ADVANCED MATERIALS
www.advmat.de

Scalable Synaptic Transistor Memory from Solution-Processed Carbon Nanotubes for High-Speed Neuromorphic Data Processing

*Jingfang Pei, Lekai Song, Pengyu Liu, Songwei Liu, Zihan Liang, Yingyi Wen, Yang Liu, Shengbo Wang, Xiaolong Chen, Teng Ma, Shuo Gao, and Guohua Hu**

Neural networks as a core information processing technology in machine learning and artificial intelligence demand substantial computational resources to deal with the extensive multiply-accumulate operations. Neuromorphic computing is an emergent solution to address this problem, allowing the computation performed in memory arrays in parallel with high efficiencies conforming to the neural networks. Here, scalable synaptic transistor memories are developed from solution-sorted carbon nanotubes. The transistors exhibit a large switching ratio of over 10^5, a significant memory window of ≈12 V arising from charge trapping, and low response delays down to tens of nanoseconds. These device characteristics endow highly stabilized reconfigurable conductance states, successful emulation of synaptic functions, and a high data processing speed. Importantly, the devices exhibit uniform characteristic metrics, e.g., with a 1.8% variation in the memory window, suggesting an industrial-scale manufacturing capability of the fabrication. Using the memories, a hardware convolution kernel is designed and parallel image processing is demonstrated at a speed of 1 M bit per second per input channel. Given the efficacy of the convolution kernel, a promising prospect of the memories in implementing neuromorphic computing is envisaged. To explore the potential, large-scale convolution kernels are simulated and high-speed video processing is realized for autonomous driving.

1. Introduction

Neural networks, computational frameworks with nonlinear mapping, parallel data processing, and adaptive learning capabilities are a core information processing technology in machine learning and artificial intelligence.[1,2] However, their implementation can demand substantial computational resources to deal with the extensive multiply-accumulate operations for reliable results. This poses challenges for the computing hardware, particularly, the von Neumann computers where the data has to be moved constantly between the processing and memory units for computation—the slow access to the memory limits the computation speed, while the constant data movement consumes energy.[3] Neuromorphic computing emerges as a promising solution to address this issue. In this architecture, using emerging nonvolatile memory arrays, the data can be processed in high parallelism to perform the matrix multiply-accumulate operations following the Ohm's and Kirchhoff's laws, conforming to the neural networks.[4] Recent advances show

J. Pei, L. Song, P. Liu, S. Liu, Y. Wen, Y. Liu, G. Hu
Department of Electronic Engineering
The Chinese University of Hong Kong
Shatin, N. T., Hong Kong SAR 999077, China
E-mail: ghhu@ee.cuhk.edu.hk

Z. Liang, X. Chen
Department of Electrical and Electronic Engineering
Southern University of Science and Technology
Shenzhen 518055, China

Y. Liu
Shun Hing Institute of Advanced Engineering
The Chinese University of Hong Kong
Shatin, N. T., Hong Kong SAR 999077, China

S. Wang, S. Gao
School of Instrumentation and Optoelectronic Engineering
Beihang University
Beijing 100191, China

T. Ma
Department of Applied Physics
Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong SAR 999077, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adma.202312783

DOI: 10.1002/adma.202312783

that synaptic transistor memories with reconfigurable conductance states and synaptic functions are a promising memory technology for neuromorphic computing.[5–9] Particularly, with a transistor configuration, the memories can allow data programming and read operations with a higher precision, greater freedom, and lesser crosstalk in the memory arrays.[10]

Synaptic transistor memories can be typically realized with electrochemical,[6,11] ferroelectric,[7,9] and floating-gate mechanisms.[5] Among them, the electrochemical transistor memories with low voltage operation and linear weight (i.e., conductance state) updating capabilities draw considerable interest. However, the need to introduce ions, typically liquid electrolytes, into the gate dielectric is problematic for integration with the complementary metal-oxide-semiconductor (CMOS) manufacturing processes. The ferroelectric and floating-gate transistor memories, on the other hand, are CMOS-compatible and allow fast switching and stable weight updating. However, the memory in these transistors is achieved using multilayer structures, which often require complex fabrication processes that can incur high production cost. Different from these synaptic transistor memories, devices based on nanomaterials and a charge trapping mechanism appear to be promising.[12] For instance, carbon nanotubes with high carrier concentrations, high carrier mobilities, and the largest specific surface area are reported to exhibit active surface chemistry and this can be exploited to develop charge trapping electronic devices with prominently regulatable conductance states.[13,14] Indeed, carbon nanotubes have been recently widely reported in charge trapping synaptic transistor memory fabrication.[15–19] However, the memory mechanism may require a detailed examination and, at the current development stage, the fabrication needs optimization for scalable high-performance devices, particularly towards their implementation of neuromorphic hardware. The major obstacle from lab to fab is at the engineering level, including the sorting of carbon nanotubes, the development of large-scale films, the scalable device fabrication, and the compatibility of the fabrication to the current semiconductor processes.[20] The implementation of neuromorphic hardware also requires careful engineering considerations for designing multiply-accumulate computing arrays.[6,21]

Herein, in this work, we develop scalable synaptic transistor memories from solution-sorted semiconducting single-walled carbon nanotubes (s-SWCNTs). The memories exhibit a large switching ratio of over 10^5, a significant memory effect with a memory window of ≈12 V, and low response delays down to tens of nanoseconds. Our analysis reveals that the memory arises from the charge trapping in the sorting polymer. These above device characteristics endow highly-stabilized reconfigurable conductance states, successful emulation of biological synaptic functions, and a high data processing speed. Leveraging solution-processing and photolithographic patterning, the fabrication is scalable and demonstrates uniform memory characteristic metrics, e.g., with a 1.8% variation in the memory window. Capitalizing on the device characteristics and the fabrication scalability, we design memory arrays to function as a hardware convolution kernel, and perform parallel convolution image processing at a speed of 1 M bit per second per input channel. Given the efficacy of the hardware convolution kernel, we simulate large-scale kernels for high-speed edge detection and noise reduction of videos to explore the potential of our memories in practical neuromorphic computing applications.

