensing.
Finally, MoS₂ based non-volatile and electrically programmable synaptic arrays (Fig. 1g) imitate the visual cortex (Fig. 1c) where learning and inference take place. Note that all components of our SNN hardware are derived based on monolayer MoS₂ FETs integrated with programmable analog memory gate-stack (Al₂O₃/Pt/TiN/p⁺⁺-Si) (Fig. 1e) allowing in-memory computing and sensing overcoming the bottleneck of von Neumann architecture. Furthermore, as we will demonstrate, each module can be reconfigured to adapt to different learning environments. Fig. 1h-k, respectively, show an example experimental demonstration of pattern illumination using a light emitting diode (LED), sensory transduction using MoS₂ phototransistor, spike encoding using MoS₂ encoding module, and unsupervised learning using MoS₂ synapses.

MoS₂ used in this study was obtained from 2D crystal consortium (2DCC) [28] grown epitaxially on a sapphire substrate using MOCVD technique at 1000 °C. Carbon-free and high-temperature growth ensures high film quality, which is critical for low-power operation of the SNN hardware. The MoS₂ film was transferred from the growth substrate on to another substrate with a back-gate stack that comprised of atomic layer deposition (ALD) grown 50 nm Al₂O₃ on Pt/TiN/p⁺⁺-Si for the FET fabrication (Fig. 1e). As we will elucidate later, this gate stack resembles the floating-gate architecture used in FLASH memory devices and in spite of being present globally allows local programming of individual MoS₂ synapses. Details on monolayer MoS₂ synthesis, film transfer, and fabrication of the back-gate stack and MoS₂ synapses can be found in the Methods section.

Fig. 1l shows the transfer characteristics, i.e. source to drain current (I_DS) as a function of the back-gate voltage (V_BG) at different drain biases (V_DS) and Fig. 1m shows the output characteristics i.e. I_DS versus V_DS for different V_BG for a representative MoS₂ FET with 1 μm channel length, 5 μm
channel width, and a stack of 40 nm Ni/30 nm Au as the source and drain contacts. MoS₂ FET shows unipolar, n-type characteristics with excellent current on/off ratio of ~10⁶, subthreshold slope (SS) of ~ 283 mV/decade over 3 orders of magnitude change in $I_{DS}$, electron field effect mobility of ~12 cm²/V-s obtained from peak transconductance, and relatively high on current of ~10 μA/μm at $V_{DS} = 5$ V for an inversion charge carrier density of ~3×10¹²/cm². These numbers are on par with the state-of-the-art literature on large area grown MoS₂. These MoS₂ FETs allow us to realize peripheral circuits including the encoding module as discussed later. Fig. 1n shows the transfer characteristics of the MoS₂ FET in dark and under the illumination of a blue LED. Clearly, the device can be used as a photodetector. Note that, instead of LASER illumination, conventionally used to study photoresponse in monolayer MoS₂ [38], we have used LED to provide optical stimuli since it represents more realistic lighting ambience where most neuromorphic sensors will be deployed. Finally, Fig. 1o shows the programmability of our MoS₂ FET to achieve analog conductance states by applying electrical voltage spikes to the back-gate terminal which form the basis for synaptic learning as we will discuss in detail later. Overall, the platform offers all capabilities including sensing, computing, and non-volatile storage that are key to develop a fully integrated, reconfigurable, and biomimetic hardware SNN system.

**MoS₂ based neuromorphic sensor:** Monolayer MoS₂ based phototransistors have been studied extensively in the recent years including our own work [20, 38–42]. The phototransduction mechanism in MoS₂ PT is typically attributed to two mechanisms: photocarrier generation in the MoS₂ channel and photogating effect arising due to charge trapping/detrapping mechanisms at the MoS₂/gate-dielectric interface. Here we exploit the photogating effect in MoS₂ PT for direct encoding of analog optical stimuli into spike trains (Fig. 2).
Figure 2. MoS\(_2\) based neuromorphic sensor. a) Transfer characteristics of monolayer MoS\(_2\) PT at \(V_{DS} = 1\) V before and after illumination from the blue LED with input currents ranging from \(I_{LED} = 0.5\) mA (low-brightness) to \(I_{LED} = 50\) mA (high-brightness) at different \(V_{BG} = V_{write}\) for 100 ms. For illuminations in the on-state \((V_{write} = 2.0\) V\) and in the subthreshold regime \((V_{write} = 0.5\) V\), there are no visible shift in the device characteristics post-illumination. This can be ascribed to photocarrier generation in the MoS\(_2\) channel, which are swept across by the applied \(V_{DS}\) and hence there is no persistent photocurrent beyond the optical exposure. However, for illuminations in the off-state \((V_{write} = -1.5\) V\) and \(-2.5\) V\) photocarrier trapping at the MoS\(_2\)/dielectric interface leads to the shift in the device threshold voltage \((V_{TH})\). The detrapping mechanism can be rather slow and can take hours to several days, which is why the shift is visible post-illumination. Higher \(I_{LED}\) and more negative \(V_{write}\) naturally result in more trapping and hence larger shifts. b) Analog valued and continuous time input optical stimuli from the blue LED. Corresponding \(I_{DS}\) sampled every 100 ms with \(V_{BG}\) toggling between \(V_{read} = 0\) V and \(V_{write} = 1.0\) V\), d) \(V_{write} = -1.5\) V\), e) \(V_{write} = -2.0\) V\) and f) \(V_{write} = -2.5\) V\). The magnitude of the \(I_{DS}\) spikes increases monotonically during sampling for any given \(I_{LED}\) and \(V_{write}\) owing to continuous carrier trapping resulting in gradual \(V_{TH}\) shift. g) Number of spikes \((N_{spike})\) and h) total spiking duration \((\tau_{spike})\) as a function of \(I_{LED}\) and \(V_{write}\). A spike is counted when \(I_{DS} > I_{ST}\), where \(I_{ST} = 3\) mA is the spiking threshold (dotted red line in e-f). i) Average energy consumption per spike \((E_{spike})\) for different \(I_{LED}\) and \(V_{write}\).
Fig. 2a shows the transfer characteristics at $V_{DS} = 1$ V for a representative monolayer MoS$_2$ PT before and after illumination from the blue LED with input currents ranging from $I_{LED} = 0.5$ mA (low-brightness) to $I_{LED} = 50$ mA (high-brightness) at different $V_{BG} = V_{write}$ for 100 ms. See Supplementary Information 1 for the optical images showing corresponding LED brightness levels. Two distinct types of photoresponse are observed in Fig. 2a. For $V_{write} > 0$ V, i.e. illuminations in the on-state ($V_{write} = 2.0$ V) and in the subthreshold regime ($V_{write} = 0.5$ V) of the MoS$_2$ FET, there are no visible shift in the device characteristics post-illumination irrespective of the brightness level of the LED ($I_{LED}$). This can be ascribed to photocarrier generation in the MoS$_2$ channel, which are swept across by the applied $V_{DS}$ and hence there is no persistent photocurrent beyond the optical exposure. However, for $V_{write} < 0$ V, i.e. illuminations in the off-state ($V_{write} = -1.5$ V and $V_{write} = -2.5$ V) of the MoS$_2$ FET, there are significant shifts in the device characteristics post-illumination. This is a feature of photogating effect where photocarrier trapping at the MoS$_2$/dielectric interface leads to the shift in the device threshold voltage ($V_{TH}$). The detrapping mechanism can be rather slow and can take hours to several days, which is why the $V_{TH}$ shift is visible post-illumination. Higher $I_{LED}$ and more negative $V_{write}$ naturally result in more trapping and hence larger $V_{TH}$ shifts.

