TWO-DIMENSIONAL MATERIAL-BASED ENERGY-EFFICIENT ELECTRONIC DEVICES TOWARDS NEUROMORPHIC COMPUTING

CHEN HONGYE

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Applied Physics

Two-Dimensional Material-Based Energy-Efficient Electronic Devices Towards Neuromorphic Computing

Chen Hongye

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

August 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that the thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the next.

_____(Signed)

CHEN Hongye (Name of student)

Abstract

The human brain is considered a highly parallel and energy-efficient computing system, able to process complex cognitive tasks. Compared to the human brain, artificial neural networks (ANN) based on transistors and random access-resistive memories (RRAMs) face critical challenges with energy efficiency in the way of further scaling of neural network. Two-dimensional (2D) materials, including transition metal dichalcogenides (TMDs), hexagonal boron nitride (h-BN), and $CuInP_2S_6$, have been regarded as one of the potential candidates for low-power and high-performance electronic devices in neuromorphic computing. Differing from bulk materials, the atomically thin 2D materials could achieve low-temperature fabrication, low-power switching, and electrostatic gate tunability. This thesis focuses on enhancing energy efficiency through the design and demonstration of devices utilizing 2D materials, including (1) selectors to suppress leakage current in RRAM array; (2) dual-gate transistor based on high-k dielectric as dendristor to calculate the synaptic input and achieve a complete dendritic artificial neuron network.

The first work focuses on the suppression of leakage current in RRAM array. The RRAM array exhibits great potential for information storage and computing due to its small cell area (~ $4F^2$). Nevertheless, in the RRAM array, there exists the sneak current through unselected low resistance state cells when the unselected word lines and bit lines are biased with half of the operation voltage in the V/2 scheme, which will contribute to energy consumption and useless data. To address this issue, we introduce the two-terminal highly nonlinear $MoS_2/WSe_2/MoS_2$ n-p-n selector to suppress the sneak current, which performs a high current density of 2×10^3 A cm^{-2} due to the low Schottky barrier height of Au/MoS_2

while maintaining a high nonlinearity of above 200 based on the punch-through mechanism. Furthermore, we successfully integrate this n-p-n selector with bipolar h-BN memory and demonstrate a two-terminal all 2D material-based 1S1R architecture, whose maximum crossbar size was estimated to be 6.5 Kbit. This work contributes to a framework for advancing 3D crossbar array memory devices.

Secondly, the artificial dendrite device was proposed to complete dendritic artificial neuron model through the nonlinear computation ability of dendrite. Neuromorphic computation in ANN typically uses a point neuron model, which only consists of synapses and soma and ignores the computational function of dendrites. In biology, dendrites nonlinearly integrate the synaptic inputs and control the somatic membrane potential, which plays a key role in the accuracy and energy-efficiency improvement of neuron network. Therefore, a more complete neural network with artificial dendrite has been proposed. The $Cu_{0.67}Ag_{0.33}InP_2S_6$ (CAIPS) as 2D high-k dielectric and MoS_2 as channel was used in the gate-all-around transistor, termed as dendristor. The adaptation to van der Waals layered materials and dual-gate design enables the dendristor with high performance, resulting from the clean interface between CAIPS and MoS_2. Meanwhile, this device performs superlinear/sublinear integration of synaptic input and synaptic plasticity with low energy consumption. We further introduce the dendristor device into the neuron system and achieve multisensory perception.

In conclusion, we investigate the 2D material-based energy-efficient electronic devices towards neuromorphic computing. With the development of modern electronics with high performance, our work will provide new insight into the development of energy-efficient electronic devices for neuromorphic computing.

Publications

1. **H.Y. Chen**, T. Q. Wan, Y. Zhou, J.M. Yan, C.S. Chen, Z.H. Xu, S.G. Zhang, Y. Zhu, H.Y. Yu*, and Y. Chai*, "Highly nonlinear memory selectors with ultrathin MoS2/WSe2/MoS2 heterojunction," Advanced Functional Materials, vol. 34, no. 15, p. 2304242, 2024.

Acknowledgments

My four-year Ph.D. journey passed quickly, which was an unforgettable and significant experience in my life. During the four years at the Hong Kong Polytechnic University, I have deepened my knowledge and expertise in my field of study, developed an international perspective, and cultivated the competence and confidence required for the next chapter of my life. I am sincerely grateful to every supervisor, engineer, and lab-mate who supported me throughout these four years, and I appreciate their contributions wholeheartedly.

I would like to express my special gratitude to my supervisors, Prof. Yang Chai from the Hong Kong Polytechnic University and Prof. Hongyu Yu from the Southern University of Science and Technology. I could not complete my Ph.D. study without their profound scientific knowledge and insightful scientific perspectives. It is truly a great honor to be their student, and what I have learned from them is an invaluable treasure in my life. Not only have they provided me with excellent research environment, but they have also taught me valuable life lessons through their words and actions. I sincerely extend my deepest respect and gratitude to them.

I am immensely grateful to the colleagues who have accompanied me throughout my journey. I would like to extend my special thanks to Mr. Jianmin Yan, Dr. Guang Zeng, Dr. Yue Zhou, Dr. Fuyou Liao, Dr. Xi Zhang, Ms Tianqing Wan, Mr. Zhihang Xu, Mr. Changsheng Chen. They gave kind help and important advice in device fabrication, material preparation, and device measurement, respectively. Also, I give gratitude to Mr. Duanzijing Liu, Dr. Jiaqi He, Mr. Songge Zhang, Ms. Sijie Ma, Mr. Jiaxin Fan, Ms. Ziyin

Huang, and Ms. Xinyu Zhang for their kind help and assistance. I am grateful for their unwavering support and help. I also thank the office staff of the PolyU and Sustech for their professional service and support throughout my study life.

I would like to express my heartfelt gratitude and appreciation to my parents and family for their unwavering support and endless love throughout my journey. Throughout my study journey, I was influenced unconsciously by my father's optimistic attitude and his diligence and surrounded by my mother's selfless love and unwavering support, which motivated me to persevere even in tough time. They always are my strongest spiritual pillar. I would like to express my gratitude to everyone and every experience that I encountered during my growth journey. Last, I want to thank myself for being positive, optimistic, and resilient.

Table of Contents

List of Figures

Chapter 1 Introduction

1.1 Background

In the past decade, great interest has focused on deep learning of neural networks, which play a crucial role in various fields such as image object detection and recognition, natural language processing, speech recognition, and face verification.[1] Among these applications, the memory-access efficiency of computing systems failed to support data-centric workloads. The requirement of future applications in computing and deep learning overcome the process capabilities of existing electronic systems.

To enhance the energy efficiency, several strategies were proposed. One of them is the monolithic three-dimensional (3D) integration due to its shorter interconnects and reduced footprint, which presents high potential in the scaling electronic system.[2] The 1S1R as one of upper tiers in monolithic 3D integration requires highly nonlinear selector to suppress the leakage current and improve energy efficiency. In addition, there still exists a significant gap between deep neural networks and human-level intelligence. Dendrites acting as independent nonlinear processing and signaling units in biological neural networks easily be neglected in artificial neural networks.[3] The artificial dendrite device allows multi-layer information processing and coincidence detection within a single neuron, which exhibits more advantages in energy consumption and recognition accuracy.

As the transistor and RRAM scale down to nanometer, the ANN faces significant challenges in achieving low energy efficiency and device scaling simultaneously. For transistors, the short-channel effects are amplified as the channel length decreases, which leads to increased off current and lower drain-induced barriers.[4] In addition, as the

dielectric becomes thinner, the quantum mechanical effect contributes to the gate leakage current.[5] For RRAM, the reliability of devices was reduced due to the amplified effect of defect.[6] Therefore, as the devices scale down, due to traditional bulk material limitations, energy consumption will increase. To address these issues, my research explores the potential of two-dimensional (2D) materials, which consist of a single layer or a few layers of atoms arranged in a 2D plane that exhibit unique and remarkable properties in a wide range of applications, such as flexible electronics, sensors, energy storage, and catalysis. The non-dangling bond surface of the 2D materials allows for top-down preparation of ultrathin 2D materials from their bulk counterpart by mechanical exfoliation, as well as construction of complex heterojunctions by stacking different 2D materials on top of each other in an arbitrary order. In addition, due to the non-dangling bond interface, the 2D materials decrease the scattering in the transistor. Also, their excellent electrostatic exhibits good gate control and small subthreshold, which leads to low energy consumption.[7]

1.2 Two-dimensional Material

Two-dimensional (2D) materials with a single layer or a few layers of atoms arranged in a 2D plane exhibit exceptional properties that differ from their bulk counterparts.[8,9] The bulk semiconductors have surface states within the band gap, which arise from the termination of crystal periodicity or the metal-induced gap states resulting from the chemical bonding between metal and semiconductor. These surface states or gap states can trap charges, lead to band bending, and change Schottky barrier height (SBH) regardless of metals with different work functions. This phenomenon is commonly known as Fermi-level pinning. Due to the Fermi level pinning, the unwanted high Schottky barrier significantly increases the contact resistance and affects the efficient carrier injection.

However, the presence of non-dangling bonds on the surface of 2D semiconductors allows for the elimination of Fermi level pinning when the exfoliated 2D layers are dry transferred on metal.[10,11] Based on the above advantages, 2D materials are widely exploited for potential applications in memories, capacitors, sensors, energy storage, and catalysis.[4,12,13]

1.2.1 Transition Metal Dichalcogenides

Transition metal dichalcogenides (TMDs), layered 2D semiconductors that consist of chalcogen atoms and metal atoms, have been widely explored (**Figure 1.1a, b**).[14] Due to the absence of dangling bonds, TMD can be exfoliated into a few layers easily and integrated with other materials. TMD as channel can be grown into atomic scale and maintain electrostatics. Moreover, unlike bulk material, the TMD maintains high mobility even when the thickness decreases.

Moreover, the layer-dependent properties of TMDs have recently attracted a great deal of attention. For example, in several semiconducting TMDs, there is a transition from an indirect band gap in the bulk to a direct gap in the monolayer: for WSe_2 the bulk indirect bandgap of 1.33 eV increases to a direct bandgap of 1.65 eV in single-layer form (**Figure 1.1c**). Most TMD materials such as MoS_2 and WS_2 are n-type semiconductors, but WSe_2 features a p-type charge carrier which is one of the most frequently used p-type 2D semiconducting materials upon contact with metal. Thereby, WSe_2 has shown great interest in electronic engineering for device design.

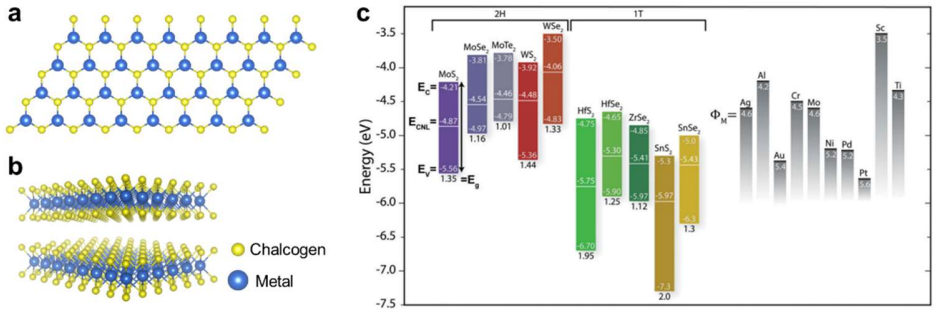


Figure 1.1 (a) Top view and (b) side view of 2D TMD material. Planes of transition metal atoms (blue) are sandwiched between and are bonded to 6 chalcogen (yellow) atoms. (c) Band alignment of various 2D materials and elemental contact metals. Schematic of bandgap, electron affinity, ionization potential, and charge neutrality level (ECNL) values for multilayer 2D materials along with the work functions (ΦM) of contact metals.[15] The electron affinity and ionization potentials are given by the energy of the conduction band minima, EC, and valence band maxima, EV, respectively versus the vacuum level. ECNL is the energy at which metal-induced gap states are predicted to form. Note that the band alignment between the metals and 2D materials does not take into account any Fermi-level pinning. The metals' work functions are either polycrystalline or averaged values for different crystal planes.