2. Results

2.1. Scalable s-SWCNTs Synaptic Transistor Memories

We start the synaptic transistor memory fabrication with sorting of the s-SWCNTs from the raw arc-discharged carbon nanotubes (see Experimental Section) (**Figure 1a–c**). The raw carbon nanotubes mix semiconducting and metallic phases. Briefly, we disperse the raw carbon nanotubes via tip sonication, and use poly[N-(1-octylnonyl)-9-hcarbazole-2,7-diyl] (PCz) to sort the s-SWCNTs. The sorting polymer has been proven to interact efficiently and selectively with the s-SWCNTs, facilitating their dispersion in solvents and leaving the metallic ones undispersed.[20,22] We evaluate the s-SWCNT sorting using UV–Vis–NIR absorption spectrum. As shown in Figure 1d, the spectrum exhibits sharp absorption peaks at 400-600 and 800–1200 nm. The peaks correspond to the characteristic peaks of s-SWCNTs, proving a successful sorting of the s-SWCNTs. We deposit the solution-sorted s-SWCNTs using a dip coating process. As the substrate is vertically lifted from the s-SWCNT solution at a well-controlled speed, the solvent ($CHCl_3$) evaporates rapidly, giving a uniform deposition of the s-SWCNTs on the substrate. Figure 1a shows the deposition of a s-SWCNT thin-film on a 4 in. silicon wafer. See Video S1 (Supporting Information) for the dip coating deposition process. Scanning electron microscopic (Figure 1a) and atomic force microscopic (Figure S1, Supporting Information) images prove the uniformity of the s-SWCNT thin-film, and allow estimation of the thin-film (with a thickness of <10 nm) density as approximately 20 s-SWCNTs μm^{-1}.

We then fabricate an array of 20 × 20 synaptic transistor memories using the s-SWCNT thin-film via CMOS process compatible standard photolithographic patterning. The transistor adopts a bottom-gate bottom-contact configuration, where the s-SWCNTs are employed as the channel (Figure 1b). Figure 1c shows an optical microscope image of the memory array. Due to the <10 nm thickness, the s-SWCNTs thin-film appears transparent and is barely visible on the wafer. As demonstrated in Figure 1e,f, the transistors exhibit typical p-type transfer characteristics with a switching ON/OFF ratio of $>10^5$, a large on-state current of 30 μA at a drain voltage of 1 V, and, importantly, a substantial memory effect with a large memory window. Particularly, due to the high carrier mobilities of carbon nanotubes, the device exhibits a high switching speed, with the switching times and response delays down to tens of nanoseconds at both pulsed gate voltage signals and pulsed drain–source voltage signals, as shown in Figure S2 (Supporting Information). To quantify the memory effect, we define the difference between the gate voltages that give a source-drain current of 1 μA as the memory window ΔV, and this gives a ΔV of ≈12 V at a gate voltage sweep range of -9 to 9 V. Note that a larger memory window can be achieved with a larger operating gate voltage (see Figure S3, Supporting Information), but an unreasonably large operating voltage can make it challenging to match the supply voltage, and can increase the energy consumption. Meanwhile, a large operating voltage can be detrimental for the operation of the devices and even lead to device breakdown. Here, we define the difference between the

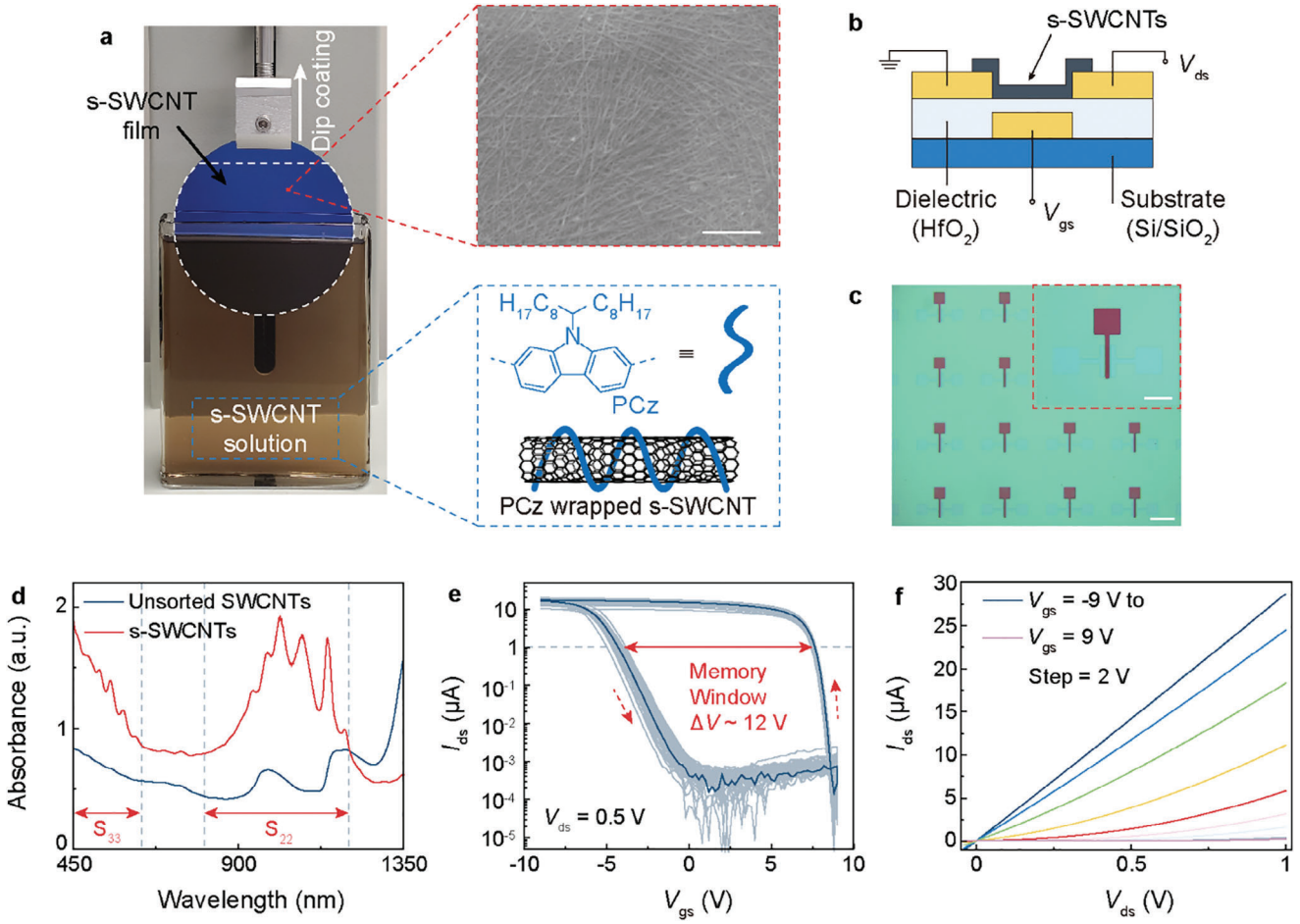

Figure 1. s-SWCNT synaptic transistor memory fabrication. a) Dip coating deposition of s-SWCNT film on a 4 in. wafer, with the subfigures showing scanning electron microscopic image of the film and the dispersion of s-SWCNTs using sorting polymer PCz. b) Schematic device configuration. c) Optical microscopic image of a device array, with the inset showing the zoomed-in microscopic image of one device. d) UV–Vis–NIR absorption spectra of the CNT solutions, with the S_{33} and S_{22} peaks proving successful sorting of the s-SWCNTs. e) The transfer characteristics from 50 randomly selected devices, showing a uniform memory window of ≈12 V. f) The output characteristic of a typical device. Scale bars: a) 500 nm, c) 200 μm for the whole image and 100 μm for the zoomed-in image, respectively.