We exploit the unique photogating effect in MoS$_2$ PT for direct encoding of the optical stimulus using spike-count and spike-timing based encoding. Fig. 2b shows analog valued and continuous time input optical stimuli from the blue LED and Fig. 2c-f show the corresponding $I_{DS}$ sampled every $\tau_s = 100$ ms with $V_{BG}$ toggling between $V_{read} = 0$ V and respective $V_{write}$. Some key observations can be made from the results: 1) the magnitude of the $I_{DS}$ spikes increases monotonically during sampling for any given $I_{LED}$ and $V_{write}$ owing to continuous carrier trapping.
resulting in gradual $V_{TH}$ shift, 2) the magnitude of the $I_{DS}$ spikes for any given $I_{LED}$ increases with increasing magnitude of $V_{write}$ i.e. more negative $V_{write}$ since more trap states are available at the MoS$_2$/dielectric interface resulting in greater $V_{TH}$ shift, 3) the time-lag for the occurrence of $I_{DS}$ spike of a predefined magnitude scales inversely with $I_{LED}$ for any given $V_{write}$ i.e. spikes of similar magnitude occur later for lower $I_{LED}$ and vice versa as shown using the dotted lines, and 4) $I_{DS}$ spike of similar magnitude occurs earlier for more negative $V_{write}$ for any given $I_{LED}$. These observations are summarized in Fig. 2g-h, respectively, showing the number of spikes ($N_{spike}$), and the total spiking duration ($\tau_{spike}$) as a function of $I_{LED}$ and $V_{write}$. Note that a spike is counted when $I_{DS} > I_{ST}$, where $I_{ST} = 3$ nA is the spiking threshold (dotted red line in Fig. 2c-f). See Supplementary Information 2 for $N_{spike}$, and $\tau_{spike}$ plotted as a function of $I_{LED}$ and $V_{write}$ for different $I_{ST}$ values. Note that $V_{write}$ and $I_{ST}$ are two design parameters for the MoS$_2$ based neuromorphic sensing module that can be adjusted for adaption to different lighting conditions. For example, setting $I_{ST}$ to 2 nA and illuminating the MoS$_2$ PT at $V_{write} = -2.5$ V allow more precise signal transduction under low light conditions mimicking the scotopic vision offered by the rod photoreceptor cells i.e. spike-counts are large and spiking duration is long even for lower $I_{LED}$ values (Supplementary Figure 2a-b) [43]. Similarly, setting $I_{ST}$ to 8 nA and illuminating the MoS$_2$ PT at $V_{write} = -2.0$ V ensures better precision under bright light conditions similar to the photopic vision offered by the cone photoreceptor cells (Supplementary Figure 2c-d). However, if $I_{ST} = 3$ nA and $V_{write} = -1.5$ V, a better dynamic range can be accomplished as seen in Fig. 2g-h. Nevertheless, the monotonic dependence between $I_{LED}$ and $N_{spike}$ and $I_{LED}$ and $\tau_{spike}$ observed under various operating conditions constitutes the foundation for spike-count and spike-timing based encoding and learning using our MoS$_2$ based biomimetic SNN platform. See Supplementary Video 1 for time-evolution of spike-count and spike-timing based encoding of an example
illumination pattern using different $V_{\text{write}}$ and $I_{\text{ST}} = 3 \text{nA}$. Finally, Fig. 2i shows the average energy consumption per spike \( E_{\text{spike}} = \frac{1}{N_{\text{spike}}} \sum_{i=1}^{N_{\text{spike}}} I_{DS} - V_{DS} \tau_s \) for different $I_{\text{LED}}$ and $V_{\text{write}}$. Even for the brightest LED illumination at the most negative $V_{\text{write}}$ that corresponds to high magnitude $I_{DS}$ spikes, $E_{\text{spike}} \sim 1.5 \text{nJ}$, which suggests energy efficient spike-encoding by our MoS$_2$ PT.

**MoS$_2$ based neuromorphic encoder:** While the above demonstration shows the conversion of external optical stimuli into corresponding spike trains, further reshaping of these spikes is necessary for the implementation of spike-count and spike-timing based learning algorithms. This is because unlike the neuronal spikes in BNN which have constant amplitude, the amplitude of the spikes from MoS2 neuromorphic photosensor are dependent on $I_{\text{LED}}$, i.e. brighter LED illuminations not only invoke earlier and a greater number of spikes, but also the strength of spikes are significantly higher compared to those obtained from dimmer LED illuminations. In order to resolve this difference between BNN and our MoS$_2$ based SNN, we have designed MoS$_2$ based neuromorphic circuit modules as discussed below (Fig. 3).

Fig. 3a shows the circuit diagram for spike-count based encoding module and Fig. 3b-f show the reshaping of $I_{DS}$ spikes obtained from the MoS$_2$ PT biased at $V_{\text{write}} = -1.5 \text{ V}$ for the input brightness level of $I_{\text{LED}} = 10 \text{ mA}$ into corresponding programming voltage spikes ($V_p$). See **Supplementary Information 3** for similar conversion results for different LED brightness levels.

First, the $I_{DS}$ spikes (Fig. 3b) are converted to voltage spikes, $V_{C1}$ (Fig. 3c), by using a switch capacitor cell (SCC1) that comprises of a capacitor ($C_1$) and a switch ($S_1$). The switch, $S_1$, is a MoS$_2$ FET, which allows charging and discharging of $C_1$ for every $I_{DS}$ spike by switching between
off-state and on-state. The magnitude of the $V_{C1}$ spikes are determined by $V_{C1} = \frac{I_{DS}t_c}{C_1}$, where $C_1 \approx 350$ pF, and charging time, $t_c = 100$ ms. Note that the $V_{C1}$ spikes, as expected, follow the trend in $I_{DS}$. Next, these $V_{C1}$ spikes are applied to the drain terminal of another MoS$_2$ FET ($S_2$), which is referred to as the spike reshaping cell (SRC) and the output current spikes ($I_{S2}$) are recorded (Fig 3d). The magnitude of the $I_{S2}$ spikes saturate at $\sim 30$ nA owing to the phenomenon of current saturation in MoS$_2$ FET. See Supplementary Information 4 for the output characteristics of $S_2$ which show current saturation for $V_{C1} > 0.8$ V, which corresponds to $I_{DS} = I_{ST} = 3$ nA. Finally, the $I_{S2}$ spikes are converted to voltage spikes, $V_{C2}$ (Fig. 3e), following $V_{C2} = \frac{I_{S2}t_c}{C_2}$ by using another switch capacitor cell (SCC2) comprising of $C_2 = 350$ pF and $S_3$. Finally, these $V_{C2}$ spikes with inverted polarity, $V_{P}$ (Fig. 3f), are used for spike-count based programming of non-volatile MoS$_2$ synapses for unsupervised learning.

