A Bio-inspired and Low-power 2D Machine Vision with Adaptive Machine Learning and Forgetting

Akhil Dodda
Pennsylvania State University https://orcid.org/0000-0002-6022-4028
Darsith Jayachandran
Pennsylvania State University https://orcid.org/0000-0001-7849-1995
Shiva Subbulakshmi Radhakrishnan
Pennsylvania State University
Saptarshi Das (sud70@psu.edu)
Pennsylvania State University https://orcid.org/0000-0002-0188-945X

Article

Keywords: Adaptive Neurobiological Mechanisms, Artificial Intelligence, Energy Gap, Monolayer 2D Phototransistor Array, Memory Gate-stack, von Neumann Bottleneck

Posted Date: February 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-258246/v1

A Bio-inspired and Low-power 2D Machine Vision with Adaptive Machine Learning and Forgetting

Akhil Dodda[1], Darsith Jayachandran[1], Shiva Subbulakshmi Radhkrishnan[1], and Saptarshi Das[2,3,4,]*

[1]Department of Engineering Science and Mechanics, Pennsylvania State University, University Park, PA 16802, USA

[2]Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA

[3]Materials Research Institute, Pennsylvania State University, University Park, PA 16802, USA

Abstract: Natural intelligence has many dimensions, and in animals, learning about the environment and making behavioral changes are some of its manifestations. In primates vision plays a critical role in learning. The underlying biological neural networks contain specialized neurons and synapses which not only sense and process the visual stimuli but also learns and adapts, with remarkable energy efficiency. Forgetting also plays an active role in learning. Mimicking the adaptive neurobiological mechanisms for seeing, learning, and forgetting can, therefore, accelerate the development of artificial intelligence (AI) and bridge the massive energy gap that exists between AI and biological intelligence. Here we demonstrate a bio-inspired machine vision based on large area grown monolayer 2D phototransistor array integrated with analog, non-volatile, and programmable memory gate-stack that not only enables direct learning, and unsupervised relearning from the visual stimuli but also offers learning adaptability under photopic (bright-light), scotopic (low-light), as well as noisy illumination conditions at miniscule energy expenditure. In short, our

"all-in-one" hardware vision platform combines "sensing", "computing" and "storage" not only to overcome the von Neumann bottleneck of conventional complementary metal oxide semiconductor (CMOS) technology but also to eliminate the need for peripheral circuits and sensors.

An intelligent system, natural or artificial, is one that monitors its environment, learns, or remembers key information, and adapts to changes, as necessary. Animals do this seamlessly, often with very limited resources and in challenging ecological conditions. Their success can be attributed to the underlying biological neural networks (BNNs) that not only correlate and collocate the neural primitives for 'sensing', 'computing', and 'storage', which drastically reduces the energy expenditure for many difficult tasks but also learn and adapt ensuring the survival of the species, even in most resource constrained environments.

The world we 'know', is a result of the perception enabled by our sensory organs. Information embedded in the outside world, takes multiple sensory pathways and associated transformations before it reaches the brain, which then processes it to give out a wide variety of outcomes, sensations, and above all, aids in learning and memory formation. In primates (including humans), vision constitutes a major portion of information input, more than all the other sensory inputs combined. Hence, a major part of the brain is devoted for processing visual stimuli highlighting the importance of visual system in learning [1].

Drawing inspiration from the biological intelligence observed in visual animals, machine learning and machine vision are pushing the limits of artificial intelligence (AI) in our everyday life, from

defeating professional players in the game of "Go" [2] to driving autonomous vehicles in crowded streets [3]. Recent years have seen significant progress in artificial neural networks (ANNs) [4], which are high-level abstractions of the biological neural networks (BNNs), i.e. neurons connected to other neurons through synapses. Software ANNs and their different incarnations such as deep neural networks (DNN) [5], convolution neural networks (CNNs) [6], and recently bio-realistic and event driven spiking neural networks (SNNs) [7] have shown remarkable success in multiple applications including image processing, pattern classification, and solving complex optimization problems. Their hardware implementation have primarily relied on the Si-based conventional complementary metal-oxide-semiconductor (CMOS) technology [8-10]. However, unlike BNNs, where the "computing" primitives, i.e. neurons, and "storage" units, i.e. synapses are collocated, the von Neumann architecture used by Si CMOS separates "compute" from "memory" leading to orders of magnitude higher energy expenditure compared to what the brain requires for similar tasks [11-13]. Non von Neumann computing architectures based on field programmable gate arrays (FPGAs) [14], and resistive random-access memory (RRAM) [15-20] bridge the gap between "memory" and "compute" and offer energy efficient alternatives for hardware implementation of ANNs. However, these in-memory compute architectures mostly rely on CMOS-based peripheral sensors and circuits adding area and energy overhead [12, 13, 21-23].

Early machine vision, on the other hand primarily relied on developing better photodetectors and in some instances integrating CMOS-based preprocessing modules for image filtering, feature extraction, light-level adaptation, noise elimination etc. bridging the gap between "sensing" and "compute" [24-28]. More recent developments in the area of brain-inspired machine vision also attempts to combine "sensing" with "memory" [13, 23, 29]. While some of these demonstrations

exploit oxide-based [30] [31] [32] and organic memristors [33], perovskites [34], etc., two-dimensional (2D) semiconducting monolayers are receiving significant attention owing to their superior photosensitivity [35, 36], gate-tunability [37], scalability [38], and integration with memory [39-44]. Some show machine learning in hardware based on back-propagation algorithm implemented using software ANNs, whereas others preprocessed the images using software and demonstrated tasks like MNIST digit recognition or filtering using hardware ANNs . Interestingly, none integrates "sensing", "compute", "memory" for adaptive machine vision and unsupervised machine learning using a single hardware platform.

Finally, while learning has been a topic of extensive research, the importance of forgetting in learning has not received adequate attention. Most researchers considered forgetting as a passive brain process that allows unused memories to disappear over time. However, this decades-old hypothesis has now been challenged by a new radical idea that suggests that the brain is built not to remember, but to forget [45]. In other words, forgetting is an active brain process that plays an important role in biological learning. As we will elucidate, it is possible to develop new unsupervised learning rules for hardware ANNs based on forgetting.

In the light of the above discussion, it is imperative that the next generations of AI will benefit from an integrated hardware platform that combines machine vision with machine learning mimicking the adaptive neurobiological architectures for seeing, learning, and forgetting. Here we accomplish the same by integrating monolayer MoS_2 phototransistor array with analog, non-volatile, and programmable memory gate-stack bridging the gap between "sensing", "compute", and "storage". In short, we combine analog optical memory observed in 2D phototransistors with

analog electrical memory enabled by the back-gate stack for direct learning and unsupervised relearning from the visual stimuli. Our biomimetic hardware vision platform also enables adaptive learning under photopic (bright-light), scotopic (low-light), as well as noisy illumination conditions at miniscule energy expenditure bridging the energy gap between AI and natural intelligence. Finally, our "all-in-one" vision platform not only overcomes the von Neumann bottleneck of CMOS-based ANNs but also eliminates the need for CMOS-based peripheral sensors and circuit components.

Fig 1 illustrates the biological resemblance and functional capabilities of our hardware vision platform. Fig. 1a-d show the visual BNN in humans and associated neural primitives for seeing and learning. Information in the outside world is conveyed through images which are projected by the lens system of the eyes into the retina (Fig. 1a). Retina contains a counterintuitive stack of neurons where multiple preprocessing steps take place, and above all, the photoreceptors, i.e. rods and cones (Fig. 1b) in retina enable vision in a wide range of illumination levels. Images under photopic illuminations are mostly sensed with cone photoreceptors, whereas, rod cells facilitate vision under scotopic illuminations. This ability of the retina is called adaptation allowing humans to perceive information under different illumination conditions [46]. Photoreceptors also transduce visual information into electrical impulses and, with the help of other cells in the retina, pass on to the visual cortex in the brain (Fig 1c). The visual cortex contains a vast network of neurons which take part in learning. While the neuroscience of learning is still a topic of active research, it is widely accepted that learning leads to strengthening of connections between the associated neurons