switching-state currents at a gate voltage of 0 V as the ΔG margin, and this gives a ΔG margin switching ratio exceeding 10^5. This memory effect suggests the capability of our synaptic transistor memories for emulating biological synapses in neuromorphic computing. Notably, the characterization of randomly selected 50 devices in this device array exhibits a good uniformity in the device performance characteristic metrics, for example, the memory window shows a variation of 1.8%, as shown in Figure 1e. This demonstrates the capability of our fabrication approach for industrial-scale manufacturing, where a variation of <30% is acceptable.[23]

2.2. Origin of the Memory

As demonstrated above, our transistors demonstrate a substantial memory window. Preceding studies on s-SWCNT based synaptic transistor memories suggest the memory effect can arise from three major charge-trapping mechanisms, including: i) oxygen doping—with the largest specific surface area, s-SWCNTs can conveniently absorb the ambient oxygen and moisture and as such, the carriers in s-SWCNTs can easily get entrapped within the absorbed oxygen;[24,25] ii) charge trapping in the gate oxides—driven by the gate voltage, the carriers in s-SWCNTs can get injected into and entrapped by the defects present at the surface of the gate oxides;[17,26] and iii) charge trapping in the sorting polymers— the sorting polymers with long alkyl chains oriented outward can tightly wrap s-SWCNTs via $\pi-\pi$ interactions and as a result, the sorting polymers may serve as semiconducting acceptors to trap the holes from s-SWCNTs.[27,28] Here we plot the possible charge trapping locations in these mechanisms in **Figure 2a**, along with the corresponding energy band diagram in Figure 2b. Both the O-vacancy in the HfO_2 dielectric layer and the O_2/H_2O attachment to the s-SWCNTs can induce defect energy levels in the energy band of the s-SWCNTs-based metal-oxide-semiconductor (MOS) structure. These defect energy levels can capture carriers and act as charge traps that can lead to memristive effects in the transistors. Besides, the bonding polymer wrapping the s-SWCNTs can

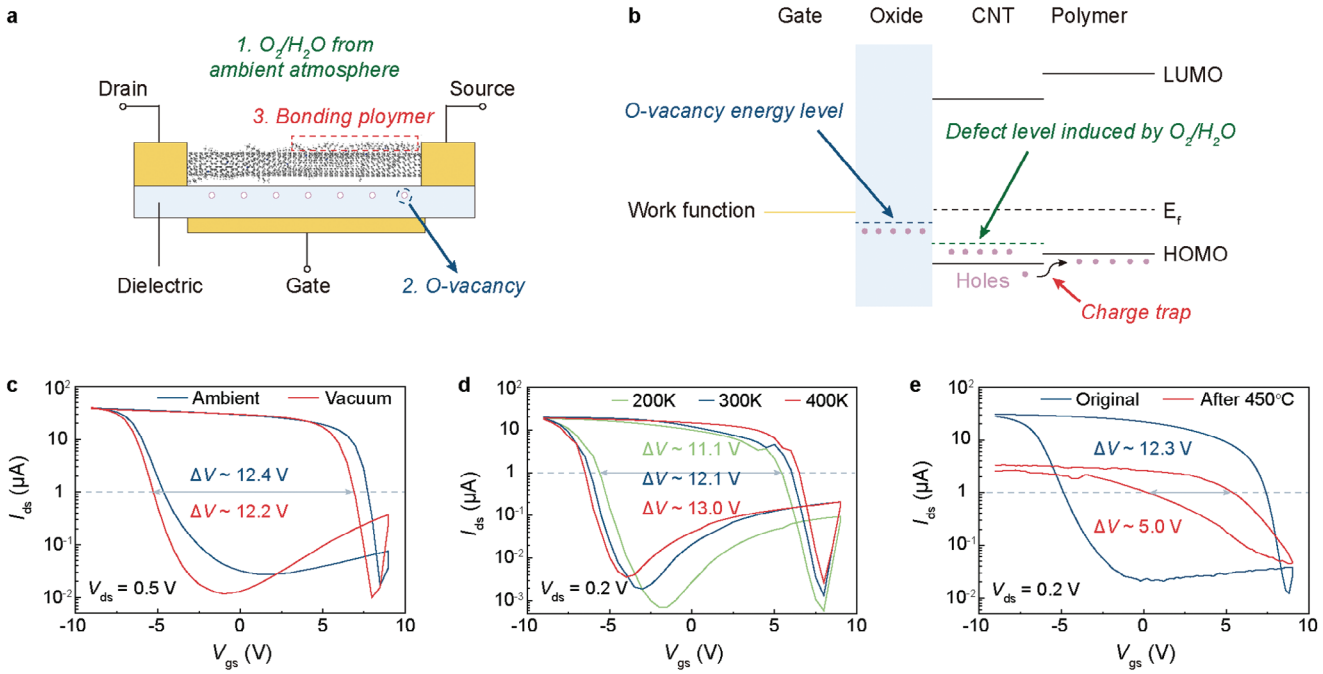

Figure 2. Memory mechanism. a) Schematic device configuration and the possible memory mechanisms, including oxygen doping, charge trapping in the gate oxides, and charge trapping in the sorting polymers. b) The corresponding energy band diagram for the possible memory mechanisms. c–e) The transfer characteristics of a typical synaptic s-SWCNT transistor memory device under the different characterization conditions, showing the memory window is significantly decreased after the decomposition of the sorting polymer PCz.

form localized junction structures that can lead to carrier transitions across the energy levels. Particularly, we note that these three mechanisms can all be the origins of the memristive effect.