As we will discuss later, the magnitude of $V_{P}$ spikes play a critical role in determining the learning rates. Interestingly, our encoding module offers tremendous design flexibility since the magnitude of $C_1$, $C_2$, and the biasing of $S_2$ can be adjusted for obtaining desired magnitude of $V_{P}$ spikes. For example, Supplementary Information 4 shows that by changing the gate-bias for $S_2$, different magnitudes of $I_{S2}$ spikes and subsequently, $V_{C2}$ and $V_{P}$ spikes can be obtained for the same set of $V_{C1}$ spikes. This can be exploited for adaptive learning under scotopic conditions by encoding lower number of spikes using higher magnitude $V_{P}$. The reconfigurability of the encoding module 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.
Figure 3. MoS$_2$ based neuromorphic encoders. a) Circuit diagram for spike-count based encoding module comprising of switch capacitor cell 1 (SCC1), spike reshaping cell (SRC), switch capacitor cell 2 (SCC2), and voltage inverter cell (VIC). The input to the module are graded $I_{DS}$ spikes obtained from the MoS$_2$ PT and output of the module are corresponding programming voltage spikes ($V_P$), which are relayed to the learning module based on non-volatile MoS$_2$ synapses. b) Input $I_{DS}$ spikes from the MoS$_2$ PT biased at $V_{write} = -1.5$ V for $I_{LED} = 10$ mA, and corresponding output of c) SCC1 ($V_{C1}$), d) SRC ($I_{D2}$), e) SCC2 ($V_{C2}$), and f) VIC ($V_P$). g) Circuit diagram for spike-timing based encoding module comprising of integrator cell (IC), spike reshaping cell (SRC), resistor cell (RC), and voltage inverter cell (VIC). h) Input $I_{DS}$ spikes from the MoS$_2$ PT biased at $V_{write} = -1.5$ V for $I_{LED} = 10$ mA, and corresponding output of i) IC ($V_{C4}$), j) SRC ($I_{D4}$), k) RC ($V_{D5}$), and l) VIC ($V_P$). $S_1$, $S_2$, $S_3$, $S_4$, and $S_5$ are MoS$_2$ FETs, $C_1 = C_2 = 350$ pF, and $C_4 = 20$ nF. See Supplementary Information 4-7 for more detail on the encoding modules. Average encoding energy expenditure per spike by each cell for m) spike-count and n) spike-timing based encoding module for different $I_{LED}$. 

Fig. 3g shows the circuit diagram for spike-timing based encoding module and Fig. 3h-l show the reshaping of $I_{DS}$ spikes obtained from the MoS$_2$ PT biased at $V_{\text{write}} = -1.5$ V for $I_{LED} = 10$ mA into $V_p$. See Supplementary Information 5 for similar conversion results for different $I_{LED}$. In this case, $I_{DS}$ spikes (Fig. 3h) are converted to analog voltage $V_{C4}$ (Fig. 3i), by using an integrator cell (IC) that comprises of a capacitor, $C_4 = 20$ nF, following $V_{C4} = \frac{1}{C_4} \int I_{DS} dt$. Higher capacitance value ensures that the maximum magnitude for $V_{C4}$ does not exceed 15 V. Next, $V_{C4}$ is applied to the drain terminal of a MoS$_2$ FET ($S_4$), i.e. the SRC and the output current ($I_{S4}$) is recorded (Fig 3j). The magnitude of $I_{S4}$ saturate at $\approx 10$ $\mu$A for $V_{C4} > 1.5$ V, which corresponds to $I_{DS} = I_{ST} = 3$ nA. Note that higher current levels are required for $I_{S4}$ since $I_{S4}$ is converted to $V_{SS}$ (Fig. 3k) using another MoS$_2$ FET ($S_5$), which is used as a linear resistor following $V_{SS} = R_{SS} I_{S4}$ and hence both $S_4$ and $S_5$ must be operated in their on state (see Supplementary Information 6). $S_5$ is also referred to as the resistor cell (RC). Finally, $V_{SS}$ with inverted polarity, $V_p$ (Fig. 3l) is used for spike-timing based unsupervised learning using MoS$_2$ synapses. As demonstrated in Supplementary Information 5, $V_p$ reaches saturation earlier for brighter illuminations and vice versa. The total programming duration is important for spike-timing based learning, which can be adjusted by adjusting the magnitude of $C_4$ i.e. faster versus slower charging to $V_{C4}$ for same input $I_{DS}$ spike train obtained from the sensing module. Alternatively, the amplitude of $V_p$ can be adjusted by changing the gate-bias for $S_4$ (see Supplementary Information 6) for adaptive learning under scotopic conditions by encoding shorter spike durations using higher magnitude $V_p$.

Finally, Fig. 3m-n, respectively, show the average encoding energy expenditure per spike by each cell for spike-count and spike-timing based encoding module for different $I_{LED}$. Note that, while specific biasing conditions for the cells can alter their respective energy expenditure, overall, the
encoding modules are energy efficient. Also note that the spike-count based encoding module consumes orders of magnitude smaller energy compared to spike-timing based encoding module owing to the higher operating currents required by $S_4$ and $S_5$. This can be reduced by scaling down the width of these MoS$_2$ FETs and/or by increasing their channel lengths. Nevertheless, our reconfigurable MoS$_2$ based encoding modules eliminate the need for CMOS-based thresholding circuits and offer a monolithic sensing and encoding solution for the biomimetic hardware SNN platform.

**Electrically programmable analog and non-volatile MoS$_2$ synapses:** Here we show that our monolayer MoS$_2$ FETs can be used as non-volatile synapses with analog conductance states programmable by applying electrical voltage spikes to the back-gate terminal. These MoS$_2$ synapses allow both spike-count as well as spike-timing based programming and can achieve both potentiation and depression analogous to chemical synapses in BNNs laying the foundation for unsupervised learning and inference (Fig. 4).

Fig 4a shows the potentiation of a representative MoS$_2$ electrical synapse from a low conductance state (LCS) after the application of a fixed number of programming spikes ($N_{\text{spike}} = 16$) of different amplitudes of negative polarity ($V_p$) with each spike applied for $t_{\text{spike}} = 100$ ms. Fig. 4b show the post-potentiated conductance states ($G_p$) measured at $V_{BG} = 0$ V as a function of $N_{\text{spike}}$ for different $V_p$. As expected, lower number of $N_{\text{spike}}$ invoke lower potentiation, i.e. smaller change in $G_p$ and *vice versa*, which can be exploited for spike-count based learning. Similarly, Fig 4c shows the depression of a potentiated MoS$_2$ synapse i.e. from high conductance state (HCS) after the application of a fixed number of programming spikes ($N_{\text{spike}} = 16$) of different amplitudes
of positive polarity ($V_D$) with each spike applied for $t_{\text{spike}} = 100$ ms. Fig. 4d show the post-depressed conductance states ($G_D$) measured at $V_{BG} = 0$ V as a function of $N_{\text{spike}}$ for different $V_D$. As expected, lower number of $N_{\text{spike}}$ invoke lower depression and *vice versa*, which can be exploited for spike-count based forgetting. Note that learning and forgetting capabilities enable unsupervised relearning using same synapses. Also note that smaller number of $N_{\text{spike}}$ can achieve higher potentiation/depression if encoded using higher $V_{P/D}$. As mentioned earlier, this aspect can be exploited to achieve learning plasticity. For example, under scotopic condition the photocurrent spikes from the neuromorphic sensor can be encoded into higher magnitude programming spikes by the neuromorphic encoder to achieve necessary potentiation of the MoS$_2$ synapses allowing a learning rate that is similar to the photopic condition. This will be illustrated further in the subsequent sections.

