Title

Synaptic Resistor Circuit based on Al Oxide and Ti Silicide for Neuromorphic Computing

Author

GAO, DAWEI

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Synaptic Resistor Circuit based on Al Oxide and Ti Silicide for Neuromorphic Computing

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Materials Science and Engineering

By

Dawei Gao

2024

ABSTRACT OF THE DISSERTATION

Synaptic Resistor Circuit Based on Al Oxide and Ti Silicide for Neuromorphic Computing

by

Dawei Gao

Doctor of Philosophy in Materials Science and Engineering

University of California, Los Angeles, 2024

Professor Yong Chen, Chair

The human brain, a remarkably efficient system operating on minimal power, excels at processing complex sensory data and learning from experience. Inspired by this biological marvel, this dissertation reports the development of dedicated neuromorphic computing hardware, specifically focusing on a 20 by 20 synaptic resistor circuit based on aluminum oxide and titanium silicide material. By mimicking the brain's synaptic connections, these devices offer low-power, non-volatile memory and the ability to learn and adapt in real time, demonstrating the potential of the synaptic resistor circuit through a successful application in drone navigation.

These findings pave the way for energy-efficient and adaptable neuromorphic computing systems with broad implications for artificial intelligence.

The dissertation of Dawei Gao is approved.

Ximin He

Qibing Pei

Mau-Chung Frank Chang

Yong Chen, Committee Chair

University of California, Los Angeles

2024

This thesis is dedicated to my family, whose love and support inspired me to pursue my dreams.

TABLE OF CONTENTS

LIST OF FIGURES

Figure 17. Drone flying experiments with **a,** a synstor circuit (left), a human operator (middle), and an ANN running in digital computers (right). In the real-time flights by the synstor circuit, the deviations of the drone position from its target position, $\boldsymbol{s}$, are detected by a camera, and converted by an interface circuit to voltage pulses, $\boldsymbol{x}$, which are input into a synstor circuit (Methods). By following the signal-processing algorithm $\boldsymbol{I} = \boldsymbol{w}\,\boldsymbol{x}$ (Equation 1) the synstors generate output currents, $\boldsymbol{I}$, that trigger actuation signals, $\boldsymbol{y}$, via an interface circuit to drive the drone and modify $\boldsymbol{s}$. Concurrently, backpropagating voltage pulses, $\boldsymbol{z}$, triggered at the output electrodes encounter the $\boldsymbol{x}$ pulses to modify the synstor conductance matrix, $\boldsymbol{w}$, by following the learning algorithm $\boldsymbol{w} = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x}$ (Equation 2). Human operators see the analog values of $\boldsymbol{s}$ displayed on a monitor, process and learn from the signals concurrently in neurobiological circuits, and drive the drone by triggering actuation signals, $\boldsymbol{y}$, with different keys on a keyboard, but initially they do not know which key triggers which actuation signal. The ANN on an edge computer receives sensing signals, $\boldsymbol{s}$, and sends actuation pulses, $\boldsymbol{y}$, to drive the drone on the basis of the synaptic weight matrices $\boldsymbol{w}(n)$ in the ANN in its nth round of trial. The data, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{w}(n)$ are saved in memory, and sent to a Google cloud computer, on which a reinforcement learning algorithm is executed to modify $\boldsymbol{w}(n)$ to $\boldsymbol{w}(n+1)$. Then $\boldsymbol{w}(n+1)$ is saved in memory, and sent back to the edge computer for the $(n+1)th$ round of trial. **b,** The deviations of the drone position from its target position, $\boldsymbol{s}$**,** are shown as a function of time t**. c,** Objective

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Prof. Yong Chen, for his invaluable guidance, support, and encouragement throughout my PhD journey. His expertise, patience, and unwavering belief in my abilities have been instrumental in my academic growth and success. I am truly fortunate to have had the opportunity to work under such an inspiring advisor.

I am also deeply grateful to my current and past colleagues, Dr. Chris Shaffer, Dr. Dhruva Nathan, Dr. Rahul Shenoy, Zixuan Rong, Dr. Atharva Deo, Dr. Jungmin Lee, Dr. Yong Hei, Yingrui Zhang, for their collaboration and support. Their contributions to our research have been invaluable, and I have learned a great deal from working with them.

To my dad, Xiuhua Gao and my mom, Lijuan Xie, I owe an immense debt of gratitude. Your unconditional love, support, and sacrifices have made this journey possible. Your encouragement has always been my driving force.

Finally, I would like to thank my PhD committee members, Prof. Ximin He, Prof. Qibing Pei, Prof. Frank Chang, Prof. Yong Chen, for their time, expertise, and valuable feedback. Their insightful comments and suggestions have significantly improved the quality of my research.

The research provided in this thesis is funded by the Air Force Office of Scientific Research (AFOSR) under the programs “Brain Inspired Networks for Multifunctional Intelligent Systems in Aerial Vehicles (FA9550-19-0213),’’ and “Multi-layer Self-Programming Neuromorphic Integrated Circuits for Deep Learning’’ (FA9550-20-1-0230).

The material in this thesis is either published, submitted or in preparation for publication.

VITA

2019	B.S. Materials Science and Engineering University of Illinois, Urbana-Champaign
2019-2024	PhD. Candidate and Graduate Student Researcher Materials Science and Engineering Department University of California, Los Angeles

PUBLICATIONS

Gao, D., Shenoy, R., Yi, S., Lee, J., Xu, M., Rong, Z., Deo, A., Nathan, D., Zheng, J.G., Williams, R.S. and Chen, Y. Synaptic Resistor Circuits Based on Al Oxide and Ti Silicide for Concurrent Learning and Signal-Processing in Artificial Intelligence Systems. *Advanced Materials*, p.2210484 (2023)

Xu, M., **Gao, D**., Zheng, J. G., & Chen, Y. (2023). Quantification of Si, Al, Ti and O Composition in Si/Al Oxide Based Synaptic Resistor Circuits. *Microscopy and Microanalysis*, *29*(1), 240–241. (2023)

Asapu, S., Pagaduan, J.N., Zhuo, Y., Moon, T., Midya, R., **Gao, D**., Lee, J., …, Yang, J.J. Large remnant polarization and great reliability characteristics in W/HFZRO/W ferroelectric capacitors. *Frontiers in Materials*, *9*, p.969188 (2022)

Nathan, D., Deo, A., Haughn, K., Yi, S., Lee, J., **Gao, D**., Shenoy, R., Xu, M., Tran, I.C., …, Chen, Y. Si-based self-programming neuromorphic integrated circuits for intelligent morphing wings. *Journal of Composite Materials*, *56*(30), pp.4561-4575 (2022)

Lee, J.*, Shenoy, R.*, Deo, A., Yi, S., **Gao, D**., …, and Chen, Y. HfZrO-based synaptic resistor circuit for super-Turing intelligent system. (*Equal contribution, Submitted)

Deo, A., Lee, J., Liu, A., **Gao, D.**, …, Chen, Y. An Intelligent self-driving car with energy-efficient synaptic resistor circuits for fast real-time learning based on HFZRO. (Submitted)

Deo, A., Lee, J., **Gao, D.**, …, Chen, Y. An Intelligent non-Turing reusable spaceship system on HfZrO-based energy-efficient synaptic resistor circuits with fast real-time learning. (Submitted)

Deo, A., Lee, J., Shenoy, R., **Gao, D.**, …, Chen, Y. Intelligent morphing wing based on

synaptic resistor circuits with real-time learning. (Submitted)

Zhao, R., Wang, T., Moon, T., Xu, Y., Sud, P., Zhuo, Y., Midya, R., Asapu, S., **Gao, D.**, ..., Williams, R. S., Chen, Y., Yang, J. J. Compact all-round spiking artificial neuron (Submitted)

Lee, J**.**, Yi, S., **Gao, D**., and Chen, Y. Associative memory circuit based on synaptic resistors for concurrent inference and learning. (In preparation)

1. Introduction

Artificial Intelligence (AI) has seen an extensive integration into diverse sectors, with revolutionizing capabilities in image recognition[1,2], natural language processing[3,4], and beyond. Central to this rapid technological evolution is the exponential miniaturization of Complementary Metal-Oxide-Semiconductor (CMOS) transistors[5], a progression succinctly described by Moore's law[6]. This principle states that the number of transistors on a microchip doubles approximately every two years, simultaneously with reduced production costs. Such advancements have not only catalyzed the development of increasingly powerful and efficient microchips but have also significantly extended the capabilities and applications of AI systems.

However, this growth is not without its challenges. As microchips become smaller and more potent, the global energy demands for computing have surged, threatening to impose unsustainable burdens on our power production systems. Estimates suggest that if current trends persist, computing could consume a significant portion of global power output within the next decade[7], highlighting an urgent need for more energy-efficient computing architectures.

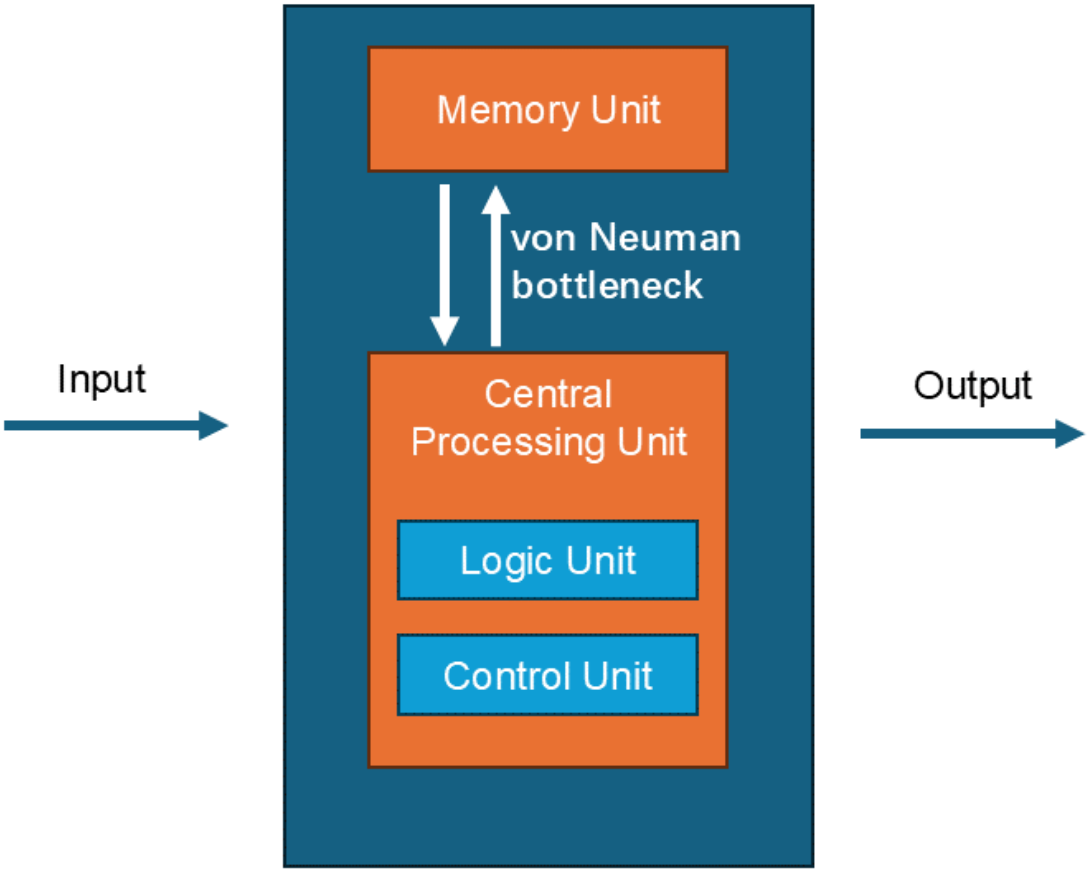

Figure 1. Schematic representation of the von Neumann bottleneck. This diagram illustrates the data flow constraint between the central processing unit (CPU) and the memory unit via a narrow bus system, which limits the rate of data transfer and impacts system performance.

A pivotal hindrance to achieving a more energy-efficient future is encapsulated in the "von Neumann bottleneck" (**Figure 1**). This bottleneck represents a fundamental inefficiency in traditional computing architectures, where separate processor and memory components necessitate constant data transfers during processing, thereby consuming excessive time and energy[8]. This separation markedly reduces the efficiency and speed of computational operations,

an issue that becomes particularly critical in AI applications involving rapid processing of large datasets. Addressing the von Neumann bottleneck is thus increasingly crucial as we push the boundaries of AI capabilities.

In addition, the ability of AI systems to learn and adapt in real-time is increasingly crucial in environments characterized by rapid and unpredictable changes[9–11]. In fields ranging from autonomous vehicle navigation[12–14] to real-time financial market analysis[15,16], the conditions under which decisions must be made can shift with little warning. Traditional computational systems, constrained by pre-programmed responses and batch processing, struggle to keep pace with such dynamism. Real-time learning enables systems to adjust their algorithms on the fly, optimizing responses as new data becomes available without the need for human intervention. This capability not only enhances the responsiveness and flexibility of AI systems but also significantly improves their reliability and effectiveness in complex, real-world applications. By incorporating real-time learning, systems can better simulate human-like adaptability and decision-making processes, making them more suited to navigate and operate within the fluid and ever-changing realities of modern environments.

Emerging as a paradigm shift in computational theory, the concept of super-Turing computing[17–19]—inspired directly by the neural processes of the human brain—promises to transcend these limitations. Unlike traditional Turing machines, which operate linearly and within the confines of predefined algorithms and sequential data processing[20], super-Turing computing adopts the complex and dynamic nature of biological systems. This approach enables continuous data processing, real-time learning, and adaptation—traits inherent to human cognition but absent in conventional computers.

Inspired by the complex biological neural networks of the human brain, which manage intricate processing and learning tasks, neuromorphic computing seeks to emulate these energy-efficient and adaptive processes. Despite technological leaps enabling supercomputers, such as Fugaku[21] and Frontier[22], to achieve computational speeds potentially surpassing the human brain, such systems still consume substantially more power and lack the efficiency and adaptability of biological counterparts. This observation underscores the limitations of traditional computing architectures and highlights the significance of neuromorphic computing.

Neuromorphic computing, a pioneering approach within the field of AI, leverages the structural and functional principles of the human brain to design and implement novel hardware architectures. By integrating memory and processing units akin to neurons and synapses, these systems reduce the data transfer demands that exacerbate the von Neumann bottleneck[23–26], thereby enhancing computational efficiency and speed. The inherent advantages of neuromorphic systems—particularly in terms of energy efficiency and processing speed—are substantial[23,27–31]. They consume significantly less power than traditional architectures, operate efficiently on sparse data typical of sensory inputs, and perform parallel processing, which enables real-time interactions and decision-making crucial for applications in robotics, autonomous vehicles, and real-time data analysis.

This dissertation introduces an innovative energy-efficient neuromorphic computing hardware design—a crossbar circuit composed of synaptic resistors (synstors), each incorporating a silicon channel integrated with an aluminum oxide memory layer and titanium silicide Schottky contacts. Operating without prior training, this circuit concurrently performs signal processing and learning in real time. In dynamically changing aerodynamic environments, it guides drones to specified targets, outperforming human controllers. The learning speed,

overall performance, power efficiency, and environmental adaptability of this synstor circuit significantly surpass those of traditional artificial neural networks (ANNs) operated on computers, thus demonstrating a promising direction for sustainable and intelligent computing solutions.

1.1. Concurrent Signal Processing and Learning in the Brain

The human brain is a complex network of ~100 billion neurons and ~100 trillion synapses[32–34]. Each neuron in the brain connects to about a thousand others through synapses that are sites of electrochemical communication. These synapses are dynamic spaces where neurotransmitters modulate neural signals through the opening and closing of ion channel gates in the neural membrane[35]. In biological neural networks, the synaptic weight represents the strength of connection, which depends on the density of neurotransmitter receptors and the volume of neurotransmitters released. Thus, the weight in a biological context controls how effectively a neuron can influence its neighbor.

In the brain, signal processing and learning are the two key functions that operate concurrently in real time. The signal processing function of the brain happens when an action potential triggers the release of neurotransmitters across the synaptic cleft, initiating a cascade of ionic flows through channels in the post-synaptic neuron's membrane. This electrochemical process results in the generation of a new action potential, which then propagates the signal to further neural connections[36]. Meanwhile, learning occurs through changes in the synaptic strength (weight) based on the frequency and timing of these signals. This phenomenon, known as synaptic plasticity[34,35], allows the brain to store memories and adapt behaviors based on experience.

The integration of signal processing and learning enables the brain to function with remarkable efficiency[14,37], processing complex stimuli and learning from environmental interactions in real time. This dual functionality is fundamental to the brain's ability to interpret sensory data, make decisions, and execute complex motor commands—all while consuming minimal energy compared to current artificial systems.

1.2. Prior Work in Neuromorphic Computing Platform

Neuromorphic computing is a research area focused on the development of integrated circuits and systems that emulate the architecture of biological neural networks. As transistor technology pushes towards more advanced nodes, the energy efficiency of transistor circuits is capped around 10^{12} floating point operations per second per watt (FLOPS/W)[38–41]. However, neuromorphic computing circuits such as synaptic transistors, resistive random-access memory (ReRAM), and phase-change memory (PCM) resistors, often utilizing the in-memory computing and analog parallel computing to achieve remarkable power efficiencies between 10^{10} and 10^{14} FLOPS/W[42–50].

However, current neuromorphic computing circuits face several challenges. First, current neuromorphic circuits suffer from "read disturb"[51,52]. The circuits require high voltages for writing to adjust conductance matrices during the execution of learning algorithms and low voltages for reading to preserve these matrices during inference and reduce power consumption. Unlike brain, current neuromorphic circuits are unable to concurrently carry out signal processing and learning due to the difference in voltage domain.

Furthermore, variability among devices results in a higher error rate for analog neuromorphic circuits during learning algorithm execution compared to digital circuits based on

transistors[53]. For accuracy, it is necessary to first run learning algorithms on digital computers to establish optimal conductance matrices for the neuromorphic circuits. These matrices are subsequently fine-tuned through sequential reading and writing processes, which necessitates the use of separate memory and logic circuits and the transmission of signals between them. This separation sometimes nullifies the advantage of in-memory computing and reduces the energy efficiency of learning processes to levels that are similar to or even less than those achieved by traditional digital circuits.

The significant energy and time required to accurately find the optimal values of a large conductance matrix using learning algorithms from large datasets with high-dimensional variables exponentially increase with the complexity and size of the data due to the "curse of dimensionality"[54]. This substantial overhead poses a major challenge for edge computers in AI systems intended for real-time, on-site learning. As a result, learning algorithms are typically run on central high-speed computers offsite, utilizing “big data” and consuming considerable amounts of power and time to develop optimal conductance matrix[55].