To discern the origin of the memory effect of our synaptic transistor memories, we first conduct tests of a typical device in both ambient and vacuum conditiones; Figure 2c. Note that the vacuum condition is approximately 3×10^{-5} mbar and is expected to remove the oxygen and moisture that can lead to oxygen doping. As shown, the device under vacuum exhibits a primarily similar memory effect compared to ambient, but with a faster switching rate and a larger switching ratio. This suggests that oxygen doping does indeed exert an impact on the memory effect, specifically, an adverse competing impact with charge trapping in the gate oxide HfO_2 and/or the sorting polymer PCz. We then conduct tests of the device in temperatures ranging from 200 to 400 K; Figure 2d. As shown, as the temperature is increased, the device switching ratio is slightly decreased, and the memory window remains largely unchanged. A memory from charge trapping that is induced by surface defects can be easily affected by the operating temperature.[27] The largely unchanged memory window under the different temperature conditions in our device suggests that the surface defects from the dielectric oxide do not play a critical role in the memory. Lastly, to investigate charge trapping in the sorting polymers, we anneal the device at 450 °C under nitrogen for 1 h to decompose the sorting polymer PCz (Figure S4, Supporting Information). As shown in Figure 2e, after the decomposition of the sorting polymer PCz, the device exhibits a significant reduction in the memory window, proving that the memory effect is primarily from charge trapping in the sorting polymer. Note that the on-state current decreases by about 1 decade, which can be attributed to the disruption of carrier transport through the percolative s-SWCNTs networks by the carbonization of the sorting polymer PCz. The off-state current, determined by the overall conduction of s-SWCNTs and the carbonized PCz, increases by ≈3 times. Based on the above discussion, we understand that PCz not only efficiently sorts the s-SWCNTs but also provides a major contribution to the memory effect of the device.

2.3. Synaptic Functions

In brain computations, the action potentials are transmitted between the neurons through the synapses for information transmission and processing—the action potentials are weighted by the synaptic weights and, in turn, modulate the synaptic weights (**Figure 3a**).[29,30] Our synaptic transistor memories present greatly varying memory windows at the different cycling gate voltages, as shown in Figure 3b. This proves that the conductance states of our memories can be conveniently modulated by the gate voltage V_{gs}, emulating the neural process where the synaptic weights are modulated by the spike potentials. A successful emulation of the synaptic functions is pivotal for neuromorphic computing. Here we explore the synaptic functions of our memories by configuring the pulsed gate voltage V_{gs} as illustrated in Figure 3a.

Figure 3c presents the inhibitory post-synaptic current (IPSC) from a typical synaptic transistor memory device at a single pulsed gate voltage signal V_{gs} of varying amplitudes. Estimating from the current output, the device shows a prominent