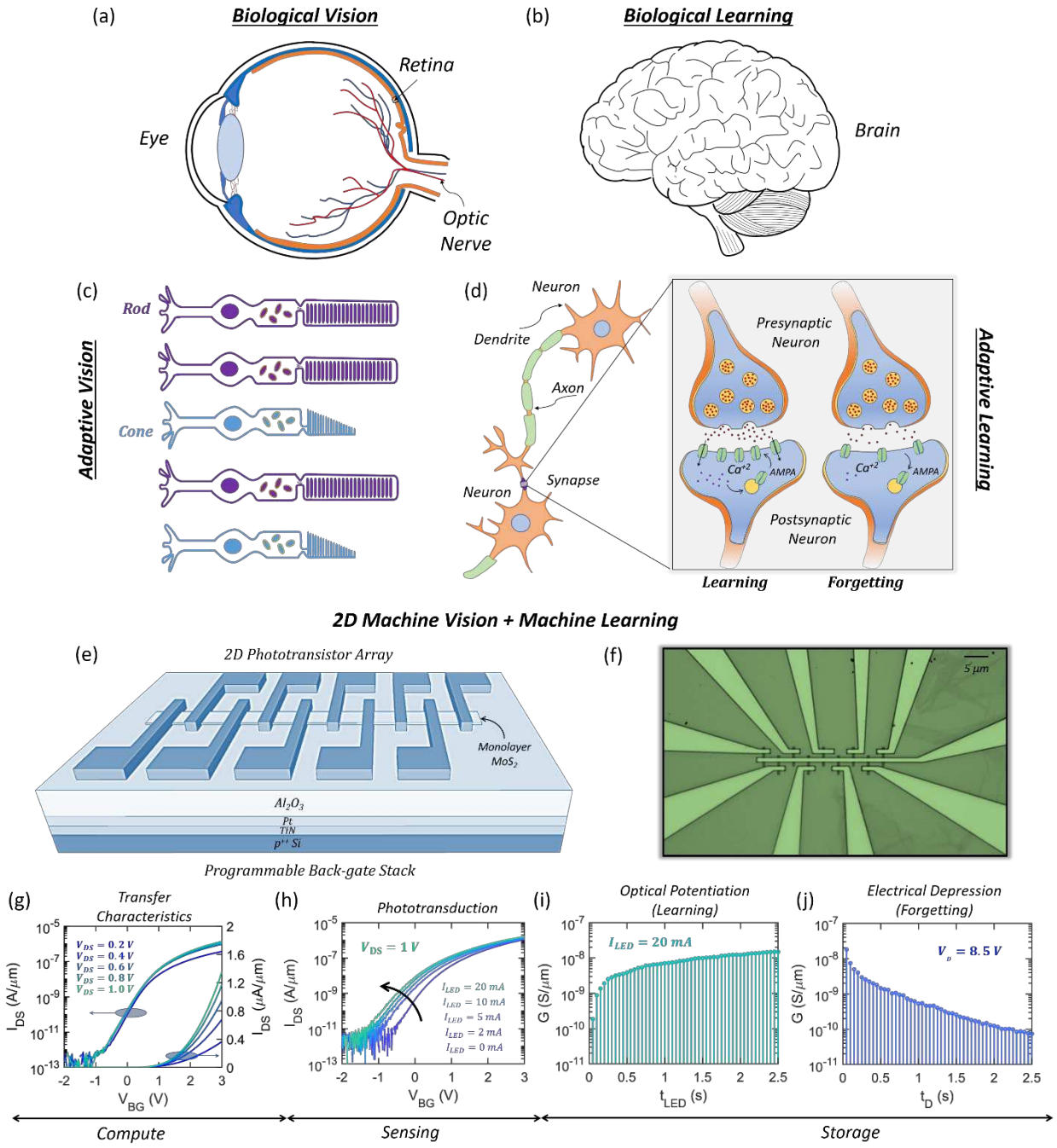

Figure 1. Bio-inspired 2D vision system. *Elementary components of biological visual neural network,* ***a)*** *eyes enabling biological vision and* ***b)*** *visual cortex in the brain enabling biological learning.* ***c)*** *Photoreceptors in the eyes enable phototransduction and adaptation. Rods facilitate scotopic and cones enable photopic vision.* ***d)*** *Synapses strengthen or weaken to learn or forget e.g., by controlling the number of AMPA receptors in the postsynaptic neuron when the presynaptic neuron releases glutamate neurotransmitter.* ***e)*** *Schematic and* ***f)*** *false color microscope image of the artificial vision system comprising of 9×1 2D phototransistor arrays integrated with a programmable back-gate stack. The platform enables functionalities like phototransduction, visual adaptation, synaptic plasticity, direct learning, unsupervised relearning, and learning under dynamic noise exploiting forgetting.* ***g)*** *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}) in dark,* ***h)*** *phototransduction under different levels of illumination from a blue light emitting diode (LED),* ***i)*** *optical potentiation induced learning or increase in device conductance (G), and* ***j)*** *electrical depression induced forgetting or decrease in G measured at V_{BG} = 0 V in a representative 2D phototransistor.*

through a process known as synaptic plasticity (Fig. 1d) [47]. For example, long-term potentiation or memory formation leads to increase in the number of AMPA receptors in the postsynaptic neuron when the presynaptic neuron uses glutamate as the neurotransmitter [48]. Similarly, forgetting leads to weakening of connection strengths through reduction in the number of AMPA receptor. The mathematical construct of synaptic plasticity determines the biological learning rule, most of which can be categorized as unsupervised in the context of machine learning, although evidence of reward-based or reinforcement learning can also be found [49].

Fig. 1e-f, respectively, show the schematic and optical image of the 2D vision platform that integrates 2D phototransistor array with non-volatile and programmable floating back-gate architecture. The back-gate stack comprised of atomic layer deposition (ALD) grown 50 nm Al_2O_3 on Pt/TiN/p^{++}-Si. As we will elucidate later, photo-induced carrier trapping at the MoS_2/Al_2O_3 interface and its dependence on the back-gate bias enable analog optical memory, which leads to direct and adaptive learning from visual stimuli under different illumination conditions. On the other hand, analog electrical memory enabled by the programmable back-gate stack allows for the realization of biological equivalent of forgetting, which facilitates unsupervised relearning as well as learning under noisy illumination. Combined, the stack enables adaptive vision and learning through in-memory sensing and computing resembling the visual BNN in humans.

Monolayer MoS_2 used in this study was grown *via* MOCVD technique using carbon-free precursors at 1000 ^{0}C on a epitaxial sapphire substrate to ensure high film quality. Following the growth, the film was transferred on to the back-gate stack for the fabrication of the phototransistor array (See the ***Methods*** section for further details on synthesis, film transfer, and device

fabrication). Fig. 1g 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}) for a representative monolayer 2D phototransistor measured in dark (***Supplementary Figure 1a-c***, respectively, show the low device-to-device variation across the array, output characteristics, and back-gate hysteresis in a representative phototransistor). Fig. 1h shows the response of the phototransistor to different levels of illumination from a blue light emitting diode (LED). Note that, instead of LASER illumination, used in most studies [35], we have chosen LED since it represents more natural lighting conditions where artificial vision sensors will be deployed. Finally, Fig. 1i-j, respectively, show optical potentiation induced learning or increase in device conductance (G) and electrical depression induced forgetting or decrease in G measured at V_{BG} = 0 V in a representative phototransistor constructing the hardware foundation artificial machine intelligence.

Direct and analog learning from visual stimuli using 2D phototransistor array

ANNs require involvement of photosensitive materials with unique properties to perform machine vision operations such as analog sensing and adaptation. Direct-bandgap monolayer 2D materials with their superior photosensitivity, and electrostatic gate tunability are, therefore, natural choices for next generations of bio-inspired machine vision platforms. Additionally, the atomically thin body nature of 2D monolayers allow aggressive dimension scaling and hence enable high integration density as reported recently [38, 50]. Moreover, some of the early criticism of 2D materials have also been successfully addressed through the realization of low contact resistance [51], high ON current [52], integration of ultra-thin and high-k gate dielectric [53], and wafer scale growth [54, 55] making them a technologically viable option. Demonstration of 2D-based microprocessors [56], analogue operational amplifier [57], and RF electronics components [58]

support this claim. Finally, unlike silicon CMOS, 2D materials enable flexible [59] and printable [60] electronic circuits adding value towards 2D-based biomimetic and neuromorphic hardware platforms [61-63].

Illumination of a 2D semiconducting channel in a phototransistor will generate photocarriers, which under an electrical bias will drift towards the respective electrodes adding to the already existing dark current in the device. The illumination intensity will determine the change in conductance of the channel, allowing one to leverage this property for analog machine vision sensors. In most cases, after the stimuli is removed, the conductance regains the initial state, without remembering the change induced by the visual stimuli. This is a limitation for many machine vision demonstrations, necessitating peripheral circuit elements to store the new conductance value induced by the optical stimuli [35]. This challenge is overcome with the property called 'optical memory' or persistent photoconductivity, which allows materials/devices to remain in the new conductance state even after the visual stimuli is removed. In 2D based vision sensors, this is accomplished mainly with trapping of photocarriers by the trap states at the 2D semiconductor-oxide interface [35, 41-43, 63, 64]. The trapped charges will alter the threshold voltage of the device, changing the conductance measured at a given V_{BG}. The retention of this optically induced conductance state depends mainly on the detrapping time and may range from several hours to days. This is called photo-gating effect and is leveraged in our 2D based machine vision to demonstrate direct learning from visual stimuli.