While these optimized algorithms excel in specific, well-defined tasks such as arithmetic calculations, games, standard image and speech recognition, and operating self-driving cars under certain conditions, they fall short in more unpredictable settings. Their performance significantly decreases when dealing with tasks outside their trained domains, such as recognizing nonstandard images or speech, managing autonomous vehicles in unpredictable conditions, operating robotic systems in complex environments, and conducting intricate medical diagnoses. Current AI systems lack the broad adaptability and general intelligence of the human brain to effectively respond to dynamic, unpredictable challenges.

Existing neuromorphic circuits also face problems with leakage currents, which typically require complex peripheral circuits to manage[56,57]. Sneak currents arise when unintended directional flows occur within the crossbar array, driven by different potentials applied to synaptic elements, pulling current from synapses with high conductance. This issue is particularly problematic in devices like memristors that demand large currents for switching. To mitigate this, neuromorphic devices are ideally designed to have minimal or zero current during switching and display a non-linear current-voltage curve with low currents at reading voltages, to minimize energy loss and sneak currents. To tackle this, a selector such as one or more transistors[56,57] is incorporated to each synaptic element to allow external programming of conductance matrix. However, this adds complexity and increases the power demands of the system.

To avoid the use of a selector, a typical approach is to apply a half-voltage to the row and column that align with the targeted device, ensuring that only the intended device receives the full tuning voltage. Consequently, it is vital that the conductance of the neuromorphic device is not adjusted at this half-voltage level. To alleviate the demands on external circuitry, the response of the neuromorphic device to the tuning voltage should ideally be a nonlinear function

1.3. Research Goals

The goal of this research is to design and develop a novel neuromorphic device that integrates three key synaptic behavior—signal processing, learning, and memory functions—into a single synstor, and integrate these synstors into a neuromorphic circuit in crossbar fashion, where outputs connect to silicon-based artificial neuron circuits for nonlinear signal processing.

To process signals effectively, the synstors should be arranged in a crossbar circuit layout that facilitates vector-matrix multiplication tasks between the input vector and the synaptic weight matrix. This configuration enables synaptic weight adjustments during the learning phase when there may be a temporal alignment between the output and input vectors. By setting the device's conductance below that of typical neuromorphic devices and eliminating the need for a selector device, it is possible to avoid sneak currents. The synstor's current remains considerably lower compared to other two-terminal neuromorphic devices like memristors because the voltage is applied across insulating oxide layers, not across conductive or semiconductive paths. The synstor, or synaptic resistor, needs to display some particular traits to operate efficiently. It incorporates nonvolatile, analog, tunable memory critical for real-time and efficient signal processing and learning. For optimal functionality, it's essential to fine-tune the device's materials and geometry so that the conductance state is modifiable with the same voltage levels utilized for learning function. This adjustment enables concurrent and efficient signal processing and learning operations. Moreover, synstors should allow for modifications via correlative Hebbian learning, which utilizes the input of pulse signals and the backpropagation of output pulse signals. By integrating these synaptic functions within a singular device and adopting pulse operations, the overhead on extensive peripheral circuits for weight programming can be significantly reduced, leading to a substantial decrease in energy consumption.

The dissertation aims to demonstrate the circuit’s effectiveness and adaptability through a practical learning demonstration involving controlling a drone in a complex aerodynamic environment. The performance of these synstor circuits will be benchmarked against human neurobiological circuits and artificial neural networks (ANN) in comparable environments to assess real-time learning capabilities against traditional offline learning approaches. The goal is

to show that without any prior training, the synstor circuits concurrently executed signal-processing and learning in real time to fly drones toward a target position in an aerodynamically changing environment faster than human controllers, and with learning speed, performance, power consumption, and adaptability to the environment significantly superior to an ANN running on computers.

2. The Synstor Circuit based on Al Oxide and Ti Silicide

This whole section has been derived from the published papers[14,37,58], which are used here to complete the dissertation.

Previously, we developed a synstor based on carbon nanotubes (CNTs) to emulate a synapse with memory, learning, and signal-processing functions.[23,37] The device is detailed in **Figure 2**. It consists of a channel formed by a network of randomly oriented p-type semiconducting carbon nanotubes. These CNTs are connected through Schottky junctions to aluminum input and output electrodes. Additionally, the device includes a memory stack featuring a HfO2/TiO2/HfO2 charge trap heterojunction structure. An aluminum bottom electrode serves as a reference potential, completing the sophisticated design of this innovative synaptic resistor.

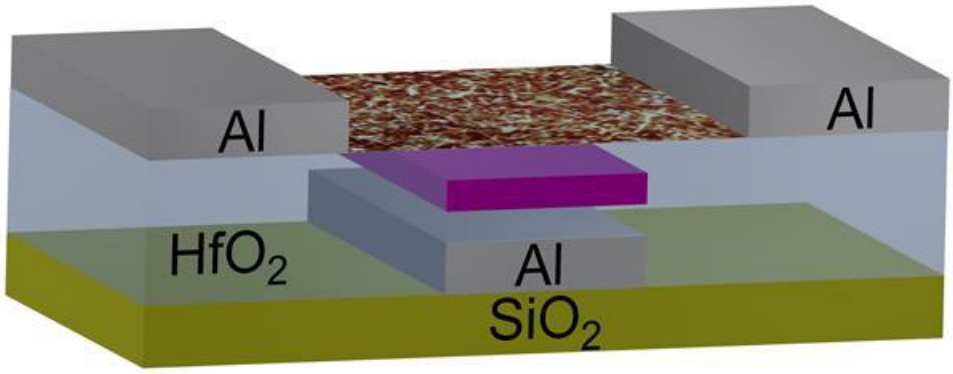

Figure 2. The device structure of a CNT-based synstor[37]. The synstor has Al input and output electrodes, connected via Schottky junctions to a random network of p-type semiconducting carbon nanotubes. The device has an Al reference electrode and a HfO2/TiO2/HfO2 charge trap heterojunction as a switching layer.

However, the CNT-based synstor circuits were not uniform because of variations in the conductance of and the metal contacts to the CNTs[14,37], which significantly worsens their performance. In this work, we fabricated a crossbar synstor circuit (**Figure 3a**) by integrating a Si channel with an Al oxide memory layer and $TiSi_{\gamma}$ Schottky contacts in each synstor. The voltage pulses triggered at M input electrodes, $\boldsymbol{x}$, induced currents, $\boldsymbol{I}$, at N output electrodes via synstors to execute signal-processing algorithm,

$$\boldsymbol{I} = \boldsymbol{w}\,\boldsymbol{x} \tag{1}$$

where $\boldsymbol{w}$ denotes a conductance matrix of synstor circuit. Concurrently, $\boldsymbol{w}$ is modified by following the synaptic learning algorithm,

$$\dot{\boldsymbol{w}}(t) = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x} \tag{2}$$

where $\boldsymbol{z}$ denotes a function of the voltage pulse vector triggered at the N output electrodes (Equation 3, in Section 3.1.1) to drive a drone in an unpredictably changing environment with learning speed, performance, adaptability, and power consumption significantly superior to those of ANNs running on digital computers. Without any prior training, the synstor circuits concurrently executed signal-processing and learning in real time to fly drones toward a target position in an aerodynamically changing environment faster than human controllers, and with learning speed, performance, power consumption, and adaptability to the environment significantly superior to an ANN running on computers.

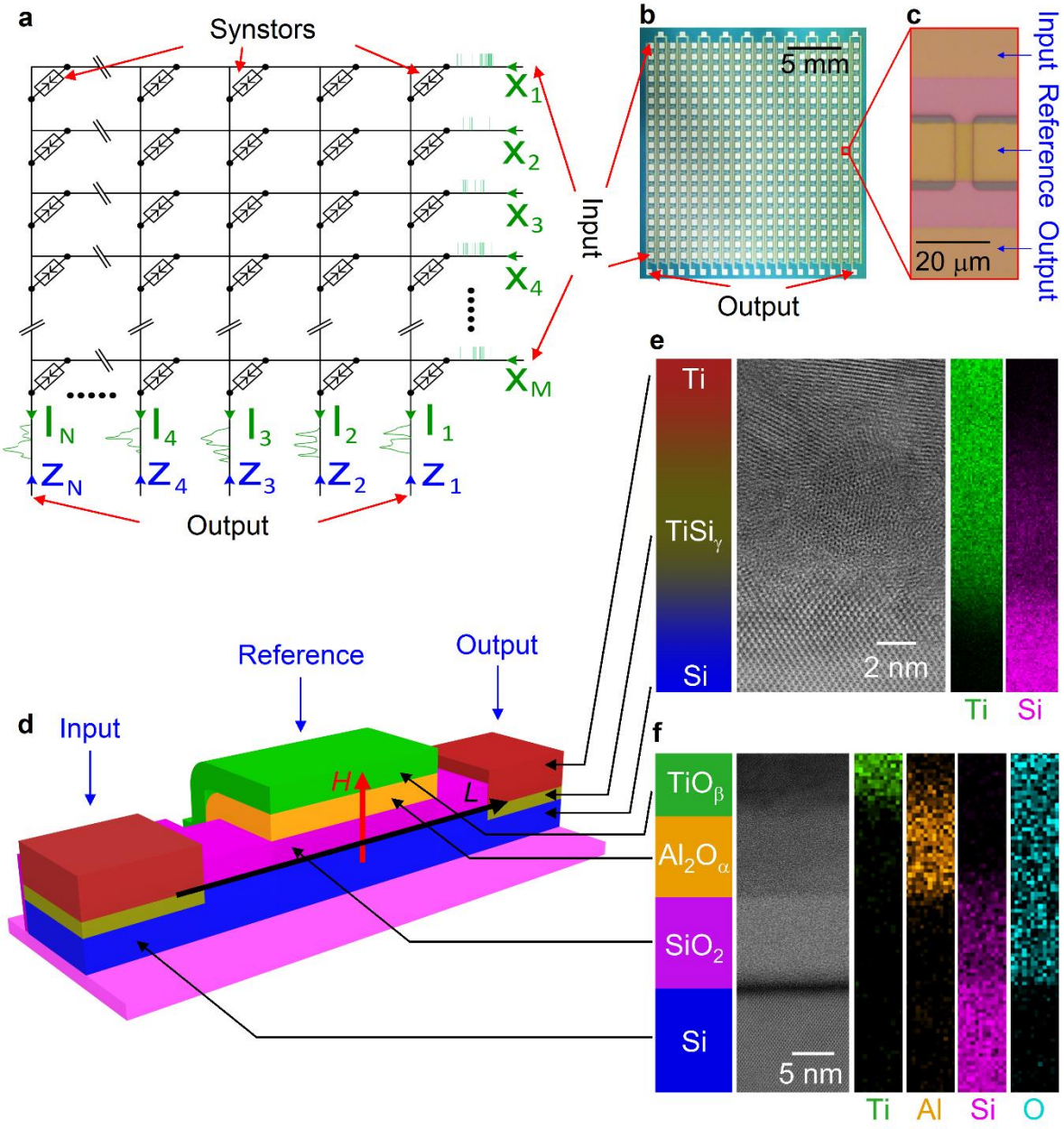

Figure 3. The structure of a synstor circuit. a) A schematic diagram showing a neuromorphic circuit composed of $M \times N$ synstors connected with M input electrodes and N output electrodes. $\boldsymbol{x}$ denotes a vector with element x_m as the voltage pulses applied to the mth input electrode, $\boldsymbol{z}$ denotes a vector with element z_n as the voltage pulses applied to the nth output electrode, and $\boldsymbol{I}$ denotes a vector with element I_n as the current induced at the nth output electrode. b) A top-view

optical image shows a 20×20 synstor crossbar composed of 20 rows of Ti input electrodes, 20 columns of Ti output electrodes, and 20 columns of TiO_β reference electrodes. c) A top-view optical image and d) a 3D schematic illustration of a synstor composed of a Si channel surrounded by a SiO_2 layer. A TiO_β reference electrode grown on an Al_2O_α layer on the SiO_2 layer on the Si channel are displayed. The Si channel is contacted at each end with a $TiSi_\gamma$ layer and Ti input and output electrodes. e) A cross-sectional high-resolution TEM image (middle), and EDX chemical maps of Ti and Si elements (right) across the Ti/$TiSi_\gamma$/Si heterojunction are displayed. f) A cross-sectional high-resolution TEM image (left), and EDX chemical maps of Ti, Al, Si, and O elements (right) across the $TiO_\beta/Al_2O_\alpha/SiO_2$/Si heterojunction are displayed. The regions of the different layers in $TiO_\beta/Al_2O_\alpha/SiO_2$/Si and Ti/$TiSi_\gamma$/Si heterojunctions are schematically indicated in e) and f).

2.1. Methods

2.1.1. Synstor Circuit Fabrication Process

The fabrication process of a synstor circuit is shown in **Figure 4**. The synstor was fabricated on a silicon-on-insulator (SOI) chip with a 220 nm-thick p-type Si layer with a boron doping concentration of $10^{15}/cm^3$ on a 3 µm-thick buried oxide layer on a 725 µm-thick Si wafer. A photoresist layer was patterned on the Si surface as an etching mask through ultraviolet photolithography. As shown in **Figure 4a**, a 5 µm-wide Si channel was etched by reactive ion etching (RIE; Oxford Plasmalab 80 Plus RIE). The photoresist was stripped with acetone, isopropanol, and de-ionized water. The silicon-on-insulator wafer was cleaned with the standard RCA cleaning process. As shown in **Figure 4b**, the surface of the Si channel was oxidized in a thermal oxidation furnace at 1000 °C in O_2 for 2 minutes to grow a 10 nm-thick SiO_2 layer. A

photoresist layer was patterned by ultraviolet photolithography on the surface of the SiO_2 layer, and the photoresist patterns were used as an etching mask to etch the SiO_2 layer by RIE (Oxford Plasmalab 80 Plus RIE; 1:1 CHF_3:Ar, 100 W) in the contact areas for input/output electrodes (**Figure 4c**). A 300 nm-thick Ti layer was deposited by electron beam evaporation (CHA Industries Mark 40), and lifted off by a pre-patterned photoresist layer to form Ti input/output electrodes in the contact areas (**Figure 4d**). A 40 nm-thick Al_2O_3 sacrificial layer was grown on the chip by atomic layer deposition (Fiji Ultratech ALD). The chip was annealed in forming gas (5% H_2 in N_2) at 470 °C for 30 minutes to generate a titanium silicide ($TiSi_\gamma$) layer sandwiched between the Si channel and Ti input/output electrodes (**Figure 4e**). The Al_2O_3 sacrificial layer protected the materials from undesirable oxidation during the annealing. After the annealing, the sacrificial layer was selectively etched away with a 3% tetramethylammonium hydroxide aqueous solution. A 40 nm-thick titanium layer was deposited on a 10 nm-thick titanium oxide layer on a 10 nm-thick aluminum oxide layer (Al_2O_α) on the SiO_2 layer on the Si channel by e-beam evaporation (CHA Industries Mark 40), and the titanium, titanium oxide, and Al_2O_α layers were lifted off by a pre-patterned photoresist layer, as shown in **Figure 4f**. The evaporation resulted in an Al_2O_α layer with a high defect density, owing to the differing vapor pressures of aluminum and oxygen. This non-stoichiometric Al_2O_α layer is desirable for forming a charge storage layer for memory and learning. The chip was then annealed in forming gas (5% H2 in N2) at 470 °C for 30 minutes. During the deposition and annealing, the titanium oxide layer also reacted with the Al_2O_α layer and the titanium layer, thus forming a 50 nm-thick metallic titanium oxide layer (TiO_β) as a reference electrode. The chip was passivated by a parylene-C film deposited by thermal evaporation (SCS PDS-2010 Parylene). A final photolithography step and RIE processes (Technics FRIE; 500 mTorr O_2, 100 W) were used to etch the parylene-C layer in the contact pads of the input/output and

reference electrodes. The extensive contact pads, input/output, and reference electrodes in the circuit are not shown in **Figure 4**.

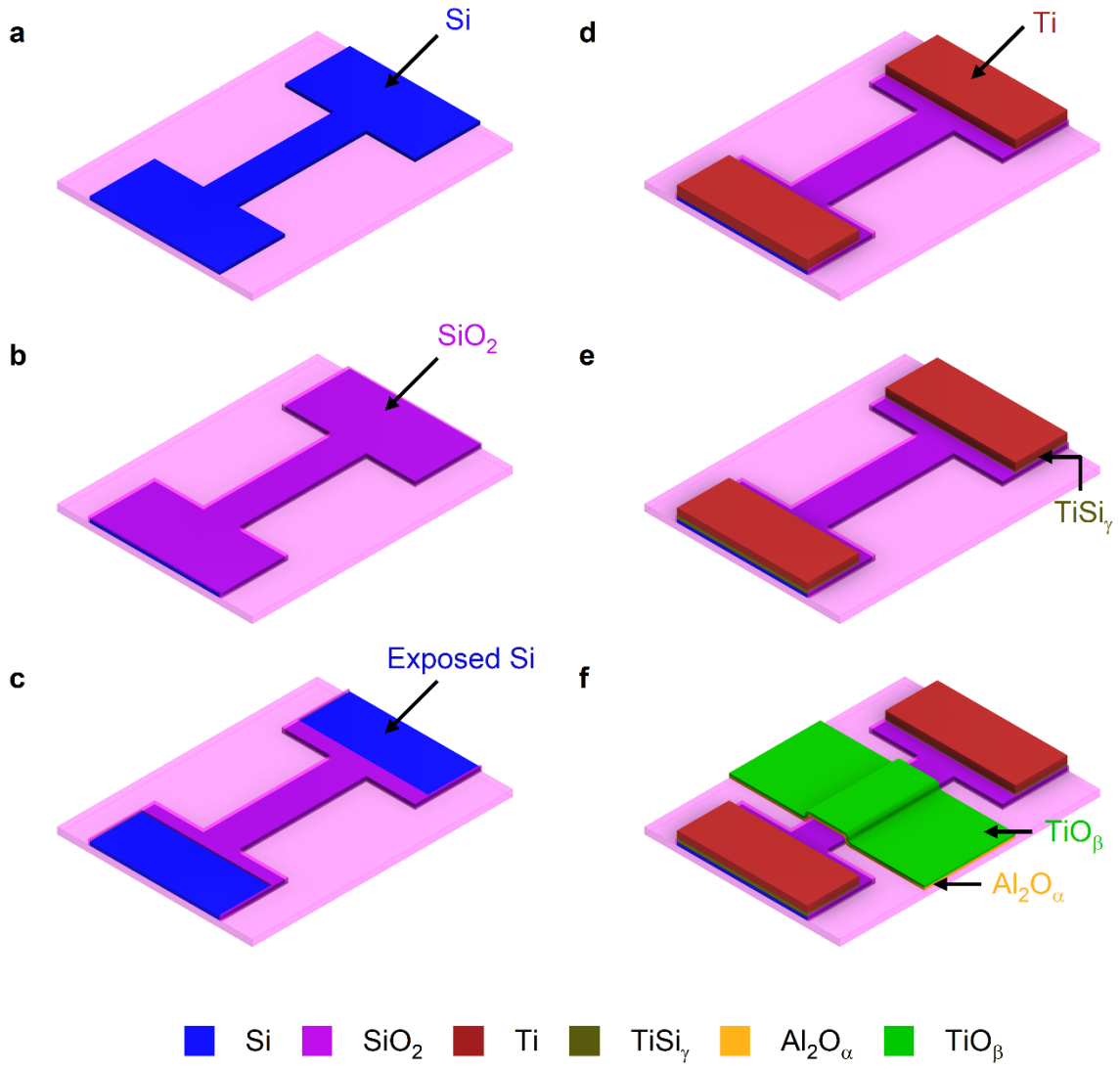

Figure 4. The synstor circuit fabrication process. a) A Si channel is etched on a SiO_2 surface. b) The Si surface is oxidized and covered by a SiO_2 layer. c) The SiO_2 layer in contact areas is etched. d) Ti input and output electrodes are deposited in the contact areas, and e) are annealed to form a titanium silicide ($TiSi_\gamma$) layer sandwiched between the Si channel and Ti

input/output electrodes. f) A titanium oxide (TiO_β) reference electrode is deposited on an aluminum oxide (Al_2O_α) memory layer on the SiO_2 layer on the Si channel.