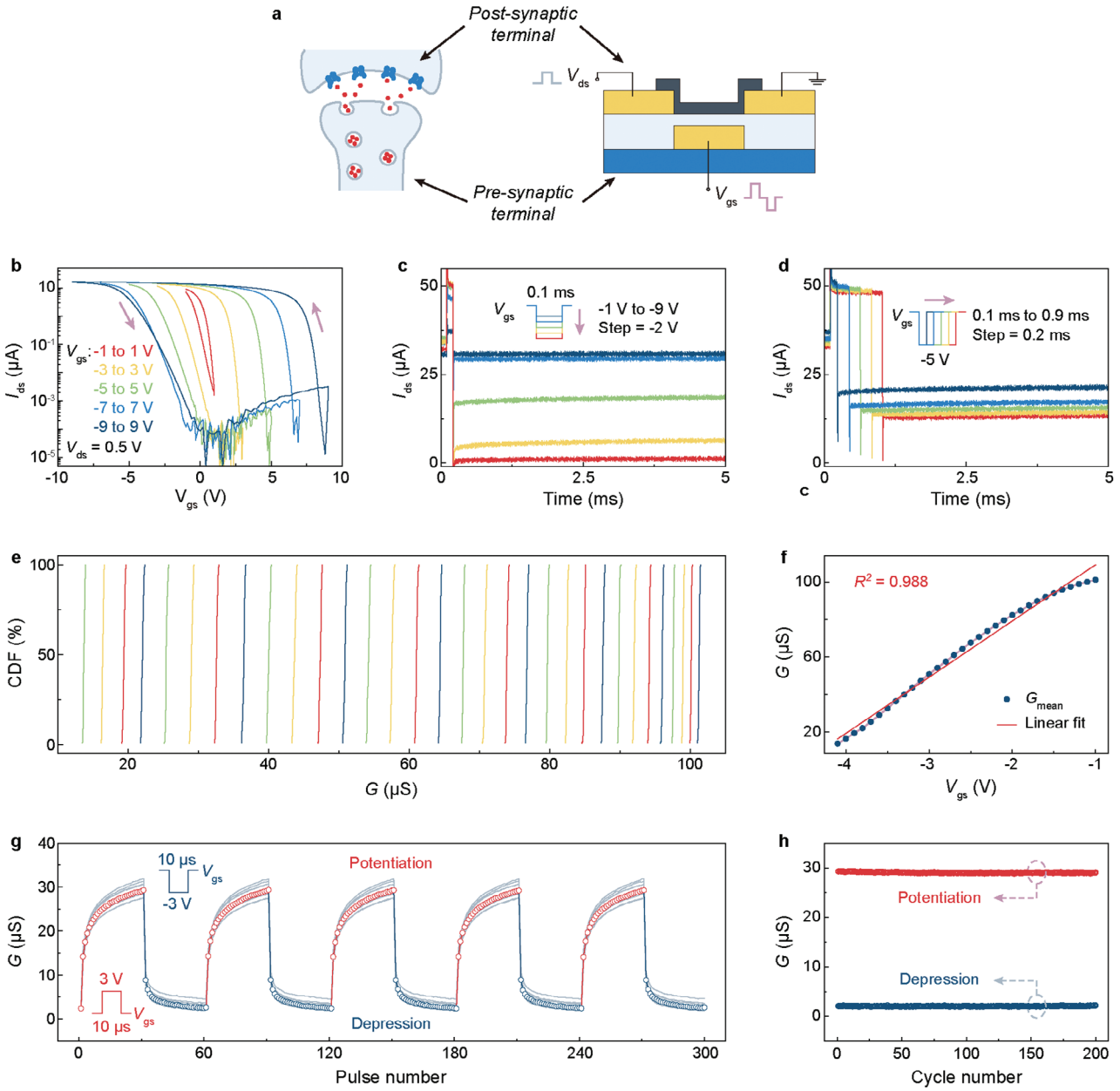

Figure 3. Synaptic functions. a) Schematic biological synapse, and the pre- and post-synaptic stimulation of the memory devices for synaptic operation. b) The transfer characteristics from a typical device with sweep gate voltage cycles V_{gs}. Single pulse responses (inhibitory post-synaptic current) from the device with a single pulse V_{gs} of c) different amplitudes and d) pulse durations. e) Cumulative distribution functions (CDF) of selected 32 conductance states measured from the device, showing highly stabilized reconfigurable conductance states with linear gate tunability. See also Figure S7 (Supporting Information) for the complete CDF results. f) Linearity evaluation of the conductance states in (e). The dots are the mean values of the conductance states, and the line is the fitted curve. g) Repetitive pulse responses from the devices, showing the long-term potentiation (LTP) and depression (LTD). 30 positive pulses followed by 30 negative pulses are applied to the gate in the test. 5 potentiation–depression cycles of 10 typical devices are shown here. See Figures S10 and S11 (Supporting Information) for the complete cycle-to-cycle and device-to-device potentiation-depression test cycles under repetitive pulses. h) Cycle-to-cycle on (red) and off (blue) conductance states from the 200 complete potentiation-depression test cycles. The drain–source voltage V_{ds} in (b–h) is 0.5 V DC.

amplitude-dependent conductance state updating behavior and, specifically, with an increased amplitude of negative input the inhibitory post-synaptic current shows a more significant inhibitory effect. Conversely, a larger positive gate voltage configures the device to a higher conductance state. See Figure S5a (Supporting Information) for the exhibitory post-synaptic current (EPSC). Similarly, at single pulsed gate voltage signal V_{gs} of varying pulse durations, the device exhibits a slightly lesser duration-dependent conductance state updating behavior (Figure 3d and Figure S5b,c, Supporting Information). In addition, as

demonstrated in Figure S6 (Supporting Information), the conductance states exhibit a stable long-term retention of 200 s. This proves the non-volatile characteristic of our devices, suggesting their use in neuromorphic hardware development.

The investigations above prove a conductance state updating behavior and that the conductance states are more readily updated by the amplitude of the gate voltage. The conductance state updating arises from charge trapping and a larger driving voltage can lead to more pronounced charge trapping. Based on this understanding, we therefore conduct tests of the device with single pulsed gate voltage signals with a gradually increasing amplitude (0 to -5 V, with -0.1 V step) to investigate the stabilized conductance states that the device can be configured to. By statistically analyzing 100 switching cycles from the device at each of the gate voltage amplitudes, we estimate the statistical cumulative distribution function (CDF) of the highly-stabilized conductance states from the device. As shown in Figure 3e (see also Figure S7a, Supporting Information), we present 50 highly stabilized conductance states from the device in this test. Particularly, the linear regression fitting of the conductance states as shown in Figure 3f (see also Figure S7b, Supporting Information) verifies that over 30 of the 50 highly stabilized conductance states are uniformly distributed in 5–100 μS region, proving a linear gate voltage amplitude-dependent conductance state updating behavior. In addition, notably, as demonstrated in Figure S8a,b (Supporting Information), the device proves highly stable and durable conductance state updating behavior by single pulsed voltage signals.

Besides the single pulsed voltage signals, we investigate the conductance state updating of the device with pulsed gate voltage signal trains; Figure 3g (see also Figure S9a,b, Supporting Information). As shown, the devices establish both long-term potentiation (LTP) and long-term depression (LTD), and a larger gate voltage amplitude gives the device a more pronounced and rapid conductance state updating and thus, LTP and LTD. Besides, the conductance state can gradually saturate with the number of the pulses, similar to the synaptic weight saturation in biology.[29] We attribute the conductance state saturation to the saturation of charge trapping in the sorting polymer PCz under the pulsed signals. Notably, the on- and off-conductance state updating using repetitive pulsed gate voltage signal trains is highly stable (Figure 3h). Besides, as demonstrated, the cycle-to-cycle and device-to-device variation in all the tested conductance states is <4% (Figure S10, Supporting Information) and <30% (Figure S11, Supporting Information), respectively.

Due to the inhomogeneity of charge trapping, charge trapping memory devices often face challenges in realizing linearity and symmetry in weight updating that are important for neuromorphic hardware operations.[31] Our devices exhibit both limited linearity and symmetry (Figure S9c, Supporting Information). To overcome this problem, we use single pulsed voltage signals to update the weight, given that our devices allow for highly stable and durable conductance state updating behavior. Here we update the weight (i.e., the conductive state) following a two-step strategy—1) resetting the devices with one single gate pulse; and 2) programming the conductive states with one single gate pulse, where the amplitudes of the pulse can be linearly determined by the required conductive states (as demonstrated in Figure 3f). Note the weight programming process is ≈1 ms to ensure stable and accurate weight programming, i.e., stable and accurate control over the charge trapping process. After the weight is programmed, data processing is achieved by allowing the programmed devices to process the data as presented in pulsed signals via the source-drain terminals. Due to the small response delay of the devices (Figure S2, Supporting Information), the data processing can be configured at a high speed of 1 M pulse bits per second, i.e., a speed of 10^{-6} s per pulse. As demonstrated in Figure S8c,d (Supporting Information), our devices allow highly stable and durable pulsed signal processing at a configurated conductance state. Following the weight updating and data processing strategy, the energy consumption is down to ≈2.16 fJ for one weight programming, and ≈103 pJ for one bit data processing, respectively (see Note S1, Supporting Information). A linear endurable conductance state updating behavior with robust data processing capability, as well as the considerably low energy consumption of our devices is pivotal for practical neuromorphic hardware implementation.[32] Note that an alternative way for weight update is via spike-timing-dependent plasticity (STDP); see Figure S12 (Supporting Information).