Fig. 2a-b, respectively, show the post illumination transfer characteristics of a representative phototransistor at V_{DS} = 1 V measured in dark after 10 s exposure to different LED brightness

levels (I_{LED}) with V_{write} = 2 V and V_{write} = -3 V applied to the back-gate during the illumination. Note that for V_{write} = 2 V, i.e. illuminations in the on-state of the phototransistor, there are little-to-no change in the device characteristics post-illumination irrespective of the brightness level of the LED (I_{LED}). As mentioned above, under these operating conditions, the photocarriers generated in the semiconducting channel drift towards the respective electrodes under the applied drain bias (V_{DS} = 1 V) resulting in non-persistent photoconductivity beyond the optical illumination. However, for V_{write} = -3 V, i.e. illuminations in the off-state of the phototransistor, noticeable shifts are observed in the threshold voltage (V_{TH}) of the device post-illumination. This can be ascribed to the photo-gating effect described above. Fig. 2c shows the evolution of post-illumination conductance (G) of the phototransistor measured at V_{BG} = 0 V as a function of the LED illumination time (t_{LED}) for different I_{LED}. Since the photocarrier trapping leads to negative V_{TH} shift, G increases with increasing t_{LED} and I_{LED} as more carriers are available for trapping. Fig. 2d shows the retention characteristics for G corresponding to different I_{LED} suggesting that the learned conductance states are stable over time.

Leveraging the merits offered by the photo-gating effect described above we demonstrate direct learning from optical stimuli (Fig. 2e and ***Supplementary Video 1***). An analog image of size 3×3, comprising of 4 different LED illumination intensities, I_{LED}= 2 mA, 5 mA, 10 mA and 20 mA is presented pixel by pixel to the phototransistors in the 9×1 array at V_{write}= -3 V. The conductance states (G) are read at V_{BG} = 0 V successively for 50 epochs by sampling I_{DS} every 500 ms. All devices start from the same conductance state, $G \approx$ 1 nS at V_{BG} = 0 V. During each epoch, devices learn the input image by updating G. As expected, devices exposed to brighter intensities reach higher G compared to the devices exposed to lower intensities due to the

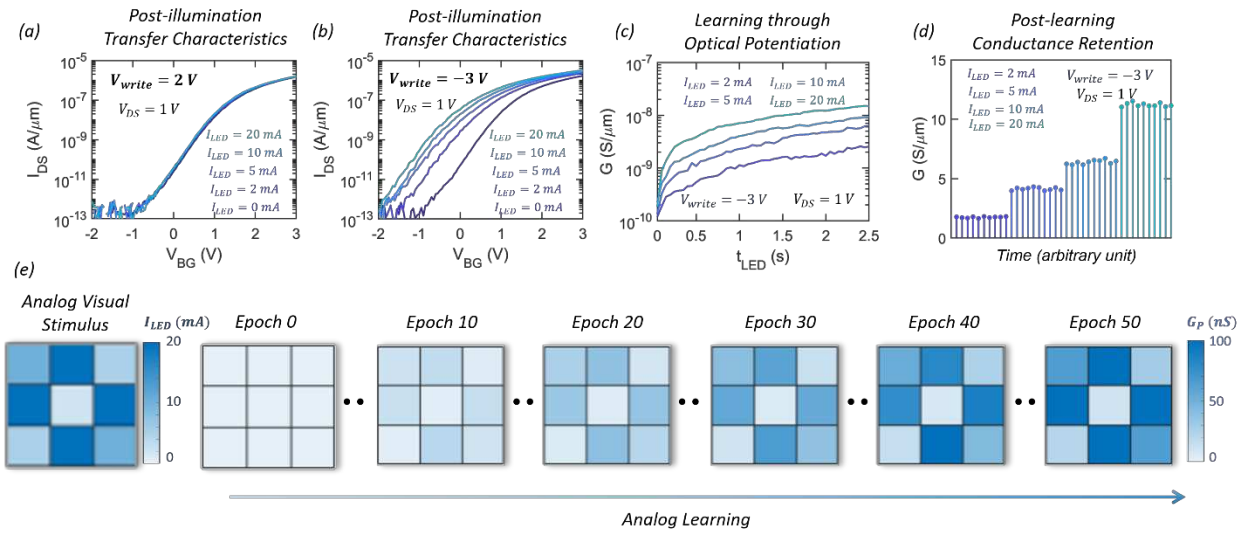

Figure 2. Analog vision and learning. *Post illumination transfer characteristics of a representative phototransistor at V_{DS} = 1 V measured in dark after 10 s exposure to different LED brightness levels (I_{LED}) with* ***a)*** *V_{write} = 2 V and* ***b)*** *V_{write} = -3 V applied to the back-gate during the illumination.* ***c)*** *Evolution of post-illumination conductance (G) of the phototransistor measured at V_{BG} = 0 V as a function of the LED illumination time (t_{LED}) for different I_{LED}. The monotonic increase in G facilitate direct learning from visual stimuli.* ***d)*** *Retention of the learned conductance states after the removal of optical stimuli.* ***e)*** *Heatmaps of the input image with each pixel corresponds to an I_{LED} value and the output images with each pixel showing the G of corresponding phototransistor in the 9×1 array, measured at V_{BG} = 0 V at different epochs. The input image is learned in 50 epochs. This shows the analog vision and learning capabilities of the photoresponsive 2D array structure.*

difference in photo-gating effect illustrated in Fig. 2c. As a result the heatmap of G (Fig. 2e) mimics the contrast present in the input image at the end of the 50 epochs suggesting direct learning by the 9×1 phototransistor array from the analog visual stimuli. The total learning energy expenditure per pixel after 50 epochs was found to be miniscule, ~ 50 nJ.

Adaptive Learning

As mentioned earlier, a key feature of BNN is neuroplasticity that enables adaptation to learning under changing environmental conditions. Here we demonstrate similar functionalities achievable in the phototransistor array allowing learning under both scotopic and photopic illuminations. While biological vision requires significant amount of time from minutes to hours to adapt to scotopic conditions, our artificial vision can be reconfigured such that similar learning rates are achievable irrespective of the illumination conditions. Fig. 3a shows the time evolution of G

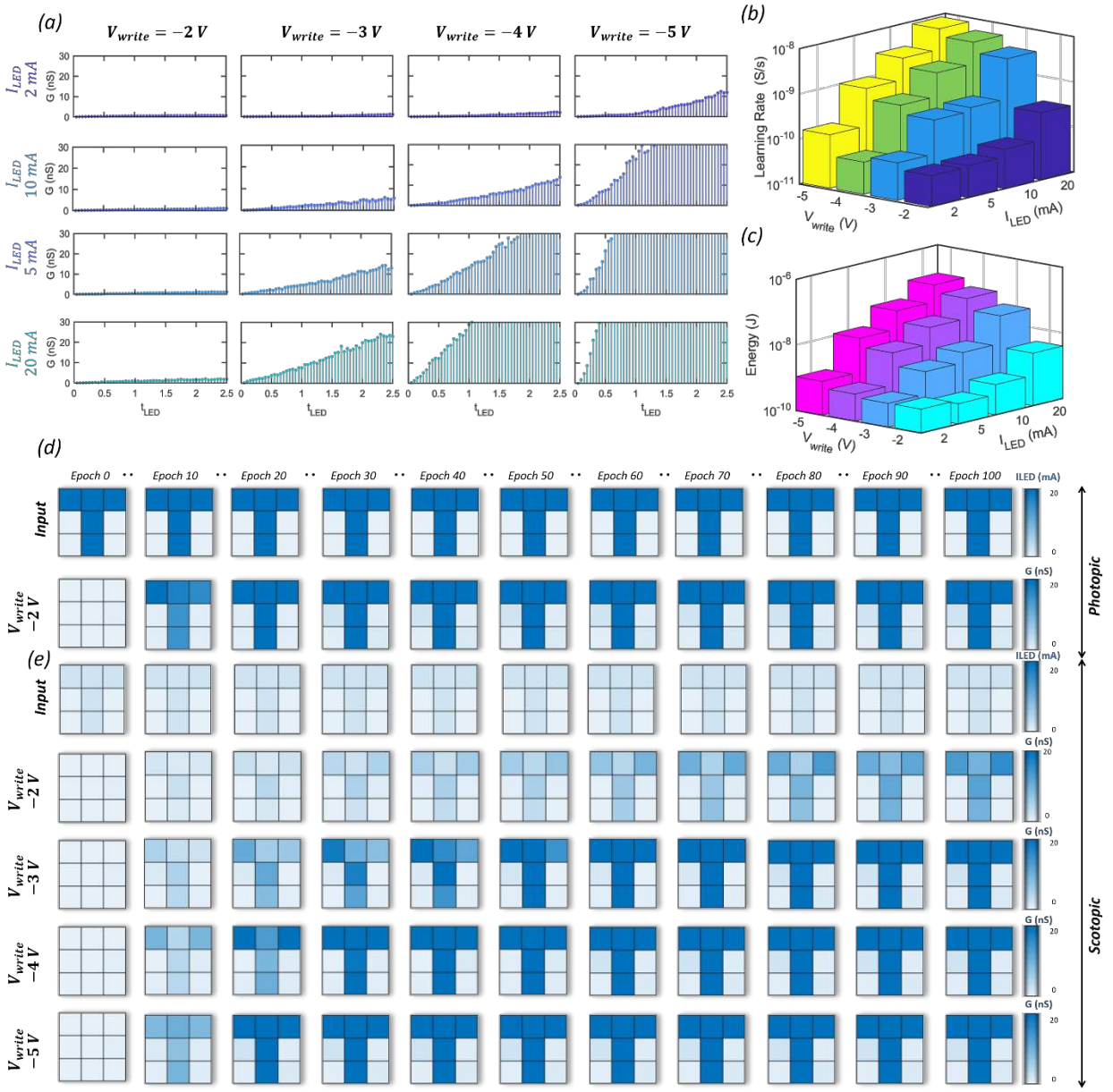

Figure 3. Learning with adaptive machine vision. ***a)*** *Time evolution of G measured at V_{BG} = 0 V, every 500 ms,* ***b)*** *Learning rate, which we define as the rate of increase in G and* ***c)*** *average learning energy expenditure per epoch as a function of I_{LED} and V_{write}.* ***d)*** *Learning under photopic illumination (I_{LED} = 20 mA). Heatmaps of the input image (top row) and the output images showing the heatmap of G measured at V_{BG} = 0 V at different epochs. The photopic "T" is learned within 10 epochs for illumination at V_{write} = -2 V.* ***e)*** *Learning under scotopic illumination (I_{LED} = 2 mA). The same experiment as in (d) is repeated for V_{write} = -2, -3, -4, and -5 V. While the photoresponsive array does not learn scotopic "T" in 50 epochs for V_{write} = -2 V, V_{write} = -5 V enables learning in just 10 epochs. These experiments highlight the importance of adaptive machine vision in learning.*