In the case of ultrathin WSe_2, besides the conventional CMOS-compatible process, the surface charge transfer has been successfully employed as an alternative approach to manipulate the carrier concentration. One interesting research by Zhou et al. utilizing the layer-dependent transport behaviors of WSe_2 provided another dimension for carrier modulation. After contacted with evaporated Ni electrode, the transport characteristics can change from p-type (less than 4.5 nm), ambipolar (6.5 nm) to n-type (20 nm) with the

increase of WSe_2's thickness.[16] Besides, WSe_2-based field effect transistor with deposited Au electrodes presents tunable transport behavior with the thickness. Further research has shown that low work function (WF) metals, such as Ti, Ag, and In, can achieve small n-type SBH with monolayer WSe_2.[17] This phenomenon results from the SBH, which comes from the interfaces of metal and semiconductors and varies between metals with different WF and the electron affinity of semiconductor.

In addition, the fabrication method of metal electrodes on WSe_2 also influences the transport behavior of WSe_2. Duan et al. manifested that WSe_2 exhibits p-type when it is integrated on transferred Au electrodes due to clean interface, while performing n-type with evaporated Au electrodes.[18] This distinction arises from the Fermi level pinning wherein the active Au ions can destroy the surface of WSe_2 and lead to the dangling bond. Although the dry transfer method of metal electrodes offers an interface without chemical bonding and disorder, it induces the formation of vdWs gap, which significantly increases contact resistance. To address this issue, one approach involves inserting oxides or h-BN layers as the tunneling layer, effectively reducing the contact resistance. In contrast to oxide tunneling layers, the thickness of the h-BN layer can be precisely controlled by the layer numbers.[19] The edge contact presents an alternative approach to eliminate the effects of the vdWs gap in metal/semiconductor interfaces. Bulk metals are directly brought into contact with the edge of 2D TMDs minimizing the van der Waals (vdWss) gap, or even forming the covalent bonds.[20] Realizing the logic circuits based on TMD material requires controllable doping and contact engineering. It remains an open question to extend conventional doping techniques to ultrathin 2D semiconductors.

1.2.2 Hexagonal Boron Nitride

Hexagonal boron nitride (h-BN) is a layered material whose structure is similar to graphene.[21] 2D h-BN flake can be prepared from top to bottom using mechanical or liquid methods or grown from bottom to top using traditional processes such as chemical vapor deposition (CVD) and magnetron sputtering.[22,23]

Due to the strong thermal and chemical stability of h-BN, h-BN is commonly used as insulation layer or gate dielectric.[24] For metal/insulation layer/metal (MIM) structure memristor, h-BN with atomic thickness as the resistive switching medium is less prone to dielectric breakdown and has low leakage current and low operating voltage, which can achieve low power consumption.[25] Moreover, the h-BN is compatible with traditional processes, which gives the potential to achieve small size and large array of memristors. In addition, h-BN as a tunneling layer can reduce the Schottky barrier height of metal/semiconductor and achieve high mobility field-effect transistors. [26,27]

1.2.3 2D Dielectric Materials

The ferroelectric materials exhibit a spontaneous and macroscopic polarization, which can be switched by an external electric field below the Curie temperature (T_c).[28] T_c is the temperature at which a ferroelectric material undergoes a phase transition from a ferroelectric phase to a paraelectric phase. The specific Curie temperature depends on the composition of the ferroelectric material. Among all the 2D ferroelectric materials, $CuInP_2S_6$ (CIPS) stands out as a representative material due to its room-temperature ferroelectricity. In addition, the out-of-plane ferroelectricity and ionic conductivity of CIPS originate from the spatial instability of the Cu^+ cation.

CIPS belongs to the family of transition metal thio/selenophosphates (TPS), a broad class of vdWss layered solids in which metal cations are incorporated within the lattice structure of thiophosphate $(P_2S_6)^{4-}$ anions. As illustrated in **Figure 1.2a, b**, the crystal structure of CIPS is defined by the sulfur framework in which the metal cations and P–P pairs fill the octahedral voids. Within each layer, the Cu, In, and P–P pairs form triangular patterns. Notably, since the site change between Cu and P–P pair from one layer to another, a complete unit cell composed two adjacent layers to fully describe the material's symmetry.

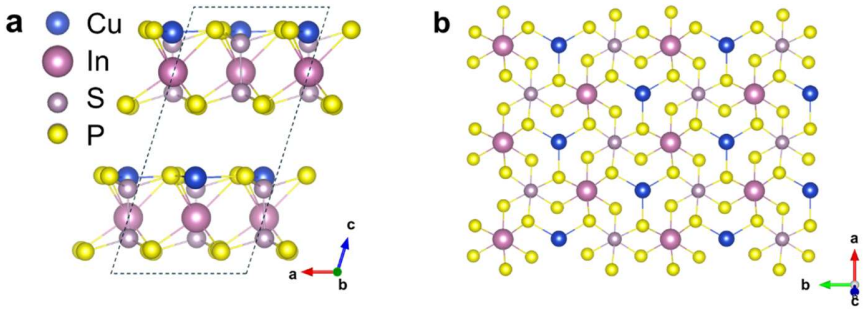


Figure 1.2 Crystal structure of CIPS viewed from (a) b axis, (b) layer normal (c^* axis), respectively.

The 2D ferroelectrics CIPS has relative high dielectric constant of 50 and high conductivity at room temperature.[29] Recent research found inducing non-ferroelectric phase by doping, such as chemical alloy, is an effective way to tune the ferroelectric phase transition characteristics and further increase the dielectric at room temperature. Previous research works indicate that compared to CIPS, its analog, $AgInP_2S_6$, has different physical properties. The substitution of the Cu cation with the bigger Ag cation results in the variation of some physical properties, such as the loss of ferroelectricity, higher air stability, larger dielectric constant, the changes in crystal inversional symmetry, and the conversion

of band gap from direct to indirect. Therefore, suppressions of the ferroelectric ordering were found in cation substitution cases, including Cu partially replaced by Ag.[30]

1.3 Electronic Device in Application

1.3.1 Monolithic 3D Integration

Nowadays, the memory-access efficiency of current computing systems cannot support data-centric workloads. The demands of future applications in computing and deep learning overwhelm the projected capabilities of current electronic systems. To satisfy these requirements, three main challenges need to be overcome for electronic systems: compute wall, memory wall, and connectivity wall.[31-33] Monolithic 3-D integration opens the possibility of realizing such future systems through leveraging fine-grained and dense vertical connectivity to enable massively concurrent accesses between computing and memory units. Such efficient implementations are achieved by several component nanotechnologies, such as energy-efficient CNFET-based logic circuits, high-density resistive random-access memory (RRAM), and dense metal interconnects. Wu et al. reported monolithic 3D system utilizing CNFETs and RRAM.[34] A brain-inspired hyperdimensional (HD) computing system is achieved, enabling classify different languages. And the accuracy could reach 98% under over 20,000 sentences per language pair.

Monolithic 3D integration is a process, whereby layers of circuits are fabricated on top of each other on the same substrate by high-density inter-layer vias (ILVs) to connect between tiers of circuits, which significantly increase the density of vertical connectivity.[35] In addition, in conventional 3D chip stacking, the active devices layer appears only on the bottom layer of the chip due to its high-temperature fabrication process. (>1000 ℃ for

steps such as dopant activation annealing).[36] In contrast, monolithic 3D integration, commonly consisting of inter-layer vias, multiple upper tiers, and a bottom tier, can be arbitrarily vertically interleaved. (**Figure 1.3**) Hence, monolithic 3D integration with low power consumption is a promising candidate for ultra-dense integration of traditionally off-chip components (e.g. sensors, memories) with energy-efficient computation units—all within a single chip. [37,38] One of the key challenges for Monolithic 3D ICs is low-temperature fabrication process for upper tiers of circuits (<400 °C) to not damage the lower layers of computing, memory, or metal interconnects. Further efforts have been attempted to achieve low temperature, such as exploiting polysilicon as channels or using low temperature fabrication process.[39]

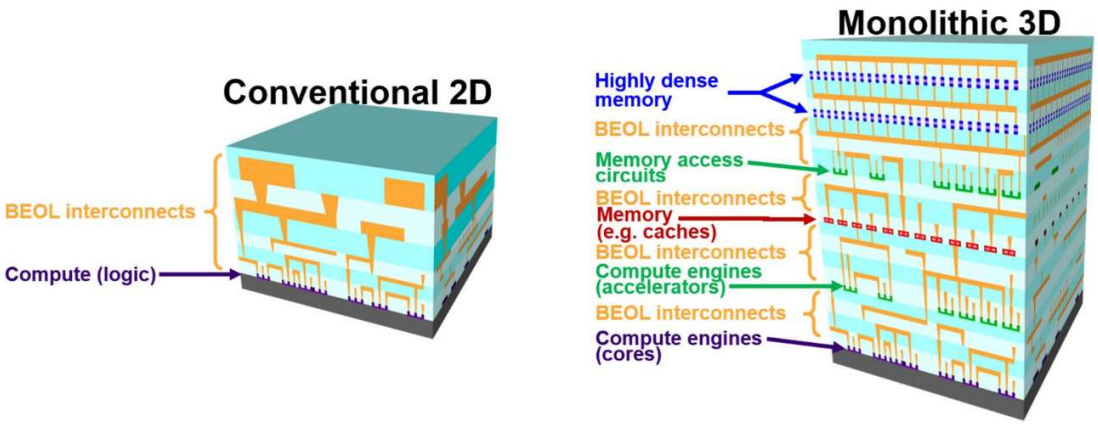


Figure 1.3. Conventional 2-D IC (left) vs. a monolithic 3-D IC (right).

RRAM as one of upper tiers is a promising candidate for monolithic 3D integration due to its high-capacity, non-volatile data storage, and low-temperature fabrication process (<300 °C).[40] Nevertheless, the RRAM cross-point array has not been integrated since the inevitable sneak current path occurs during write and read operations.[41] To reduce the sneak currents associated with RRAM array, Many cell structures, such as 1 transistor–1 RRAM

cell (1T1R), and 1 diode–1 RRAM cell (1D1R), have been used.[42] Wu et al. demonstrated monolithic 3D integration of RRAM array using an IGZO access transistor for each HfO_2-based RRAM (1T1R) to avoid sneak current and programming disturbance.[43] The system suggested basic functionality of XNOR operation as in-memory computing for binary neural network (BNN) artificial intelligence applications. However, the transistor as selector device has larger feature size (~ $6F^2$). Ji et al. put forward flexible non-volatile memory cell array which consists of all-organic one diode–one resistor (1D1R).[44] Therefore, by carefully combining with diverse selectors, the advances of multiple technologies can be integrated with monolithic 3D ICs while simultaneously avoiding their flaws.