2.4. Hardware Convolution Kernel

Convolutional neural networks (CNNs) are a widely used algorithmic framework designed to deal with grid-structured data in, for instance, image and video processing.[21] The underlying mechanism of CNNs relies on data computation using convolution kernels,[33] as illustrated in **Figure** 4a. Briefly, the kernel with an array of weights that determine certain feature filtering functions scans through the input grid, and then feeds the dot product of the input with the weights into the output. However, extensive convolution matrix multiply-accumulate operations can be required for reliable results, posing challenges for the current computing hardware. Designing hardware kernels using memory arrays for direct convolution matrix multiply-accumulate operations of the input, therefore, holds a great interest in addressing the heavy computation issue and accelerating the efficiency of CNNs.[2]

To explore the feasibility of our synaptic transistor memories in implementing convolution kernels, we design the kernel and convolution operation workflow as illustrated in Figure 4b. In this convolution operation workflow, to conveniently conduct convolution multiply-accumulate operations using fundamental Ohm's and Kirchhoff's principles, a vertical string of 9 synaptic transistor memory devices is integrated as a 3 × 3 convolution kernel. When in operation, the conductance states of the devices are programmed via the gate voltage signals to set certain feature filtering functions, and the input encoded in the form of voltages is then applied to the drain terminals of the devices for the convolution matrix multiply-accumulate operations. In this convolution operation, the input voltage is weighted by the conductance state of the corresponding device node following the Ohm's law, and the currents from the 9 nodes are summed up as the current output following the Kirchhoff's current law. For an easy read out of the current output, the current output is amplified and transformed into a voltage output. However, note that as the conductance state of our synaptic transistor memories cannot be preprogrammed as

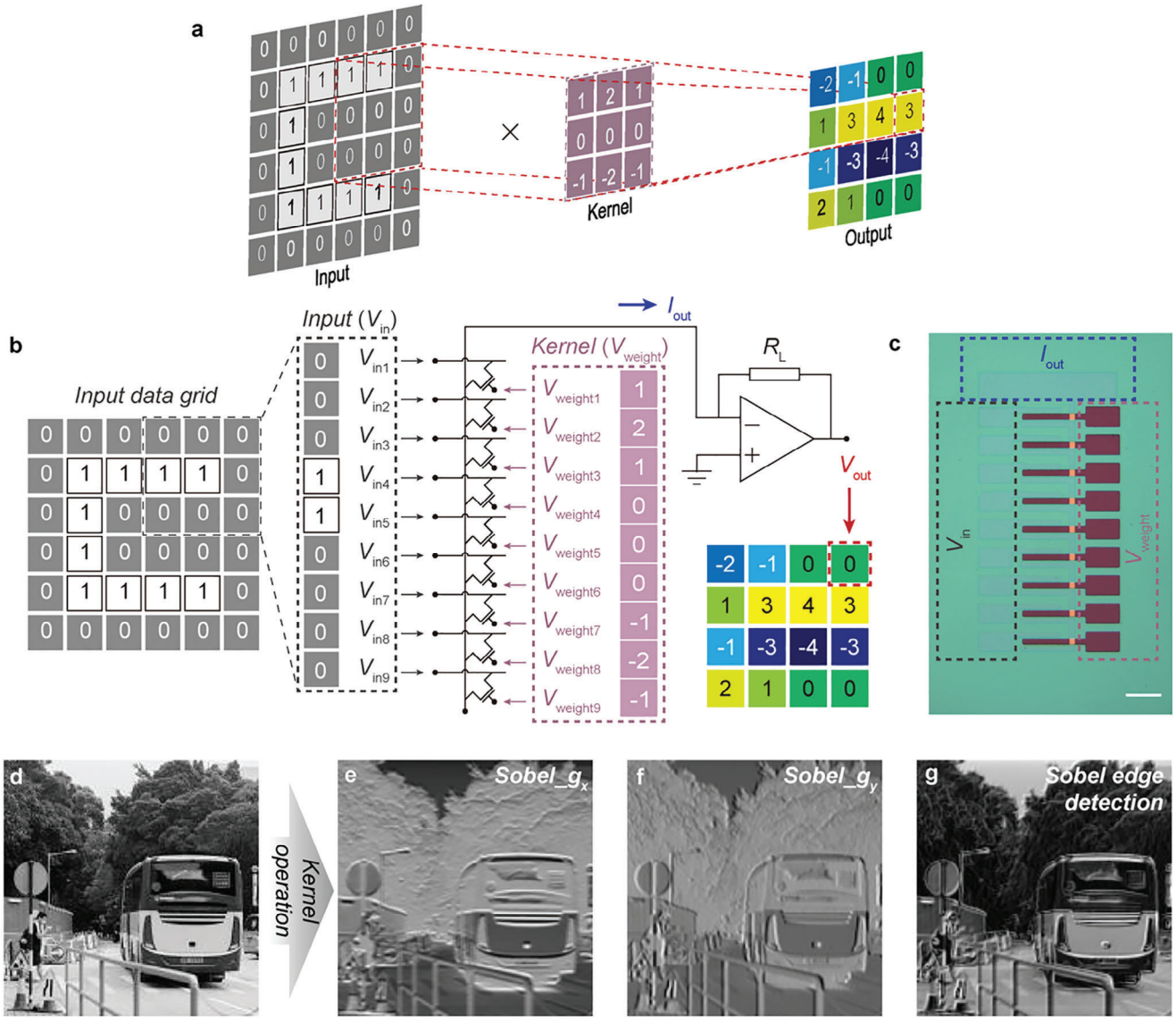

Figure 4. Hardware convolution kernel. a) Schematic convolution operation flow. b) Schematic implementation of hardware convolution kernel using the synaptic transistor memory array, and the operational flow of the hardware convolution kernel. The kernel weights for filtering functions are programmed with the gate voltage signals applied to the transistors, and the input data grid is encoded as voltage signals and applied to the drains of the transistors. The voltage output determined from the summed up weighted currents is fed to the output layer. c) Optical microscopic image of the hardware convolution kernel for convolution operation. Scale bar: 100 μm. d) Grayscale input photo of 160 × 160 pixels. Convolution Sobel *x*-directional e) and *y*-directional f) edge detection operation output from the original image. g) Convolution Sobel edge detection operation combined from e) and f) using the function $g_{Sobel} = \sqrt{g_x{}^2 + g_y{}^2}$.

negative, the input can be encoded as negative instead through an operational amplifier. This can allow the hardware convolution kernel to perform a wider range of convolution matrix multiply-accumulate operations, accommodating both positive and negative contributions to the output. Figure 4c shows an optical microscopic image of a hardware convolution kernel fabricated with our synaptic transistor memories.

To explore the feasibility of using our hardware convolution kernel for efficiency acceleration of CNNs, we perform convolution image processing. Figure 4d shows the original image of 160 × 160 pixels for the test. As an example, to perform edge detection (one of the key convolution image processing operations), we program the conductance states of the devices to function as the Sobel operator.[34,35] Specifically, we perform Sobel operator operations on the image in the horizontal (*x*-axis) and vertical (*y*-axis) directions, respectively (Figure 4e,f). Combining these two Sobel operator operations, as shown in Figure 4g, we achieve successful Sobel edge detection. As demonstrated, the edges of the car in the image are clearly revealed. This proves that our synaptic transistor memories can indeed implement convolution kernels for accelerating the efficiency of CNNs. Notably, arising from the fast-switching characteristic of the synaptic transistor memories, the

hardware convolution kernel allows parallel data processing with a speed of 1 M bit per second per input channel. Note that the edge detection performance achieved by our hardware convolution kernel is comparable to the state-of-the-art reports achieved by memristors[36,37] and transistors,[7,38] however, with simpler circuit designs and/or weight programming and data processing strategy.