measured at V_{BG} = 0 V, every 500 ms, corresponding to different I_{LED} and V_{write}. It is observed that for any given I_{LED}, the rate of change in G increases as the magnitude of V_{write} increases, i.e. as V_{write} becomes more negative. This is another consequence of photogating effect as more trap

states become available for carrier capture at the 2D/dielectric interface when the phototransistor is biased in the deep off-state. Fig. 3b shows the learning rate, which we define as the rate of increase in G as a function of I_{LED} and V_{write}. Note that the learning rates obtained for photopic illuminations (e.g. I_{LED}= 20 mA) at a lower negative V_{write} can also be achieved for scotopic illuminations (e.g. I_{LED}= 2 mA) at a higher negative V_{write}. Fig. 3c shows the average learning energy expenditure per epoch as a function of I_{LED} and V_{write}. Fig 3d-e and ***Supplementary Video 2*** show experimental demonstration of adaptive learning using the 9×1 phototransistor array under photopic and scotopic illuminations, respectively. We have used 3×3 pixel images of the letter "T" as the visual stimuli. The photopic and scotopic "T"s were realized by using I_{LED} = 20 mA and I_{LED} = 2 mA, respectively. The photopic "T" is learned within only 10 epochs (Fig. 3d), whereas learning scotopic "T" is incomplete even after 50 epochs at V_{write}= -2 V as shown using time evolution of the heatmap of G measured at V_{BG} = 0 V (Fig. 3e). We consider the learning to be complete when G = 100 nS. However, by using more negative V_{write}, e.g. V_{write} = -5 V, the scotopic "T" can be learned within 10 epochs highlighting the capability of our vision platform to adapt to the learning ambience. The learning energy expenditure for the photopic and scotopic "T"s were found to be ~ 3.6 nJ/pixel at V_{write}= -2 V and ~ 1.1 nJ/pixel at V_{write}= -5 V, respectively.

Adaptive forgetting

Forgetting has traditionally been considered to be a passive brain process, which ensures unused information fade over time so that neural resources can be reallocated for storing more important and newer 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 internet of things (IoT) edge devices deployed in remote locations, forgetting can play an active role in smart learning.

Forgetting is enabled in our phototransistors by exploiting the non-volatile and analog programmability of our back-gate dielectric stack. Fig. 4a shows the transfer characteristics of the phototransistor when positive programming voltages (V_{D}) of increasing amplitude are applied to the back-gate, each for a total duration of 100 ms. During programming, the source and drain terminals are kept grounded. Also note that before programming the device is set at a high conductance state. Transfer characteristics clearly show a positive shift in V_{TH} with increasing magnitude of V_{D}. This change in V_{TH} can be attributed to the back-gate dielectric stack which closely resembles the floating gate architecture in non-volatile flash memory (See ***Supplementary Information 2*** for a detailed explanation of programmable memory using energy band diagrams). In short, the p^{++}-Si/TiN/Pt interface in the stack is characterized by a tunnel barrier (TB), whereas, the gate dielectric, i.e. 50 nm Al_2O_3, acts as an oxide barrier (OB). The OB is much wider and taller compared to the TB. When a positive programming voltage is applied to the control gate (CG), i.e. p^{++}-Si, carriers tunnel from the p^{++}-Si into the Pt/TiN floating gate (FG) and remain trapped even after the programming voltage is removed. These trapped charges on the FG screens the electric field from the CG, which is reflected through positive shift in the V_{TH}. By increasing the amplitude and duration of the programming pulse, the amount of charge injected into the FG can be tuned which allows precise control of V_{TH}.

Fig. 4b shows the evolution of the device conductance (G) measured at V_{BG} = 0 V as a function of number of epochs i.e. programming pulses (N_D) of pulse width t_D = 100 ms of different amplitude