Monolithic 3D integration has great potential to overcome bottleneck of data-intensive applications. In contrast to 3D stacking based on TSV, monolithic 3D integration can offer significant energy-efficiency advantages. However, it requires low-temperature fabrication for upper tiers of circuits (< 400℃), to maintain the pre-fabricated lower tiers and prevent interconnect diffusion. The fabrication process of RRAM can align with monolithic 3D compatible due to their low-temperature fabrication. With rapid advancement in this field, increasing efforts are paid to achieve demonstrations of monolithic 3D integration, which will expand both the system complexity and the breadth of potential applications.

1.3.2 Dendritic Computing

Biological Dendrites

A neuron in the biological nerve system consists of dendrites, a soma, and an axon, with each neuron being able to handle thousands of different synaptic inputs unequally.[45] The neuron receives thousands of postsynaptic potentials, which are segregated into

multiple dendritic branches. Postsynaptic potentials received within a certain time interval from the same or different synapses are integrated into the dendrites nonlinearly. Both spatial and temporal information are encoded in the dendritic computing units. Then, the soma integrates these potentials further, determining whether to fire spikes based on integrated amplitude over a specific threshold and transmitting them via the axon [46]. **Figure 1.4** illustrates the key computing features of the synapse, dendrite, and soma. The synapse represents a plastic weight, and the soma provides integration and spike-firing functions, while the nonlinear integration and filtering of postsynaptic potentials on different dendritic branches provide the network with the ability to process complex information.

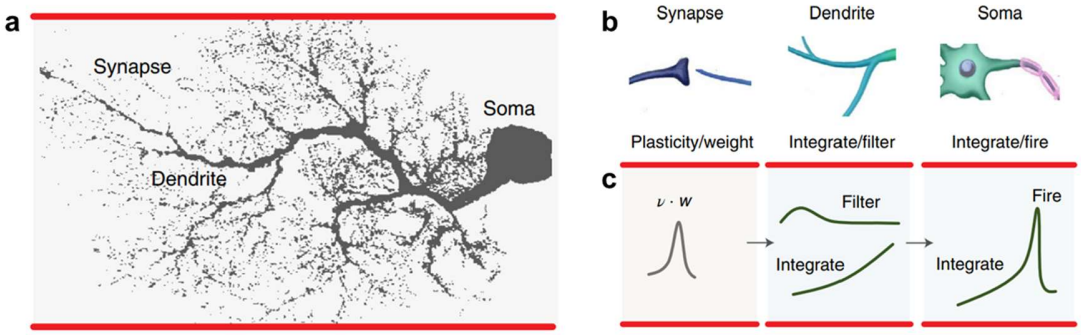


Figure 1.4 Overview of the biological neural networks with dendrites [47]. (a) An image of a brain neuron of a mouse where dendrites serve as connections between the soma and synapses. The spontaneous postsynaptic potentials arriving are nonlinearly integrated in the dendrites. (b) Comparison of key computing features of three important functional elements: synapse for plasticity and weight, dendrite for integrate and filter, and soma for integrate and fire. (c) Illustration of their typical computing function.

Recent researchers confirmed that dendrites are no longer treated as passive conductors of information to the soma.[48,49] The active properties of the dendritic tree are crucial for

determining how electrical signals propagate. One of the active properties of the dendrites: nonlinear integration, is thought to increase the computational ability of neurons [50]. The dendrite can tune local signals within a branch, either to amplify them or to dampen them, with such nonlinear processing again being kept local by the unfavorable impedance matches at branch points. Dendritic integration can be defined by comparing the measured excitatory post-synaptic potentials (EPSPs) resulting from the simultaneous activation of the synaptic inputs and the arithmetic sum of individual EPSPs. The dendritic nonlinear integration can be classified into three operations: (1) linear, where the measured EPSP equals the arithmetic sum of individual EPSPs; (2) supralinear, where the measured EPSP exceeds the arithmetic sum of individual EPSPs; and (3) sublinear, where the measured EPSPs is less than the arithmetic sum of individual EPSPs. These signals' nonlinear integration and insignificant background information filter functions are one of the main reasons why biological neural networks can handle very complex tasks and yet consume little energy. Incorporating dendritic computing units into artificial neural networks is thus highly desired.

Artificial Dendrites Device

Over the past few years, significant research has drawn attention to biomimetic electronic devices, such as artificial synapses and soma.[51,52] The electrical-stimulated single-input synaptic devices have made significant progress. However, in biological systems, the dendrites of a single neuron receive multiple inputs from other neurons through hundreds of synapses and process these inputs by unique dendritic integration. Hence, electrical-based multi-input artificial neural devices are vital for the future construction of multifunctional and ultra-low power brain-like computing networks.

Increasing research focused on artificial dendrite devices. A metal oxide-based dendritic device has been developed to emulate the functions of dendrites, such as integrating spatial and temporal information in a nonlinear manner and denoise insignificant background information.[47] Compared to the conventional neural network without artificial dendrite, this neural network chip exhibited less power consumption and higher recognition accuracy based on the filtering and integration function of artificial dendrite. However, this device didn't exhibit the nonlinear integration function of dendrites.

Recently, multiterminal neurotransmitters have been reported to emulate the nonlinear integration of the dendritic tree through multiple in-plane gates.[53] The conductivity property of the channel is controlled by these in-plane gates, which are coupled by strong lateral protonic/electronic capacitive coupling effect through a proton-conducting solid-state electrolyte film. In addition, based on the nonlinear integration function of dendrites, the Boolean function (OR, AND, and XOR) can be realized by defining a somatic spike threshold.[54,55] The transistor performed supralinear under three or fewer light stimuli and sublinear under four light stimuli. Moreover, a neuron with nonlinear dendritic compartments can implement the set of linearly non-separable functions, which encompasses a much larger fraction of all computations. Thus, both supralinear and sublinear compartments unlock access to all the possible computations. If we vary synaptic weights, the thresholds, and the nonlinear dendritic operations, we can use Boolean analysis to examine the different functions this model can implement. A functionally salient neuronal computation that requires dendritic nonlinearities is the association (or binding) of two features of an object (for example, their shape and color). This is known as the feature binding problem.

Besides the multiple input signal process, recent researches prove that the artificial dendrite can improve the efficiency and training accuracy of neural networks.[56] Recently, a scalable and stackable artificial dendrite neural circuit has been demonstrated, which shows great promise for a multilayer network system that mimics feature binding situation.[57,58]

1.4 Objectives and Structure of Thesis

1.4.1 Objectives of Thesis

In contrast to the human brain, artificial neural networks (ANN) based on transistors and RRAMs encounter significant challenges in energy efficiency when further scaling neural networks. Therefore, applications for energy efficiency have attracted much attention in recent years. This thesis specializes in improving energy efficiency through the design and implementation of devices utilizing 2D materials.

2D materials such as transition metal dichalcogenides (TMDs), hexagonal boron nitride (h-BN), and $CuInP_2S_6$ are considered promising candidates for low-power and high-performance electronic devices in neuromorphic computing. Unlike bulk materials, atomically thin 2D materials offer advantages including low-temperature fabrication, low-power switching, and electrostatic gate tunability.

The thesis investigates the applications of how to enhance energy efficiency based on two configurations. (1) 1S1R structure can address the leakage current through unselected low-resistance state cells. To achieve energy-efficient and large-scale memory array in monolithic 3D integration, a high nonlinear selector fabricated under low temperature is highly demanded. (2) dendritic neural networks involving dendritic nonlinear computation

can solve real-world classification challenges and handle datasets utilized in machine learning applications. Based on dendritic function, more efficient utilization of resources can be achieved, which can enhance energy efficiency. Therefore, the multi-gate devices emulating the dendritic function should be proposed.

1.4.2 Structure of Thesis

The chapters of this thesis are organized as follows:

Chapter 1: Introduction. This chapter points out the significance of addressing the energy consumption issue and introduces the main 2D materials used, such as TMD, h-BN, and CIPS. Then, the applications in electronics, like 1S1R in monolithic 3D integration and dendritic neural computing, are demonstrated. Last, the organization of the thesis is introduced.

Chapter 2: The main growth and characterization methods. The ferroelectric materials were grown through chemical vapor transport and characterized by scanning microwave impedance microscopy.

Chapter 3: Highly nonlinear memory selectors with ultrathin $MoS_2/WSe_2/MoS_2$ heterojunction. The advancement of RRAM arrays requires highly nonlinear selectors with high current density to address a specific memory cell and control leakage current through the unselected cell. Enabling monolithic 3D integration of RRAM array demands selector devices with small footprint and low-temperature processing for ultrahigh-density data storage. Hence, an ultrathin two-terminal n-p-n selector utilized 2D TMDs by low-temperature dry transfer method, where the van der Waals contact can reduce the Fermi level pinning and preserve the intrinsic transport

behavior of TMDs. The fine-tuning of Schottky barrier heights and the manipulation of WSe_2 thickness in $MoS_2/WSe_2/MoS_2$ n-p-n selector based on a punch-through transport results in high nonlinearity (≈ 230) and high current density (2×10^3 A cm^{-2}). Then, the n-p-n selectors are integrated into bipolar h-BN memory and the maximum crossbar size of the 1S1R is calculated, which has great potential in future monolithic 3D integration.

Chapter 4: Dendrite Integration Mimicked on high-k-based Dual Gate Transistors. Nowadays, state-of-the-art artificial neural networks exhibit higher energy consumption than the human brain. The advancements of synapse and soma devices are notable, while little research focuses on the multi-input and nonlinear integration of dendrite devices. The lack of these critical functions limited the performance of neural networks in terms of flexibility, energy efficiency, and complex task handling. To counter this, a dual-gate transistor with innovative $Cu_{0.67}Ag_{0.33}InP_2S_6$ as gate dielectric is designed. The low-temperature dry transfer fabrication process ensures the stability of ferroelectric materials under ambient conditions. The ferroelectric properties of CAIPS gate dielectric enable these transistors not only to emulate synaptic functions but also to realize the nonlinear integration of dendrites.

Chapter 5: Conclusion and outlook. In this chapter, the innovations and shortcomings of the research work in this thesis are systematically discussed. Meanwhile, future research based on low-energy consumption applications is proposed.

1.5 References

1 Fuyou Liao, F. Z. Y. C. Neuromorphic vision sensors: Principle, progress and perspectives. *Journal of Semiconductors*, doi:10.1088/1674-4926/42/1/013105 (2021).

2 Vinet, M. *et al.* 3D monolithic integration: Technological challenges and electrical results. *Microelectronic Engineering* **88**, 331-335, doi:10.1016/j.mee.2010.10.022 (2011).

3 Acharya, J. *et al.* Dendritic Computing: Branching Deeper into Machine Learning. *Neuroscience* **489**, 275-289, doi:10.1016/j.neuroscience.2021.10.001 (2022).

4 Cao, W. *et al.* The future transistors. *Nature* **620**, 501-515, doi:10.1038/s41586-023-06145-x (2023).

5 Liu, Y. *et al.* Promises and prospects of two-dimensional transistors. *Nature* **591**, 43-53, doi:10.1038/s41586-021-03339-z (2021).

6 Hong, S. in *2012 IEEE Asian Solid State Circuits Conference (A-SSCC).* 5-8.

7 Chen, R. S. *et al.* MoS2 Transistor with Weak Fermi Level Pinning via MXene Contacts. *Adv. Funct. Mater.*, doi:10.1002/adfm.202204288 (2022).

8 Zeng, S., Liu, C. & Zhou, P. Transistor engineering based on 2D materials in the post-silicon era. *Nature Reviews Electrical Engineering*, doi:10.1038/s44287-024-00045-6 (2024).