2.1.2. Capacitor Fabrication Process

As shown in **Figure 5**, capacitors with the same structure as the synstor with a TiO_β layer grown on an Al_2O_α layer on a SiO_2 layer on a Si layer on a SiO_2 layer were fabricated through the same fabrication process described above. Al electrodes were then deposited on the Si and TiO_β layers by e-beam evaporation, and patterned by photolithography to form Al/Si ohmic contact and Al/TiO_β metallic contacts in the capacitor.

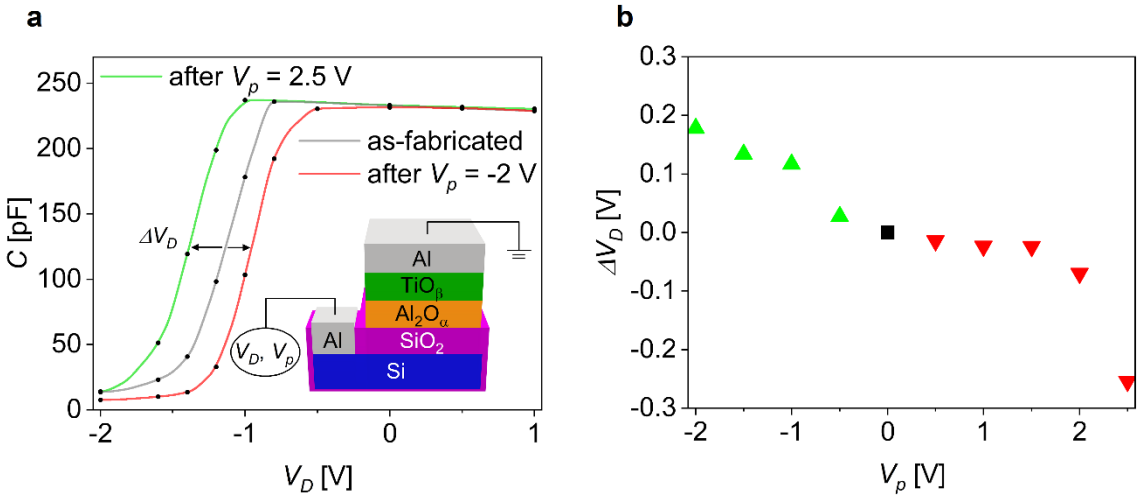

Figure 5. Capacitance-voltage testing. a) The capacitance (C) of a capacitor is plotted versus a voltage bias, V_D, applied on the bottom Al electrode with respect to its grounded top Al reference electrode on the as-fabricated capacitor (grey line), after the capacitor experienced $-2.0\ V$ voltage pulses for 10 s (red line), and the capacitor experienced 2.5 V voltage for 10 s (green

line). The structure of the capacitor is shown in the insert. The shift of the $C - V_D$ curves with respect to those of the as-fabricated capacitor is marked as ΔV_D. b) After the capacitor experienced voltage pulses with an amplitude, V_p, ΔV_D is displayed versus V_p, under the conditions of $-2\ V \leq V_p < 0$ (green triangles), $0 < V_p < 2.5\ V$ (red triangles), and $V_p = 0$ (square).

2.1.3. Material Characterization and Electrical Tests of Synstors

The structures, compositions, and defect densities of the materials in synstors were characterized by TEM, EDX analysis performed with a JEOL Grand ARM TEM operated at 300 kV with dual silicon drift detectors. The EDX Cliff-Lorimer sensitivity factors of Ti, Al, Si, O are calibrated by known TiO_2, Al_2O_3, and SiO_2 standards. The reference electrodes of the synstors and control devices were always grounded during the tests. Current-voltage characteristics were measured with a Keithley 4200 semiconductor parameter analyzer. The electrical voltage pulses applied to the input and output electrodes of the devices and circuits were generated by FPGA (National Instruments, cRIO-9063), computer-controlled modules (National Instruments, NI-9264), and a Tektronix AFG3152C waveform/function generator. Currents flowing through the synstors were measured by a semiconductor parameter analyzer, computer-controlled circuit modules (National Instruments, NI-9205 and NI-9403), and oscilloscope (Tektronix TDS 3054B). Testing protocols were programmed (NI LabVIEW) and implemented in an embedded FPGA (Xilinx), a microcontroller, and a reconfigurable I/O interface (NI CompactRIO). Synstors in the circuit were tuned to their lowest (or highest) conductance values before modifying their conductance values by applying a train of paired 10 ms-wide x and z pulses concurrently with a

width of 10 ms and an amplitude of -2 V (or 2.5 V) on each synstor until their conductance values reached the targeted analog values closely.

2.1.4. Electrical tests of the capacitor

The AC capacitance of the as-fabricated capacitor was measured by applying an AC voltage with an amplitude of $200\ mV$ and sweeping a DC voltage, V_D, from $1\ V$ to $-2\ V$ on the bottom electrode (**Figure 5**). After application of 10 s V_p voltage pulses with an amplitude of $-2.0\ V \leq V_p \leq 2.5\ V$ for 10 s on the bottom electrode, a shift in the $C - V_D$ curves with respect to those of the as-fabricated capacitor, ΔV_D, was induced by the modification of charge density, $\Delta\rho$, at the Al_2O_α/ SiO_2 interface of the $TiO_\beta/Al_2O_\alpha/SiO_2/Si$ heterojunction[59,60]. After application of voltage V_p, ranging between $-2.0\ V$ and 0 (or 0 and $2.5\ V$) on the bottom Al electrode, ΔV_D was measured by sweeping V_D from $-1.6\ V$ to $-0.8\ V$. $\Delta\rho$ was derived by $\Delta\rho = C_{SiO2}\Delta V_D$ with $C_{SiO2} = 3.5 \times 10^{-7}\ F/cm^2$ as the capacitance per area of the SiO_2 layer (**Figure 6h**).

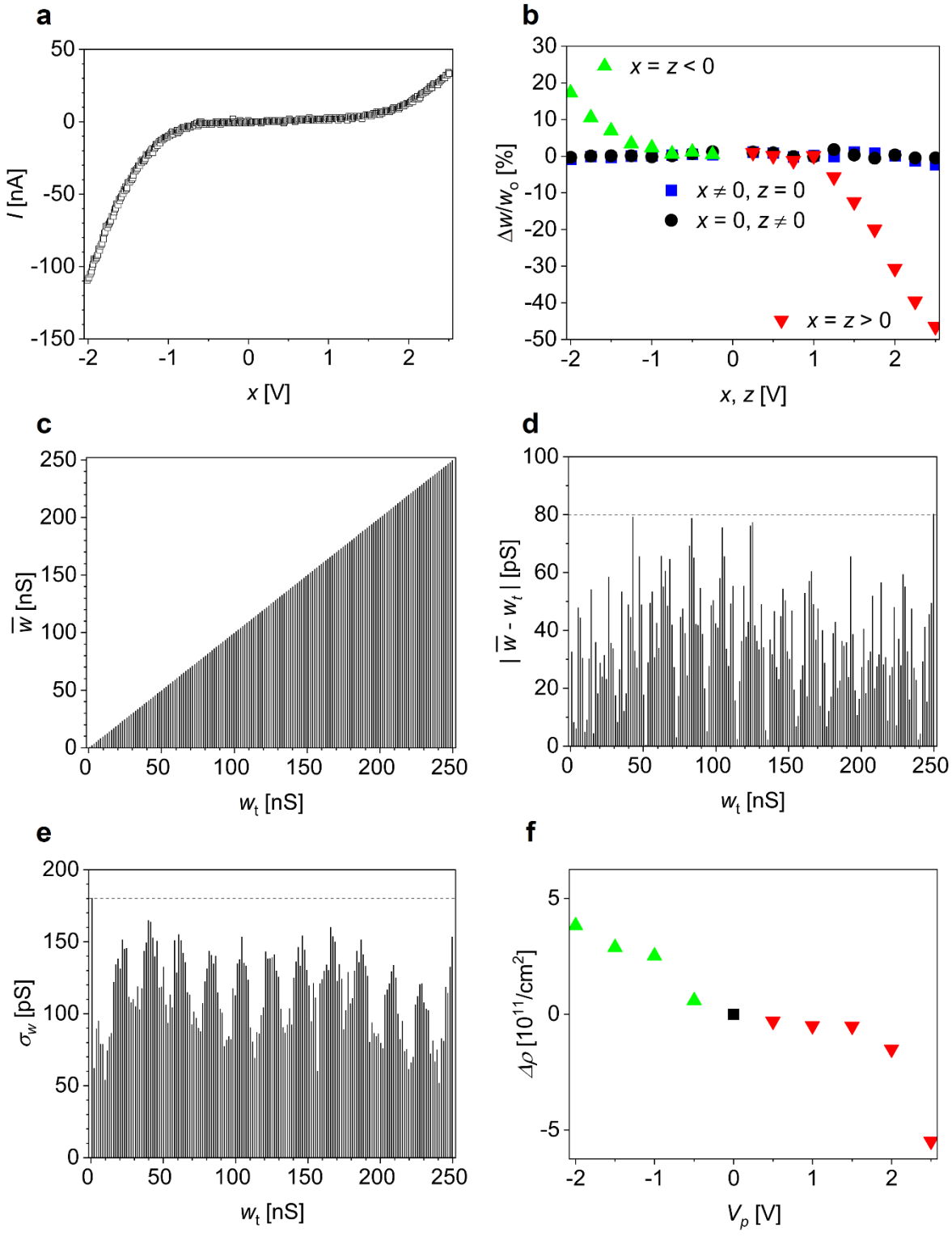

Figure 6. Synstor characterization. a) The current, I, on the output electrode of a single synstor is displayed versus the amplitude of the voltage pulses, x, applied on the input electrode. b) The

percentage change in the synstor conductance, $(\Delta w/w_0) \times 100\%$, is plotted versus various x and z pulses concurrently applied on the input and output electrodes under the conditions of $-2\ V \leq x = z < 0$ (green triangles), $0 < x = z \leq 2.5\ V$ (red triangles), $-2\ V \leq x \leq 2.5\ V$ and $z = 0$ (blue squares), and $-2\ V \leq z \leq 2.5\ V$ and $x = 0$ (black circles). c) The average analog conductance values, $\overline{w}$, d) the absolute values of the differences between $\overline{w}$ and w_t, $|\overline{w} - w_t|$, and e) the standard deviations of w, σ_w, are plotted against the targeted analog conductance values of synstors, w_t. f) The change in the charge density in the Al_2O_α layer, $\Delta\rho$, is plotted against V_p, the amplitude of the voltage pulses applied to a capacitor with the same $TiO_\beta/Al_2O_\alpha/SiO_2/Si$ heterojunction of the synstor under the conditions of $-2\ V \leq x = z < 0$ (green triangles), $0 < x = z \leq 2.5\ V$ (red triangles), and $x = z = 0\ V$ (black squares).

2.1.5. Computer-aided Device and Circuit Simulation

On the basis of the measured properties of synstors and their circuits, we have designed and calculated the performance of synstors and their circuits with a Technology Computer-Aided Design (TCAD) simulator (Spectre, Cadence and Sentaurus Device, Synopsys).[37] The device simulator performed numerical calculations of the synstor physics by solving Poisson's equation describing the electrostatics and drift-diffusion carrier transport under a set of boundary conditions defined by the device structure. Quasi-stationary simulations were conducted under various voltage biases on the input/output electrodes of the synstors with respect to the grounded reference electrodes. The band diagrams of the synstors were extracted from the simulations, and the electronic properties of the synstors were analyzed through simulation. According to the TCAD simulation, when the synstor channel length was scaled down to 30 nm, the synstor still functioned

properly for signal-processing and learning (**Figure 7**). We have designed and simulated a crossbar circuit of synstors with a channel length scaled to 30 nm The M input electrodes, N output electrodes, reference electrodes, and Si channels of the circuits were designed on the basis of the TSMC 28 nm process. The circuit was required to satisfy the following conditions to execute signal-processing and learning concurrently: (i) To avoid the phase shifts of input pulses, $\boldsymbol{x}$, and backpropagating pulses, $\boldsymbol{z}$, on the input and output electrodes, $M^2R_eC_e, N^2R_eC_e \ll t_d$, where M represents the number of input electrodes, N represents the number of output electrodes, t_d denotes the pulse duration, R_e denotes the resistance of the electrode per synstor unit, and C_e denotes the capacitance of the electrode per synstor unit. In the synstor circuit, $C_e \approx 10^{-17}\ F$, $R_e \approx 0.54\ \Omega$, and $t_d = 10^{-8}\ s$; thus, $M, N \ll 10^4$. (ii) To avoid significant voltage variations at the input and output electrodes induced by the sneak currents passing through the synstors, the worst-case voltage variation on an input or output electrode induced by the maximal sneak currents, $\Delta V_{Max}\langle w\rangle\ R_eM < k_BT/q \approx 26\ mV$, where q is the charge of an electron, k_B is the Boltzmann constant, and T is room temperature. In the synstor circuit with a channel length of 30 nm, $\langle w\rangle = 1.9$ nS, $V_a = 1.2\ V$, and $R_e \approx 0.54\ \Omega$; thus, $M, N \lesssim 10^4$. Therefore, a crossbar circuit can be scaled up to 10^8 synstors (with $M, N = 10^4$). A large scale synstor crossbar circuit with 10^8 synstors, $M, N = 10^4$, an average synstor conductance $\langle w\rangle = 1.9$ nS, an analog signal input rate $f_i = 50\ MHz$, an average pulse magnitude $V_a = 1.1\ V$, and a pulse duration $D_p = 0.01$ could execute signal-processing ($\boldsymbol{I} = \boldsymbol{w}\ \boldsymbol{x}$, Equation 1) and learning algorithms ($\dot{\boldsymbol{w}} = \alpha\ \boldsymbol{z} \otimes \boldsymbol{x}$, Equation 2) concurrently with a computing speed of $25\ petaOPS$, a power consumption of $\sim$2.3 mW, while occupying an area of $\sim$0.8 mm^2.

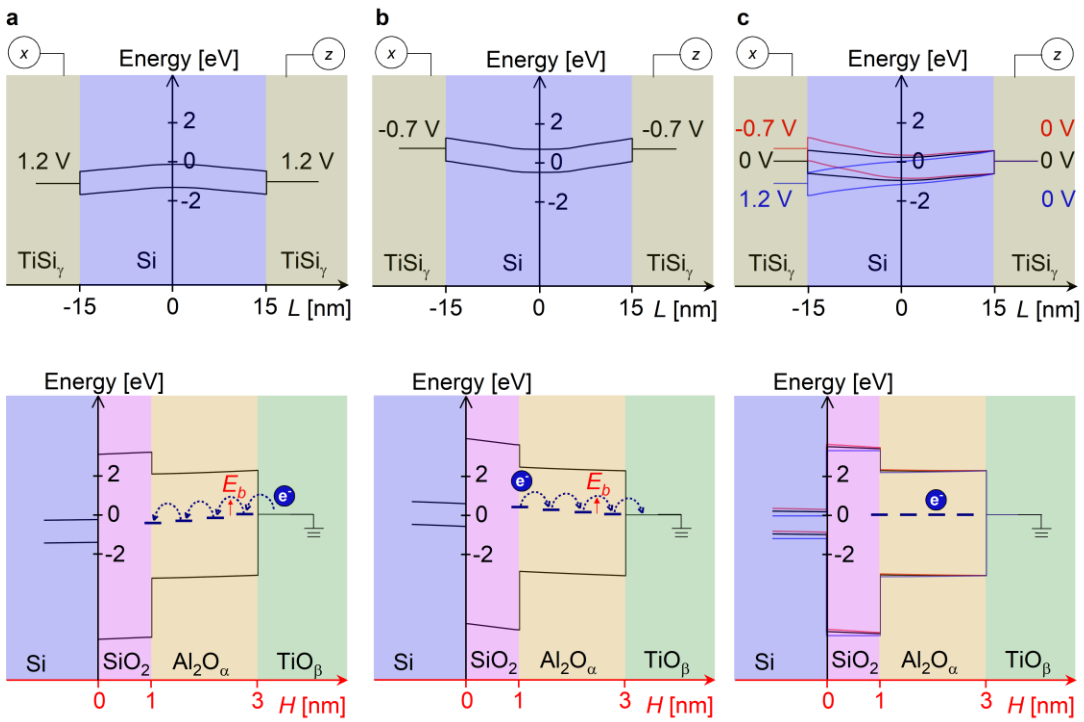

Figure 7. The energy-band structures of a synstor with a channel length of 30 nm. The simulated energy-band structures of the $TiSi_\gamma$ input electrode, Si channel, and $TiSi_\gamma$ output electrode along the L axis in **Figure 3d** (top), and of the grounded TiO_β reference electrode, Al_2O_α charge storage layer, SiO_2 dielectric layer, and Si channel along the H axis in **Figure 3d** (bottom) of a synstor with a channel length of 30 nm under the conditions of a) $x = z = 1.2\ V$, b) $x = z = -0.7\ V$, and c) the potential at the input electrode $x = -0.7\ V$, the potential at the output electrode $z = 0$ (red lines), $x = z = 0$ (black lines), and $x = 1.2\ V$ and $z = 0$ (blue lines). The scales represent the lateral distance from the input electrode to the output electrode (top row), and the vertical distance from the Si channel to the TiO_β reference electrode (bottom row). The band edges and Fermi levels are marked by solid lines. Electrons (e^-) are driven by a large potential difference across the Al_2O_α layer to hop across the oxygen vacancies in the Al_2O_α layer, as illustrated by blue dashed lines in a) and b)

2.2. Results and Discussion

As shown in **Figure 3b**, a 20×20 synstor crossbar circuit composed of 20 rows of Ti input electrodes, 20 columns of Ti output electrodes, and 20 columns of TiO_β reference electrodes was fabricated on a silicon-on-insulator substrate according to the process described in the Methods and **Figure 4**. The structure of the active area of a synstor is shown in **Figure 3c** and **Figure 3d**. A Si channel with a thickness of 220 nm, a width of 5 μm, a length of 40 μm, and a boron doping concentration of $10^{15}/cm^3$ was made from a single-crystal Si layer on a 3 μm-thick SiO_2 layer. A 14 μm-wide, 50 nm-thick TiO_β layer was grown on a 10 nm-thick Al_2O_α layer on a 10 nm-thick SiO_2 layer on the Si channel. The device structure was determined by transmission electron microscopy (TEM). As shown in **Figure 3e**, **Figure 8a**, and **Figure 9**, a polycrystal $TiSi_\gamma$ layer is sandwiched between the Si channel and Ti input/output electrodes. The compositions of the Ti/$TiSi_\gamma$/Si heterojunction were analyzed by energy-dispersive X-ray spectroscopy (EDX). The $TiSi_\gamma$ layer with average $\gamma = 0.7$ is a metallic silicide, which formed a Schottky contact to the Si channel.[61] As explained by the device simulation in the following, the Schottky barrier facilitated the concurrent signal-processing and learning functions of the synstor. The compositions of the $TiO_\beta/Al_2O_\alpha/SiO_2$/Si heterojunction were analyzed with TEM and EDX (**Figure 3f**, **Figure 4**, **Figure 10**, and **Figure 11**). The Al_2O_α layer with $\alpha = 2.4$ is oxygen-deficient. Electrons can be trapped in the oxygen vacancies in the Al_2O_α layer,[62–64] which functions as a charge storage layer. The TiO_β layer with $\beta = 1.5$ is metallic[65], and functions as a reference electrode, as well as a protection layer to preserve the high density of oxygen vacancies in the Al_2O_α layer.