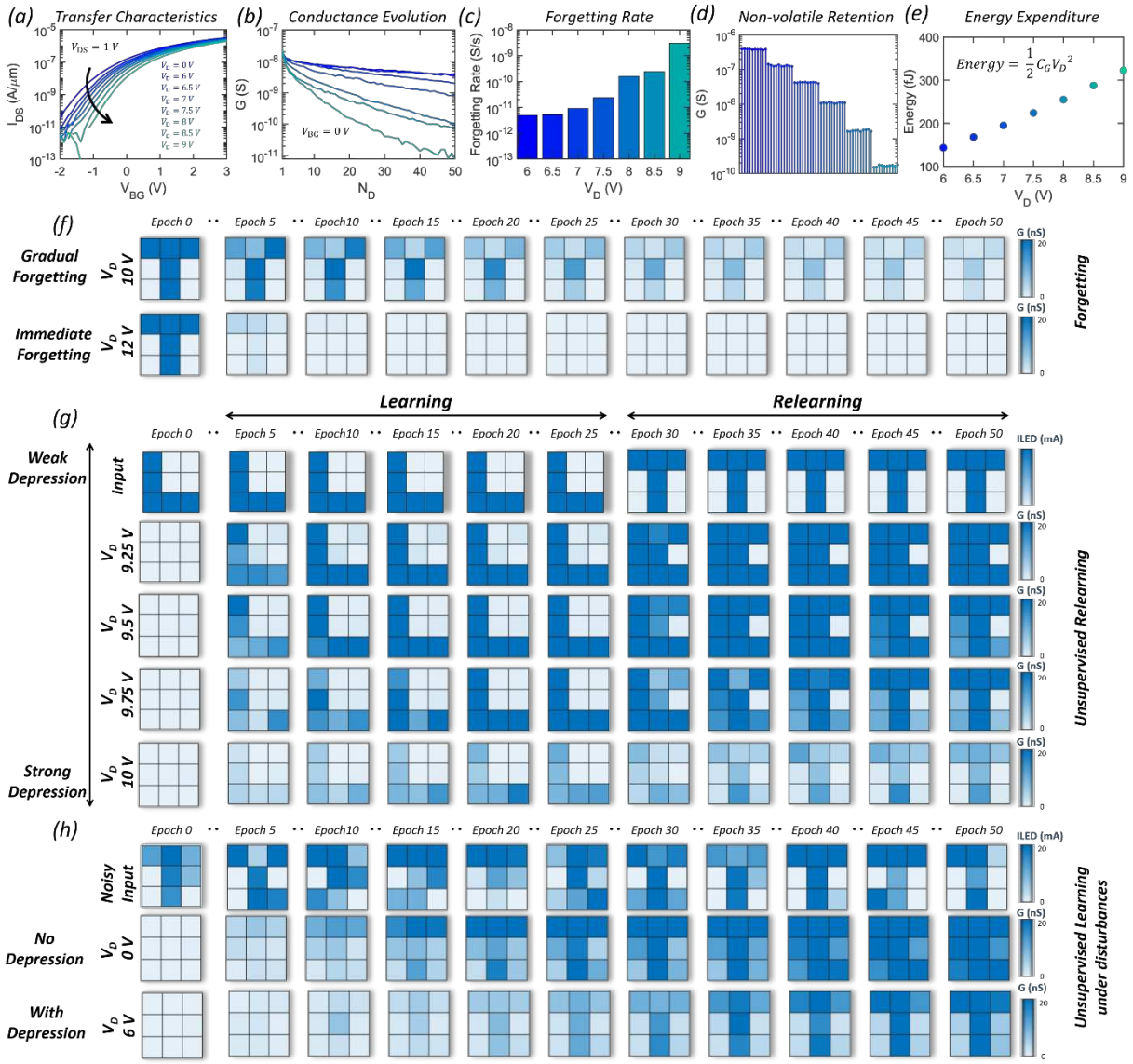

Figure 4. Forgetting for unsupervised relearning and learning under random disturbances. a) *Transfer characteristics of a representative phototransistor when positive programming voltages (V_D) of increasing amplitude are applied to the back-gate, each for a total duration of 100 ms.* ***b)*** *Device conductance (G) measured at V_{BG} = 0 V as a function of number of programming pulses (N_D) of pulse width t_D = 100 ms of different amplitudes (V_D).* ***c)*** *Forgetting rate, defined as the rate of decrease in G for different V_D.* ***d)*** *Non-volatile retention of the post forgotten G measured at V_{BG} = 0 V.* ***e)*** *Forgetting energy expenditure for different V_D.* ***f)*** *Heatmaps of G show that smaller amplitudes of V_D, e.g., V_D= 10 V (top-row) enable gradual forgetting of the learnt letter 'T' while higher amplitudes of V_D, e.g., V_D = 12 V (bottom-row) causes immediate forgetting.* ***g)*** *Unsupervised relearning enabled by adaptive forgetting. Heatmaps show learning a letter 'L' (first 25 epochs) and relearning of another letter 'T' (final 25 epochs) by the phototransistor array with different forgetting rates. Here, each epoch consists of optical potentiation and electrical depression (application of V_D) enabling relearning and eliminating the need for supervision. The top-row shows the input images (I_{LED}). Each row below shows the corresponding heatmaps of G for different amplitudes of V_D. Under optimum potentiation and depression (e.g., V_D = 9.75 V), the phototransistor array can learn, forget, and relearn the input patterns, unsupervised.* ***h)*** *Learning under noisy conditions with and without depression. Top-row shows an example of noisy input images of the letter "T". Bottom-rows show the corresponding heatmaps of G with V_D = 0 V and V_D = 6 V. This shows the importance of forgetting in learning under noisy illuminations.*

(V_{D}) for a total duration of 2.5 s. Note that since the electrical programming *via* back-gate results in positive shift in the V_{TH}, G decreases with increasing N_D and V_D and hence can be exploited as synaptic depression or forgetting. However, unlike biological forgetting where humans have limited control, forgetting rate, which we define as the rate of decrease in G in our vision platform can be precisely controlled through V_D (Fig. 4c). Forgetting is also permanent as shown in Fig. 4d using non-volatile retention of post forgotten G measured at V_{BG} = 0 V. Fig. 4e shows the forgetting energy expenditure for different V_D, calculated as $^1/_2\, C_G {V_{\mathrm{D}}}^2$, where C_G = 8 fF is the gate capacitance and is in the range of hundreds of femto Joules. The conductance (G) heatmaps shown in Fig. 4f and ***Supplementary Video 3*** illustrate the forgetting of a learned "T" pattern at different rates enabled by V_{D}. Clearly at a higher magnitude of V_{D}, e.g. V_{D} = 12 V, FET array forgets the letter "T" almost immediately (~ 10 epochs), whereas, lower magnitudes of V_{D} enable gradual forgetting (> 50 epochs). The energy expenditure for immediate and gradual forgetting were found to be ~ 0.75 pJ/pixel and 2.5 pJ/pixel, respectively.

Importance of forgetting in unsupervised relearning

Next, we elucidate the role of forgetting in unsupervised relearning. In this instance each epoch consists of two cycles, optical potentiation for learning followed by electrical depression for forgetting. We have considered 3×3 pixel images of the letters "L" and "T" for learning and relearning, respectively, with each being presented for 25 epochs. As before, all devices were initially programmed to a low conductance state with $G \approx 1$ nS. Fig. 4g and ***Supplementary Video 4*** show the time evolution of heatmap of G during learning and unsupervised relearning for various strengths of electrical depression achieved by using different V_{D}. Some of the key observations include: 1) If the strength of electrical depression is too weak (e.g. V_{D} = 9.25 V) compared to the

optical potentiation, devices will never forget the learned pattern, and relearning becomes futile. 2) If the strength of electrical depression is too strong (e.g. V_D = 10 V) compared to the optical potentiation then it becomes difficult to learn any input pattern. 3) Under optimum potentiation and depression (e.g. V_D = 9.75 V), the phototransistor array can learn, forget, and relearn the input patterns, unsupervised. These demonstrations illustrate the critical role of forgetting in unsupervised learning.

Forgetting for learning under random disturbances

Next, we show that forgetting plays an even more significant role when learning under noisy illumination conditions. The human visual system possesses the remarkable ability to identify important features in an image even in the presence of disturbances. For example, the brain can extract information in poor weather conditions such as mist, storm, blizzard, and other impediments to perfect vision. Avoidance of dynamic noise, on the other hand is a challenging task for machine vision and it must rely on sophisticated computer algorithms for the elimination of the same. Hardware implementations of such algorithms are naturally energy hungry. However, as we demonstrate below, forgetting can help learning under such dynamic noise.

Fig. 4g and ***Supplementary Video 5*** show the time evolution of the heatmap of G measured at V_{BG} = 0 V while learning from 3×3 pixel noisy images of the letter "T" with and without electrical depression or forgetting for 50 epochs. In the absence of forgetting, the phototransistors which are expected to remain in the low conductance state also get randomly potentiated due to noise making it impossible to learn the pattern as seen at the end of the 50 epochs. In contrast, when forgetting is enabled, the random and occasional potentiation in the unintended phototransistors are

compensated through regular electrical depression allowing the phototransistor array to learn the letter "T" under noisy illumination. As expected, learning takes longer under noisy condition since forgetting reduces the learning rate. Nevertheless, this demonstration highlights the importance of forgetting in perceiving information obscured by noise and offers a new route towards the development of in-memory vision sensors that can seamlessly operate under environmental disturbances.

Conclusion

In summary, we have experimentally demonstrated a gamut of possibilities offered by a 2D based in-memory optoelectronic platform for next-generation machine intelligent systems. To the best of our knowledge, this is the first demonstration which exploits analog and non-volatile optical and electrical memory to integrate adaptive machine vision with unsupervised machine learning. Direct analog learning from visual stimuli was experimentally demonstrated followed by learning under different illumination conditions elucidating the advantages of adaptive machine vision. Finally, the role of forgetting in unsupervised relearning and in learning under noisy conditions are shown as two examples, both of which are challenging to achieve in a neural network which cannot forget. Remarkably, the energy expenditure by our hardware vision platform was found to be in the ranges of tens of nano Joules for different learning and forgetting examples. Our findings highlight the benefits of in-memory computing and sensing for hardware acceleration of low-power and biomimetic machine intelligence.

Methods

Film growth: Monolayer MoS_2 was obtained from 2D crystal consortium (2DCC) [54]. It 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 [65]. Molybdenum hexacarbonyl ($Mo(CO)_6$) and hydrogen sulfide (H_2S) 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_2S 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_2S 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[66].

Back-gate stack fabrication: Direct replacement of thermally oxidized SiO_2 with a high-κ

dielectric such as Al_2O_3 grown *via* atomic layer deposition (ALD) is a logical choice to scale the effective oxide thickness (EOT). However, we found that Al_2O_3/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 [67]. 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 [68]. 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 Al_2O_3 surface with a rms roughness of 0.7 nm.

Fabrication of monolayer MoS_2 FET: We have fabricated the back-gated field effect transistors on a 50nm alumina (Al_2O_3) acting as a gate oxide and a stack of Pt/TiN/p^{++} Si as a back-gate electrode. First, MOCVD grown MoS_2 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 (SF_6) 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.

AUTHOR INFORMATION

Corresponding Author

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

Author Contributions

A.D, D. J, and S.D conceived the idea and designed the experiments. A.D, D. J, and S.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 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.

References

[1] J. P. Frisby and J. V. Stone, *Seeing: The Computational Approach to Biological Vision*: MIT Press, 2010.

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche*, et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature,* vol. 529, pp. 484-489, 2016/01/01 2016.

[3] M. Lechner, R. Hasani, A. Amini, T. A. Henzinger, D. Rus, and R. Grosu, "Neural circuit policies enabling auditable autonomy," *Nature Machine Intelligence,* vol. 2, pp. 642-652, 2020/10/01 2020.

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

[5] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing,* vol. 234, pp. 11-26, 2017.

[6] D. Mishkin, N. Sergievskiy, and J. Matas, "Systematic evaluation of convolution neural network advances on the imagenet," *Computer Vision and Image Understanding,* vol. 161, pp. 11-19, 2017.

[7] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural Networks,* vol. 111, pp. 47-63, 2019.

[8] 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.

[9] 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.

[10] 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.

[11] S. Salahuddin, K. Ni, and S. Datta, "The era of hyper-scaling in electronics," *Nature Electronics,* vol. 1, pp. 442-450, 2018.

[12] D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices," *Nature Electronics,* vol. 1, pp. 333-343, 2018.

[13] Y. Chai, "In-sensor computing for machine vision," ed: Nature Publishing Group, 2020.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang*, et al.*, "Fully hardware-implemented memristor convolutional neural network," *Nature,* vol. 577, pp. 641-646, 2020.

[22] S. Tanachutiwat, M. Liu, and W. Wang, "FPGA Based on Integration of CMOS and RRAM," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 19, pp. 2023-2032, 2010.

[23] F. Zhou and Y. Chai, "Near-sensor and in-sensor computing," *Nature Electronics,* vol. 3, pp. 664-671, 2020.

[24] K. Kyuma, E. Lange, J. Ohta, A. Hermanns, B. Banish, and M. Oita, "Artificial retinas—fast, versatile image processors," *Nature,* vol. 372, pp. 197-198, 1994.