9 Wu, P., Zhang, T., Zhu, J., Palacios, T. & Kong, J. 2D materials for logic device scaling. *Nat Mater* **23**, 23-25, doi:10.1038/s41563-023-01715-w (2024).

10 Chen, R.-S., Ding, G., Zhou, Y. & Han, S.-T. Fermi-level depinning of 2D transition metal dichalcogenide transistors. *Journal of Materials Chemistry C* **9**, 11407-11427, doi:10.1039/d1tc01463c (2021).

11 Pal, A. *et al.* Two-dimensional materials enabled next-generation low-energy compute and connectivity. *MRS Bull.*, doi:10.1557/s43577-022-00270-0 (2022).

12 Li, M. *et al.* Imperfection-enabled memristive switching in van der Waals materials. *Nature Electronics*, doi:10.1038/s41928-023-00984-2 (2023).

13 Zhou, F., Chen, J., Tao, X., Wang, X. & Chai, Y. 2D Materials Based Optoelectronic Memory: Convergence of Electronic Memory and Optical Sensor. *Research (Wash D C)* **2019**, 9490413, doi:10.34133/2019/9490413 (2019).

14 Zhao, Y. *et al.* Doping, Contact and Interface Engineering of Two-Dimensional Layered Transition Metal Dichalcogenides Transistors. *Adv. Funct. Mater.* **27**, 1603484, doi:10.1002/adfm.201603484 (2017).

15 Schulman, D. S., Arnold, A. J. & Das, S. Contact engineering for 2D materials and devices. *Chem Soc Rev* **47**, 3037-3058, doi:10.1039/c7cs00828g (2018).

16 Zhou, C. *et al.* Carrier Type Control of WSe2 Field-Effect Transistors by Thickness Modulation and MoO3 Layer Doping. *Adv. Funct. Mater.* **26**, 4223-4230, doi:10.1002/adfm.201600292 (2016).

17 Liu, W. *et al.* Role of metal contacts in designing high-performance monolayer n-type WSe2 field effect transistors. *Nano Lett* **13**, 1983-1990, doi:10.1021/nl304777e (2013).

18 Kong, L. *et al.* Doping-free complementary WSe2 circuit via van der Waals metal integration. *Nat Commun* **11**, 1866, doi:10.1038/s41467-020-15776-x (2020).

19 Jeong, H. *et al.* Metal-Insulator-Semiconductor Diode Consisting of Two-Dimensional Nanomaterials. *Nano Lett* **16**, 1858-1862, doi:10.1021/acs.nanolett.5b04936 (2016).

20 Wang, Y. & Chhowalla, M. Making clean electrical contacts on 2D transition metal dichalcogenides. *Nature Reviews Physics* **4**, 101-112, doi:10.1038/s42254-021-00389-0 (2021).

21 Villena, M. A. *et al.* Variability of metal/h-BN/metal memristors grown via chemical vapor deposition on different materials. *Microelectronics Reliability* **102**, doi:10.1016/j.microrel.2019.113410 (2019).

22 Knobloch, T. *et al.* The performance limits of hexagonal boron nitride as an insulator for scaled CMOS devices based on two-dimensional materials. *Nature Electronics* **4**, 98-108, doi:10.1038/s41928-020-00529-x (2021).

23 Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499**, 419-425, doi:10.1038/nature12385 (2013).

24 Wang, C. H. *et al.* Ultrathin Three-Monolayer Tunneling Memory Selectors. *ACS Nano* **15**, 8484-8491, doi:10.1021/acsnano.1c00002 (2021).

25 Wu, X. *et al.* Thinnest Nonvolatile Memory Based on Monolayer h-BN. *Adv Mater* **31**, e1806790, doi:10.1002/adma.201806790 (2019).

26 Wang, X. *et al.* Van der Waals negative capacitance transistors. *Nat Commun* **10**, 3037, doi:10.1038/s41467-019-10738-4 (2019).

27 Britnell, L. *et al.* Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett* **12**, 1707-1710, doi:10.1021/nl3002205 (2012).

28 Wu, J. *et al.* High tunnelling electroresistance in a ferroelectric van der Waals heterojunction via giant barrier height modulation. *Nature Electronics* **3**, 466-472, doi:10.1038/s41928-020-0441-9 (2020).

29 Liu, F. *et al.* Room-temperature ferroelectricity in CuInP2S6 ultrathin flakes. *Nat Commun* **7**, 12357, doi:10.1038/ncomms12357 (2016).

30 Dziaugys, A., Banys, J., Macutkevic, J., Sobiestianskas, R. & Vysochanskii, Y. Dipolar glass phase in ferrielectrics: CuInP2S6 and Ag0.1Cu0.9InP2S6 crystals. *physica status solidi (a)* **207**, 1960-1967, doi:10.1002/pssa.200925346 (2010).

31 Wulf, W. A. & McKee, S. A. Hitting the memory wall: implications of the obvious. *SIGARCH Comput. Archit. News* **23**, 20–24, doi:10.1145/216585.216588 (1995).

32 Villa, O. *et al.* in *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. 830-841.

33 Shulaker, M. M. *et al.* Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature* **547**, 74-78, doi:10.1038/nature22994 (2017).

34 Wu, T. F. *et al.* Hyperdimensional Computing Exploiting Carbon Nanotube FETs, Resistive RAM, and Their Monolithic 3D Integration. *IEEE Journal of Solid-State Circuits* **53**, 3183-3196, doi:10.1109/jssc.2018.2870560 (2018).

35 Batude, P. *et al.* in *2011 International Electron Devices Meeting.* 7.3.1-7.3.4.

36 Griggio, F. *et al.* in *2018 IEEE International Reliability Physics Symposium (IRPS).* 6E.3-1-6E.3-5.

37 Zhang, Y., Ye, J., Matsuhashi, Y. & Iwasa, Y. Ambipolar MoS2 thin flake transistors. *Nano Lett* **12**, 1136-1140, doi:10.1021/nl2021575 (2012).

38 Aly, M. M. S. *et al.* Energy-Efficient Abundant-Data Computing: The N3XT 1,000x. *Computer* **48**, 24-33, doi:10.1109/MC.2015.376 (2015).

39 Tiemin, Z., Min, C., Saraswat, K. C. & Plummer, J. D. A vertical submicron polysilicon thin-film transistor using a low temperature process. *IEEE Electron Device Letters* **15**, 415-417, doi:10.1109/55.320986 (1994).

40 Wong, H. S. & Salahuddin, S. Memory leads the way to better computing. *Nat Nanotechnol* **10**, 191-194, doi:10.1038/nnano.2015.29 (2015).

41 Jiang, Z. *et al.* in *2018 IEEE Symposium on VLSI Technology.* 107-108.

42 Sun, L. *et al.* Self-selective van der Waals heterostructures for large scale memory array. *Nat Commun* **10**, 3161, doi:10.1038/s41467-019-11187-9 (2019).

43 Wu, J., Mo, F., Saraya, T., Hiramoto, T. & Kobayashi, M. A Monolithic 3-D Integration of RRAM Array and Oxide Semiconductor FET for In-Memory Computing in 3-D Neural Network. *IEEE Transactions on Electron Devices* **67**, 5322-5328, doi:10.1109/ted.2020.3033831 (2020).

44 Ji, Y. *et al.* Flexible and twistable non-volatile memory cell array with all-organic one diode-one resistor architecture. *Nat Commun* **4**, 2707, doi:10.1038/ncomms3707 (2013).

45 Merolla, P. A. *et al.* A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668-673, doi:doi:10.1126/science.1254642 (2014).

46 Wang, J. & Zhuge, F. Memristive Synapses for Brain-Inspired Computing. *Advanced Materials Technologies* **4**, doi:10.1002/admt.201800544 (2019).

47 Li, X. *et al.* Power-efficient neural network with artificial dendrites. *Nat Nanotechnol* **15**, 776-782, doi:10.1038/s41565-020-0722-5 (2020).

48 Cazé, R., Tran-Van-Minh, A., Gutkin, B. & Digregorio, D. *Demonstration that sublinear dendrites enable linearly non-separable computations*. (2023).

49 Ralf, W., William B. Kristan, Jr. & David, K. Supralinear Summation of Synaptic Inputs by an Invertebrate Neuron: Dendritic Gain Is Mediated by an "Inward Rectifier" K⁺ Current. *The Journal of Neuroscience* **19**, 5875, doi:10.1523/JNEUROSCI.19-14-05875.1999 (1999).

50 Tran-Van-Minh, A. *et al.* Contribution of sublinear and supralinear dendritic integration to neuronal computations. *Front Cell Neurosci* **9**, 67, doi:10.3389/fncel.2015.00067 (2015).

51 Chen, J. *et al.* Optoelectronic graded neurons for bioinspired in-sensor motion perception. *Nature Nanotechnology* **18**, 882-888, doi:10.1038/s41565-023-01379-2 (2023).

52 Duan, Q. *et al.* Spiking neurons with spatiotemporal dynamics and gain modulation for monolithically integrated memristive neural networks. *Nat Commun* **11**, 3399, doi:10.1038/s41467-020-17215-3 (2020).

53 Gao, W. T. *et al.* Dendrite Integration Mimicked on Starch-Based Electrolyte-Gated Oxide Dendrite Transistors. *ACS Appl Mater Interfaces* **10**, 40008-40013, doi:10.1021/acsami.8b16495 (2018).

54 Fu, Y. *et al.* in *2021 IEEE International Electron Devices Meeting (IEDM)*. 12.16.11-12.16.14.

55 He, W. *et al.* A multi-input light-stimulated synaptic transistor for complex neuromorphic computing. *Journal of Materials Chemistry C* **7**, 12523-12531, doi:10.1039/c9tc03898a (2019).

56 Meng, J. *et al.* Ionic Diffusive Nanomemristors with Dendritic Competition and Cooperation Functions for Ultralow Voltage Neuromorphic Computing. *ACS Nano*, doi:10.1021/acsnano.4c00424 (2024).

57 Schemmel, J., Kriener, L., Müller, P. & Meier, K. in *2017 International Joint Conference on Neural Networks (IJCNN)*. 2217-2226.

58 Baek, E. *et al.* Neuromorphic dendritic network computation with silent synapses for visual motion perception. *Nature Electronics*, doi:10.1038/s41928-024-01171-7 (2024).

Chapter 2 Material Growth and Characterization Method

2.1 Chemical Vapor Transport (CVT)

The Chemical Vapor Transport (CVT) method is considered one of the effective ways to synthesize high-quality single crystals or thin films.[1] It is a process used in materials science and crystal growth to transport and deposit materials through vapor phase reactions. In CVT, a source material or precursor is vaporized and transported by a transport agent (such as Ar or I_2) to a cooler region where it undergoes chemical reactions or condenses to form a desired material.[2] Usually, it would take at least seven days to grow large and high-quality single crystals. In addition, the process parameters, such as temperature, pressure, and gas composition, are carefully controlled to achieve the desired material properties and crystal quality. In most cases, the source materials are put at the high-temperature side and single crystals would be obtained at the low-temperature side. **Figure 2.1** shows the scheme of CVT experiments for single crystal growth.

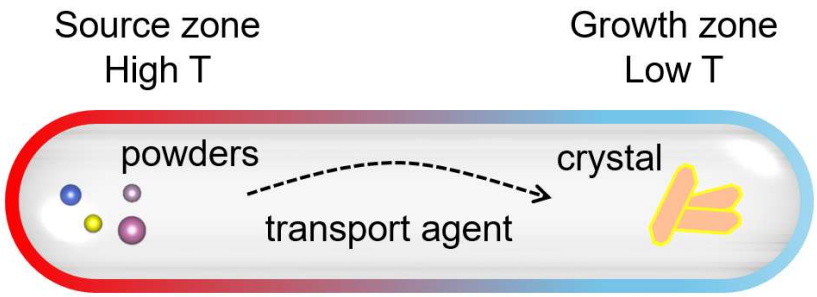


Figure 2.1 Scheme of CVT experiments for single crystal growth.