3. Conclusion

In this work, we have successfully realized scalable fabrication of charge trapping s-SWCNT synaptic transistor memories. The memories demonstrate highly-stabilized reconfigurable conductance states, successful emulation of the biological synaptic functions, and a high data processing speed. Notably a high uniformity in the device performance characteristic metrics is proved, suggesting the capability of our device fabrication for industrial-scale manufacturing. These above device characteristics allow us to design and implement hardware convolution kernel using the memories for high-speed convolution image processing. We show the key convolution operation for edge detection using the Sobel operators. Given the efficacy of the convolution kernel, we envisage the use of the hardware convolution kernel in real-time high throughput information processing. As demonstrations, we simulate large-scale convolution kernels for high-speed edge detection and noise reduction of videos (Videos S2 and S3, Supporting Information) to explore the potential of our synaptic transistor memories in practical neuromorphic computing applications in, for instance, autonomous driving.

4. Experimental Section

Preparation of s-SWCNTs Solution: s-SWCNTs solution is prepared following a reported method.[22] Briefly, 1 mg mL^{-1} arc-discharged SWCNTs (Carbon Solutions Inc.) and 2 mg mL^{-1} poly[N-(1-octylnonyl)-9Hcarbazole-2,7-diyl] (PCz, Macklin) are mixed and dispersed in toluene for 1 h using tip sonicator (Fangxu NE-1000Z) at 1000 W × 40% power. The mixture is then centrifuged (Hettich Universal 320R) at 20 000 g for 2 h to remove the undispersed bundles and 90% of the supernatant is collected. The supernatant is vacuum filtered with 0.45 µm polytetrafluoroethylene (PTFE) filter membrane, washed with tetrahydrofuran (THF) to remove the residual polymer, and then redispersed in $CHCl_3$ with another 0.5 h tip sonication.

Fabrication of s-SWCNT Synaptic Transistor Memory Array: The SiO_2/Si substrate is cleaned with deionized water, acetone, and isopropanol with bath sonication for 10 min, respectively. The source, drain, and bottom gate electrodes (5 nm Ti and 15 nm Au) are patterned by photolithography, deposited by electron beam evaporation (EBE, IVS EB-600), and then obtained by a lift-off process. The dielectric oxide layer (20 nm HfO_2) is deposited with atomic layer deposition (ALD). The channel layer (i.e., s-SWCNTs) is deposited with the dip coating process (lowering speed 500 µm s^{-1}, lifting speed 100 µm s^{-1}, repeating for 15 times). The s-SWCNT layer is patterned and etched by oxygen plasma to reveal the channel. The as-deposited s-SWCNT film is baked at 150 °C for 1 h to improve the contact with electrodes before etching.

Electrical Characterizations: The electrical properties and synaptic functions are measured by Tektronix Keithley 4200A-SCS equipped with model 4225-RPM remote pulse modules. The multistate conductance states, endurance, and the convolution kernel tests are conducted using Siglent SDG7032A arbitrary waveform generator and SDS2354X digital storage oscilloscope, in conjunction with an operational amplifier (TL082CP) and an external resistor that matches the signal. Note that the setup allows parallel data processing with only 3 channels. Due to this limitation, the hardware convolution kernel tests are performed by summing up the 3 channel results from the separate tests.

Large-Scale Convolution Edge Detection: The Sobel operator is adopted for edge detection. It uses two 3 × 3 kernels, consisting of 0, ±1, and ±2 as weights. To implement it, the device conductance of 10, 50, and 100 µS is mapped to the positive weight of 0, 1, and 2, respectively. The negative weights are realized by inverting the input.

Large-Scale Convolution Noise Reduction: The neural network for noise reduction is based on a lightweight dataset-free model, Zero Shot Noise2Noise, with only a two-layer convolutional network.[39] Implementing noise reduction involves two stages, model training and model conversion. For model training, the network is trained in the same way as training a typical DNN. After training, the model is converted into a memristive DNN using the MemTorch framework,[40] where non-ideal device characteristics are considered. For instance, the 32 relatively stable, linear conductance states in Figure 3e,f are adopted as the parameter finite number of discrete conductance states in the model conversion. See Table S1 for the detailed parameters in model training and conversion. The denoising performance is evaluated by the peak signal-to-noise ratio (PSNR). The PSNR score of each frame before and after denoising is marked in Video S3 (Supporting Information). Despite the high frame resolution (1280 × 720) and noise level (50, Gaussian noise), an average PSNR of ≈20 dB indicates a good denoising performance. The simulation is carried out in Python 3 and is available upon request.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

G.H.H. acknowledges funding from CUHK (4055115) and SHIAE (RNE-p3-21), Y.L. from SHIAE (RNE-p3-21), J.F.P. and Y.Y.W. from RGC (24200521), T.M. from PolyU (P0042991), X.C. from NSFC (62275117), and S.G. from NSFC (62171014).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

carbon nanotubes, convolution kernel, neuromorphic computing, synaptic memory, thin-film transistors

Received: November 27, 2023
Revised: May 13, 2024
Published online: October 29, 2024

[1] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[2] X. Xu, Y. Ding, S. X. Hu, M. Niemier, J. Cong, Y. Hu, Y. Shi, *Nat. Electron.* **2018**, *1*, 216.