Fig 4e-h show the spike-timing based potentiation and depression of MoS$_2$ synapses. Fig 4e shows the potentiation of MoS$_2$ synapse from LCS after the application of single spike of constant magnitude $V_P = -9$ V for different $t_{\text{spike}}$ and Fig. 4f show the post-potentiated $G_P$ measured at $V_{BG} = 0$ V as a function of $t_{\text{spike}}$ for different $V_P$. Similarly, Fig 4g shows the depression of MoS$_2$ synapse from HCS after the application of single spike of constant magnitude $V_D = 14$ V for different $t_{\text{spike}}$ and Fig. 4h show the post-depressed $G_D$ measured at $V_{BG} = 0$ V as a function of $t_{\text{spike}}$ for different $V_D$. Here, shorter $t_{\text{spike}}$ invokes lower potentiation/depression and *vice versa*, which can be used for spike-timing based unsupervised learning/forgetting. Note that similar to spike-count based learning/forgetting, higher potentiation/depression can be achieved for shorter spike durations when encoded using higher magnitude of $V_{P/D}$ enabling spike-timing based learning plasticity under scotopic condition.
Figure 4. Electrically programmable analog and non-volatile MoS$_2$ synapses. a) Potentiation of a MoS$_2$ synapse from low conductance state (LCS) after the application of a fixed number of programming spikes ($N_{\text{spike}} = 16$) of different amplitudes of negative polarity ($V_p$) with each spike applied for $t_{\text{spike}} = 100$ ms. b) Post-potentiated conductance states ($G_p$) measured at $V_{BG} = 0$ V as a function of $N_{\text{spike}}$ for different $V_p$. c) Depression of a MoS$_2$ synapse from high conductance state (HCS) after the application of a fixed number of programming spikes ($N_{\text{spike}} = 16$) of different amplitudes of positive polarity ($V_D$) with each spike applied for $t_{\text{spike}} = 100$ ms. d) Post-depressed conductance states ($G_D$) measured at $V_{BG} = 0$ V as a function of $N_{\text{spike}}$ for different $V_D$. e) Potentiation of MoS$_2$ synapse from LCS after the application of single spike of constant magnitude $V_p = -9$ V for different $t_{\text{spike}}$. f) Post-potentiated $G_p$ measured at $V_{BG} = 0$ V as a function of $t_{\text{spike}}$ for different $V_p$. g) Depression of MoS$_2$ synapse from HCS after the application of single spike of constant magnitude $V_D = 14$ V for different $t_{\text{spike}}$. h) Post-depressed $G_D$ measured at $V_{BG} = 0$ V as a function of $t_{\text{spike}}$ for different $V_D$. i) Retention for 10 representative potentiated ($G_p$) states. j) Programming energy expenditure for potentiation. k) Retention for 10 representative depressed ($G_D$) states. l) Programming energy expenditure for depression.
The underlying mechanism behind the spike-count and spike-timing based potentiation and depression of MoS$_2$ synapses can be explained using the shift in $V_{TH}$ observed in the transfer characteristics of MoS$_2$ FETs. The $V_{TH}$ shift is attributed to our back-gate stack that resembles floating gate (FG) architecture used in non-volatile flash memory [44] (see Supplementary Information 7 for the energy band diagram of the back-gate stack). The p$^{++}$-Si/TiN/Pt interface in the stack is characterized by a Schottky barrier (SB), whereas, the gate dielectric, i.e. 50 nm Al$_2$O$_3$, acts as an oxide barrier (OB). The OB is much wider and taller compared to the SB. When programming voltage spikes are applied to the control gate (CG), i.e. p$^{++}$-Si, carriers tunnel from the p$^{++}$-Si into the Pt/TiN floating gate (FG) and remains trapped even after the release of the spike. These trapped charges on the FG screen the electric field from CG and thereby shifts the $V_{TH}$. The total amount of charge injected into the FG, and hence shift in $V_{TH}$ of the MoS$_2$ FET can be controlled by the amplitude, duration, and polarity of $V_{P/D}$. Furthermore, once programmed, the MoS$_2$ synapses continue to remain in the programmed states as evident from the retention measurements displayed in Fig. 4i and Fig. 4k for 10 representative potentiated ($G_P$) and depressed ($G_D$) conductance states, respectively. Note that the retention of conductance states in trained synapses is key for achieving high inference accuracy. Finally, Fig. 4j and Fig. 4l, respectively, show the programming energy expenditure per spike ($E_{P/D}$) as a function of $V_{P/D}$ for potentiation and depression of MoS$_2$ synapses, calculated using $E_{P/D} = \frac{1}{2} C_G V_{P/D}^2$, where $C_G$ is the total gate capacitance of the MoS$_2$ FET. It is worth mentioning that in spite of global back-gate geometry the potentiation and depression of MoS$_2$ synapses can be achieved locally, i.e. independent of each other (see Supplementary Information 8).
Spike-based unsupervised learning and inference using MoS$_2$ synapses: In this section, we demonstrate spike-count and spike-timing based unsupervised learning, forgetting, and relearning using MoS$_2$ synapses under various synaptic conditions (Fig. 5).

Fig. 1g shows the optical image and Fig. 5a shows the schematic of a fully connected 2-layer SNN with 9 presynaptic input neurons and 1 postsynaptic output neuron for learning and inferring patterns from 3×3 pixelated images. Fig. 5b shows the training and retraining schedule consisting of total $M = 40$ epochs, with each epoch having two cycles: potentiation and depression. During the potentiation cycle, the pattern to be learned is presented to the SNN, whereas during the depression cycle all 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 SNN can forget previously learned pattern and relearn new patterns. Fig. 5c-d, respectively, show the spiking profiles used for spike-count and spike-timing based learning. For each type of learning, we consider three configurations of the SNN: 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 ($V_D$) are 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 3×3 images is encoded with $N_{spike} = 10$ if it is bright and $N_{spike} = 0$ if it is dark. For spike-timing based learning the strength of potentiation and depression are adjusted using the spike duration, i.e. $t_{spike} = 800$ ms for strong and $t_{spike} = 100$ ms for weak potentiation/depression and for the pattern to be learned, each pixel in the 3×3 images are encoded with respective $t_{spike}$ (weak/strong) if it is bright and $t_{spike} = 10$ ms if it is dark.
Figure 5. Spike-based unsupervised learning using MoS₂ synapses. a) Schematic of a 2-layer SNN with 9 presynaptic neurons and 1 postsynaptic neuron for learning and inferring patterns from 3×3 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 SNN, whereas during the depression all synapses are uniformly depressed. Spiking profiles used for c) spike-count and d) spike-timing based learning. For each type of learning, three SNN 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 (\( V_D \)) are adjusted using the spike magnitude and spike duration for spike-count and spike-timing based learnings, respectively. The time evolution of colormap of synaptic weights i.e., the conductance states of the 9 synapses during e) spike-count and f) spike-timing based learning. For each type of learning all synapses are initialized either in a high conductance state (HCS) with \( G_{TCS} = 100 \) nS, or a low conductance state (LCS) with \( G_{LCS} = 100 \) pS (also see the Supplementary Video 3 and 4). Learning of the left diagonal followed by relearning of the right diagonal when potentiation and depression are both strong for g) spike-count and h) spike-timing based learnings (also see the Supplementary Video 5 and 6). Two sets of 9×1 synapses with synaptic weights for i) “Yes” postsynaptic neuron learned using the actual pattern and j) “No” postsynaptic neuron learned using the inverse of the pattern. The output currents from the “Yes” and “No” neurons are integrated using capacitors (\( C_{Yes/No} \)) to obtain \( V_{Yes} \) and \( V_{No} \) to determine the winner.
Fig. 5e-f, respectively, show the time evolution of colormap of synaptic weights i.e. the conductance states of the 9 synaptic devices during the spike-count and spike-timing based learning cycles. For each type of learning all synapses are initialized either in a high conductance state (HCS) with $G_{HCS} = 100$ nS, or a low conductance state (LCS) with $G_{LCS} = 100$ pS (also see the Supplementary Video 2 and 3). 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 LCS is fast, but forgetting and hence relearning from 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 Fig. 5g-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-timing based learnings, respectively (also see the Supplementary Video 4 and 5). Our findings indicate that the relative strengths of potentiation and depression play critical role in learning using SNN. This is similar to BNN, where autism, or autism spectrum disorder (ASD), which includes a broad range of conditions such as challenges with learning social skills, repetitive behaviors, etc. are related to dysregulation or deficit in long term depression in several mouse models [45, 46]. Therefore, our hardware SNN platform offers a unique opportunity to bridge the gap between neuroscience of learning and machine learning. The energy consumption by the learning module under different synaptic conditions are tabulated in Supplementary Information 9. The energy expenditures are on the orders of few nano Joules per epoch highlighting low-power learning in our integrated SNN platform.
For inference, we have used a 9×2 fully connected neural network implemented using two sets of 9×1 synapses as shown in Fig. 5i-j. The synaptic connections between the 9 presynaptic neurons and the “Yes” postsynaptic neuron are trained with the actual pattern, whereas the corresponding synaptic connections between the 9 presynaptic neurons and the “No” postsynaptic neuron are trained with the inverse of the pattern to obtain the respective conductance maps \( G_{i-\text{Yes}/\text{No}} \). Any input pattern from the LED is sensed similarly using the MoS\(_2\) based neuromorphic photosensor and transduced into \( I_{DS} \) spikes which are then fed to the neuromorphic encoding modules. For spike-count based inference, the output voltage spikes \( V_{ij}, i = 1,2,3,...,8,9; j = 1,2,3,...,N_{\text{spike}} \) obtained at the output of the encoding module corresponding to each pixel of the 3×3 image are applied to the drain terminals of the 9 presynaptic neurons. The output currents from the common source terminal i.e. post-synaptic “Yes” and “No” neurons are integrated using capacitors \( (C_{\text{Yes}/\text{No}}) \) to obtain \( V_{\text{Yes}} \) and \( V_{\text{No}} \) as shown in Fig. 5i-j following Eq.1.