2.2 Scanning Microwave Impedance Microscopy

Atomic force microscopy can detect the surface of the sample with high scanning speed and resolution according to the change in the force between the needle tip of piezoelectric detectors and samples.[3] Derived from AFM, scanning microwave impedance microscopy (sMIM) was first proposed by Shen et al. from Stanford University,[4] and was applied to the detection of dielectric constant.[5] For different frequencies of electromagnetic waves, the polarization phenomena are different. At low frequencies, the electric field change period is much longer than the relaxation time. The real dielectric constant (ε_r) is close to the static dielectric constant (ε_s); At high frequencies, polarization cannot keep with change in the electric field, and the loss coefficient increases, dissipating in the form of heat. Regarding to the dielectric properties of samples, using the microwave frequency (300 MHz – 300 GHz) can achieve strong penetration and explore internal information of the sample. In addition, the polarization affected by microwave frequency includes short relaxation time polarization, instead of long relaxation time polarization (spatial charge distribution).

Differing from conventional microwave near-field equipment mainly analyzes local sample properties by measuring signal frequency shift and related factor changes, the sMIM can detect local electrical properties of the sample by measuring in-phase and out-of-phase signals. The new type of near-field microprobe with strip lines based on Si_3N_4 was prepared to improve the quality of signal transmission. Due to the addition of a strip-line structured microwave conduction circuit in the probe, its tip radius (50 ± 10 nm) is

slightly larger than that of a conventional AFM probe, resulting in a slightly lower resolution in surface morphology. The co-axial probe of sMIM is coated by TiW/Au on both sides as shielding and reflex and consists of the rectangular cantilever and pyramid tip (**Figure 2.2a, b**). The sMIM delivers a microwave signal at a few GHz to the tip apex to interact with the sample (**Figure 2.2c**) and consequently probes its local electrical properties by analyzing the reflected microwave response.

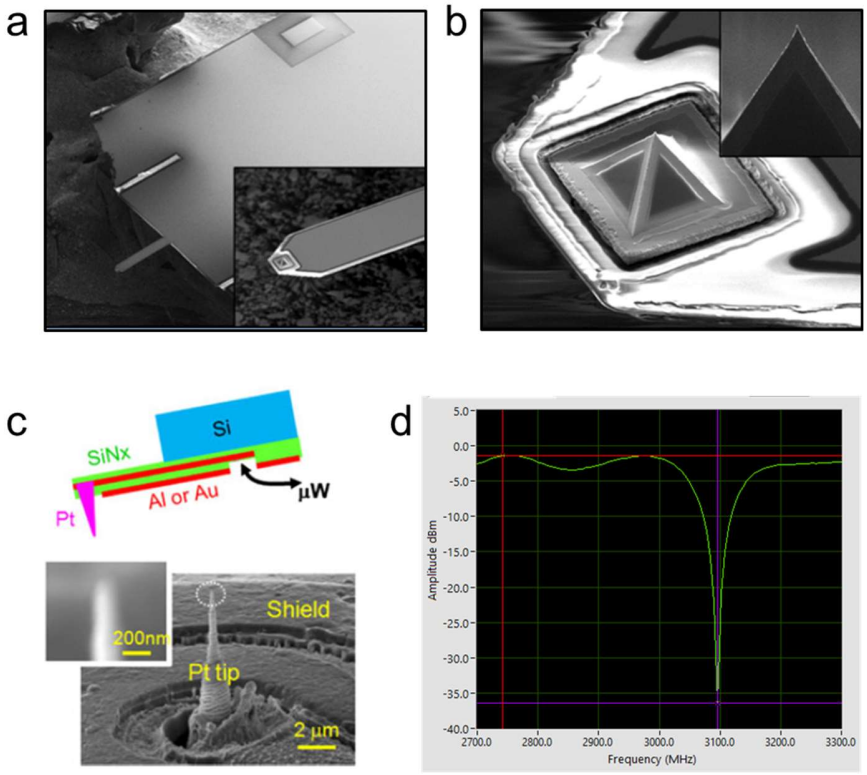


Figure 2.2 The electron microscopy images of (a) the cantilever with 150 um length and (b) the pyramid tip of a microwave impedance microscope at different magnifications. (c) the cross-section of probe. (d) the detection of microwave signal (~ 3 GHz) delivered to the tip.

The schematic diagram of the AFM-derived sMIM setup is illustrated in **Figure 2.3**. During sMIM operation, the 3 GHz microwave signal is transmitted to the tip apex through a directional coupler to interact with the sample. The reflected microwave signal, which contains the information on the impedance of the tip-sample system, passes through a background signal suppression circuit, a power amplifier, and finally a mixer. Then, the signal, taking the fine-tuned microwave source phase as reference, is demodulated and processed into two DC outputs, that is, real (sMIM-R) and imaginary (sMIM-C) components of the tip-sample complex impedance, respectively.

Before characterizing the sample, the sMIM signal should be modulated. The sMIM-C signal rises consistently with the film conductivity, with the highest sensitivity observed between the insulating and conductive states. Conversely, the sMIM-R signal, reflecting the effective reduction in tip-sample interaction, diminishes towards zero at low and high sample conductivities, peaking at an intermediate conductivity level of 5 S m^{-1}.

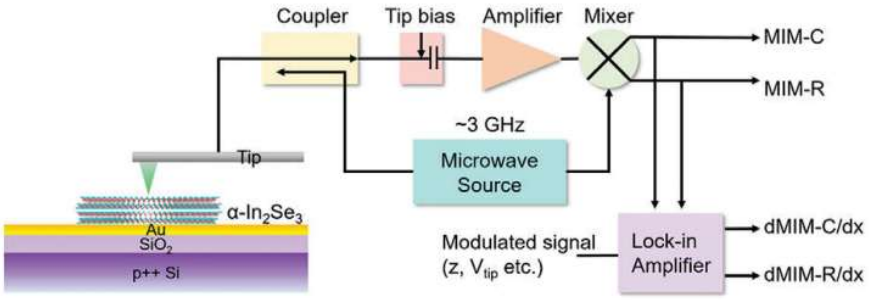


Figure 2.3 Experimental schematic of the sMIM setup.[6]

Based on the sMIM-C and sMIM-R signals, the local permittivity, conductivity, and impedance of the sample can be detected (**Figure 2.4**). In addition, the subsurface information can be obtained through sMIM, such as doping situations.

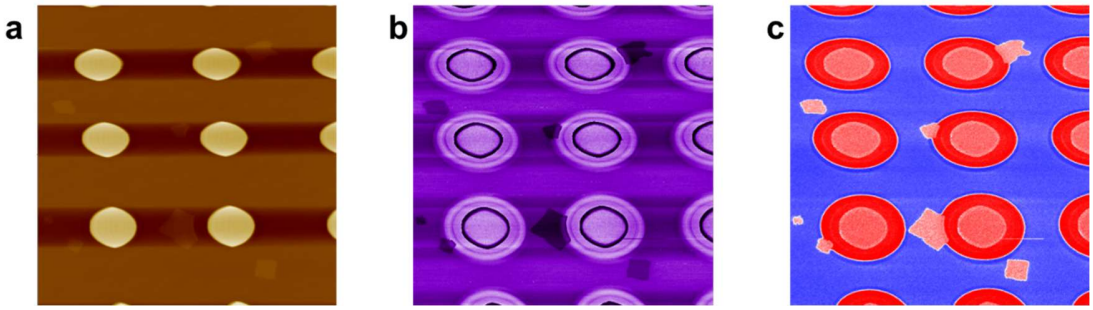


Figure 2.4. (a) AFM image of the sample. Scale bar, 5 μm. (b) the sMIM-C signal indicating the doping situation of the subsurface. Scale bar, 5 μm. (c) the sMIM-C signal indicating the doping type of subsurface. N-doping marked as blue; p-doping marked as red. The color saturation represents doping concentration. Scale bar, 5 μm

2.3 Reference

1 Xu, X. & Liu, K. Progress and perspective on the growth of two-dimensional single crystals. *Science Bulletin* **67**, 1410-1412, doi:https://doi.org/10.1016/j.scib.2022.06.012 (2022).

2 Binnewies, M., Glaum, R., Schmidt, M. & Schmidt, P. Chemical Vapor Transport Reactions - A Historical Review. *Zeitschrift für anorganische und allgemeine Chemie* **639**, 219-229, doi:10.1002/zaac.201300048 (2013).

3 Rao, M. *et al.* Thousands of conductance levels in memristors integrated on CMOS. *Nature* **615**, 823-829, doi:10.1038/s41586-023-05759-5 (2023).

4 Lai, K. *et al.* Tapping mode microwave impedance microscopy. *Rev. Sci. Instrum.* **80**, doi:10.1063/1.3123406 (2009).

5 Lai, K. *et al.* Nanoscale Electronic Inhomogeneity in In2Se3 Nanoribbons Revealed by Microwave Impedance Microscopy. *Nano Lett.* **9**, 1265-1269, doi:10.1021/nl900222j (2009).

6 Wang, L. *et al.* A Scanning Microwave Impedance Microscopy Study of α-In2Se3 Ferroelectric Semiconductor. *Adv. Funct. Mater.*, doi:10.1002/adfm.202316583 (2024).

Chapter 3 Highly Nonlinear Memory Selectors with Ultrathin $MoS_2/WSe_2/MoS_2$ Heterojunction

Resistive random access memory (RRAM) crossbar arrays require the highly nonlinear selector with high current density to address a specific memory cell and suppress leakage current through the unselected cell. 3D monolithic integration of RRAM array requires selector devices with a small footprint and low-temperature processing for ultrahigh-density data storage. Here, an ultrathin two-terminal n-p-n selector with 2D transition metal dichalcogenides (TMDs) is designed by a low-temperature transfer method. The van der Waals contact between transferred Au electrodes and TMDs reduces the Fermi level pinning and retains the intrinsic transport behavior of TMDs. The selector with a single type of TMD exhibits a trade-off between current density and nonlinearity depending on the barrier height. By tuning the Schottky barrier height and controlling the thickness of p-type WSe_2 in $MoS_2/WSe_2/MoS_2$ n-p-n selector for a punch-through transport, the selector shows high nonlinearity (≈ 230) and high current density (2×10^3 A cm^{-2}) simultaneously. The authors further integrate the n-p-n selectors with a bipolar hexagonal boron nitride memory and calculate the maximum crossbar size of the 2D material-based one-selector one-resistor (1S1R) according to a 10% read margin, which offers the possible realization of future 3D monolithic integration.