[3] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, D. S. Modha, *Science* **2014**, *345*, 668.
[4] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, H. Bhaskaran, *Nature* **2021**, *589*, 52.
[5] L. Danial, E. Pikhay, E. Herbelin, N. Wainstein, V. Gupta, N. Wald, Y. Roizin, R. Daniel, S. Kvatinsky, *Nat. Electron.* **2019**, *2*, 596.
[6] E. J. Fuller, S. T. Keene, A. Melianas, Z. Wang, S. Agarwal, Y. Li, Y. Tuchman, C. D. James, M. J. Marinella, J. J. Yang, A. Salleo, A. A. Talin, *Science* **2019**, *364*, 570.
[7] M.-K. Kim, I.-J. Kim, J.-S. Lee, *Sci. Adv.* **2022**, *8*, eabm8537.
[8] J. Cui, F. An, J. Qian, Y. Wu, L. L. Sloan, S. Pidaparthy, J.-M. Zuo, Q. Cao, *Nat. Electron.* **2023**, *6*, 292.
[9] S. Kamaei, X. Liu, A. Saeidi, Y. Wei, C. Gastaldi, J. Brugger, A. M. Ionescu, *Nat. Electron.* **2023**, *6*, 658.
[10] S. Dai, Y. Zhao, Y. Wang, J. Zhang, L. Fang, S. Jin, Y. Shao, J. Huang, *Adv. Funct. Mater.* **2019**, *29*, 1903700.
[11] X. Ji, B. D. Paulsen, G. K. K. Chik, R. Wu, Y. Yin, P. K. L. Chan, J. Rivnay, *Nat. Commun.* **2021**, *12*, 2480.
[12] F. S. Yang, M. Li, M. P. Lee, I. Y. Ho, J. Y. Chen, H. Ling, Y. Li, J. K. Chang, S. H. Yang, Y. M. Chang, K. C. Lee, Y. C. Chou, C. H. Ho, W. Li, C. H. Lien, Y. F. Lin, *Nat. Commun.* **2020**, *11*, 2972.
[13] H. F. Haneef, A. M. Zeidell, O. D. Jurchescu, *J. Mater. Chem. C* **2020**, *8*, 759.
[14] Y. Hu, L.-M. Peng, L. Xiang, H. Zhang, *Acc. Mater. Res.* **2020**, *1*, 88.
[15] J. Zhao, F. Liu, Q. Huang, T. Lu, M. Xi, L. Peng, X. Liang, *Nano Res.* **2021**, *14*, 4258.
[16] W. Talsma, H. van Loo, S. Shao, S. Jung, S. Allard, U. Scherf, M. A. Loi, *Adv. Intell. Syst.* **2020**, *2*, 2000154.
[17] I. Sanchez Esqueda, X. Yan, C. Rutherglen, A. Kane, T. Cain, P. Marsh, Q. Liu, K. Galatsis, H. Wang, C. Zhou, *ACS Nano* **2018**, *12*, 7352.
[18] S. Kim, J. Yoon, H. D. Kim, S. J. Choi, *ACS Appl. Mater. Interfaces* **2015**, *7*, 25479.
[19] S. Kim, Y. Lee, H. D. Kim, S. J. Choi, *Nanoscale* **2020**, *12*, 2040.
[20] L. Liu, J. Han, L. Xu, J. Zhou, C. Zhao, S. Ding, H. Shi, M. Xiao, L. Ding, Z. Ma, C. Jin, Z. Zhang, L.-M. Peng, *Science* **2020**, *368*, 850.
[21] P. Lin, C. Li, Z. Wang, Y. Li, H. Jiang, W. Song, M. Rao, Y. Zhuo, N. K. Upadhyay, M. Barnell, Q. Wu, J. J. Yang, Q. Xia, *Nat. Electron.* **2020**, *3*, 225.
[22] J. Gu, J. Han, D. Liu, X. Yu, L. Kang, S. Qiu, H. Jin, H. Li, Q. Li, J. Zhang, *Small* **2016**, *12*, 4993.
[23] J. Noh, M. Jung, Y. Jung, C. Yeom, M. Pyo, G. Cho, *Proc. IEEE* **2015**, *103*, 554.
[24] W. Kim, A. Javey, O. Vermesh, Q. Wang, Y. Li, H. Dai, *Nano Lett.* **2003**, *3*, 193.
[25] M. Qu, H. Li, R. Liu, S. L. Zhang, Z. J. Qiu, *Nat. Commun.* **2014**, *5*, 3185.
[26] R. S. Park, M. M. Shulaker, G. Hills, L. Suriyasena Liyanage, S. Lee, A. Tang, S. Mitra, H. S. Wong, *ACS Nano* **2016**, *10*, 4599.
[27] S. Wang, Q. Huang, J. Zhao, Y. Zhou, X. Zhao, P. Yao, X. Liu, X. Liang, *ACS Appl. Energy Mater.* **2019**, *2*, 8253.
[28] F. A. Lemasson, T. Strunk, P. Gerstel, F. Hennrich, S. Lebedkin, C. Barner-Kowollik, W. Wenzel, M. M. Kappes, M. Mayor, *J. Am. Chem. Soc.* **2011**, *133*, 652.
[29] H. Yu, H. Wei, J. Gong, H. Han, M. Ma, Y. Wang, W. Xu, *Small* **2021**, *17*, e2000041.
[30] F. Chen, Y. Zhou, Y. Zhu, R. Zhu, P. Guan, J. Fan, L. Zhou, N. Valanoor, F. von Wegner, E. Saribatir, I. Birznieks, T. Wan, D. Chu, *J. Mater. Chem. C* **2021**, *9*, 8372.
[31] S. Kim, M. Lim, Y. Kim, H. D. Kim, S. J. Choi, *Sci. Rep.* **2018**, *8*, 2638.
[32] M. K. Kim, Y. Park, I. J. Kim, J. S. Lee, *iScience* **2020**, *23*, 101846.
[33] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, H. Qian, *Nature* **2020**, *577*, 641.
[34] N. Kanopoulos, N. Vasanthavada, R. L. Baker, *IEEE J.Solid-State Circuits* **1988**, *23*, 358.
[35] O. R. Vincent, O. Folorunso, in *Proceedings of Informing Science & IT Education Conf. (InSITE)*, Vol. *9*, Informing Science Institute, Macon, United States, **2009**, *9*, 97.
[36] C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Dávila, C. E. Graves, Z. Li, J. P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, *Nat. Electron.* **2017**, *1*, 52.
[37] J. Tian, H. J. Zhou, H. Bao, J. Chen, X. D. Huang, J. C. Li, L. Yang, Y. Li, X. S. Miao, *IEEE Trans. Electron Devices* **2022**, *69*, 6043.
[38] W. Li, S. Zhou, X. Xia, Y. Wang, K. Yang, T. Hao, X. Zhang, Q. Yang, Z. Ni, J. Jiang, J. Si, F. Zhang, Z. Zhang, *Adv. Funct. Mater.* **2023**, *33*, 2305724.
[39] Y. Mansour, R. Heckel, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Piscataway, NJ **2023**, 14018.
[40] C. Lammie, W. Xiang, B. Linares-Barranco, M. R. Azghadi, *Neurocomputing* **2022**, *485*, 124.