\[
V_{\text{Yes}} = \frac{t_{\text{spike}}}{C_{\text{Yes}}} \sum_{i=1}^{9} \sum_{j=1}^{N_{\text{spike}}} G_{i-\text{Yes}} V_{ij} \quad \text{and} \quad V_{\text{No}} = \frac{t_{\text{spike}}}{C_{\text{No}}} \sum_{i=1}^{9} \sum_{j=1}^{N_{\text{spike}}} G_{i-\text{No}} V_{ij} \tag{1}
\]

For spike-timing 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 3×3 image with different spiking durations. In this case, \( V_{\text{Yes}} \) and \( V_{\text{No}} \) are given by Eq. 2.

\[
V_{\text{Yes}} = \frac{1}{C_{\text{Yes}}} \int_0^{t_{\text{spike}}} G_{i-\text{Yes}} V_i \quad \text{and} \quad V_{\text{No}} = \frac{1}{C_{\text{No}}} \int_0^{t_{\text{spike}}} G_{i-\text{No}} V_i \tag{2}
\]

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 total \(\binom{6}{1} + \binom{6}{2} = 21\) 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 the training with the inverse pattern resulting in an order of magnitude higher learned conductance value.

Finally, Fig. 6a-d and Supplementary video 6 show a complete demonstration of our SNN hardware from sensing to encoding to learning. Input pattern obtained by illuminating the blue LED (Fig. 6a) is directly encoded into graded spike trains using the MoS2 based phototransduction module (Fig. 6b) with spike-count reflecting (Fig. 6c) reflecting the analog nature of the input stimulus. Graded spike trains are reshaped by the encoding module into corresponding programming voltages which is subsequently used for learning the pattern via MoS2 based non-volatile synapses (Fig. 6d). For this demonstration, all synapses were initially programmed in their LCS and exact programming spike profiles obtained from the neuromorphic encoding modules were used without invoking any depression to learn the analog pattern. This demonstration highlights the fully integrated nature of our MoS2 based hardware SNN and distinguishes it from other hardware SNN architectures based on CMOS or emerging technologies such as RRAM, PCM, memristor, all-optic, as well as hybrid approaches.
Conclusion

In conclusion, we have experimentally demonstrated a fully integrated and biomimetic SNN hardware platform based on monolayer MoS$_2$ that combines sensing, encoding, unsupervised learning, and inference. We have employed both spike-count and spike-timing 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 impact

Figure 6. Complete demonstration of MoS$_2$ based fully integrated hardware SNN. a) Analog input pattern obtained by illuminating the blue LED. Temporal evolution of b) graded spike trains in MoS$_2$ based phototransduction module and corresponding c) spike-count reflecting the analog encoding of the input stimulus. d) Corresponding temporal evolution of conductance states of MoS$_2$ based non-volatile synapses using programming voltages obtained by reshaping the graded spike trains using the encoding module. All synapses were initially programmed in their LCS and exact programming spike profiles obtained from the neuromorphic encoding modules were used without invoking any depression to learn the analog pattern. This demonstration highlights the fully integrated nature of our MoS$_2$ based hardware SNN.
of relative strengths of synaptic potentiation and depression on learning and forgetting. Our accomplishments can be attributed to the unique photoresponse of monolayer MoS$_2$ based phototransistors for sensing, uniquely designed MoS$_2$ based neuromorphic circuit modules for encoding, and programmable and non-volatile MoS$_2$ synapses enabled by our floating-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 MoS$_2$ based low-power and fully integrated hardware SNN system is more bio-realistic 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 Internet of Things (IoT), but also offer a platform for adaptive leaning and for modeling plasticity-related learning disorders of the BNNs.
Methods

Film growth: Monolayer MoS$_2$ was deposited on epi-ready 2” c-sapphire substrate by metalorganic 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 [51]. Molybdenum hexacarbonyl (Mo(CO)$_6$) and hydrogen sulfide (H$_2$S) were used as precursors. Mo(CO)$_6$ maintained at 10°C and 950 Torr in a stainless-steel bubbler was used to deliver 0.036 sccm of the metal precursor for the growth, while 400 sccm of H$_2$S was used for the process. MoS$_2$ deposition was carried out at 1000°C and 50 Torr in H$_2$ ambient, where monolayer growth was achieved in 18 min. The substrate was first heated to 1000°C in H$_2$ and maintained for 10 min before the growth was initiated. After growth, the substrate was cooled in H$_2$S to 300°C to inhibit decomposition of the MoS$_2$ films.

Film transfer: After the growth of monolayer MoS$_2$ on sapphire substrate, the film is then transferred onto the FET gate dielectric substrate by wet transfer technique. Polymethyl-methacrylate (A3 PMMA) resist is spin coated onto the growth substrates encapsulating the MoS$_2$ and then immersed into the 1M NaOH solution kept at 90°C. Capillary action draws the NaOH solution to the PMMA/substrate interface, separating the hydrophobic PMMA/MoS$_2$ from the sapphire substrate. The detached film floats on the surface, which is then rinsed for multiple times in deionized water and is finally transferred on to the Alumina/ Pt/TiN/p$^{++}$Si gate dielectric stack[52].

Back-gate stack fabrication: Direct replacement of thermally oxidized SiO$_2$ with a high-$\kappa$ dielectric such as Al$_2$O$_3$ grown via atomic layer deposition (ALD) is a logical choice to scale
the effective oxide thickness (EOT). However, we found that Al₂O₃/p⁺⁺-Si interface is not ideal for back-gated FET fabrication owing to higher gate leakage current, more interface trap states and large hysteresis which negatively impact the performance of the device. Replacing Si with Pt, a large work function metal (5.6 eV) allows minimal hysteresis and trap state effects [53]. Since Pt readily forms a Pt silicide at temperatures as low as 300 °C, a 20 nm TiN diffusion barrier deposited by reactive sputtering was placed between the p⁺⁺ Si and the Pt permitting subsequent high temperature processing [54]. This conductive TiN diffusion barrier allows the back-gate voltage to be applied to the substrate, thus simplifying the fabrication and measurement procedures. The polycrystalline Pt introduces very little surface roughness to the final Al₂O₃ surface with a rms roughness of 0.7 nm.