3.1 Introduction

Resistive random access memories (RRAM) exhibit great potential for information storage and computing.[1,2] Their two-terminal structure leads to an effective cell area of $4F^2$ (F is the minimum feature size), which is the optimal architecture that enables high packing density in crossbar arrays for terabits solid-state storage.[3] In the RRAM array, the selected word line and selected bit line are applied to full operation voltage (V) and 0 voltage, respectively. As the density of RRAM array increases, one fundamental issue is inevitable: sneak currents through unselected low resistance state (LRS) cells when the unselected word lines and bit lines are biased with half of the operation voltage in the V/2 scheme, which limits the maximum array size. Researchers have invested great efforts to address the above issue, including a self-rectifying cell (SRC) with inherent nonlinear I-V characteristics or the integration of a nonlinear selector to the memory cell.[4,5] Different from SRC, one selector-one RRAM cell, including one-selector one-resistor (1S1R) or one-transistor one-resistor (1T1R), can optimize individual components separately to reach the overall device performance targets easily.[6] The 1T1R architecture with field-effect transistor as selector involves a complex and high-temperature fabrication process [7] and occupies the large footprint of $6F^2$. Conventional two-terminal selectors with silicon-based p-n heterojunction suffer from high-temperature fabrication due to in-situ doped epitaxial Si growth process with a thermal budget of over 700 °C.[8] This high-temperature diffusion process cannot satisfy the thermal stability of interconnect structure and three-dimensional (3D) monolithic integration with the sub-400 °C processing temperature requirement.[9] To suppress the sneak current, we require highly nonlinear, high-current, and small footprint selectors that can be fabricated at low temperature.

Transition metal dichalcogenides (TMDs)-based electronic devices can meet the key requirements of monolithic integration since the absence of surface dangling bonds in 2D materials allows low-temperature fabrication by the dry transfer method.[10] The transfer method can help to establish van der Waals (vdWss) metal-semiconductor junctions, reducing the Fermi level pinning and creating an interface that is essentially free from chemical disorder.[11] More importantly, this vdWss gap between the electrode and semiconductor serves as an effective tunneling barrier to reduce the current at low voltage, which can enhance the nonlinearity of selectors.[12] In addition, the excellent thermal stability of TMDs can guarantee the desirable characteristics of selector during the transfer process.[13] Since TMDs have controllable ultrathin thickness and good intrinsic carrier mobility,[14] the two-terminal TMDs-based selectors are expected to provide a sufficiently high current density (J) and large nonlinearity (NL) factor ($> 10^2$).[15]

This work demonstrates an n-p-n selector with 2D heterojunction, presenting high NL (over 200) and J ($\sim 3\times10^3$ A/cm^2). Its dry transfer process at 60 ℃ meets the requirements of 3D stacked memory integration. For vertical TMD selectors, the J and nonlinearity are closely related to the Schottky barrier height (SBH) of the metal/semiconductor interface and the depletion layer in heterojunction. The low SBH of Au/MoS_2 and the electron-hole combination in MoS_2/WSe_2 heterojunction allow us to put forward a high-performance n-p-n selector based on the punch-through mechanism. Furthermore, we successfully demonstrate a two-terminal all 2D material-based 1S1R architecture, in which an n-p-n selector and a bipolar resistive switching memory are connected in series. This work contributes to a framework for advancing 3D crossbar array memory devices.

3.2 Results and Discussion

3.2.1 The Design of 2D Heterojunction Selector

The unwanted sneak current is one of the significant issues for the RRAM crossbar array, which leads to misleading READ and WRITE operations. **Figure 3.1** shows an ideal scenario of a 1S1R array, where the selector can block the possible sneak path (black line path) through the neighboring memory cell and allow the desired current (red line path) through the selected cell (marked in black border).

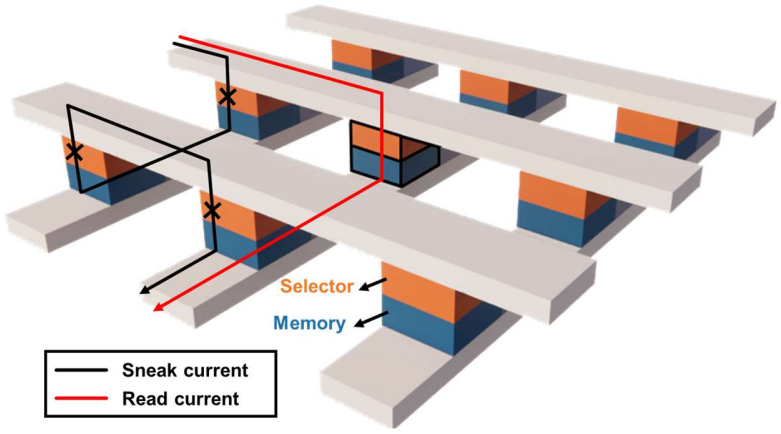


Figure 3.1 Schematic illustration of 1S1R array. Black line: sneak current path. Red line: desired read current.

The selector design with high nonlinearity and *J* can meet this requirement. **Figure3.2a** shows the electrical characteristic schematic of one RRAM cell and one ideal selector, where we expect the selector to have large *NL* over 100 and high current to support RRAM operation. Nonlinearity is defined as the ratio of the current at the read bias to the current at half the read bias (Formula in **Figure 3.2a**). **Figure 3.2b** shows the corresponding I-V characteristic schematic of a 1S1R cell with the selector and the RRAM connected in series.

When the unselected cell is under half read bias ($V_R/2$), compared to single RRAM cell, the sneak current of 1S1R cell is reduced significantly since the selector component acts as a blocking layer.

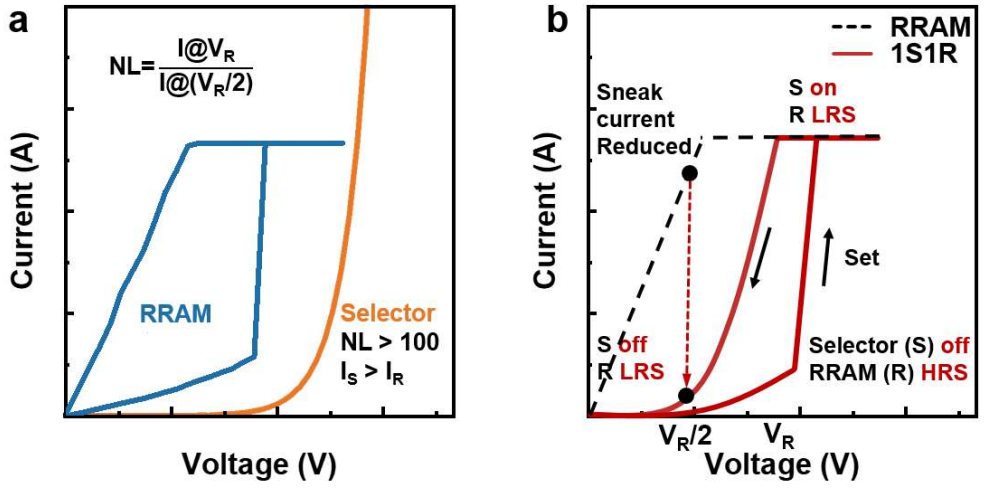


Figure 3.2 (a) The I-V characteristic schematic of the RRAM and nonlinear selector devices. V_R: Read voltage; I_S: Current of the selector; I_R: Current of RRAM cell. (b) The I-V characteristic schematic of the 1S1R cell. HRS: high resistance state; LRS: low resistance state.

We design an n-p-n selector based on TMD fabricated by the low-temperature dry transfer method. As can be seen in **Figure 3.3a**, the selector is treated as a complementary p-n diode, where two WSe_2/MoS_2 p-n diodes are connected back-to-back. The equilibrium band diagram of the designed selector suggests that the p-n barrier and charge transfer behavior exist in MoS_2/WSe_2 heterojunction. As illustrated in **Figure 3.3b**, we further prove the existence of WSe_2 and MoS_2 based on Raman spectra excited with a 532 nm laser. Pristine MoS_2 shows two peaks near 383 cm^{-1} (E^1_{2g} mode; in-plane vibration) and 403

cm^{-1} (A_{1g} mode; out-of-plane vibration).[16] On the pristine WSe_2 side, WSe_2 presents two distinct peaks centered at 247 cm^{-1} (E^1_{2g} mode) and 254 cm^{-1} (A_{1g} mode), consistent with the previous works.[17,18] Meanwhile, the layer-number sensitive mode A^2_{1g} at around 307 cm^{-1} (out-of-plane mode) is observable for multilayer WSe_2.[19] These phenomena reveal the presence of the 2D heterojunction.[20]

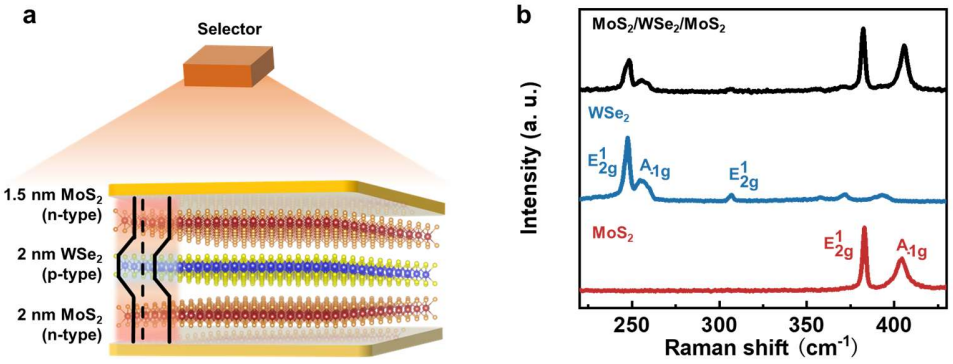


Figure 3.3 (a) Schematic illustration and band diagram of Au/MoS_2/WSe_2/MoS_2/Au selector without voltage bias. The depletion layers are marked in the orange region of the band diagram. (b) Raman spectra for MoS_2 (blue), WSe_2 (red), and MoS_2/WSe_2/MoS_2 heterojunction (black) from bottom to top plot.

Cross-section transmission electron microscopy (TEM) image in **Figure 3.4a, b** further confirm the nanoscale thickness and stacked structure, which is bilayer MoS_2, bilayer WSe_2, and trilayer MoS_2 from top to bottom. In addition, energy dispersive spectroscopy (EDS) mapping exhibits the corresponding elemental distribution mapping of the device respectively **Figure 3.4c**.

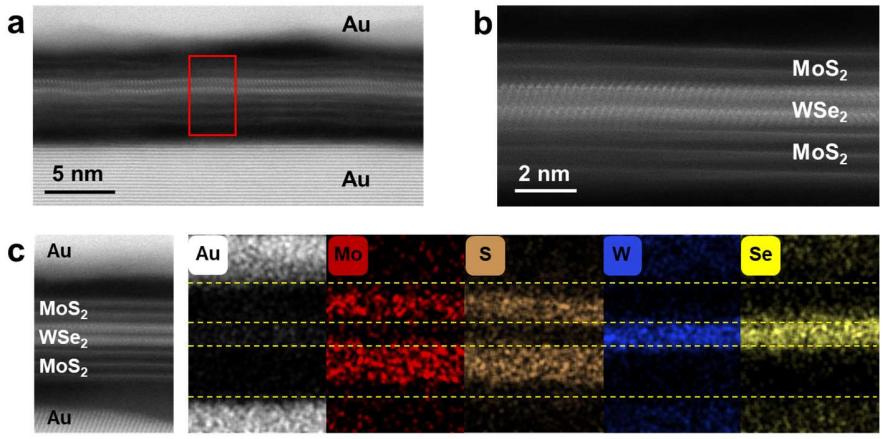


Figure 3.4 (a) Cross-sectional TEM image of the Au/MoS_2/WSe_2/MoS_2/Au device. (b) The detailed TEM image of the partial area in (a) shows the clear thickness of TMDs. (c) The corresponding elemental distribution mapping of the device.