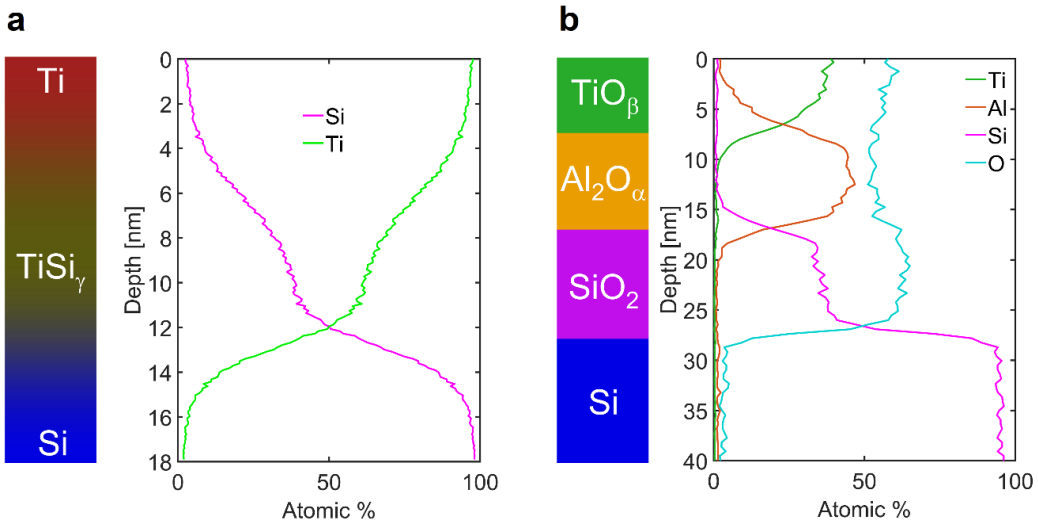

Figure 8. The EDX composition profiles. a) The EDX composition profile of the Ti/$TiSi_\gamma$/Si heterojunction. b) The EDX composition profile of the TiO_β/Al_2O_α/SiO_2/Si heterojunction.

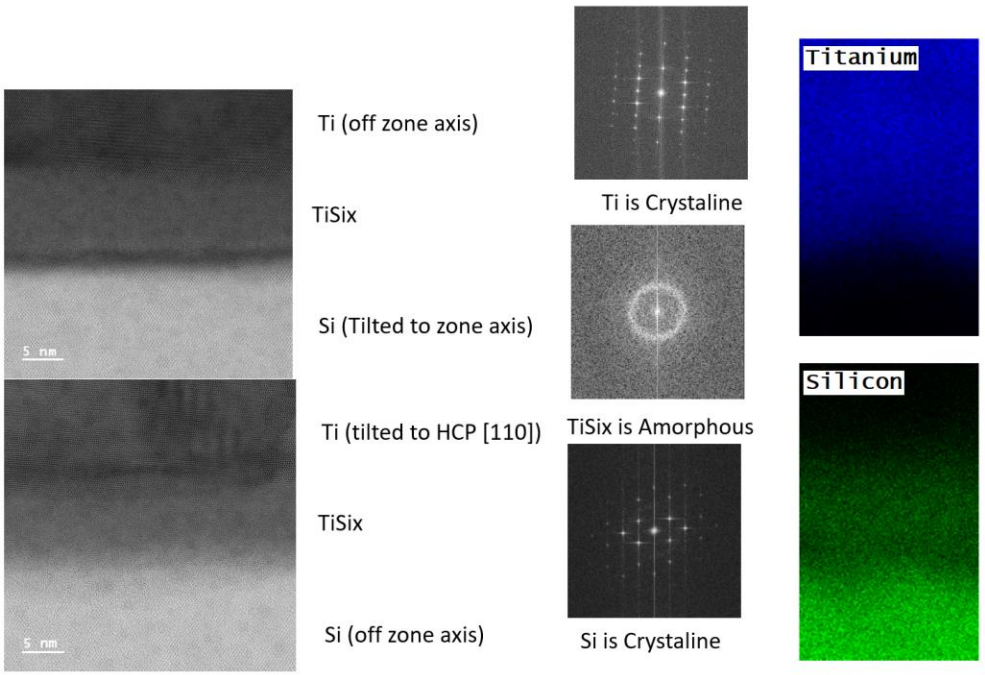

Figure 9. left: Cross-sectional TEM image of the Ti/$TiSi_\gamma$/Si heterojunctions; middle: Fast Fourier transform image of different regions; right: EDX maps.

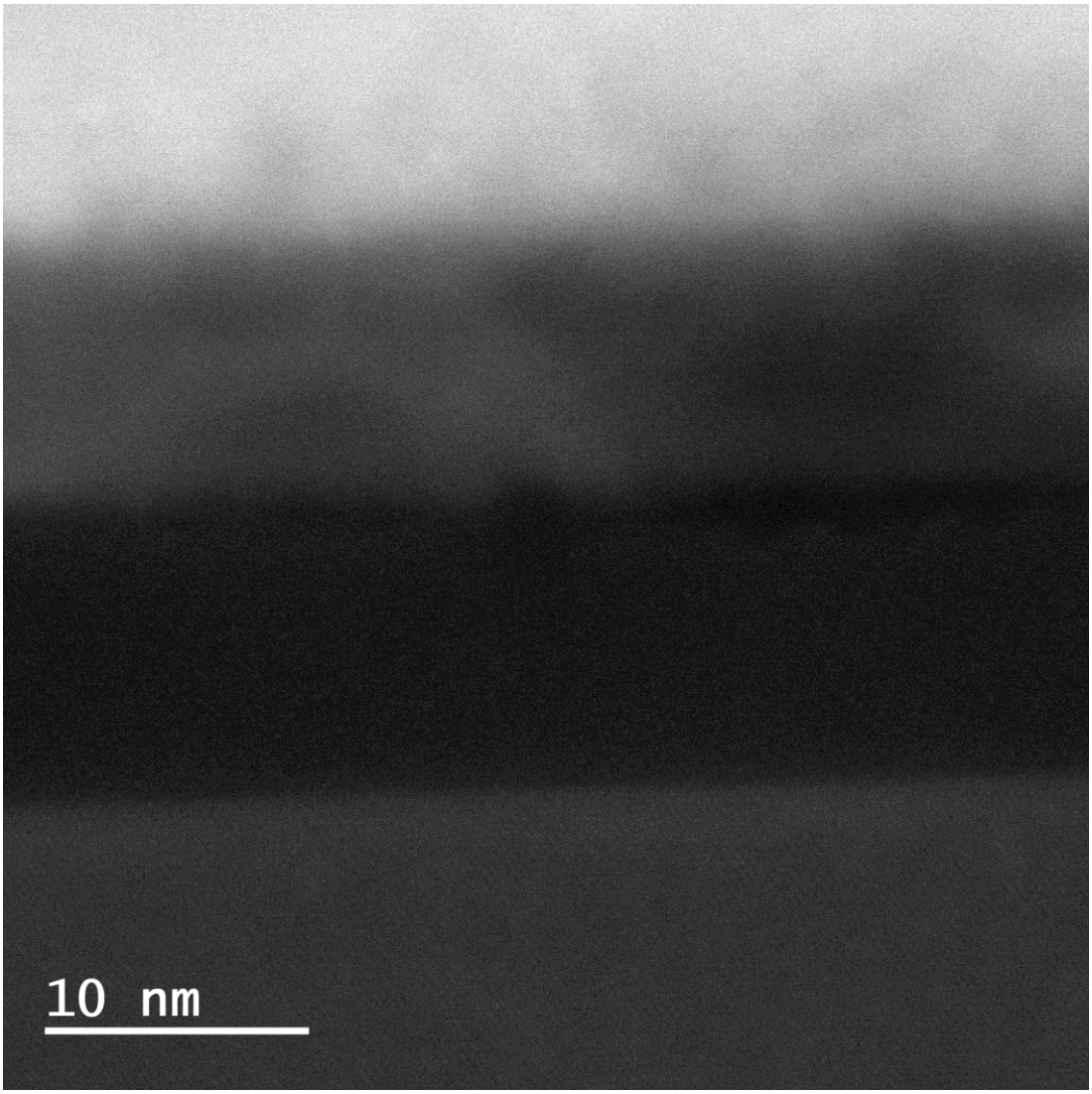

Figure 10. Cross-sectional TEM image of the $TiO_{\beta}/Al_2O_{\alpha}/SiO_2/Si$ heterojunction

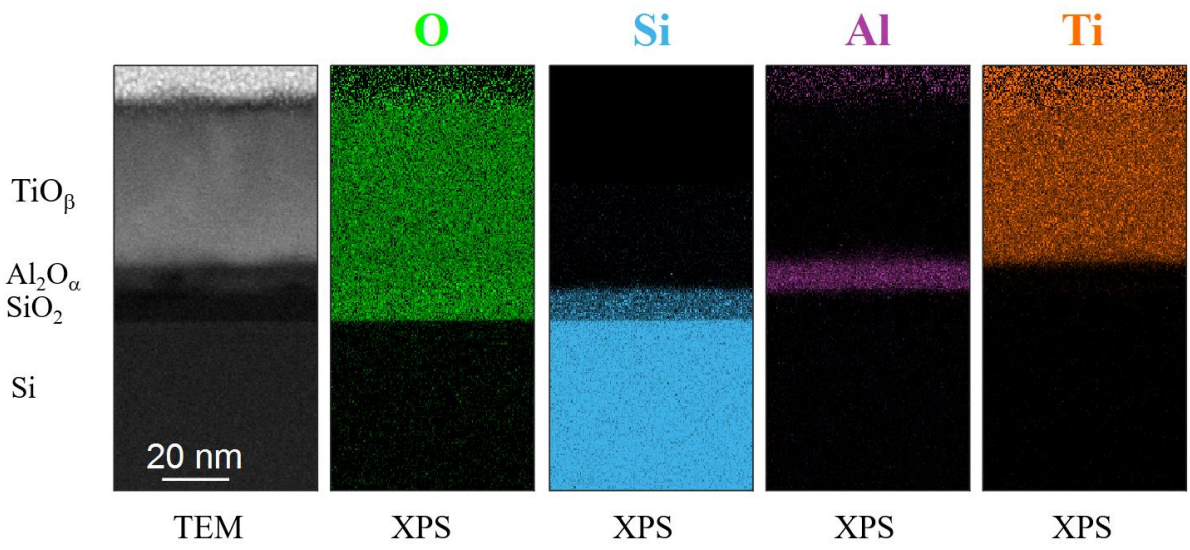

Figure 11. EDX map of the $TiO_\beta/Al_2O_\alpha/SiO_2/Si$ heterojunction.

2.2.1. Electronic Properties of Synstors

To emulate synapses grounded to the cerebrospinal fluid in neurobiological circuits, the reference electrodes of synstors were always grounded during the tests. Voltage pulses with an amplitude x were applied on the input electrode of a synstor to induce current I flowing through the synstor channel to its output electrode. **Figure 6a** shows a typical nonlinear $I - x$ relation, revealing the Schottky barriers between the Si channel and $TiSi_\gamma$ layers.[61] Each synstor in the circuit executes the signal-processing algorithm $I_n = w_{nm}(x)\, x_m$ (Equation 1). A synstor was tested by the application of 400 x and z pulses with a 1 ms width, and various amplitudes on the input and output electrodes, respectively. The typical percentage changes in w, $(\Delta w/w_0) \times 100\%$, are plotted against the amplitudes (**Figure 6b**) and numbers, and widths (**Figure 12**) of the applied pulses on the synstor. After the synstor experienced 400 paired x and z pulses with $x = z$, $\Delta w/w_0$

increased from 0 to 0.173 when the applied voltages decreased from 0 to $-2\ V$; $\Delta w/w_0$ decreased from 0 to -0.465 when x and z increased from 0 to $2.5\ V$. After the synstor experienced 400 individual x or z pulses, $|\Delta w/w(0)| < 0.023$ for $-2\ V \leq x \leq 2.5\ V$ and $z = 0\ V$, or $x = 0\ V$ and $-2\ V \leq z \leq 2.5\ V$. The magnitudes of the percentage changes in w were asymptotically increased with the increasing numbers or widths of the applied paired pulses with $x_m = z_n$ (**Figure 12**). Each synstor in the circuit executed the learning algorithm, $\dot{w}_{nm}(t) = \alpha\ z_n\ x_m$ (Equation 2) with the modification coefficient $\alpha \geq 0$ when $x_m = z_n \leq 0$, and $\alpha \leq 0$ when $x_m = z_n \geq 0$. When $z_n\ x_m = 0$, $\dot{w}_{nm} \approx 0$.

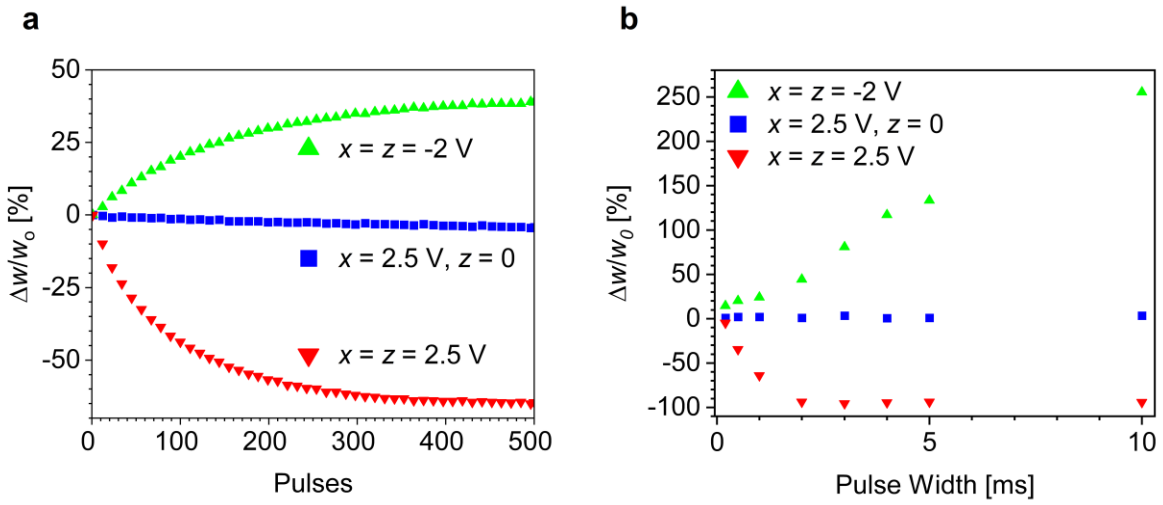

Figure 12. Pulse tests of a synstor. The percentage changes in the synstor conductance, $\Delta w/w_0$, induced by various x and z pulses with $x = z = -2\ V$ (green triangles), $x = z = 2.5\ V$ (red triangles), and $x = 2.5\ V$ and $z = 0$ (blue squares) are plotted against a) the numbers of applied pulses and b) the pulse widths after applying 400 pulses.