Fabrication of monolayer MoS₂ FET: We have fabricated the back-gated field effect transistors on a 50nm alumina (Al₂O₃) acting as a gate oxide and a stack of Pt/TiN/p⁺⁺ Si as a back-gate electrode. First, MOCVD grown MoS₂ are transferred onto the alumina sample, then the sample is spin coated with A6 PMMA and followed by electron-beam (e-beam) lithography to specify the channels and then separating them out by sulfur hexafluoride (SF₆) etch under 5 degree centigrade for 30s. After etch step, sample is rinsed in Acetone for 30 min followed by 2-propanol (IPA). To define the source and drain contacts, sample is then spin coated with methyl methacrylate (MMA) followed by A3 PMMA. Then using electron-beam lithography source and drain contacts are patterned and further developed by using 1:1 mixture of 4-methyl -2-pentanone (MIBK) and 2 propanol for 60s. 40nm of Nickel (Ni) and 30 nm of Gold (Au) are deposited/ evaporated on to the patterns using E-beam evaporation. Lift- off the evaporated materials is done by immersing the sample in Acetone for 30 min followed by 2-propanol (IPA).
**Electrical Characterization:** Electrical characterization of the fabricated devices are performed using Lake Shore CRX-VF probe station under atmospheric condition using a Keysight B1500A parameter analyzer.

**Data Availability:** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

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

[1] M. H. Hassoun, *Fundamentals of artificial neural networks*: MIT press, 1995.

[2] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, et al., "Mastering the game of Go without human knowledge," *Nature*, vol. 550, pp. 354-359, 2017/10/01 2017.

[3] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, et al., "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, pp. 668-673, 2014.

[4] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, et al., "Loihi: a neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, pp. 82-99, 2018.

[5] S. Ghosh-Dastidar and H. Adeli, "Spiking neural networks," *International journal of neural systems*, vol. 19, pp. 295-308, 2009.

[6] F. Ponulak and A. Kasinski, "Introduction to spiking neural networks: Information processing, learning and applications," *Acta neurobiologiae experimentalis*, vol. 71, pp. 409-433, 2011.

[7] B. Han, A. Sengupta, and K. Roy, "On the energy benefits of spiking deep neural networks: A case study," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 971-976.

[8] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, et al., "Towards artificial general intelligence with hybrid Tianjic chip architecture," *Nature*, vol. 572, pp. 106-111, 2019.

[9] B. Glackin, T. M. McGinnity, L. P. Maguire, Q. Wu, and A. Belatreche, "A novel approach for the implementation of large scale spiking neural networks on FPGA hardware," in *International Work-Conference on Artificial Neural Networks*, 2005, pp. 552-563.

[10] Z. Wang, S. Joshi, S. Savel’ev, W. Song, R. Midya, Y. Li, et al., "Fully memristive neural networks for pattern classification with unsupervised learning," *Nature Electronics*, vol. 1, pp. 137-145, 2018.

[11] M. Al-Shedivat, R. Naous, G. Cauwenberghs, and K. N. Salama, "Memristors empower spiking neurons with stochasticity," *IEEE journal on Emerging and selected topics in circuits and systems*, vol. 5, pp. 242-253, 2015.

[12] V. Milo, G. Pedretti, R. Carboni, A. Calderoni, N. Ramaswamy, S. Ambrogio, et al., "Demonstration of hybrid CMOS/RRAM neural networks with spike time/rate-dependent plasticity," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 16.8.1-16.8.4.

[13] S. Kim, M. Ishii, S. Lewis, T. Perri, M. BrightSky, W. Kim, et al., "NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning," in *2015 IEEE international electron devices meeting (IEDM)*, 2015, pp. 17.1.1-17.1.4.

[14] I. Boybat, M. Le Gallo, S. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, et al., "Neuromorphic computing with multi-memristive synapses," *Nature communications*, vol. 9, pp. 1-12, 2018.

[15] S. Ambrogio, N. Ciocchini, M. Laudato, V. Milo, A. Pirovano, P. Fantini, et al., "Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses," *Frontiers in neuroscience*, vol. 10, p. 56, 2016.

[16] P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, et al., "Face classification using electronic synapses," *Nature communications*, vol. 8, pp. 1-8, 2017.

[17] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors," *Nature materials*, vol. 12, pp. 114-117, 2013.

[18] M. Chu, B. Kim, S. Park, H. Hwang, M. Jeon, B. H. Lee, et al., "Neuromorphic hardware system for visual pattern recognition with memristor array and CMOS neuron," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 2410-2419, 2014.

[19] C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, et al., "Analogue signal and image processing with large memristor crossbars," *Nature Electronics*, vol. 1, p. 52, 2018.
[20] Z. Yin, H. Li, H. Li, L. Jiang, Y. Shi, Y. Sun, et al., "Single-layer MoS2 phototransistors," ACS nano, vol. 6, pp. 74-80, 2012.
[21] L. Wang, Y. Wang, J. I. Wong, T. Palacios, J. Kong, and H. Y. Yang, "Functionalized MoS2 nanosheet-based field-effect biosensor for label-free sensitive detection of cancer marker proteins in solution," Small, vol. 10, pp. 1101-1105, 2014.
[22] M. Park, Y. J. Park, X. Chen, Y. K. Park, M. S. Kim, and J. H. Ahn, "MoS2-based tactile sensor for electronic skin applications," Advanced Materials, vol. 28, pp. 2556-2562, 2016.
[23] A. J. Arnold, T. Shi, I. Jovanovic, and S. Das, "Extraordinary Radiation Hardness of Atomically Thin MoS2," ACS applied materials & interfaces, vol. 11, pp. 8391-8399, 2019.
[24] Q. Smets, G. Arutchelvan, J. Jussot, D. Verreck, I. Asselberghs, A. N. Mehta, et al., "Ultra-scaled MOCVD MoS2 MOSFETs with 42nm contact pitch and 250μA/μm drain current," in 2019 IEEE International Electron Devices Meeting (IEDM), 2019, pp. 23.2.1-23.2.4.
[25] A. Rai, A. Valsaraj, H. C. Movva, A. Roy, R. Ghosh, S. Sonde, et al., "Air Stable Doping and Intrinsic Mobility Enhancement in Monolayer Molybdenum Disulfide by Amorphous Titanium Suboxide Encapsulation," Nano Lett, vol. 15, pp. 4329-36, Jul 8 2015.
[26] C. D. English, K. K. H. Smithe, R. L. Xu, and E. Pop, "Approaching ballistic transport in monolayer MoS2 transistor with self-aligned 10 nm top gates," in 2016 IEEE International Electron Devices Meeting (IEDM), 2016, pp. 5.6.1-5.6.4.
[27] K. M. Price, K. E. Schauble, F. A. McGuire, D. B. Farmer, and A. D. Franklin, "Uniform Growth of Sub-5-Nanometer High-κ Dielectrics on MoS2 Using Plasma-Enhanced Atomic Layer Deposition," ACS applied materials & interfaces, vol. 9, pp. 23072-23080, 2017.
[28] 2DCC. 2d-crystal-consortium. Available: https://www.mri.psu.edu/2d-crystal-consortium/user-facilities/thin-films/list-thin-film-samples-available
[29] K. Kang, S. Xie, L. Huang, Y. Han, P. Y. Huang, K. F. Mak, et al., "High-mobility three-atom-thick semiconducting films with wafer-scale homogeneity," Nature, vol. 520, p. 656, 2015.
[30] S. Wachter, D. K. Polyushkin, O. Bethge, and T. Mueller, "A microprocessor based on a two-dimensional semiconductor," Nature communications, vol. 8, p. 14948, 2017.
[31] D. K. Polyushkin, S. Wachter, L. Mennel, M. Paur, M. Paliy, G. Iannaccone, et al., "Analogue two-dimensional semiconductor electronics," Nature Electronics, vol. 3, pp. 486-491, 2020/08/01 2020.
[32] Q. Gao, Z. Zhang, X. Xu, J. Song, X. Li, and Y. Wu, "Scalable high performance radio frequency electronics based on large domain bilayer MoS2," Nature Communications, vol. 9, p. 4778, 2018/11/14 2018.
[33] S. J. Kim, K. Choi, B. Lee, Y. Kim, and B. H. Hong, "Materials for Flexible, Stretchable Electronics: Graphene and 2D Materials," Annual Review of Materials Research, vol. 45, pp. 63-84, 2015.
[34] F. Torrisi and J. N. Coleman, "Electrifying inks with 2D materials," Nature nanotechnology, vol. 9, pp. 738-739, 2014.
[35] S. Das, A. Dodda, and S. Das, "A biomimetic 2D transistor for audiomorphic computing," Nature Communications, vol. 10, p. 3450, 2019/08/01 2019.
[36] A. Sebastian, A. Pannone, S. S. Radhakrishnan, and S. Das, "Gaussian synapses for probabilistic neural networks," Nature communications, vol. 10, pp. 1-11, 2019.
[37] A. J. Arnold, A. Razavieh, J. R. Nasr, D. S. Schulman, C. M. Eichfeld, and S. Das, "Mimicking Neurotransmitter Release in Chemical Synapses via Hysteresis Engineering in MoS2 Transistors," ACS nano, vol. 11, pp. 3110-3118, 2017.
[38] L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, and T. Mueller, "Ultrafast machine vision with 2D material neural network image sensors," Nature, vol. 579, pp. 62-66, 2020.
[39] J. R. Nasr, N. Simonson, A. Oberoi, M. W. Horn, J. A. Robinson, and S. Das, "Low-Power and Ultra-Thin MoS2 Photodetectors on Glass," *ACS nano*, 2020.