[25] H. C. Ko, M. P. Stoykovich, J. Song, V. Malyarchuk, W. M. Choi, C.-J. Yu*, et al.*, "A hemispherical electronic eye camera based on compressible silicon optoelectronics," *Nature,* vol. 454, pp. 748-753, 2008/08/01 2008.

[26] M. Ishikawa, K. Ogawa, T. Komuro, and I. Ishii, "A CMOS vision chip with SIMD processing element array for 1 ms image processing," in *1999 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC. First Edition (Cat. No. 99CH36278)*, 1999, pp. 206-207.

[27] A. Rodriguez-Vazquez, J. Fernandez-Berni, J. A. Lenero-Bardallo, I. Vornicu, and R. Carmona-Galan, "CMOS vision sensors: embedding computer vision at imaging front-ends," *IEEE Circuits and Systems Magazine,* vol. 18, pp. 90-107, 2018.

[28] L. Gu, S. Poddar, Y. Lin, Z. Long, D. Zhang, Q. Zhang*, et al.*, "A biomimetic eye with a hemispherical perovskite nanowire array retina," *Nature,* vol. 581, pp. 278-282, 2020.

[29] F. Zhou, Z. Zhou, J. Chen, T. H. Choy, J. Wang, N. Zhang*, et al.*, "Optoelectronic resistive random access memory for neuromorphic vision sensors," *Nature nanotechnology,* vol. 14, pp. 776-782, 2019.

[30] S. Chen, Z. Lou, D. Chen, and G. Shen, "An Artificial Flexible Visual Memory System Based on an UV-Motivated Memristor," *Adv Mater,* vol. 30, Feb 2018.

[31] S. M. Kwon, S. W. Cho, M. Kim, J. S. Heo, Y. H. Kim, and S. K. Park, "Environment-Adaptable Artificial Visual Perception Behaviors Using a Light-Adjustable Optoelectronic Neuromorphic Device Array," *Adv Mater,* vol. 31, p. e1906433, Dec 2019.

[32] F. Zhou, Z. Zhou, J. Chen, T. H. Choy, J. Wang, N. Zhang*, et al.*, "Optoelectronic resistive random access memory for neuromorphic vision sensors," *Nat Nanotechnol,* vol. 14, pp. 776-782, Aug 2019.

[33] S. Dai, X. Wu, D. Liu, Y. Chu, K. Wang, B. Yang*, et al.*, "Light-Stimulated Synaptic Devices Utilizing Interfacial Effect of Organic Field-Effect Transistors," *ACS Appl Mater Interfaces,* vol. 10, pp. 21472-21480, Jun 27 2018.

[34] L. Yin, W. Huang, R. Xiao, W. Peng, Y. Zhu, Y. Zhang*, et al.*, "Optically Stimulated Synaptic Devices Based on the Hybrid Structure of Silicon Nanomembrane and Perovskite," *Nano Lett,* vol. 20, pp. 3378-3387, May 13 2020.

[35] 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, Mar 2020.

[36] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, "Electronics and optoelectronics of two-dimensional transition metal dichalcogenides," *Nature nanotechnology,* vol. 7, p. 699, 2012.

[37] S. Das, J. A. Robinson, M. Dubey, H. Terrones, and M. Terrones, "Beyond Graphene: Progress in Novel Two-Dimensional Materials and van der Waals Solids," *Annual Review of Materials Research, Vol 45,* vol. 45, pp. 1-27, 2015.

[38] Q. Smets, G. Arutchelvan, J. Jussot, D. Verreck, I. Asselberghs, A. N. Mehta*, et al.*, "Ultra-scaled MOCVD MoS 2 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.

[39] G. Migliato Marega, Y. Zhao, A. Avsar, Z. Wang, M. Tripathi, A. Radenovic*, et al.*, "Logic-in-memory based on an atomically thin semiconductor," *Nature,* vol. 587, pp. 72-77, 2020/11/01 2020.

[40] 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,* vol. 3, pp. 646-655, 2020/10/01 2020.

[41] H. Jang, C. Liu, H. Hinton, M. H. Lee, H. Kim, M. Seol*, et al.*, "An Atomically Thin Optoelectronic Machine Vision Processor," *Adv Mater,* vol. 32, p. e2002431, Sep 2020.

[42] S. Seo, S. H. Jo, S. Kim, J. Shim, S. Oh, J. H. Kim*, et al.*, "Artificial optic-neural synapse for colored and color-mixed pattern recognition," *Nat Commun,* vol. 9, p. 5106, Nov 30 2018.

[43] C. Choi, J. Leem, M. S. Kim, A. Taqieddin, C. Cho, K. W. Cho*, et al.*, "Curved neuromorphic image sensor array using a MoS2-organic heterostructure inspired by the human visual recognition system," *Nat Commun,* vol. 11, p. 5934, Nov 23 2020.

[44] X. Hou, C. Liu, Y. Ding, L. Liu, S. Wang, and P. Zhou, "A Logic-Memory Transistor with the Integration of Visible Information Sensing-Memory-Processing," *Advanced Science,* vol. 7, p. 2002072, 2020.

[45] L. Gravitz, "The importance of forgetting," *Nature,* vol. 571, pp. S12-S14, 2019.

[46] R. Shapley and C. Enroth-Cugell, "Visual adaptation and retinal gain controls," *Progress in retinal research,* vol. 3, pp. 263-346, 1984.

[47] L. F. Abbott and S. B. Nelson, "Synaptic plasticity: taming the beast," *Nature neuroscience,* vol. 3, pp. 1178-1183, 2000.

[48] J. D. Shepherd and R. L. Huganir, "The cell biology of synaptic plasticity: AMPA receptor trafficking," *Annu. Rev. Cell Dev. Biol.,* vol. 23, pp. 613-643, 2007.

[49] C. Zhuang, S. Yan, A. Nayebi, M. Schrimpf, M. C. Frank, J. J. DiCarlo*, et al.*, "Unsupervised neural network models of the ventral visual stream," *Proceedings of the National Academy of Sciences,* vol. 118, 2021.

[50] A. Sebastian, R. Pendurthi, T. H. Choudhury, J. M. Redwing, and S. Das, "Benchmarking monolayer MoS2 and WS2 field-effect transistors," *Nature Communications,* vol. 12, p. 693, 2021/01/29 2021.

[51] 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.

[52] C. D. English, K. K. H. Smithe, R. L. Xu, and E. Pop, "Approaching ballistic transport in monolayer MoS<inf>2</inf> transistors with self-aligned 10 nm top gates," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 5.6.1-5.6.4.

[53] 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.

[54] 2DCC. *2d-crystal-consortium*. Available: https://www.mri.psu.edu/2d-crystal-consortium/user-facilities/thin-films/list-thin-film-samples-available

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] F. Torrisi and J. N. Coleman, "Electrifying inks with 2D materials," *Nature nanotechnology,* vol. 9, pp. 738-739, 2014.

[61] S. Das, A. Dodda, and S. Das, "A biomimetic 2D transistor for audiomorphic computing," *Nature Communications,* vol. 10, p. 3450, 2019/08/01 2019.

[62] A. Sebastian, A. Pannone, S. S. Radhakrishnan, and S. Das, "Gaussian synapses for probabilistic neural networks," *Nature communications,* vol. 10, pp. 1-11, 2019.

[63] 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.

[64] C.-Y. Wang, S.-J. Liang, S. Wang, P. Wang, Z. a. Li, Z. Wang*, et al.*, "Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor," *Science Advances,* vol. 6, p. eaba6173, 2020.

[65] 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.

[66] 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.

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

[68] 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.

Figure Captions

Figure 1. Bio-inspired 2D vision system. Elementary components of biological visual neural network, **a)** eyes enabling biological vision and **b)** visual cortex in the brain enabling biological learning. **c)** Photoreceptors in the eyes enable phototransduction and adaptation. Rods facilitate scotopic and cones enable photopic vision. **d)** Synapses strengthen or weaken to learn or forget e.g., by controlling the number of AMPA receptors in the postsynaptic neuron when the presynaptic neuron releases glutamate neurotransmitter. **e)** Schematic and **f)** false color microscope image of the artificial vision system comprising of 9×1 2D phototransistor arrays integrated with a programmable back-gate stack. The platform enables functionalities like phototransduction, visual adaptation, synaptic plasticity, direct learning, unsupervised relearning, and learning under dynamic noise exploiting forgetting. **g)** 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}) in dark, **h)** phototransduction under different levels of illumination from a blue light emitting diode (LED), **i)** optical potentiation induced learning or increase in device conductance (G), and **j)** electrical depression induced forgetting or decrease in G measured at V_{BG} = 0 V in a representative 2D phototransistor.

Figure 2. Analog vision and learning. Post illumination transfer characteristics of a representative phototransistor at V_{DS} = 1 V measured in dark after 10 s exposure to different LED brightness levels (I_{LED}) with **a)** V_{write} = 2 V and **b)** V_{write} = -3 V applied to the back-gate during the illumination. **c)** Evolution of post-illumination conductance (G) of the phototransistor measured at V_{BG} = 0 V as a function of the LED illumination time (t_{LED}) for different I_{LED}. The monotonic increase in G facilitate direct learning from visual stimuli. **d)** Retention of the learned

conductance states after the removal of optical stimuli. **e)** Heatmaps of the input image with each pixel corresponds to an I_{LED} value and the output images with each pixel showing the G of corresponding phototransistor in the 9×1 array, measured at V_{BG} = 0 V at different epochs. The input image is learned in 50 epochs. This shows the analog vision and learning capabilities of the photoresponsive 2D array structure.