In a learning process, 400 synstors in the circuit were tuned to 167 targeted analog values, w_t, evenly separated by $\Delta w_t = 1.5\ nS$ by applying a train of 10 ms wide $2.5\ V$ x and $-2\ V$ z pulses

concurrently. The synstor conductance values were tuned to different targeted values, w_t (**Figure 6c**) with the differences between the average conductance values of the synstors and w_t, $|\overline{w} - w_t|$, less than $80\ pS$ or 5.3% of Δw_t (**Figure 6d**), and the standard deviations of w, σ_w, less than $180\ pS$ or 12% of Δw_t (**Figure 6e**), or $6\sigma_w$ less than 72% of Δw_t. To examine the nonvolatile memory of the synstors, they were modified to different initial analog conductance values, $w(0)$, in the range of $0 - 230\ nS$ by applying paired positive or negative pulses, and their conductance values, w, were measured versus time, t, over $10^6\ s$ at a temperature of $T = 40\ °C$ (**Figure 13a**) and $85\ °C$ (**Figure 13b**). As shown in **Figure 13a**, the variations of the memorized synstor conductance values over $1.2 \times 10^6\ s$ at $40\ °C$ were less than $\pm 6.0\ nS$. The $w\ vs.\ t$ data were fitted by $w(t) = w(\infty) + [w(0) - w(\infty)]e^{-\eta t}$ (solid lines) with $w(\infty)$ and η as the fitting parameters, where $\eta = \eta_0\ e^{-E_B/kT}$, E_B represents the potential barrier to trap electrons in the oxygen vacancies in the Al_2O_α layer,[62–64] and k denotes the Boltzmann constant.[59] From the average η determined from the synstors tested at $40\ °C$ and $85\ °C$, the average $E_B = 0.83\ eV$. The endurance testing of the synstors was performed by modification of their conductance to high and low values, w_H or w_L, by iterative application of paired positive or negative pulses. No deterioration in device conductance was observed over 1,200,000 modification cycles (**Figure 14**). The endurance, accuracy, and uniformity of the analog conductance modification of the Si and Al oxide based synstors were significantly superior to those of CNT-based synstors developed previously in our laboratory (**Figure 15**).[37]

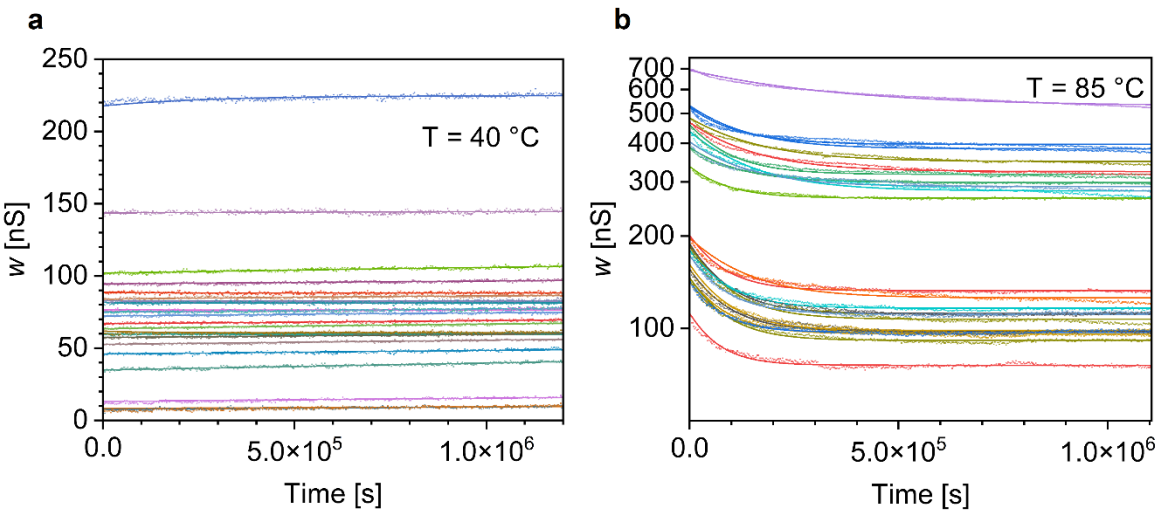

Figure 13. Memory testing of synstors. After synstors were modified to different initial analog conductance values, the conductance value of each synstor, $w(t)$, measured at a) $40\ °C$ and b) $85\ °C$ is shown by dots in different colors as a function of time t over $10^6\ s$, and is best-fitted by $w(t) = [w(0) - w(\infty)]e^{-kt} + w(\infty)$ (solid lines in different colors).

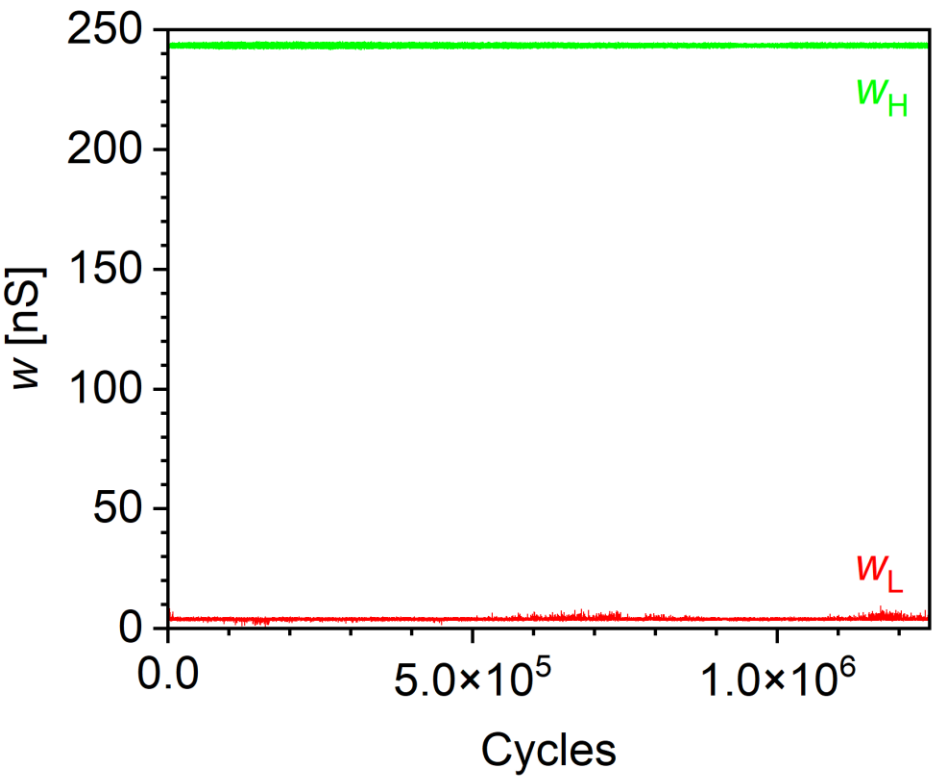

Figure 14. Endurance testing of synstor. A synstor was modified to its high and low conductance values, w_H and w_L, iteratively over 1,200,000 cycles. In each cycle, w_H and w_L are shown as green and red dots, respectively.

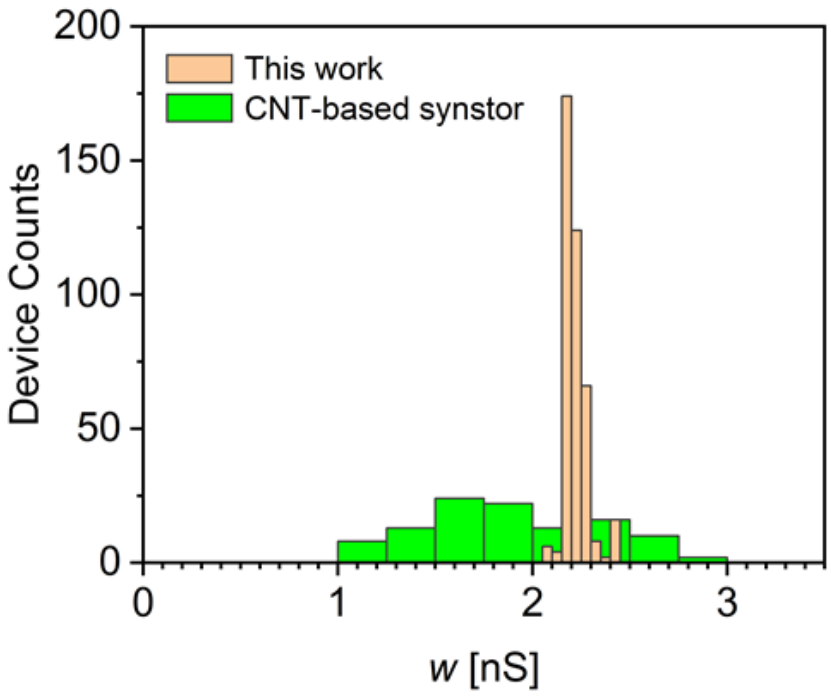

Figure 15. Variations of the conductance values of the CNT- and Si-based synstors. The distributions of conductance values $\boldsymbol{w}$ of the CNT-based synstors (green) and Si-based synstors (brown, this work) are plotted after the synstors were tuned toward their targeted conductance values.

The changes in charge density in the Al_2O_α layer of the synstor, $\Delta\rho$, induced by the voltage pulses with an amplitude,V_p, under $-2\,V \leq V_p \leq 2.5\,V$ were derived from capacitance-voltage tests[59,60] on a capacitor with the same $TiO_\beta/Al_2O_\alpha/SiO_2/Si$ heterojunction of the synstor (**Figure 5**), and shown in **Figure 6f**. After the capacitor experienced voltage pulses, $\Delta\rho$ increased from 0 to $3.83 \times 10^{11}\ /cm^2$ when V_p decreased from 0 to $-2\,V$, whereas $\Delta\rho$ decreased from 0 to $-5.49 \times 10^{11}\ /cm^2$ when V_p increased from 0 to $2.5\,V$ (**Figure 6f**). As shown in **Figure 16a** and **Figure 16b** in the simulated band structures of a synstor, when a pair of negative (or positive) x

and z voltage pulses with $x = z = -2.0\ V$ (or $+2.5\ V$) were applied, the electronic potential dropped across the Al_2O_α layer -0.41 e V (or 0.55 eV) was comparable to the the potential barrier for electronic hopping ($E_B = 0.83$ eV), thus resulting in depletion (or injection) of electrons through the Al_2O_α layer via Poole-Frenkel hopping among its oxygen vacancies,[66] decreasing (or increasing) the density of the electrons trapped in the oxygen vacancies in the Al_2O_α layer.[62–64] The trapped electrons in the Al_2O_α layer repelled the electrons in the Si channel; therefore, the w of a synstor experiencing x and z voltage pulses with $x = z = -2.0\ V$ (or $x = z = 2.5\ V$) increased (or decreased). The hopping rate of the electrons in the Al_2O_α layer, and thus $\dot{w}$, increased with the magnitude of the applied voltage (**Figure 6b**). As shown in **Figure 16c**, when a synstor experienced x and z voltage pulses with $x = 2.5\ V$ and $z = 0$, or $x = -2\ V$ and $z = 0$, the electronic potential mainly dropped across the Schottky junction between the Si channel and $TiSi_\gamma$ layer. The $TiO_\beta/Al_2O_\alpha/SiO_2$/Si heterojunction located beyond the Schottky junction, and the electronic potential dropped across the Al_2O_α layer (**Figure 16a**) was significantly smaller (0.058 e V or 0.097 eV) than the the potential barrier for electronic hopping ($E_B = 0.83$ eV), thus, could not drive a substantial number of electrons across the Al_2O_α layer. Therefore, these voltage pulses did not significantly modify the density of the electrons stored in the Al_2O_α layer, and no meaningful change in w was observed (**Figure 6b**).

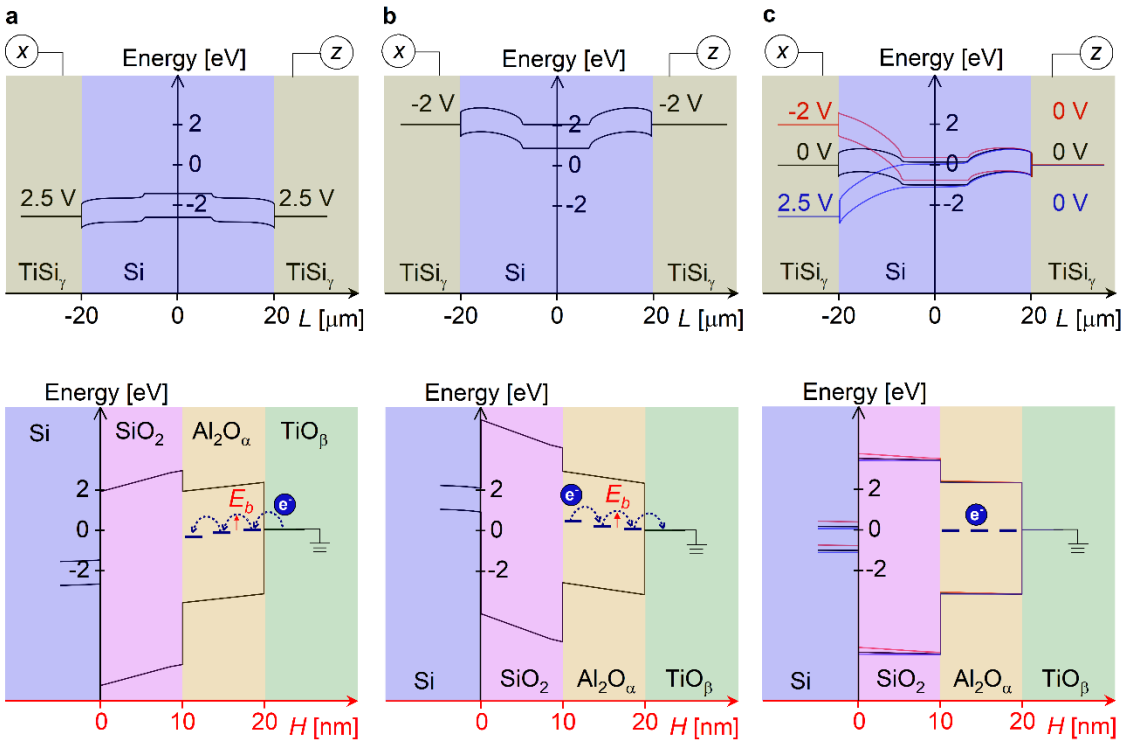

Figure 16. The energy-band structures of a synstor. The simulated energy-band structures of the $TiSi_\gamma$ input electrode, Si channel, and $TiSi_\gamma$ output electrode along the L axis in **Figure 3d** (top row), and of the grounded TiO_β reference electrode, Al_2O_α charge storage layer, SiO_2 dielectric layer, and Si channel along the H axis in **Figure 3d** (bottom row) of a synstor under the conditions of a) $x = z = 2.5\ V$, b) $x = z = -2\ V$, and c) the potential at the input electrode $x = -2\ V$, the potential at the output electrode $z = 0$ (red lines), $x = z = 0$ (black lines), and $x = 2.5\ V$ and $z = 0$ (blue lines). The scales represent the lateral distance, L, from the input electrode to the output electrode (top row), and the vertical distance, H, from the Si channel to the TiO_β reference electrode (bottom row) in **Figure 3d**. The band edges and Fermi levels are marked by solid lines. Electrons (e^-) are driven by a large potential difference across the Al_2O_α layer to hop across the oxygen vacancies in the Al_2O_α layer, as illustrated by blue dashed lines in a) and b) (bottom).

2.2.2. Simulation of a Synstor with a Channel Length of 30 nm

As shown in the simulated band structures of a synstor with a channel length of 30 nm in **Figure 7**, when a synstor experienced x and z voltage pulses with $x = 1\ V$ and $z = 0$, or $x = -1\ V$ and $z = 0$, when a synstor experienced x and z voltage pulses with $x = 2.5\ V$ and $z = 0$, or $x = -2\ V$ and $z = 0$, the electronic potential mainly dropped across the Schottky junction between the Si channel and $TiSi_\gamma$ layer. The $TiO_\beta/Al_2O_\alpha/SiO_2$/Si heterojunction located beyond the Schottky junction, and the electronic potential dropped across the Al_2O_α layer was significantly smaller (0.05 e V or 0.09 eV) than the the potential barrier for electronic hopping ($E_B = 0.83$ eV), thus, could not drive a substantial number of electrons across the Al_2O_α layer. Therefore, these voltage pulses did not modify the density of the electrons stored in the Al_2O_α layer, and no significant change in w was observed. When a pair of negative (or positive) x and z voltage pulses with $x = z = -0.7\ V$ or $1.2\ V$, the electronic potential dropped across the Al_2O_α layer (0.18 eV or -0.18 eV) was comparable to the the potential barrier for electronic hopping ($E_B = 0.83$ eV), thus resulting in depletion (or injection) of electrons through the Al_2O_α layer via Poole-Frenkel hopping its oxygen vacancies,[66] decreasing (or increasing) the density of the electrons trapped in the oxygen vacancies in the Al_2O_α layer[62–64] and at SiO_2/Al_2O_α interface. The trapped electrons in the Al_2O_α layer repelled the electrons in the inversed p-type Si channel; therefore, the w of a synstor experiencing x and z voltage pulses with $x = z = -0.7\ V$ (or $x = z = 1.2\ V$) increased (or decreased).

3. Synstor Circuit for Drone Control

This whole section has been derived from the published papers[14,37,67], which is used here to complete the dissertation.

3.1. Methods

3.1.1. A Drone Driven by a Synstor Circuit

A crossbar synstor circuit was integrated with sensors and actuators to drive a drone (**Figure 17a**, **Figure 18a**). The video images of a drone (Parrot AR Drone 2.0) from a camera (Genius WideCam F100) were processed with NI-IMAQ Toolkit for Cameras Virtual Instrument (VI). The deviations of the drone from its target position along four directions, $\boldsymbol{s}$, were converted to voltage pulses, $\boldsymbol{x}$, with an amplitude of $-2.2\ V$ or $1.1\ V$ and a duration of $10\ ns$, which were input to the synstor circuit with an analog signal input rate $f_i = 50\ MHz$ by a signal converter (FPGA, National Instruments, cRIO-9063). The firing rate of $\boldsymbol{x}$ pulses increased monotonically with increasing $\boldsymbol{s}$. $\boldsymbol{x}$ generated currents, $\boldsymbol{I} = \boldsymbol{w}\,\boldsymbol{x}$ (Equation 1), via the synstor circuit. The currents, $\boldsymbol{I}$, flowed through the synstors to interface circuits and triggered the actuation pulses, $\boldsymbol{y}$, to drive the drone, and backpropagating pulses, $\boldsymbol{z}$, to modify the synstor conductance matrix, $\boldsymbol{w}$, by following the learning algorithm, $\dot{\boldsymbol{w}} = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x}$ (Equation 2). Neuron circuits were designed and fabricated in our laboratory (a manuscript about the neuron circuits is in preparation) to emulate the functions of biological neurons. When $I < I_{th} \approx 30\ nA$, the threshold current of the neuron circuit, no actuation pulse was triggered. When $I_{th} \leq I \leq I_S \approx 60\ nA$, the saturation current of the neuron circuit, the average firing rate of the actuation pulses increased monotonically with increasing I. When $I > I_S$, the average firing rate of the actuation pulses was saturated at ~14 Hz. The output pulses from the neuron circuits were converted to actuation signals, $\boldsymbol{y}$, by a signal converter (FPGA, Xilinx, Kintex-7), and the drone was driven by the actuation signals transmitted via WiFi (**Figure 18**). In the learning process, the backpropagating pulses, $\boldsymbol{z}$, were triggered from the neuron circuits according to the following equation,

$$\boldsymbol{z} = \boldsymbol{y} * \tilde{\theta} \tag{3}$$

with $\tilde{\theta}(t) = \begin{cases} \theta(t - \tau_-) \\ -\theta(t + \tau_+) \end{cases}$, where $\tau_+ = 1180\ ms$ and $\tau_- = 20\ ms$, and $\theta(t) > 0$. The average $\tilde{\theta}$ over learning period T was $\langle \tilde{\theta} \rangle = 0$, and the average $\boldsymbol{z}$ over learning period, $\langle \boldsymbol{z} \rangle = 0$. For generate backpropagation of backpropagating pulses with $z_n = y_n * \tilde{\theta}$, when a y_n pulse is triggered at moment $t = t_n$ from the n^{th} output electrode of the synstor circuit, a train of positive (or negative) backpropagating pulses with a pulse firing rate proportional to θ within the time window $t_n - \tau_- < t < t_n - \tau_- + t_w$ at the n^{th} (or n^{th} complimentary) output electrode, and a train of negative (or positive) backpropagating pulses with a pulse firing rate proportional to θ within the time window $t_n + \tau_+ < t < t_n + \tau_+ + t_w$ were triggered at the nth (or n^{th} complimentary) output electrode, where t_w represents the duration of the time window of the pulse trains.