3.2.2 The Selector with Metal/Semiconductor Junction

Firstly, we investigate the effects of the metal-semiconductor (M-S) junction characteristics of the selectors on the current and nonlinearity. Au electrodes form a Schottky barrier with MoS_2 and WSe_2. The carrier type of MoS_2 usually exhibits n-type, while the carrier type of WSe_2 depends on layer number and contact electrode methods.[21] Since the evaporated Au electrode has chemical interaction with WSe_2 and perturbs its electrical properties, only WSe_2 FETs with below 5 layers WSe_2 exhibit p-type characteristics.[22] However, the transferred metal electrode has negligible influence on the intrinsic properties of WSe_2.[23] The transport behavior of WSe_2 FETs with transferred Au electrode maintains p-type for WSe_2 less than 12 layers (**Figure 3.5**). Thus, we confirmed that the WSe_2 in our devices exhibits p-type.

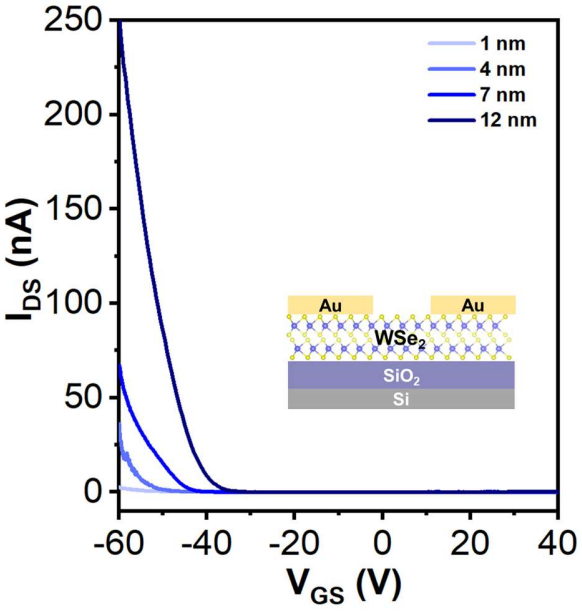


Figure 3.5 The I_{DS}–V_{GS} transfer characteristics of the WSe_2 transistor with different WSe_2 flake thickness using transferred Au electrodes (Device channel length = 3 μm). V_{DS} = 1 V. The inset is the cross-sectional schematic of the transistor.

We fabricate the Au/TMD/Au device with 6-nm TMD thickness. The thickness of TMD materials can refer to the optical microscopy image and AFM results, as shown in **Figure 3.6a, b**. The Au/MoS_2/Au device shows higher *J* of approximately 2×10^3 A/cm^2, smaller *NL* of around 8 at V_{TE} = 2 V, and turn-on voltage (V_{ON}) of 0.8 V (**Figure 3.6c**). As defined in Figure 1b, nonlinearity is the current at full-read voltage (V_R) divided by the current at V_R/2. We define the V_{ON} of the devices as the voltage at which the current reaches 10% of the maximum current.[24] In contrast, Au/WSe_2/Au device exhibits smaller *J* of about 200 A/cm^2 with larger *NL* of around 40 at V_{TE} = 2 V and V_{ON} of 1.3 V. This phenomenon is consistent with previous research.[25] The observed asymmetrical I-V curve for the WSe_2 device results from the asymmetrical metal–WSe_2 contact areas at the top and the bottom

contacts, which become moderate with decreasing thickness.[26] The different performances of WSe_2 and MoS_2 devices can be explained by several different current transport mechanisms, including thermal emission (TE), Fowler-Nordheim (FN) tunnelling, and direct tunnelling.[27] Usually, direct tunneling occurs at a very high field or ultrathin layers. Hence, when the thickness of the TMD layer is less than 2 nm, the device is shorted at low voltage (< 2 V) due to the short tunneling distance. For the device with thick TMD layer, the overall current consists of the contribution from both thermal emission current at low fields and field-assistant tunnelling current at high fields (FN tunnelling).

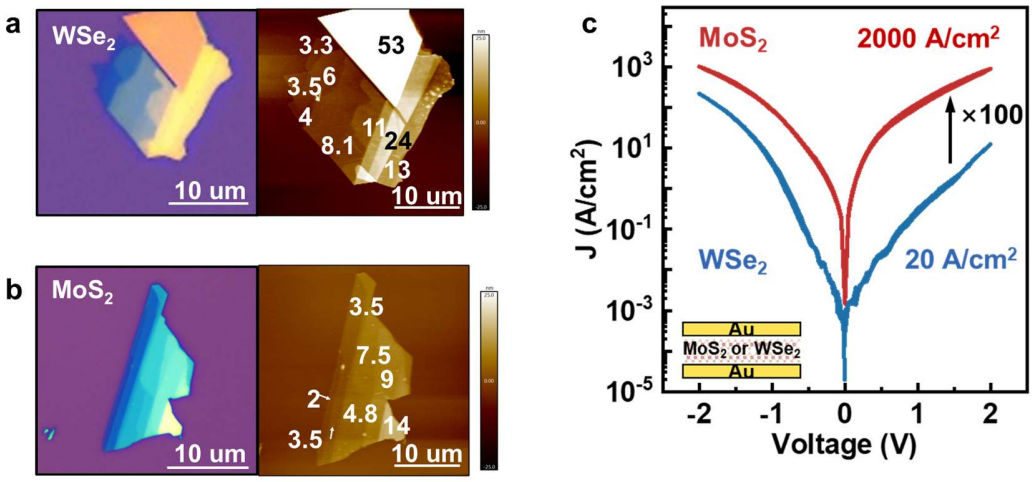


Figure 3.6 Optical microscopy image (left) and AFM image (right) of exfoliated (a) WSe_2 and (b) MoS_2 flakes, where the exact thicknesses are marked in white numbers as thickness references. (c) Current density (J) at sweep V_{TE} = ± 2 V for Au/MoS_2/Au (red line) and Au/WSe_2/Au (blue line) selectors, respectively. The inset is the cross-sectional schematic of the devices.

In Au/TMD/Au devices, Φ, SBH, plays a crucial role in different performances between WSe_2 and MoS_2 devices. As can be seen in **Figure 3.7a**, different transport mechanisms

dominate for different voltage conditions.[28] For E ≥ Φ, electrons travel from one contact electrode to another through thermal emission over the Schottky barrier between the Au electrode and the TMD. For lower energies, Φ > E > Φ − qV_{TE}, transport is facilitated by FN tunneling through barrier. As can be seen in **Figure 3.7b**, the I–V_{TE} curves of Au/WSe_2/Au and Au/MoS_2/Au devices are plotted in a log-log scale at T = 300 K. Under very small V_{TE}, as guided with black dashed lines, the current clearly shows a linear relationship with applied voltage ($I \sim V$). This behavior is consistent with the thermal emission equation with assumption barrier height (Φ) ≫ kT:[29]

$$I_{TH} = \frac{2q^2(2\pi)m}{h^3} \exp\left(-\frac{\Phi}{kT}\right) V_{TE} \quad (1)$$

where q is the electron charge, h is the Planck constant, m is free electron mass, k is the Boltzmann constant, and V_{TE} is the applied voltage on the top electrode (TE). Meanwhile, the thermal emission mechanism at low voltage has previously been confirmed by temperature-dependent measurements.[30]

According to Equation (1), the SBH is the key determining factor for the actual current level in the thermal emission region. Although the vdWss gap between the transferred Au electrode and TMDs reduces the Fermi level pinning, the surface states, originated from the fabrication process or metal-induced gap states, still contribute to the charge neutrality level (E_{CNL}) of TMDs.[31] The E_{CNL} of multilayer MoS_2 and WSe_2 are approximately 4.9 eV and 4.3 eV, respectively.[32] The bandgaps of n-type MoS_2 and p-type WSe_2 are 1.29 eV and 1.25 eV, respectively.[33] After metal and TMD are contacted, the smaller SBH for MoS_2 allows higher thermionic currents to be reached before the tunneling current dominates (**Figure 3.7d**).

The performance of WSe_2 and MoS_2 devices is also related to the Fowler Nordheim (FN) tunnelling at a large voltage. The vertical transport through insulator has been quantitatively analyzed in metal–insulator–metal by the following Equation:[34]

$$I_{FN} = \frac{q^3 V_{TE}^2}{8\pi h \Phi} \exp\left(-\frac{4(2m)^{\frac{1}{2}}\Phi^{\frac{3}{2}}}{3\hbar q V_{TE}}\right) \quad (2)$$

where m is the free electron/hole mass. **Figure 3.7c** shows I/V_{TE}^2 as a function of $1/V_{TE}$. At high fields (small $1/V_{TE}$ values), the curves can fit well with the signature of Equation (2) for both WSe_2 and MoS_2 devices, indicating that the charge transport is dominated by FN tunnelling. Meanwhile, smaller SBH leads to higher FN tunneling current. Combining with Equations (1) and (2), the relationship between *NL* and SBH in Au/TMD/Au devices can be indicated as the following Equation:

$$NL' = 4 \exp\left(\frac{4(2m)^{\frac{1}{2}}\Phi^{\frac{3}{2}}}{3\hbar q V_{TE}}\right) \quad (3)$$

where *NL'* represents the *NL* of Au/TMD/Au devices based on the FN tunneling mechanism under large bias. In summary, SBH plays a crucial role in *NL* and *J*. The transport characteristics of the vertical TMD device are dominated by thermal emission over an SBH Φ at low voltage bias and FN tunneling through at high voltage bias. The SBH between MoS_2 and Au is smaller than that of WSe_2, which leads to the higher current and lower nonlinearity of MoS_2 devices.

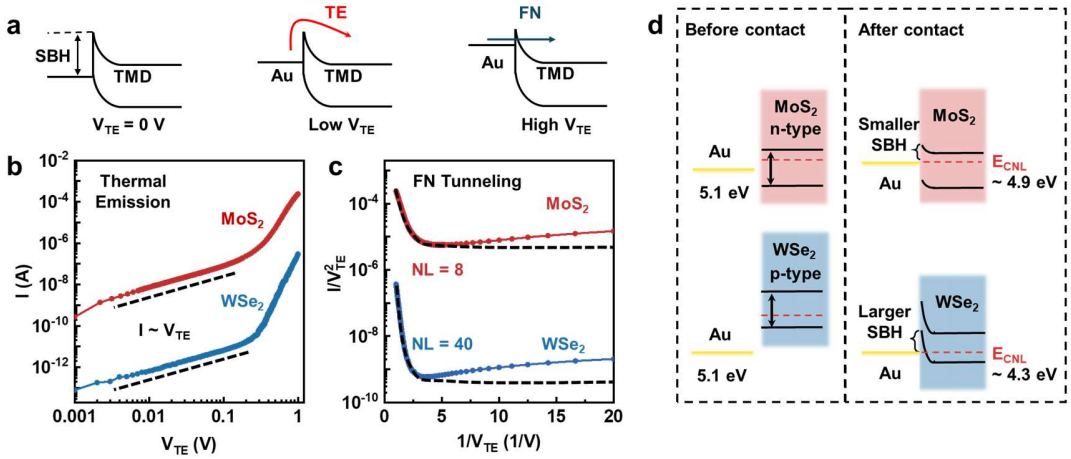


Figure 3.7 (a) The schematic energy band diagram for carrier transport based on different mechanisms under various voltage biases. SBH: Schottky barrier height; TE: Thermal Emission; FN: Fowler-Nordheim tunneling. (b) I vs. V_{TE} curves for Au/TMD/Au selectors plotted in log-log scale at 1 V. The dashed lines suggest that the current proportionally increases to the voltage bias I ~ V_{TE}. (c) $I/\boldsymbol{V_{TE}^2}$ vs. $1/V_{TE}$ plots, reconfigured from I–V_{TE} characteristics at room temperature for WSe_2 and MoS_2. The dashed curves indicate that FN tunneling is the leading quantum transport mechanism through the junctions. (d) Band alignment of multilayer MoS_2, WSe_2, and Au electrode before (left) and after (right) contact.