[40] A. Dodda, A. Oberoi, A. Sebastian, T. H. Choudhury, J. M. Redwing, and S. Das, "Stochastic resonance in MoS2 photodetector," *Nature Communications*, vol. 11, p. 4406, 2020/09/02 2020.

[41] D. Jayachandran, A. Oberoi, A. Sebastian, T. H. Choudhury, B. Shankar, J. M. Redwing, et al., "A low-power biomimetic collision detector based on an in-memory molybdenum disulfide photodetector," *Nature Electronics*, pp. 1-10, 2020.

[42] P. Han, L. S. Marie, Q. X. Wang, N. Quirk, A. El Fatimy, M. Ishigami, et al., "Highly sensitive MoS2 photodetectors with graphene contacts," *Nature Nanotechnology*, vol. 29, p. 20LT01, 2018.

[43] K. A. Zaghloul, K. Boahen, and J. B. Demb, "Contrast adaptation in subthreshold and spiking responses of mammalian Y-type retinal ganglion cells," *Journal of Neuroscience*, vol. 25, pp. 860-868, 2005.

[44] P. Cappelletti, C. Golla, P. Olivo, and E. Zanoni, *Flash memories*: Springer Science & Business Media, 2013.

[45] C. Piochon, M. Kano, and C. Hansel, "LTD-like molecular pathways in developmental synaptic pruning," *Nature neuroscience*, vol. 19, p. 1299, 2016.

[46] C. Hansel, "Deregulation of synaptic plasticity in autism," *Neuroscience letters*, vol. 688, pp. 58-61, 2019.

[47] V. Milo, G. Pedretti, R. Carboni, A. Calderoni, N. Ramaswamy, S. Ambrogio, et al., "A 4-transistors/1-resistor hybrid synapse based on resistive switching memory (RRAM) capable of spike-rate-dependent plasticity (SRDP)," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, pp. 2806-2815, 2018.

[48] W. Wang, G. Pedretti, V. Milo, R. Carboni, A. Calderoni, N. Ramaswamy, et al., "Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses," *Science Advances*, vol. 4, p. eaat4752, 2018.

[49] I. Gupta, A. Serb, A. Khiat, R. Zeitler, S. Vassanelli, and T. Prodromakis, "Real-time encoding and compression of neuronal spikes by metal-oxide memristors," *Nature communications*, vol. 7, pp. 1-9, 2016.

[50] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, pp. 208-214, 2019.

[51] Y. Xuan, A. Jain, S. Zafar, R. Lotfi, N. Nayir, Y. Wang, et al., "Multi-scale modeling of gas-phase reactions in metal-organic chemical vapor deposition growth of WSe2," *Journal of Crystal Growth*, vol. 527, 2019.

[52] A. Sebastian, F. Zhang, A. Dodda, D. May-Rawding, H. Liu, T. Zhang, et al., "Electrochemical Polishing of Two-Dimensional Materials," *ACS Nano*, vol. 13, pp. 78-86, Jan 22 2019.

[53] *CRC Handbook of Chemistry and Physics*, 98th Edition (Internet Version 2018) ed. Boca Raton, FL: CRC Press/Taylor & Francis, 2018.

[54] C. A. Crider, J. M. Poate, J. E. Rowe, and T. T. Sheng, "Platinum Silicide Formation under Ultrahigh-Vacuum and Controlled Impurity Ambients," *Journal of Applied Physics*, vol. 52, pp. 2860-2868, 1981.
AUTHOR INFORMATION

Corresponding Author

sud70@psu.edu, das.sapt@gmail.com

Author Contributions

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

Competing Interest

The authors declare no competing interests

Acknowledgement

The work was supported by Army Research Office (ARO) through Contract Number W911NF1920338. Authors also acknowledge Mr. Amritanand Sebastian and Mr. Thomas F Schranghamer for help with device fabrication. 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.
Figure Captions

Figure 1. All-in-one biomimetic hardware SNN based on MoS2 FETs. a) Example optical stimulus. b) Phototransduction pathways in eyes for spike encoding. c) Visual cortex for information processing. d) Blue LED used as external optical stimulus for MoS2 based SNN. e) MoS2 based neuromorphic phototransistor, which is equivalent to photoreceptor cells (rods and cones) in the eyes that convert external optical stimuli into corresponding graded potentials. Rods primarily enable scotopic vision whereas cones are responsible for photopic vision, both of which can be achieved using our MoS2 PT. f) MoS2 based neuromorphic encoding cells mimicking the functionality of retinal ganglion cell that encode the graded potentials into spike trains. g) Optical image of MoS2 based non-volatile and electrically programmable synaptic array. An example experimental demonstration of h) pattern illumination using LED and corresponding i) sensory transduction using MoS2 phototransistor, j) spike-count based encoding using MoS2 encoding module, and k) conductance states of MoS2 synapses following unsupervised learning. l) Transfer characteristics, i.e., source to drain current ($I_{DS}$) as a function of the back-gate voltage ($V_{BG}$) at different drain biases ($V_{DS}$), m) output characteristics i.e., $I_{DS}$ versus $V_{DS}$ for different $V_{BG}$, n) phototransduction under different LED illumination, and o) spike-based programmed states for representative MoS2 FETs. Overall, the platform offers all capabilities including sensing, computing, and non-volatile storage based on monolayer MoS2 FETs integrated with programmable analog memory gate-stack (Al$_2$O$_3$/Pt/TiN/p$^{++}$-Si) allowing in-memory computing and sensing.