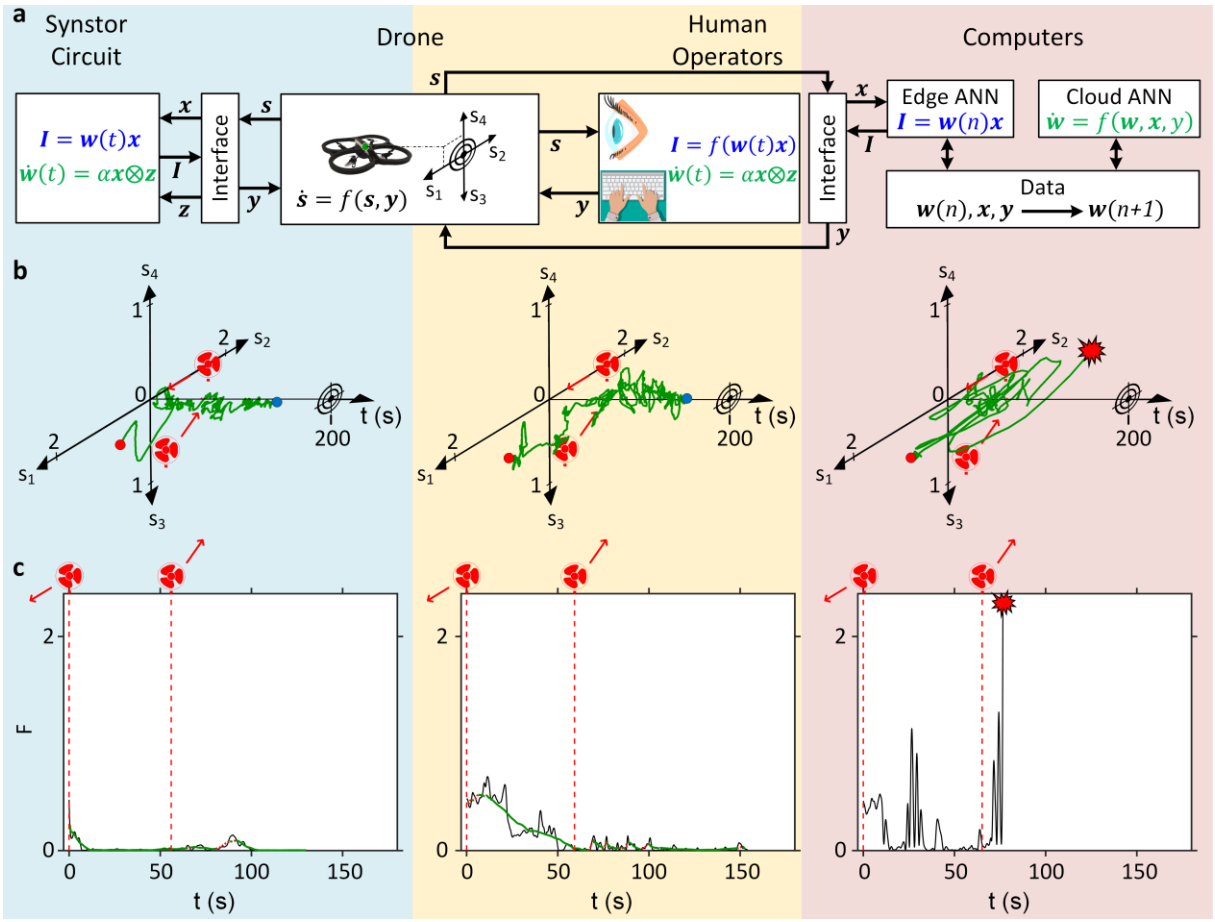

Figure 17. Drone flying experiments with **a,** a synstor circuit (left), a human operator (middle), and an ANN running in digital computers (right). In the real-time flights by the synstor circuit, the deviations of the drone position from its target position, $\boldsymbol{s}$, are detected by a camera, and converted by an interface circuit to voltage pulses, $\boldsymbol{x}$, which are input into a synstor circuit (Methods). By following the signal-processing algorithm $\boldsymbol{I = w\,x}$ (Equation 1) the synstors generate output currents, $\boldsymbol{I}$, that trigger actuation signals, $\boldsymbol{y}$, via an interface circuit to drive the drone and modify $\boldsymbol{s}$. Concurrently, backpropagating voltage pulses, $\boldsymbol{z}$, triggered at the output electrodes encounter the $\boldsymbol{x}$ pulses to modify the synstor conductance matrix, $\boldsymbol{w}$, by following the learning algorithm $\boldsymbol{\dot{w} = \alpha\, z \otimes x}$ (Equation 2). Human operators see the analog values of $\boldsymbol{s}$ displayed on a monitor, process and learn from the signals concurrently in neurobiological

circuits, and drive the drone by triggering actuation signals, $\boldsymbol{y}$, with different keys on a keyboard, but initially they do not know which key triggers which actuation signal. The ANN on an edge computer receives sensing signals, $\boldsymbol{s}$, and sends actuation pulses, $\boldsymbol{y}$, to drive the drone on the basis of the synaptic weight matrices $\boldsymbol{w}(n)$ in the ANN in its n^{th} round of trial. The data, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{w}(n)$ are saved in memory, and sent to a Google cloud computer, on which a reinforcement learning algorithm is executed to modify $\boldsymbol{w}(n)$ to $\boldsymbol{w}(n+1)$. Then $\boldsymbol{w}(n+1)$ is saved in memory, and sent back to the edge computer for the $(n+1)^{th}$ round of trial. **b,** The deviations of the drone position from its target position, $\boldsymbol{s}$**,** are shown as a function of time t**. c,** Objective functions, $F = \frac{1}{2}\boldsymbol{s}^2$ are displayed versus time t. For **b** and **c**, the data are for the experiments in which the drone was flown by the synstor circuit (left), a human controller (center), and the ANN (right). The average objective functions, $\langle F \rangle$, are shown versus t in green with $d\langle F \rangle/dt < 0$ induced by the learning and in red when $d\langle F \rangle/dt > 0$ induced by the environmental change. Dashed red lines show when the fans were turned on and arrows indicate the directions of the wind generated by fans. The red crash symbols indicate that the drone hit a wall.

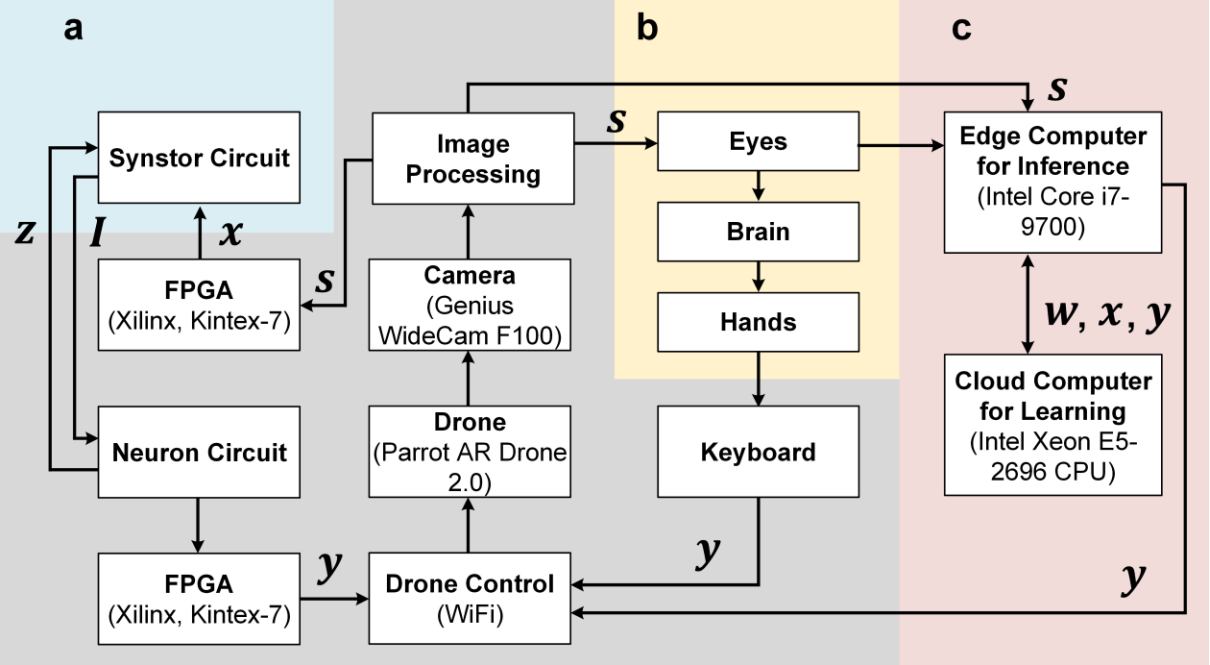

Figure 18. A schematic of experimental systems with a) a synstor circuit, b) a human operator, and c) an ANN running in digital computers.

3.1.2. Theoretical Analysis

We established a theoretical model for brain inspired AI systems based on synstor circuits that execute the signal-processing $\boldsymbol{I} = \boldsymbol{w}\,\boldsymbol{x}$ and learning $\dot{\boldsymbol{w}} = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x}$ concurrently[67]. When $\boldsymbol{s} = 0$, the performance function $F = \frac{1}{2}\boldsymbol{s}^2 = 0$, $\boldsymbol{x} = 0$, $\dot{\boldsymbol{w}} = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x} = 0$ (Equation 2), and $\boldsymbol{w}$ reaches the optimal steady value $\hat{\boldsymbol{w}} = arg\ \min_{\boldsymbol{w}} F$. When $\boldsymbol{w} \neq \hat{\boldsymbol{w}}$, $\boldsymbol{s} \neq 0$, and $F > 0$. In the learning process, the average performance function $\langle F \rangle = \frac{1}{2}\langle \boldsymbol{s}^2 \rangle$ changes with the following rate:

$$\langle \dot{F} \rangle = -\langle F \rangle / T_L + \delta F \qquad (4)$$

where $T_L > 0$ represents the learning time, and δF is associated with environment and the nonlinear terms of F. When the learning time T_L is decreased to satisfy $\delta F < \langle F \rangle / T_L$, then $\langle \dot{F} \rangle < 0$; $\langle F \rangle$ represents a Lyapunov function and is asymptotically decreased, thus resulting in modification of $\boldsymbol{w}$ toward $\widehat{\boldsymbol{w}}$ in the learning process. When $\langle F \rangle$ is decreased to $\langle F \rangle = F_e = T_L \langle \delta F \rangle$, then $\langle \dot{F} \rangle = 0$, and $\langle F \rangle$ reaches its optimal steady-state value F_e. The solution of Equation 3 gives the following:

$$\langle F(t) \rangle = (\langle F(0) \rangle - F_e)\, e^{-t/T_L} + F_e \qquad (5)$$

where $\delta F * e^{-t/T_L}$ represents the convolution between δF and e^{-t/T_L}.

3.1.3. Power Consumptions Analysis of the Synstor Circuit

When voltage pulses were applied on a crossbar synstor circuit, the average power consumption of the circuit (excluding the power consumptions of interface circuits),

$$P = \boldsymbol{I} \otimes \boldsymbol{x} = (\boldsymbol{w}\boldsymbol{x}) \otimes \boldsymbol{x} \approx w_T\, V_a^2 D_p \qquad (6)$$

where w_T denotes the total conductance of the synstors in the circuit, V_a denotes the magnitude of pulses, D_p denotes the average duty-cycle of the pulses. During the concurrent signal-processing

and learning in synstor circuit for the drone, $w_T \approx 152\ nS$, $V_a = 1.1\ V$, and $D_p = 0.087$, thus $P \approx 16\ nW$.

3.1.4. A Drone Driven by Humans

Real-time learning experiments were also performed with the drone driven by human operators under the same environment of the synstor circuit experiments with wind. The human operators (**Figure 18b**), who had no prior knowledge of the drone control system, sat behind a curtain without directly seeing the drone, but instead visually saw $\boldsymbol{s}$ values displayed on a computer monitor, and were instructed to minimize the $\boldsymbol{s}$ values. They drove the drone by triggering actuation pulses $\boldsymbol{y}$ with four keys in a keyboard, which were randomly associated with the different actuation signals y_n, and the firing rates of the pulses were proportional to the keystroke times (**Figure 17c** and **Figure 19b**).

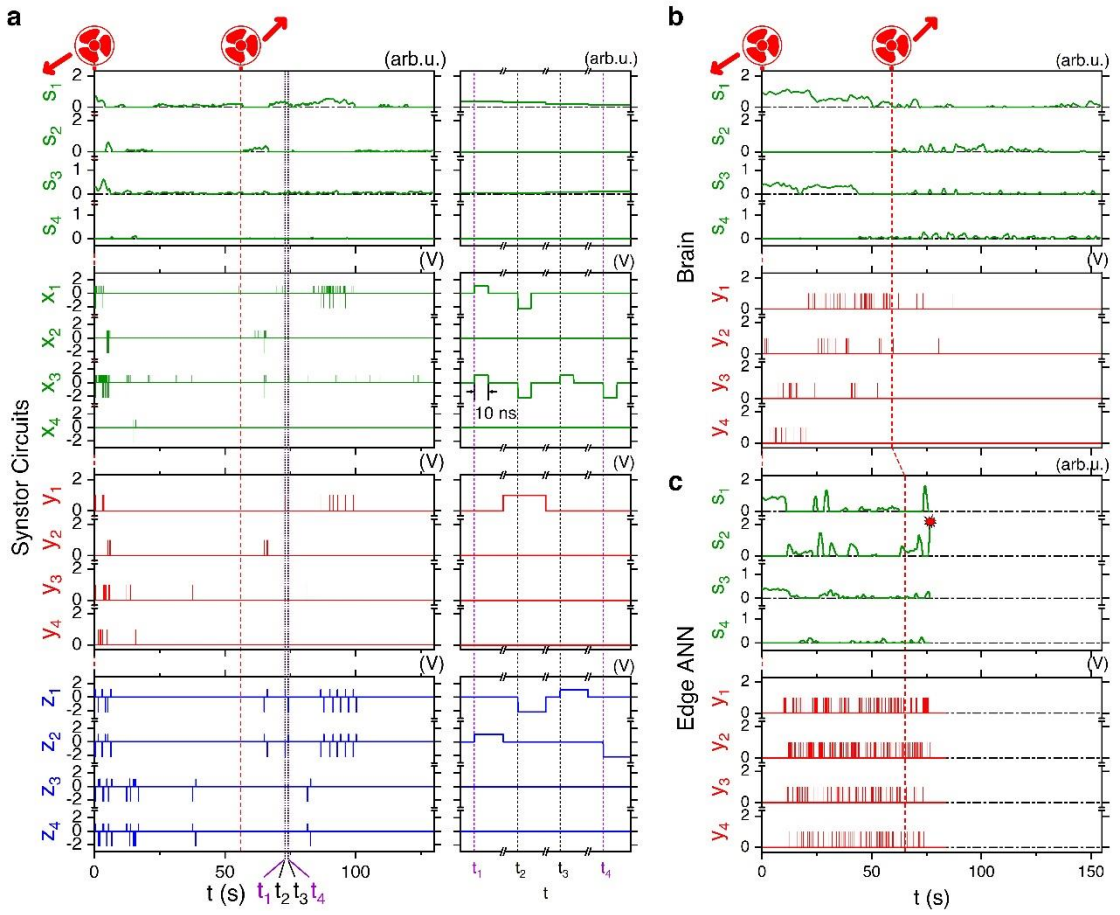

Figure 19. Experimental signals of synstor circuits, the human brain, and edge artificial neuronal network (ANN). a) In the experiment with a drone driven by a synstor circuit, the drone position s_m, input voltage pulses to a synstor circuit x_m, actuation voltage pulses to drive a drone y_n, and backpropagating pulses z_n are displayed versus time t (left) and at the moments t_1, t_2, t_3, and t_4 equal to 72.858, 73.019, 74.153, and 74.315 s, respectively (right). In the experiments with a drone driven by b) a human, and c) a computer, s_m and y_n are displayed versus time t. The arrows associated with the fan symbols indicate the changes in wind directions. The red crash symbol indicates that the drone crashed into the wall.

3.1.5. A Drone Driven by an ANN

A drone was driven by the ANN for signal-processing on an edge computer (**Figure 18c**) under the same environment as the synstor experiments with wind generated by fans or an environment without wind from fans. An onsite computer received the $\boldsymbol{s}$ signals from the interface circuit with an analog signal input rate $f_i = 50\ MHz$, the same as that in the trials of the synstor circuit, executed the signal-processing algorithm in an ANN with three layers of 137 neurons and 1152 synapses in real-time (**Figure 20**), and triggered actuation pulses, $\boldsymbol{y}$, to modify $\boldsymbol{s}$ (S8c). After the ANN controller with synaptic weight matrices $\boldsymbol{w}$ drove the drone for ~40 s (if the drone crashed with the walls, then flying time of the drone was less than 40 s), the data, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{w}$ were sent to a Google cloud computer[68] to execute a reinforcement learning algorithm[69] for ~$650\ s$ (if the flying time of the drone was less than 40 s, then the learning time was less than $650\ s$), and the modified $\boldsymbol{w}$ was sequentially sent back for the next round of signal-processing to drive the drone iteratively (**Figure 18a**). The ANN size, learning rate of 5×10^{-9} $(a.u.)$, and a discount factor of 0.999994 $(a.u.)$ were also optimized. After the Google cloud computer[68] with an Intel Xeon E5-2696 CPU executed the learning algorithm for $T_{LC} = 16020\ s$ according to the data with the drone driven by the ANN on the edge computer for $T_{IE} = 979\ s$ in an environment without wind from fans (**Figure 19c**), the $\boldsymbol{w}$ converged to the correlated synaptic weights, and F was minimized (**Figure 21f-j** and **Figure 22**). The average power consumption of the cloud computer for learning was $145\ W$[70], and the average power consumption of the edge computer for signal-processing was $65\ W$[71](excluding the power consumption of interface circuits).