Figure 3. Learning with adaptive machine vision. **a)** Time evolution of G measured at V_{BG} = 0 V, every 500 ms, **b)** Learning rate, which we define as the rate of increase in G and **c)** average learning energy expenditure per epoch as a function of I_{LED} and V_{write}. **d)** Learning under photopic illumination (I_{LED} = 20 mA). Heatmaps of the input image (top row) and the output images showing the heatmap of G measured at V_{BG} = 0 V at different epochs. The photopic "T" is learned within 10 epochs for illumination at V_{write} = -2 V. **e)** Learning under scotopic illumination (I_{LED} = 2 mA). The same experiment as in (d) is repeated for V_{write} = -2, -3, -4, and -5 V. While the photoresponsive array does not learn scotopic "T" in 50 epochs for V_{write} = -2 V, V_{write} = -5 V enables learning in just 10 epochs. These experiments highlight the importance of adaptive machine vision in learning.

Figure 4. Forgetting for unsupervised relearning and learning under random disturbances. **a)** Transfer characteristics of a representative phototransistor when positive programming voltages (V_D) of increasing amplitude are applied to the back-gate, each for a total duration of 100 ms. **b)** Device conductance (G) measured at V_{BG} = 0 V as a function of number of programming pulses (N_D) of pulse width t_D = 100 ms of different amplitudes (V_D). **c)** Forgetting rate, defined as the rate of decrease in G for different V_D. **d)** Non-volatile retention of the post forgotten G measured

at V_{BG} = 0 V. **e)** Forgetting energy expenditure for different V_D. **f)** Heatmaps of G show that smaller amplitudes of V_D, e.g., V_D= 10 V (top-row) enable gradual forgetting of the learnt letter 'T' while higher amplitudes of V_D, e.g., V_D = 12 V (bottom-row) causes immediate forgetting. **g)** Unsupervised relearning enabled by adaptive forgetting. Heatmaps show learning a letter 'L' (first 25 epochs) and relearning of another letter 'T' (final 25 epochs) by the phototransistor array with different forgetting rates. Here, each epoch consists of optical potentiation and electrical depression (application of V_D) enabling relearning and eliminating the need for supervision. The top-row shows the input images (I_{LED}). Each row below shows the corresponding heatmaps of G for different amplitudes of V_D. Under optimum potentiation and depression (e.g., V_D = 9.75 V), the phototransistor array can learn, forget, and relearn the input patterns, unsupervised. **h)** Learning under noisy conditions with and without depression. Top-row shows an example of noisy input images of the letter "T". Bottom-rows show the corresponding heatmaps of G with V_D = 0 V and V_D = 6 V. This shows the importance of forgetting in learning under noisy illuminations.