Figure 2. MoS2 based neuromorphic sensor. a) Transfer characteristics of monolayer MoS2 PT at $V_{DS} = 1$ V before and after illumination from the blue LED with input currents ranging from
I_{LED} = 0.5 \text{ mA (low-brightness)} \text{ to } I_{LED} = 50 \text{ mA (high-brightness)} \text{ at different } V_{BG} = V_{write} \text{ for 100 ms. For illuminations in the on-state } (V_{write} = 2.0 \text{ V}) \text{ and in the subthreshold regime } (V_{write} = 0.5 \text{ V}), \text{ there are no visible shift in the device characteristics post-illumination. This can be ascribed to photocarrier generation in the MoS}_2 \text{ channel, which are swept across by the applied } V_{DS} \text{ and hence there is no persistent photocurrent beyond the optical exposure. However, for illuminations in the off-state } (V_{write} = -1.5 \text{ V} \text{ and } -2.5 \text{ V}) \text{ photocarrier trapping at the } MoS_2/dielectric \text{ interface leads to the shift in the device threshold voltage } (V_{TH}). \text{ The detrapping mechanism can be rather slow and can take hours to several days, which is why the shift is visible post-illumination. Higher } I_{LED} \text{ and more negative } V_{write} \text{ naturally result in more trapping and hence larger shifts. b) Analog valued and continuous time input optical stimuli from the blue LED. Corresponding } I_{DS} \text{ sampled every 100 ms with } V_{BG} \text{ toggling between } V_{read} = 0 \text{ V and } c) V_{write} = -1.0 \text{ V, d) } V_{write} = -1.5 \text{ V, e) } V_{write} = -2.0 \text{ V and f) } V_{write} = -2.5 \text{ V. The magnitude of the } I_{DS} \text{ spikes increases monotonically during sampling for any given } I_{LED} \text{ and } V_{write} \text{ owing to continuous carrier trapping resulting in gradual } V_{TH} \text{ shift. g) Number of spikes } (N_{spike}) \text{ and h) total spiking duration } (\tau_{spike}) \text{ as a function of } I_{LED} \text{ and } V_{write}. \text{ A spike is counted when } I_{DS} > I_{ST}, \text{ where } I_{ST} = 3 \text{ nA is the spiking threshold (dotted red line in c-f). i) Average energy consumption per spike } (E_{spike}) \text{ for different } I_{LED} \text{ and } V_{write}.}

**Figure 3. MoS2 based neuromorphic encoders.** a) Circuit diagram for spike-count based encoding module comprising of switch capacitor cell 1 (SCC1), spike reshaping cell (SRC), switch capacitor cell 2 (SCC2), and voltage inverter cell (VIC). The input to the module are graded } I_{DS} \text{ spikes obtained from the MoS}_2 \text{ PT and output of the module are corresponding programming voltage spikes } (V_p), \text{ which are relayed to the learning module based on non-volatile MoS}_2.
synapses. b) Input $I_{DS}$ spikes from the MoS$_2$ PT biased at $V_{\text{write}} = -1.5$ V for $I_{LED} = 10$ mA, and corresponding output of c) SCC1 ($V_{C1}$), d) SRC ($I_{S2}$), e) SCC2 ($V_{C2}$), and f) VIC ($V_p$). g) Circuit diagram for spike-timing based encoding module comprising of integrator cell (IC), spike reshaping cell (SRC), resistor cell (RC), and voltage inverter cell (VIC). (h) Input $I_{DS}$ spikes from the MoS$_2$ PT biased at $V_{\text{write}} = -1.5$ V for $I_{LED} = 10$ mA, and corresponding output of i) IC ($V_{C4}$), j) SRC ($I_{S4}$), k) RC ($V_{SS}$), and l) VIC ($V_p$). $S_1$, $S_2$, $S_3$, $S_4$ and $S_5$ are MoS$_2$ FETs, $C_1 = C_2 = 350$ pF, and $C_4 = 20$ nF. See Supplementary Information 4-7 for more detail on the encoding modules. Average encoding energy expenditure per spike by each cell for m) spike-count and n) spike-timing based encoding module for different $I_{LED}$.

**Figure 4. Electrically programmable analog and non-volatile MoS$_2$ synapses.** a) Potentiation of a MoS$_2$ synapse from low conductance state (LCS) after the application of a fixed number of programming spikes ($N_{\text{spike}} = 16$) of different amplitudes of negative polarity ($V_p$) with each spike applied for $t_{\text{spike}} = 100$ ms. b) Post-potentiated conductance states ($G_p$) measured at $V_{BG} = 0$ V as a function of $N_{\text{spike}}$ for different $V_p$. c) Depression of a MoS$_2$ synapse from high conductance state (HCS) after the application of a fixed number of programming spikes ($N_{\text{spike}} = 16$) of different amplitudes of positive polarity ($V_D$) with each spike applied for $t_{\text{spike}} = 100$ ms. d) Post-depressed conductance states ($G_D$) measured at $V_{BG} = 0$ V as a function of $N_{\text{spike}}$ for different $V_D$. e) Potentiation of MoS$_2$ synapse from LCS after the application of single spike of constant magnitude $V_p = -9$ V for different $t_{\text{spike}}$. f) Post-potentiated $G_p$ measured at $V_{BG} = 0$ V as a function of $t_{\text{spike}}$ for different $V_p$. g) Depression of MoS$_2$ synapse from HCS after the application of single spike of constant magnitude $V_D = 14$ V for different $t_{\text{spike}}$. h) Post-depressed $G_D$ measured at $V_{BG} = 0$ V as a function of $t_{\text{spike}}$ for different $V_D$. i) Retention for 10 representative potentiated ($G_p$) states. j)
Programming energy expenditure for potentiation. k) Retention for 10 representative depressed (G_D) states. l) Programming energy expenditure for depression.

**Figure 5. Spike-based unsupervised learning using MoS2 synapses.** a) Schematic of a 2-layer SNN with 9 presynaptic neurons and 1 postsynaptic neuron for learning and inferring patterns from 3×3 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 SNN, whereas during the depression all synapses are uniformly depressed. Spiking profiles used for c) spike-count and d) spike-timing based learning. For each type of learning, three SNN 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) are adjusted using the spike magnitude and spike duration for spike-count and spike-timing based learnings, respectively. The time evolution of colormap of synaptic weights i.e., the conductance states of the 9 synapses during e) spike-count and f) spike-timing based learning. For each type of learning all synapses are initialized either in a high conductance state (HCS) with G_{HCS} = 100 nS, or a low conductance state (LCS) with G_{LCS} = 100 pS (also see the Supplementary Video 3 and 4). Learning of the left diagonal followed by relearning of the right diagonal when potentiation and depression are both strong for g) spike-count and h) spike-timing based learnings (also see the Supplementary Video 5 and 6). Two sets of 9×1 synapses with synaptic weights for i) “Yes” postsynaptic neuron learned using the actual pattern and j) “No” postsynaptic neuron learned using the inverse of the pattern. The output currents from the “Yes” and “No” neurons are integrated using capacitors (C_{Yes/No}) to obtain V_{Yes} and V_{No} to determine the winner.
Figure 6. Complete demonstration of MoS2 based fully integrated hardware SNN. a) Analog input pattern obtained by illuminating the blue LED. Temporal evolution of b) graded spike trains in MoS2 based phototransduction module and corresponding c) spike-count reflecting the analog encoding of the input stimulus. d) Corresponding temporal evolution of conductance states of MoS2 based non-volatile synapses using programming voltages obtained by reshaping the graded spike trains using the encoding module. All synapses were initially programmed in their LCS and exact programming spike profiles obtained from the neuromorphic encoding modules were used without invoking any depression to learn the analog pattern. This demonstration highlights the fully integrated nature of our MoS2 based hardware SNN.