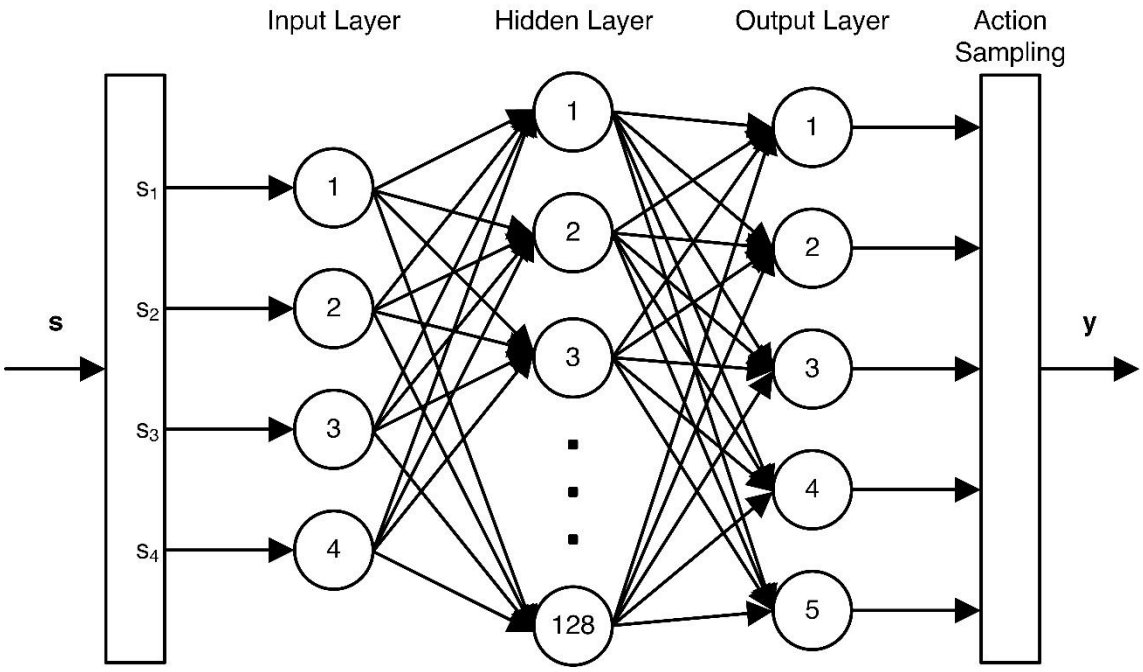

Figure 20. ANN structure. In the ANN, the sensory signals, $\boldsymbol{s}$, are transmitted to four input neurons, which send signals via synapses to 128 neurons in the hidden layer. The output signals from the neurons in the hidden layer trigger output signals from the five neurons in the output layer. The activation function in the neurons of the hidden layer is in rectified linear units (ReLU). The output signals from the neurons in the output layer are normalized with a softmax function and converted to a probability distribution of actuation signals, $\boldsymbol{y}$. The first four output signals correspond to the probability of triggering four y_n actions, and the last output signals correspond to no action. The actuation signal, $\boldsymbol{y}$, is sent out to drive the drone.

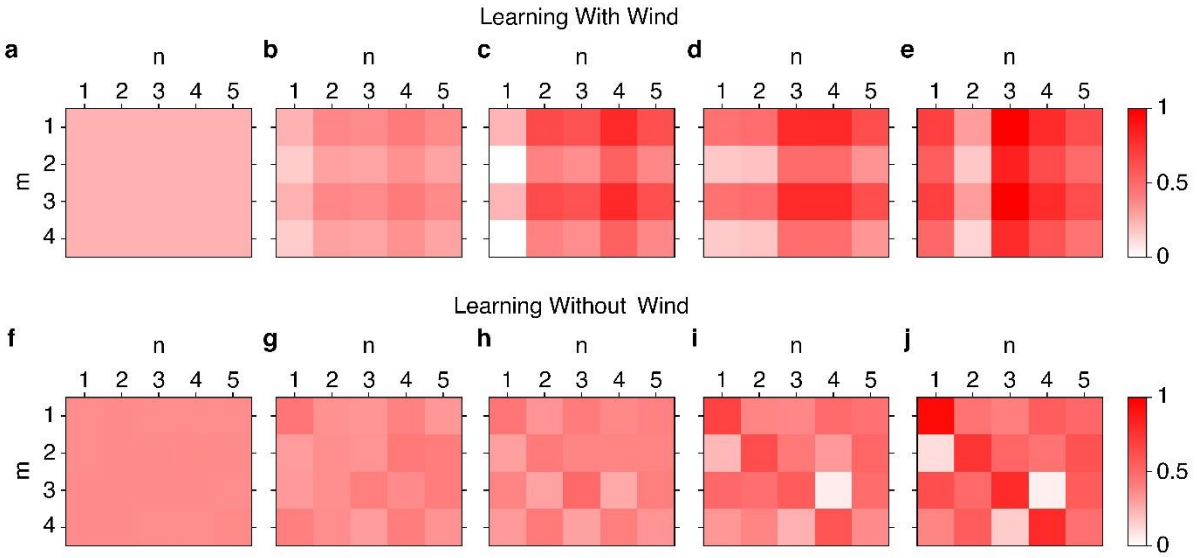

Figure 21. Heatmaps showing the evolution of synaptic weights in an ANN during learning. The product of synaptic weights $w_{nm} = \sum_{k=1}^{128} w_{nk} w_{km}$ is displayed on a normalized scale of $(w_{nm} - w_{min})/(w_{max} - w_{min})$ in the heatmaps and plotted versus the cumulative learning time a) $T_L = 0\ s$, b) $T_L = 7596\ s$, c) $T_L = 20900\ s$, d) $T_L = 37537\ s$, and e) $T_L = 54267\ s$ in the learning process from the data with the drone driven by the ANN in an environment with wind from fans, and f) $T_L = 0\ s$, g) $T_L = 9548\ s$, h) $T_L = 23162\ s$, i) $T_L = 47434\ s$, and j) $T_L = 55254\ s$ in the learning process from the data with the drone driven by the ANN in an environment without wind from fans. The w_{nm} denotes the synaptic weight of the synapses connecting the m^{th} neuron at the input layer and the n^{th} neuron at the output layer; w_{nk} denotes the weight of the synapse connecting the n^{th} neuron at the output layer and the k^{th} neuron at the hidden layer, w_{km} denotes the weight of the synapse connecting the m^{th} neuron at the input layer and the k^{th} neuron at the hidden layer; w_{min} denotes the minimal value of the effective synaptic weights, and w_{max} denotes the maximal value of the effective synaptic weights in the ANN.

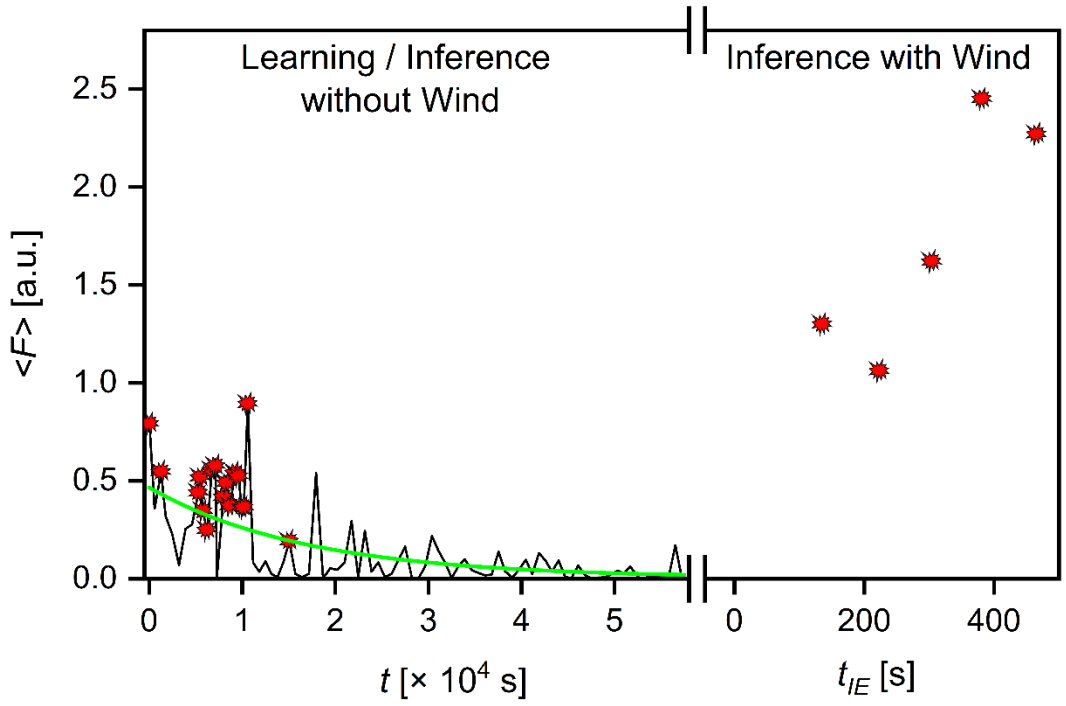

Figure 22. A performance function of a drone driven by an ANN on a computer. The average performance function $\langle F \rangle$ of a drone in each signal-processing period (points connected by black lines) is plotted versus the cumulative signal-processing and learning time t signal-processing with the drone driven in an environment without wind from the fans. The learning time $T_L = 17456\ s$ is extrapolated by best-fitting the experimental $\langle F \rangle - t_L$ curves with $\langle F(t) \rangle = (\langle F(0) \rangle - F_e)\ e^{-t_L/T_L} + F_e$ (green line, Methods, Equation 5). After the learning, the average performance function $\langle F \rangle$ is plotted versus accumulative signal-processing time T_{IE} with the drone driven in an environment with wind. The crash symbols indicate $\langle F \rangle$ values when the drone crashed into the top walls or sidewalls.

3.2. Results and Discussion

3.2.1. Concurrent Learning and Signal Processing to Fly a Drone

We performed experiments in which a synstor circuit, human operators, and a computer-based ANN learned to fly a drone toward a target position (**Figure 17a**). Two fans produced wind from opposite directions sequentially (**Figure 17b**) to generate an aerodynamically changing environment during the flight. The synstor circuit, human operators, and ANN had no prior learning experience or knowledge of the drone control system, and the synaptic weights of the synstor circuit and ANN were set to random values before each the experiments started. The deviations $\boldsymbol{s}$ of the drone from its target position were detected by a camera, and $\boldsymbol{s}$ was converted to voltage pulses, $\boldsymbol{x}$, by an interface circuit to generate currents via the synstor circuit according to the signal-processing algorithm, $\boldsymbol{I} = \boldsymbol{w}\,\boldsymbol{x}$. The currents triggered actuation signals, $\boldsymbol{y}$, via the interface circuit to drive the drone and change $\boldsymbol{s}$. Concurrently, voltage pulses, $\boldsymbol{z}$, were triggered at the output electrodes of the synstor circuit (Methods, Equation 4) to modify $\boldsymbol{w}$ according to the learning algorithm, $\dot{\boldsymbol{w}} = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x}$ (Equation 2 and **Figure 19a**). The human operators visually saw the analog values of $\boldsymbol{s}$ displayed on a monitor, processed and learnt from the signals concurrently in their neurobiological circuits, and drove the drone by triggering actuation signals, $\boldsymbol{y}$, with different keys on a keyboard, which were randomly associated with the different actuation signals to modify the $\boldsymbol{s}$ values (**Figure 17a** and **Figure 19b**). The synstor circuit and human operators flew the drones and concurrently learned how to drive the drone toward its target position, adapt to the changing environment, and minimize its performance (error) function $F = \frac{1}{2}\boldsymbol{s}^{\boldsymbol{2}}$ (**Figure 17b** and **Figure 17c**).

The synstor/synapses learning process was analyzed on the basis of a theoretical model that we published previously[67]. When the synstors/synapses concurrently execute the signal-processing and learning algorithms (Eqs. 1 and 2), the average $\langle F\rangle = \frac{1}{2}\langle \boldsymbol{s}^2\rangle$ changes with a rate, $\langle \dot{F}\rangle = -\langle F\rangle/T_L + \delta F$ (Methods, Equation 4), where T_L represents the learning time, and δF is associated with the environment and the nonlinear terms of F. The learning rate $\dot{\boldsymbol{w}} = \alpha\, \boldsymbol{z} \otimes \boldsymbol{x}$ can be increased by increasing the firing rates of the $\boldsymbol{x}$ and $\boldsymbol{z}$ pulses, shortening T_L to satisfy $-\langle F\rangle/T_L < \delta F$ and $\langle \dot{F}\rangle < 0$. When $\langle F\rangle$ is decreased to $F_e = T_L\langle \delta F\rangle$, then $\langle \dot{F}\rangle = 0$, and $\langle F\rangle$ reaches its optimal steady-state value F_e. Therefore, $\langle F(\boldsymbol{w})\rangle$ represents a Lyapunov function[72], and is asymptotically decreased. $\boldsymbol{w}$ does not need to be pre-programmed, and is spontaneously evolved toward the optimal matrix $\widehat{\boldsymbol{w}} = arg\ \min_{\boldsymbol{w}} F$ during the learning process, thus adapting to the environment by minimizing $\langle F\rangle$ toward F_e. The T_L and F_e are determined by fitting the experimental $\langle F\rangle - t$ curves with $\langle F(t)\rangle = (\langle F(0)\rangle - F_e)\, e^{-t/T_L} + F_e$, the solution of Equation 4 (Equation 5 and **Figure 17c**). Adaptability to the environment is represented by the successful rate to minimize $\langle F\rangle$ toward F_e in the experiments. The average T_L ($5.2\ s$), F_e (0.0043) and adaptability (100%) of the synstor circuit are significantly superior to the average T_L ($30.7\ s$), F_e (0.013), and adaptability (65%) of the human controllers (Table. 1).

Table 1. Comparison of signal-processing and learning processes, real-time learning functionality, adaptability to the changing environment with wind from fans, optimal steady performance function (F_e), learning time (T_L), and computing power consumption (P) of the synstor circuit, human operators, and computer-based ANN.

	Synstor Circuit	Human	Computers
Concurrent signal-processing and learning	Yes	Yes	No
Real-time learning	Yes	Yes	No
Adaptability to the changing environment	100%	65%	0%
Steady performance function $\langle F_e \rangle$ $(a.u.)$	0.0043	0.013	>2*
Learning time T_L (s)	5.2	30.7	16020**
Computing power consumption P (w) ***	1.6×10^{-8}	N.A.	65/145

*F_e>2 when the drone crashed into the wall in the environment with wind from fans.

**The learning time in an environment without wind from fans, Experimental section/Methods).

***The power consumption of the synstor circuit is for the concurrent execution of the signal-processing and learning algorithms, and the power consumptions of the computers are for the sequential execution of the signal-processing ($65\ W$) and learning ($145\ W$) algorithms (excluding the power consumption of interface circuits, Experimental section/Methods).

3.2.2. Benchmarking with Learning and Signal Processing by an ANN

Control experiments were also performed with the drone controlled by an ANN (**Figure 20**) on an edge computer, which received the $\boldsymbol{s}$ signals from the interface circuit, executed the signal-processing algorithm in the ANN, and triggered actuation pulses, $\boldsymbol{y}$, to modify $\boldsymbol{s}$ (**Figure 17a** and **Figure 19c**). After the ANN with synaptic weight matrices $\boldsymbol{w}(n)$ drove the drone for ~40 s in its n^{th} round of trial, the data, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{w}(n)$ were sent to a Google cloud computer[68] to execute a reinforcement learning algorithm[69] for ~$650\ s$, and the modified $\boldsymbol{w}(n+1)$ was sequentially sent

back for the $(n+1)^{th}$ round of signal-processing to drive the drone iteratively (**Figure 17a**). The time required to execute learning algorithm was significantly longer than that to execute the inference algorithm (**Figure 23**), and thus the learning latency precluded real-time learning by the computer. After the computers executed the signal-processing and learning algorithms sequentially for 90 rounds of trials by flying the drone in a stable environment without wind from fans, the random staring matrices $\boldsymbol{w}(1)$ was modified toward optimal matrices $\boldsymbol{w}(90)$, and F was minimized (**Figure 21f-j** and **Figure 22**). By fitting the experimental $\langle F \rangle - t$ curve (**Figure 22**), the computer learning time T_L was determined as $16{,}999\ s$, which was significantly longer than those of the synstor circuit ($5.2\ s$) and human operators ($30.7\ s$). Nevertheless, when the ANN with $\boldsymbol{w}(90)$ attempted to fly the drone in the same changing environment with the wind from the fans, as that of the synstor circuit and human operators, the drone always crashed into walls (**Figure 17c** and **Figure 22**).

The computers also executed the signal-processing and learning algorithms sequentially for 90 rounds of trials by flying the drone in the changing environment with wind from fans, the matrices could not be optimized, and the drone always crashed into walls (**Figure 21a-e**). Thus, the ANN adaptability to the changing environment with the wind from the fans was assigned as 0 (Table 1). The power consumption of the synstor circuit in concurrently executing the signal-processing and learning algorithms during the drone experiments (16 nW) was significantly superior to that of the computers in sequentially executing the signal-processing ($65\ W$) and learning ($145\ W$) algorithms (Table 1, Methods, Equation 6).

Although the synstors demonstrated here had feature sizes of micrometer scale, the device and circuit simulations show that a synstor can be scaled down to a channel length of 30 nm, and a crossbar circuit as shown in **Figure 3a** could be scaled up to 10^8 synstors with 10^4 input and

output channels (**Figure 7**). The large scale synstor circuits could execute signal-processing ($\boldsymbol{I} = \boldsymbol{w}\,\boldsymbol{x}$, Equation 1) and learning algorithms ($\dot{\boldsymbol{w}} = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x}$, Equation 2) concurrently with a computing speed of $25\ petaOPS$ ($peta\ operations\ per\ second$), a power consumption of $\sim 2.3\ mW$, while occupying an area of $\sim 0.8\ mm^2$ (Methods).