Figures

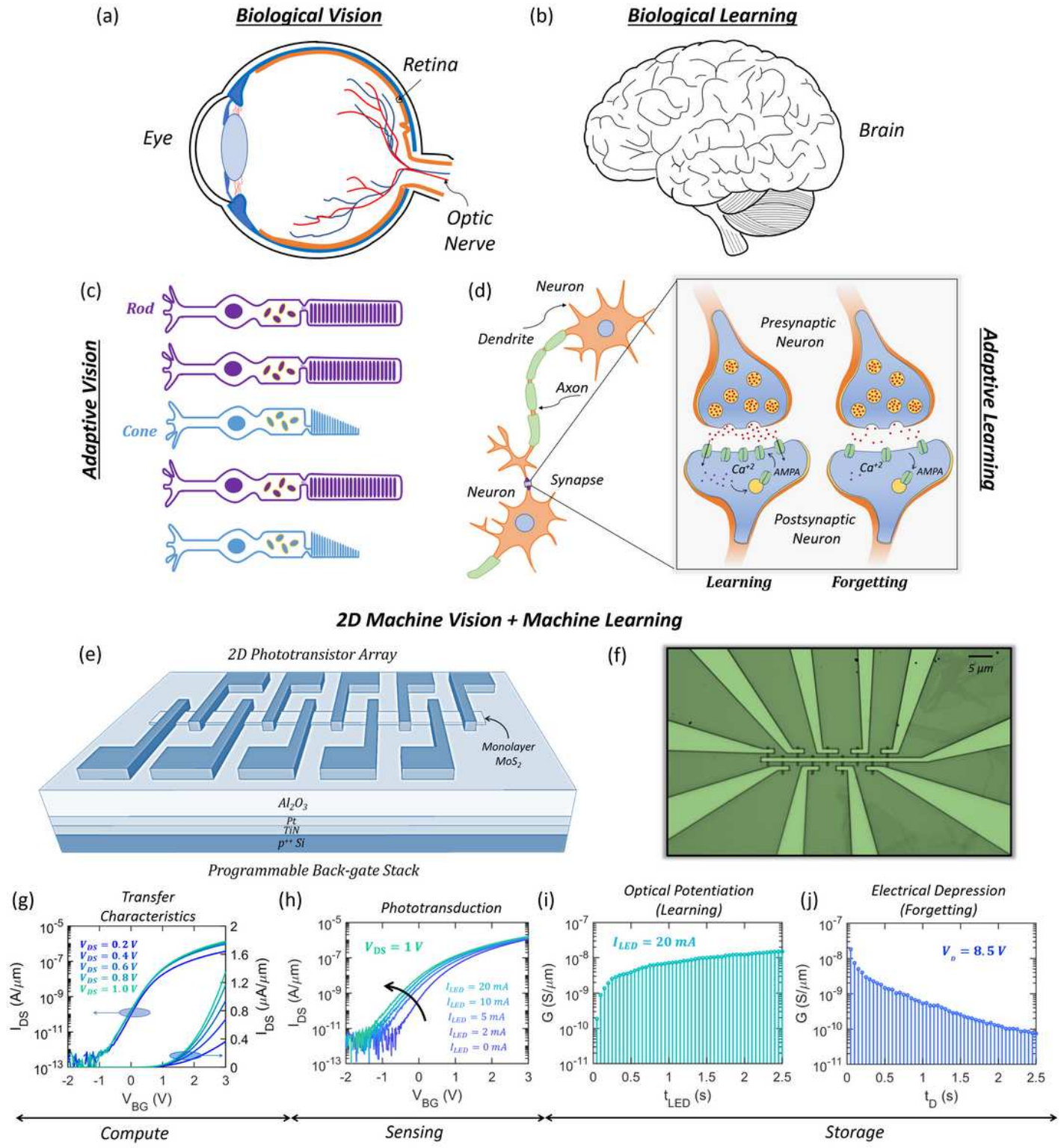

Figure 1

Bio-inspired 2D vision system. Please see manuscript .pdf for figure caption

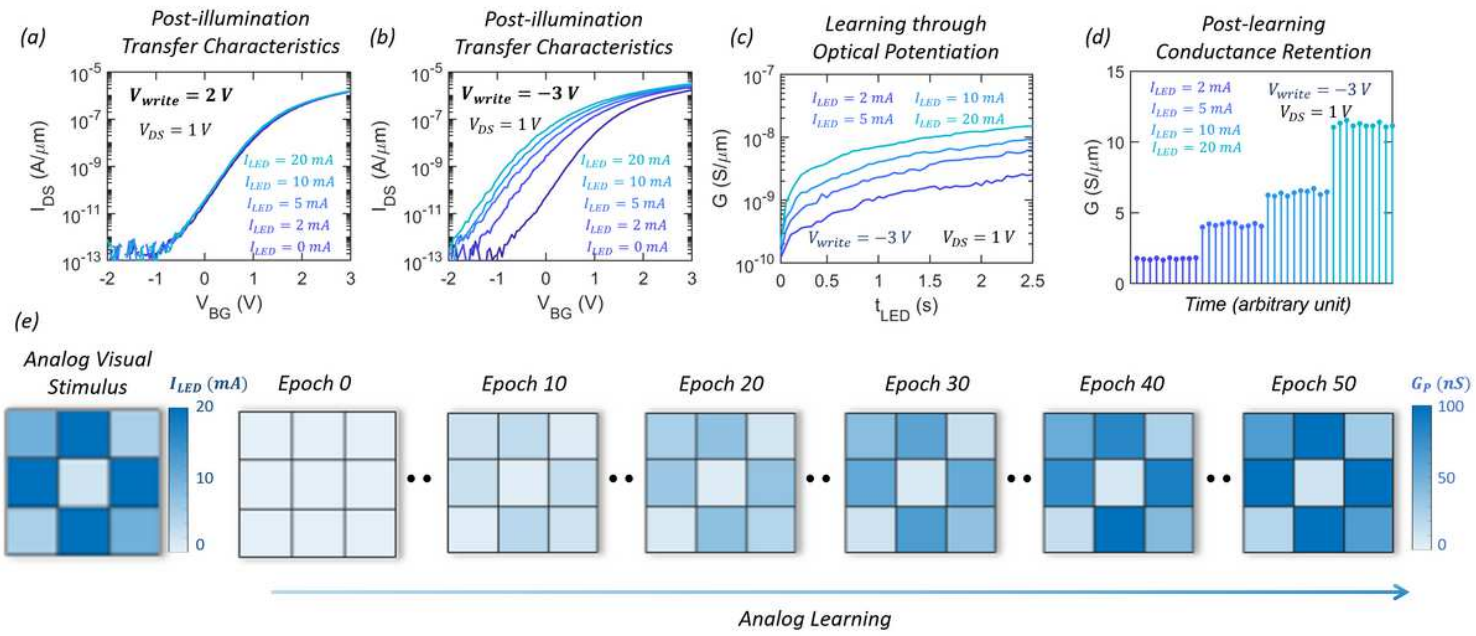

Figure 2

Analog vision and learning. Please see manuscript .pdf for figure caption

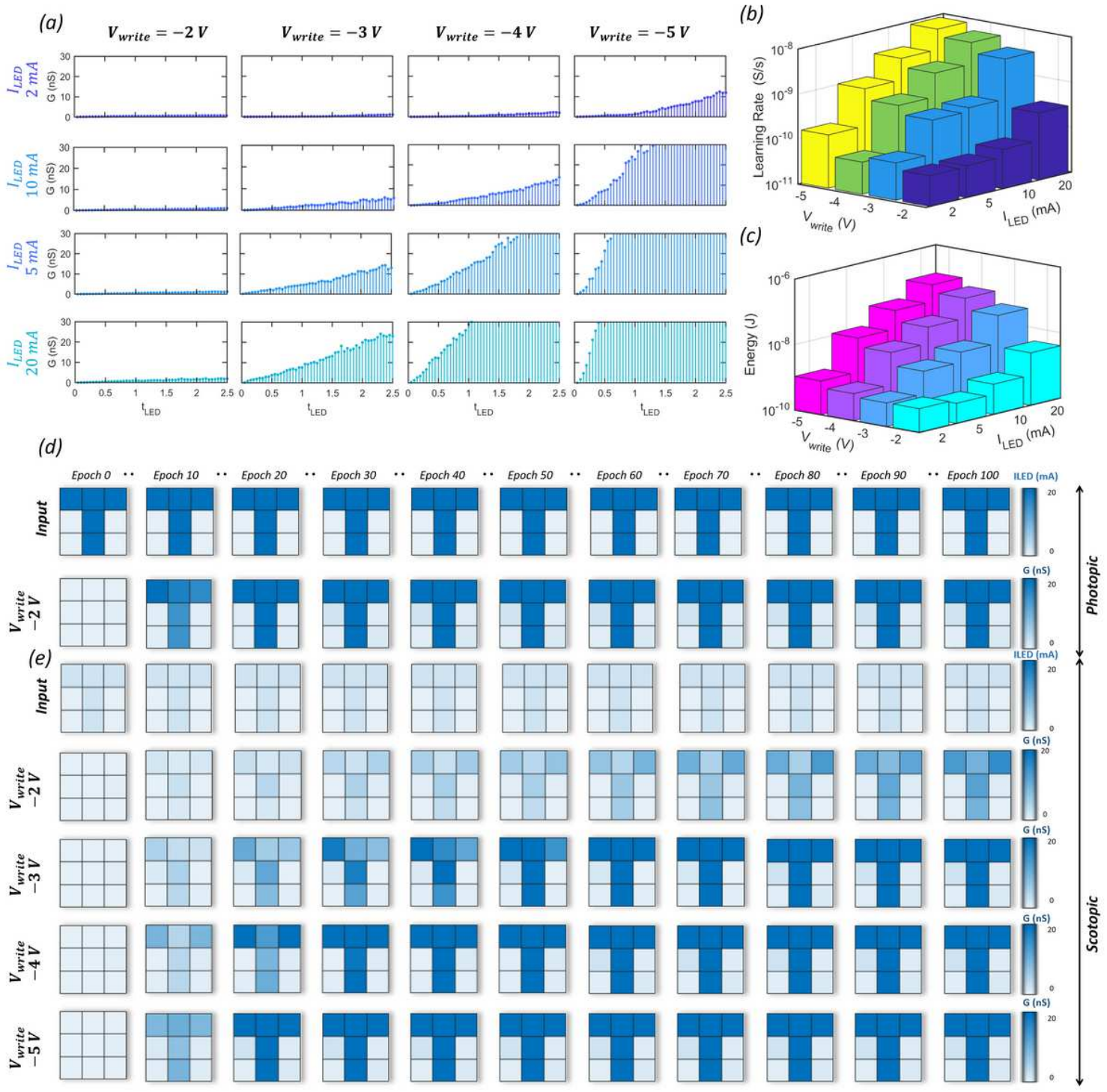

Figure 3

Learning with adaptive machine vision. Please see manuscript .pdf for figure caption

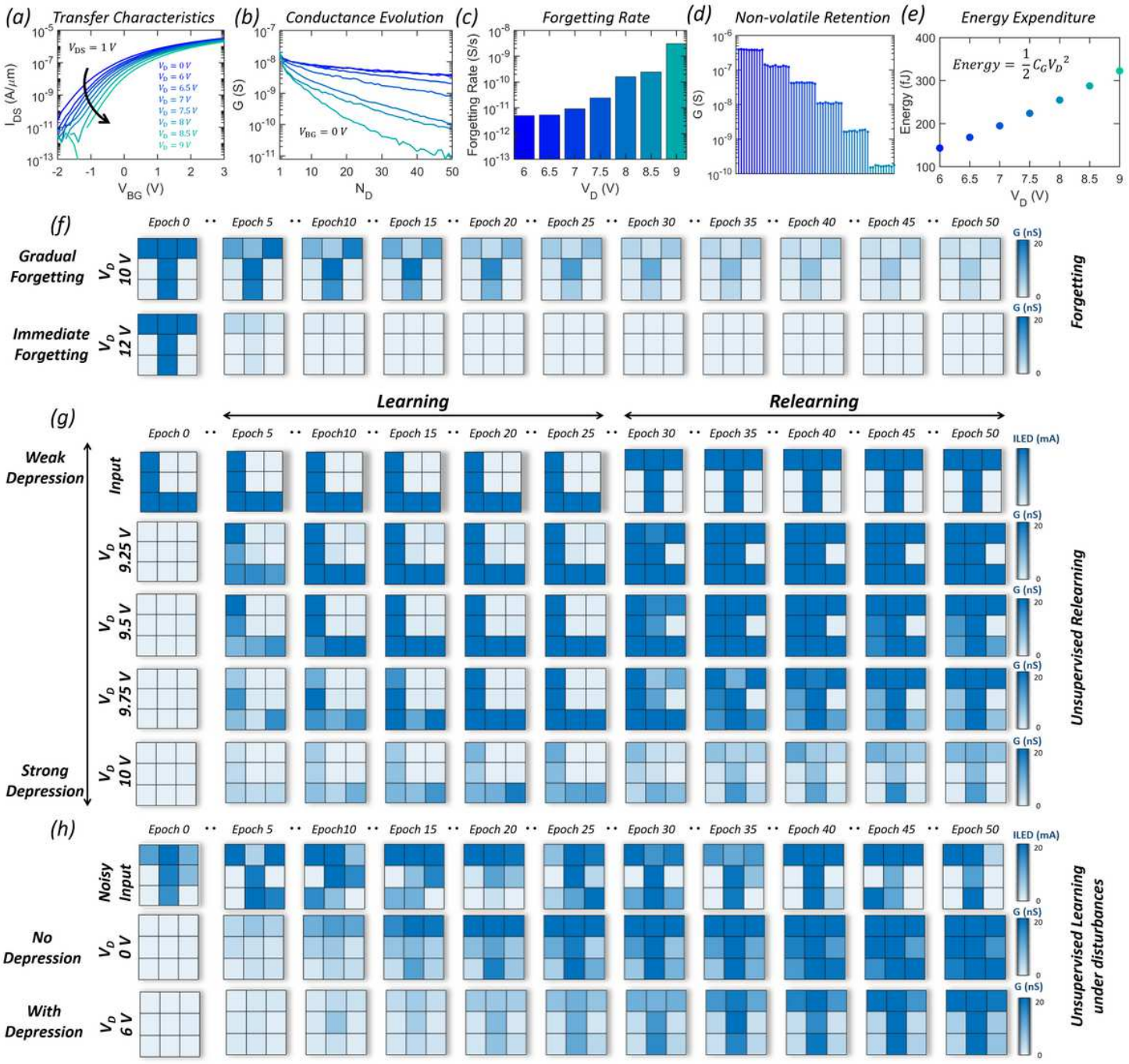

Figure 4

Forgetting for unsupervised relearning and learning under random disturbances. Please see manuscript .pdf for figure caption

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- SupplementaryVideo1.mp4

- SupplementaryVideo2.mp4
- SupplementaryVideo3.mp4
- SupplementaryVideo4.mp4
- SupplementaryVideo5.mp4