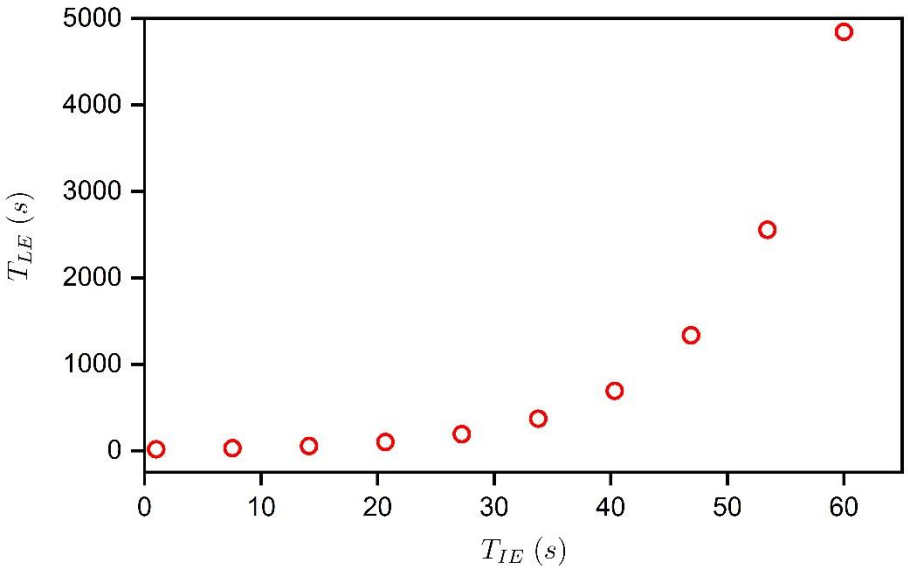

Figure 23. Learning and signal-processing times. The learning time, T_{LE}, to execute learning algorithm on a cloud computer, on the basis of the signal-processing data is shown against the signal-processing time, T_{IE}, to execute signal-processing algorithm on an edge computer.

3.2.3. Evolution of synaptic weights in an ANN during learning

After the computer executed the learning algorithm for 55267 s, on the basis of the data in the environment with the wind generated by the fans, as shown in **Figure 21a-e**, the synaptic weights $w_{nm} = \sum_{k=1}^{128} w_{nk} w_{km}$ was not converged to correlated synaptic weights w_{nm} with $m =$

n. After the computer executed the learning algorithm for 7596 s, on the basis of the data in the environment without wind generated by the fans, as shown in **Figure 21f-j**, the synaptic weights w_{nm} in the ANN gradually evolved from uncorrelated random values, and converged to correlated synaptic weights w_{nm} with $m = n$.

4. Conclusion

This section has been derived from the published paper[14], which is used here to complete the dissertation.

In summary, a crossbar circuit of synstors was fabricated by forming Schottky contacts between a Si channel/$TiSi_{0.7}$/Ti input and output electrodes laterally, and a heterojunction of the Si channel/SiO_2 dielectric layer/$Al_2O_{2.4}$ memory layer/$TiO_{1.5}$ metallic reference electrode vertically in each synstor. The synstor circuit concurrently executed the signal-processing ($\boldsymbol{I} = \boldsymbol{w}\,\boldsymbol{x}$, Equation 1) and learning ($\dot{\boldsymbol{w}} = \alpha\,\boldsymbol{z} \otimes \boldsymbol{x}$, Equation 2) algorithms in analog parallel mode.

Experiments with a drone driven toward a target position by the synstor circuit, human controllers, or an ANN on a computer were performed. Without any prior learning, the synstor circuit and human neurobiological circuits concurrently executed the signal-processing and learning algorithms in analog parallel mode, and their synaptic weight matrices $\boldsymbol{w}$ were self-evolved in the real-time learning processes to drive the drone toward its target position and adapt to the changing environment with wind from fans. The $\boldsymbol{w}$ matrices in the ANN were modified in its offline cloud learning processes, according to the signal-processing data $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{w}$. The average time of the sequential offline learning by the cloud computer ($16020\ s$) was significantly

longer than those of parallel real-time learning by the synstor circuit ($5.2\ s$) and humans ($30.7\ s$). The learning time of the synstor circuits (the computer) was shorter (longer) than the changing period of the environment ($\sim 60\ s$), therefore, the adaptability to the changing environment of the synstor circuit (100%) and humans (65%) was superior to that of the computer (0%), and the steady performance (error) functions, F_e, of the synstor circuit ($0.0043\ a.u.$) and human operator ($0.013\ a.u.$) was superior to that of the computers ($> 2\ a.u.$) in the changing environment. The power consumption of the synstor circuit in concurrently executing the signal-processing and learning algorithms during the drone experiments (16 nW) was significantly superior to that of the computers in sequentially executing the signal-processing ($65\ W$) and learning ($145\ W$) algorithms.

The fabrication processes of the Si-based synstor and transistor circuits are CMOS compatible. On the basis of the simulation, the synstor could be scaled down to a channel length of 30 nm, and a crossbar synstor circuit could be scaled up to 10^8 synstors with 10^4 input and output channels using the nanoscale fabrication techniques of the Si-based transistor circuit. Through application of backpropagating various $\boldsymbol{z}$ pulses, diverse learning algorithms in all three major machine learning paradigms (unsupervised, supervised, and reinforcement learnings)[67,73] can be executed in synstor circuits.

By circumventing sequential executions of the signal-processing and learning algorithms, high device conductance, digital computation, data movements between physically separated memory and logic circuits in computers, and data memory, synstor circuits can concurrently execute signal-processing and learning algorithms in the parallel analog mode, providing a path to establish artificial intelligent systems with a learning speed, performance, power consumption, and

adaptability to changing environments significantly superior to those of digital computers for versatile applications.

REFERENCES

1. Tian, Y. Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. *IEEE Access* **8**, 125731–125744 (2020).

2. Traore, B. B., Kamsu-Foguem, B. & Tangara, F. Deep convolution neural network for image recognition. *Ecol Inform* **48**, 257–268 (2018).

3. Vaswani, A. *et al. Attention Is All You Need*. (2017).

4. Boiko, D. A., Macknight, R. & Gomes, G. *Emergent Autonomous Scientific Research Capabilities of Large Language Models*. (2023).

5. Mocuta, A. *et al.* Enabling CMOS Scaling Towards 3nm and Beyond. in *2018 IEEE Symposium on VLSI Technology* 147–148 (2018). doi:10.1109/VLSIT.2018.8510683.

6. Moore, G. E. Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Society Newsletter* **11**, 33–35 (2006).

7. Semiconductor Research Corporation. SRC decadal plan 2020. (2020).

8. Zou, X., Xu, S., Chen, X., Yan, L. & Han, Y. Breaking the von Neumann bottleneck: architecture-level processing-in-memory technology. *Science China Information Sciences* **64**, (2021).

9. Shin, K. G. & Ramanathan, P. Real-time computing: a new discipline of computer science and engineering. *Proceedings of the IEEE* **82**, 6–24 (1994).

10. Yang, S. *et al.* Real-time neuromorphic system for large-scale conductance-based spiking neural networks. *IEEE Trans Cybern* **49**, 2490–2503 (2019).

11. Marković, D., Mizrahi, A., Querlioz, D. & Grollier, J. Physics for neuromorphic computing. *Nature Reviews Physics* vol. 2 499–510 Preprint at https://doi.org/10.1038/s42254-020-0208-2 (2020).

12. Do, T.-D., Duong, M.-T., Dang, Q.-V. & Le, M.-H. Real-Time Self-Driving Car Navigation Using Deep Neural Network. in *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)* 7–12 (IEEE, 2018). doi:10.1109/GTSD.2018.8595590.

13. Zanoni, F. D. & de Barros, E. A. A real-time navigation system for autonomous underwater vehicle. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* vol. 37 1111–1127 Preprint at https://doi.org/10.1007/s40430-014-0231-2 (2015).

14. Gao, D. *et al.* Synaptic Resistor Circuits Based on Al Oxide and Ti Silicide for Concurrent Learning and Signal Processing in Artificial Intelligence Systems. *Advanced Materials* **35**, (2023).

15. Kranthi, V. & Reddy, S. Stock Market Prediction Using Machine Learning. *International Research Journal of Engineering and Technology* 1032 (2008).

16. Albahli, S. *et al.* A Machine Learning Method for Prediction of Stock Market Using Real-Time Twitter Data. *Electronics (Switzerland)* **11**, (2022).

17. Koprinkova-Hristova, P., Mladenov, V. & Kasabov Editors, N. K. *Artificial Neural Networks*. vol. 4 (Springer International Publishing, Cham, 2015).

18. Siegelmann, H. T. Turing on Super-Turing and adaptivity. *Progress in Biophysics and Molecular Biology* vol. 113 117–126 Preprint at https://doi.org/10.1016/j.pbiomolbio.2013.03.013 (2013).

19. Wiedermann, J. Characterizing the super-Turing computing power and efficiency of classical fuzzy Turing machines. in *Theoretical Computer Science* vol. 317 61–69 (2004).

20. Turing, A. M. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* **s2-42**, 230–265 (1937).

21. Dongarra, J. *REPORT ON THE FUJITSU FUGAKU SYSTEM*. (2020).

22. Schneider, D. The Exascale Era is Upon Us: The Frontier supercomputer may be the first to reach 1,000,000,000,000,000,000 operations per second. *IEEE Spectr* **59**, 34–35 (2022).

23. Shaffer, C. M. *et al.* Self-Programming Synaptic Resistor Circuit for Intelligent Systems. *Advanced Intelligent Systems* **3**, (2021).

24. Wright, C. D., Hosseini, P. & Diosdado, J. A. V. Beyond von-neumann computing with nanoscale phase-change memory devices. *Adv Funct Mater* **23**, 2248–2254 (2013).

25. Sebastian, A., Le Gallo, M., Khaddam-Aljameh, R. & Eleftheriou, E. Memory devices and applications for in-memory computing. *Nature Nanotechnology* vol. 15 529–544 Preprint at https://doi.org/10.1038/s41565-020-0655-z (2020).

26. Verma, N. *et al.* In-Memory Computing: Advances and Prospects. *IEEE Solid-State Circuits Magazine* **11**, 43–55 (2019).

27. Yang, J. J., Strukov, D. B. & Stewart, D. R. Memristive devices for computing. *Nature Nanotechnology* vol. 8 13–24 Preprint at https://doi.org/10.1038/nnano.2012.240 (2013).

28. Rao, M. *et al.* Thousands of conductance levels in memristors integrated on CMOS. *Nature* **615**, 823–829 (2023).

29. Hu, M. *et al.* Memristor-Based Analog Computation and Neural Network Classification with a Dot Product Engine. *Advanced Materials* **30**, (2018).

30. Merced-Grafals, E. J., Dávila, N., Ge, N., Williams, R. S. & Strachan, J. P. Repeatable, accurate, and high speed multi-level programming of memristor 1T1R arrays for power efficient analog computing applications. *Nanotechnology* **27**, (2016).

31. Nathan, D. *et al.* Si-based self-programming neuromorphic integrated circuits for intelligent morphing wings. *J Compos Mater* **56**, 4561–4575 (2022).

32. von Bartheld, C. S., Bahney, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *Journal of Comparative Neurology* vol. 524 3865–3895 Preprint at https://doi.org/10.1002/cne.24040 (2016).

33. Zhang, J. Basic Neural Units of the Brain: Neurons, Synapses and Action Potential. (2019).

34. Ho, V. M., Lee, J.-A. & Martin, K. C. The Cell Biology of Synaptic Plasticity. *Science (1979)* **334**, 623–628 (2011).

35. Citri, A. & Malenka, R. C. Synaptic plasticity: Multiple forms, functions, and mechanisms. *Neuropsychopharmacology* vol. 33 18–41 Preprint at https://doi.org/10.1038/sj.npp.1301559 (2008).

36. Mihalaş, Ş. & Niebur, E. A Generalized Linear Integrate-and-Fire Neural Model Produces Diverse Spiking Behaviors. *Neural Comput* **21**, 704–718 (2009).

37. Danesh, C. D. *et al.* Synaptic Resistors for Concurrent Inference and Learning with High Energy Efficiency. *Advanced Materials* **31**, (2019).

38. Pei, J. *et al.* Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* **572**, 106–111 (2019).

39. Davies, M. *et al.* Loihi: A Neuromorphic Manycore Processor with On-Chip Learning. *IEEE Micro* **38**, 82–99 (2018).

40. Merolla, P. A. *et al.* A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science (1979)* **345**, 668–673 (2014).

41. Shawahna, A., Sait, S. M. & El-Maleh, A. FPGA-Based accelerators of deep learning networks for learning and classification: A review. *IEEE Access* vol. 7 7823–7859 Preprint at https://doi.org/10.1109/ACCESS.2018.2890150 (2019).

42. Eryilmaz, S. B. *et al.* Brain-like associative learning using a nanoscale non-volatile phase change synaptic device array. *Front Neurosci* (2014) doi:10.3389/fnins.2014.00205.

43. Ambrogio, S. *et al.* Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **558**, 60–67 (2018).

44. Goswami, S. *et al.* Decision trees within a molecular memristor. *Nature* **597**, 51–56 (2021).

45. Yao, P. *et al.* Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020).

46. Hu, M. *et al.* Memristor-Based Analog Computation and Neural Network Classification with a Dot Product Engine. *Advanced Materials* **30**, (2018).

47. Prezioso, M. *et al.* Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).

48. Diorio, C., Hasler, P., Minch, B. A. & Mead, C. A. *A Single-Transistor Silicon Synapse*. *IEEE TRANSACTIONS ON ELECTRON DEVICES* vol. 43 (1996).

49. Kim, K., Chen, C. L., Truong, Q., Shen, A. M. & Chen, Y. A carbon nanotube synapse with dynamic logic and learning. *Advanced Materials* **25**, 1693–1698 (2013).

50. Yang, J. J. *et al.* Memristive switching mechanism for metal/oxide/metal nanodevices. *Nat Nanotechnol* **3**, 429–433 (2008).

51. Yilmaz, Y. & Mazumder, P. A drift-tolerant read/write scheme for multilevel memristor memory. *IEEE Trans Nanotechnol* **16**, 1016–1027 (2017).

52. Diware, S. *et al.* Dynamic Detection and Mitigation of Read-disturb for Accurate Memristor-based Neural Networks. in 393–397 (Institute of Electrical and Electronics Engineers (IEEE), 2024). doi:10.1109/aicas59952.2024.10595966.

53. Gao, L., Ren, Q., Sun, J., Han, S. T. & Zhou, Y. Memristor modeling: Challenges in theories, simulations, and device variability. *Journal of Materials Chemistry C* vol. 9 16859–16884 Preprint at https://doi.org/10.1039/d1tc04201g (2021).

54. Altman, N. & Krzywinski, M. The curse(s) of dimensionality. *Nat Methods* **15**, 399–400 (2018).

55. Cao, K., Liu, Y., Meng, G. & Sun, Q. An Overview on Edge Computing Research. *IEEE Access* vol. 8 85714–85728 Preprint at https://doi.org/10.1109/ACCESS.2020.2991734 (2020).

56. Shi, L., Zheng, G., Tian, B., Dkhil, B. & Duan, C. Research progress on solutions to the sneak path issue in memristor crossbar arrays. *Nanoscale Advances* vol. 2 1811–1827 Preprint at https://doi.org/10.1039/d0na00100g (2020).

57. Demin, V. A., Surazhevsky, I. A., Emelyanov, A. V., Kashkarov, P. K. & Kovalchuk, M. V. Sneak, discharge, and leakage current issues in a high-dimensional 1T1M memristive crossbar. *J Comput Electron* **19**, 565–575 (2020).

58. Xu, M., Gao, D., Zheng, J. G. & Chen, Y. Quantification of Si, Al, Ti and O Composition in Si/Al Oxide Based Synaptic Resistor Circuits. *Microsc Microanal* **29**, 240–241 (2023).

59. Sze, S., Li, Y. & Ng, K. K. *Physics of Semiconductor Devices*. (2021).

60. Valik, L. *et al.* Distribution of fixed oxide charge in MOS structures with ALD grown Al2O3 studied by capacitance measurements. in *The Ninth International Conference on Advanced Semiconductor Devices and Mircosystems* 227–230 (IEEE, 2012). doi:10.1109/ASDAM.2012.6418526.

61. Liauh, H. R., Chen, M. C., Chen, J. F. & Chen, L. J. Electrical and microstructural characteristics of Ti contacts on (001)Si. *J Appl Phys* **74**, 2590–2597 (1993).

62. Linderälv, C., Lindman, A. & Erhart, P. A Unifying Perspective on Oxygen Vacancies in Wide Band Gap Oxides. *J Phys Chem Lett* **9**, 222–228 (2018).

63. Dicks, O. A., Cottom, J., Shluger, A. L. & Afanas'ev, V. V. The origin of negative charging in amorphous Al2O3 films: the role of native defects. *Nanotechnology* **30**, 205201 (2019).

64. Khosla, R. *et al.* Charge Trapping Analysis of Metal/Al2O3/SiO2/Si, Gate Stack for Emerging Embedded Memories. *IEEE Transactions on Device and Materials Reliability* **17**, 80–89 (2017).

65. Arif, A. F. *et al.* Highly conductive nano-sized Magnéli phases titanium oxide (TiOx). *Sci Rep* **7**, (2017).

66. Liu, D., Clark, S. J. & Robertson, J. Oxygen vacancy levels and electron transport in Al2O3. *Appl Phys Lett* **96**, 032905 (2010).

67. Chen, Y. Brain-Inspired Synaptic Resistor Circuits for Self-Programming Intelligent Systems. *Advanced Intelligent Systems* **3**, (2021).

68. Bisong, E. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (Apress, 2019). doi:10.1007/978-1-4842-4470-8.

69. Montague, P. R. Reinforcement Learning: An Introduction, by Sutton, R.S. and Barto, A.G. *Trends Cogn Sci* **3**, 360 (1999).

70. Intel Xeon Processor E5-2699. https://ark.intel.com/content/www/us/en/ark/products/91317/intel-xeon-processor-e52699-v4-55m-cache-2-20-ghz.html.

71. Intel Core i7-9700 CPU. https://www.cpubenchmark.net/cpu.php?cpu=Intel+Core+i7-9700+%40+3.00GHz&id=3477.

72. Khalil, H. *Control Systems, Robotics and AutomatioN*. vol. 12 (2009).

73. Chen, Z., Haykin, S., Eggermout, J. j. & Becker, S. *Correlative Learning: A Basis for Brain and Adaptive Systems*. vol. 1 (2007).