To confirm the robustness of the effect of SBH on *J* and nonlinearity, we measured over 15 devices and extracted selector parameters as a function of TMD thickness (**Figure 3.8**). The current densities of WSe_2 and MoS_2 devices, extracted at 1 V, decreased almost linearly with increasing thickness from 4 to 15 nm. The low out-of-plane mobility of WSe_2 may contribute to the steeper decrease slope as TMD thickness increases. It easily breaks down for devices below 4 nm before V_{TE} = 2 V, probably due to the avalanche multiplication in ultrathin TMD devices.[35] The nonlinearity extracted at 2 V remains stable (~8) for MoS_2

devices under various thicknesses; for the WSe_2 devices, it shows an upward trend since the SBH rises as the WSe_2 thickness increases.[36] Although the nonlinearity of thick WSe_2 device can satisfy the requirement of the ideal selector, the corresponding current is too low for the operation of RRAM cell. These data indicate a trade-off between J and nonlinearity for metal/TMD/metal devices. In other words, it is hard to achieve large J and high nonlinearity simultaneously.

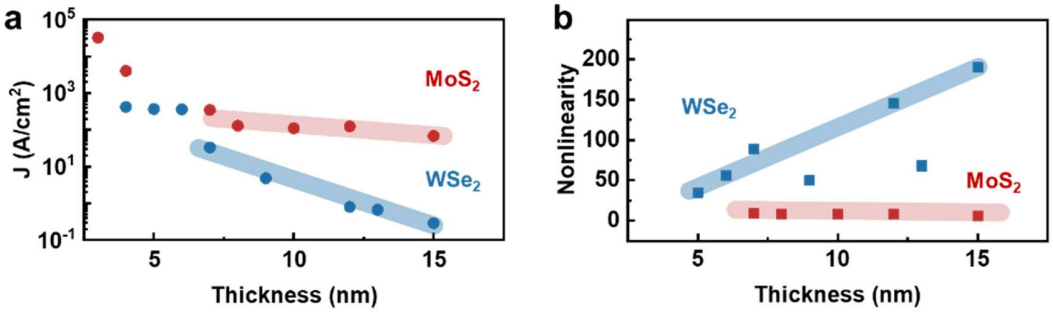


Figure 3.8 The (a) J and (b) nonlinearity of Au/TMD/Au selectors as a function of TMD thickness.

3.2.3 Electrical Characteristics of Heterojunction Selector

To suppress the current at low voltage and increase the nonlinearity of the selector, we introduce the built-in field of the MoS_2/WSe_2 p-n heterojunction (**Figure 3.9a**). As shown in **Figure 3.9b**, the device represents the I–V characteristics of Au/MoS_2 (3 nm)/WSe_2 (3 nm)/Au diode with rectification ratio ($|I_{-2V}/I_{+2V}|$) of ~ 230 and ideality factor (n) of 12 according to the following Equation (4):

$$I = I_0 \left[\exp\left(\frac{qV_{TE}}{nkT}\right) - 1\right] \quad (4)$$

Where I_0 is the reverse saturation current, q is the electron charge, V_{TE} is the voltage applied on the TE, T is room temperature (300 K), and k is the Boltzmann constant. The rectification ratio and ideality factor are similar to those reported by the previous study.[37] The unexpected ideality factor of our p-n device is higher than that of the ideal p-n junction (usually between 1 and 2), which may arise from the interface traps during the dry transfer process. These introduced traps at the interface may lead to the non-ideal forward current and the reverse leakage current, giving an unexpected rise in the ideality factor.[38,39]

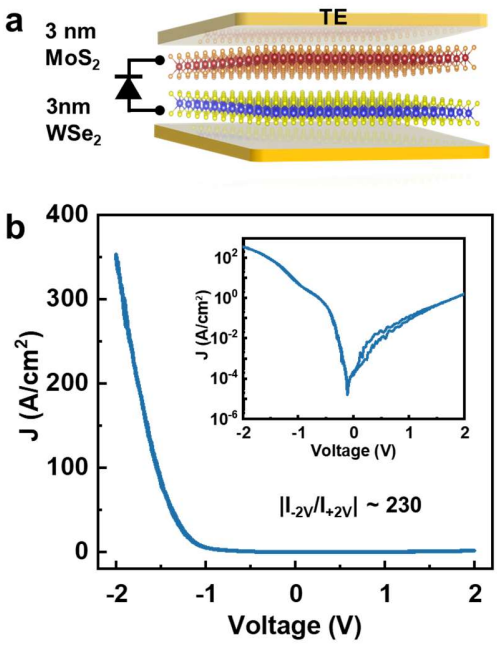


Figure 3.9 The characteristics of the heterojunction-based selector. (a) The schematic structure of the p-n diode. (b) *I-V* rectifying characteristic of the p-n diode. The inset shows *I-V* on a logarithmic scale.

When the negative bias is applied to the TE, the diffusion current from the majority carriers of TMDs and thinner depletion layer at the MoS_2/WSe_2 interface contribute to the high current (**Figure 3.10**).[40] Because of the low barrier at heterojunction, both diffusion

and drift current exist at large bias, contributing to higher current. However, the current is suppressed under positive bias due to the drift of minority carriers and increasing depletion layer. This heterojunction structure is expected to be applied in the selector to suppress the current at low bias.

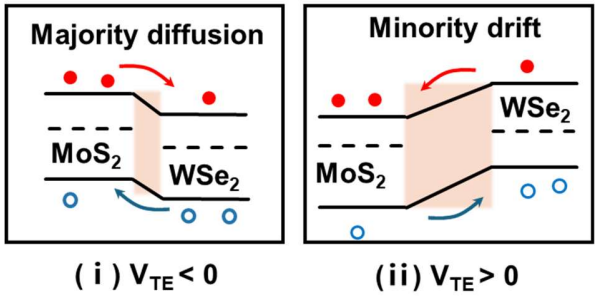


Figure 3.10 The mechanism of its rectifying behavior at (i) negative bias and (ii) positive bias.

Inspired by the characteristics of heterojunction, we designed the Au/MoS_2/WSe_2/MoS_2/Au n-p-n selector. The Au/MoS_2 (1.5 nm)/WSe_2 (2 nm)/MoS_2 (2.1 nm)/Au n-p-n selector exhibits the symmetrical current behavior with a high nonlinearity (*NL*, defined by $I_{TE}/I_{TE/2}$) of 235 and high *J* of 2.6×10^3 A/cm^2 with an area of 4.37 μm^2 (**Figure 3.11a**). Only the TMD layers at the cross-point area were kept, and other parts were etched by SF_6 Plasma. The optical microscopy image before and after SF_6 plasma can be seen in **Figure 3.11b**. The V_{ON} of 1.4 V suggests significant current suppression due to the barrier in heterojunction.[41]

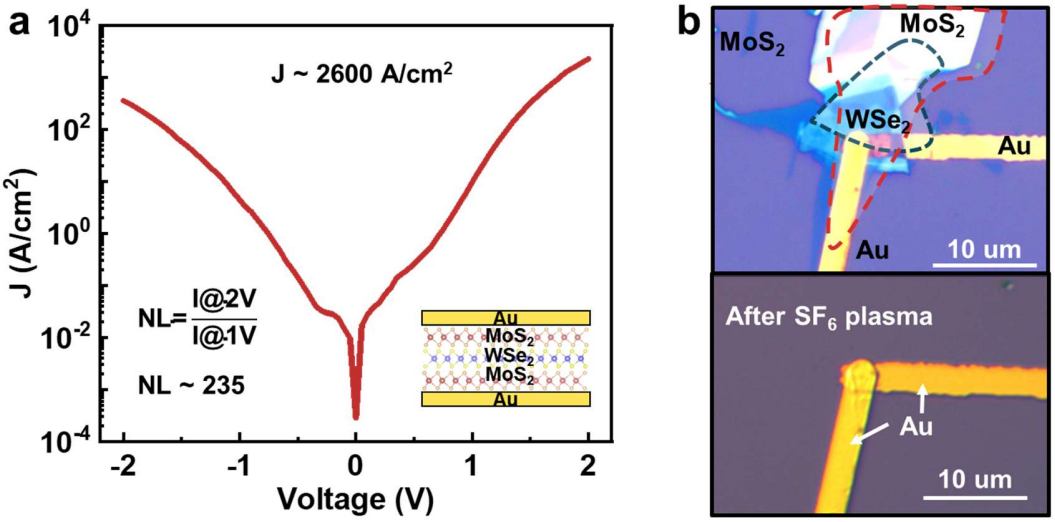


Figure 3.11 The characteristics of the n-p-n selector. (a) Current at V_{TE} = ± 2 V for the n-p-n selector. The inset: the schematic structure of the n-p-n selector. (b) The optical microscopy image of n-p-n selector before and after SF_6 plasma.

It is well-known that the punch-through mechanism easily happens in thin-base bipolar transistors resulting from the depletion-layer coupling.[42] By using a depletion model for the conventional p−n junction heterostructures, the depletion layer widths located in the MoS_2 and WSe_2 sides can be roughly estimated:[43]

$$x_p = \left[\frac{2N_d\varepsilon_1\varepsilon_2\varphi_b}{qN_a(\varepsilon_1N_a+\varepsilon_2N_d)}\right]^{0.5} \text{ and } x_p = \left[\frac{2N_a\varepsilon_1\varepsilon_2\varphi_b}{qN_d(\varepsilon_1N_a+\varepsilon_2N_d)}\right]^{0.5} \quad (5)$$

where N_d and N_a are the electron and hole doping concentrations, ε_1 and ε_2 are the dielectric constants of MoS_2 and WSe_2, and φ_b is the built-in potential at the heterojunction. Since the carrier concentration of WSe_2 (N_a ~ 10^{17} cm^{-3}) is typically 10 times smaller than that of MoS_2 (N_d ~ 10^{18} cm^{-3}),[44] which leads to a thicker depletion layer located on the WSe_2 side according to Equation (4). Once the effective middle layer width (W_{eff} in **Figure 3.12a**) of WSe_2 is thin enough, the punch-through mechanism can be facilitated, which

leads to a high current. Meanwhile, the electron-hole recombination in the effective WSe_2 region happens under low voltage bias, which can effectively suppress the current.

Hence, we hypothesize that the bidirectional high-nonlinear behavior with large current in the thin-WSe_2 n-p-n selector arises from the punch-through mechanism. As the increasing applied bias raises the barrier height of the p-n junction on one side, the depletion layer of one MoS_2/WSe_2 side can reach another region of the MoS_2/WSe_2 heterojunction, which means each side of MoS_2 is connected (**Figure 3.12b**). In this situation, the carriers can cross through the WSe_2 layer to reach the MoS_2 layer by thermionic emission or tunneling. The punch-through mechanism results in the similar *J* for Au/MoS_2/Au and Au/MoS_2/WSe_2/MoS_2/Au devices, which can be seen in **Figure 3.12c**.

In addition, the barrier height can be inferred by the yellow dashed line extracted from the FN tunneling curve in **Figure 3.12d**. The Au/MoS_2/Au device has a large $1/V_{TE}$, corresponding to the smaller barrier height (φ_{SBH}), and the Au/MoS_2/WSe_2/MoS_2/Au has a higher barrier potential ($\varphi_{SBH} + \varphi_{p\text{-}n}$), which results from the built-in field in heterojunction. As a result, due to the punch-through, the n-p-n selector presents a large current and nonlinearity.

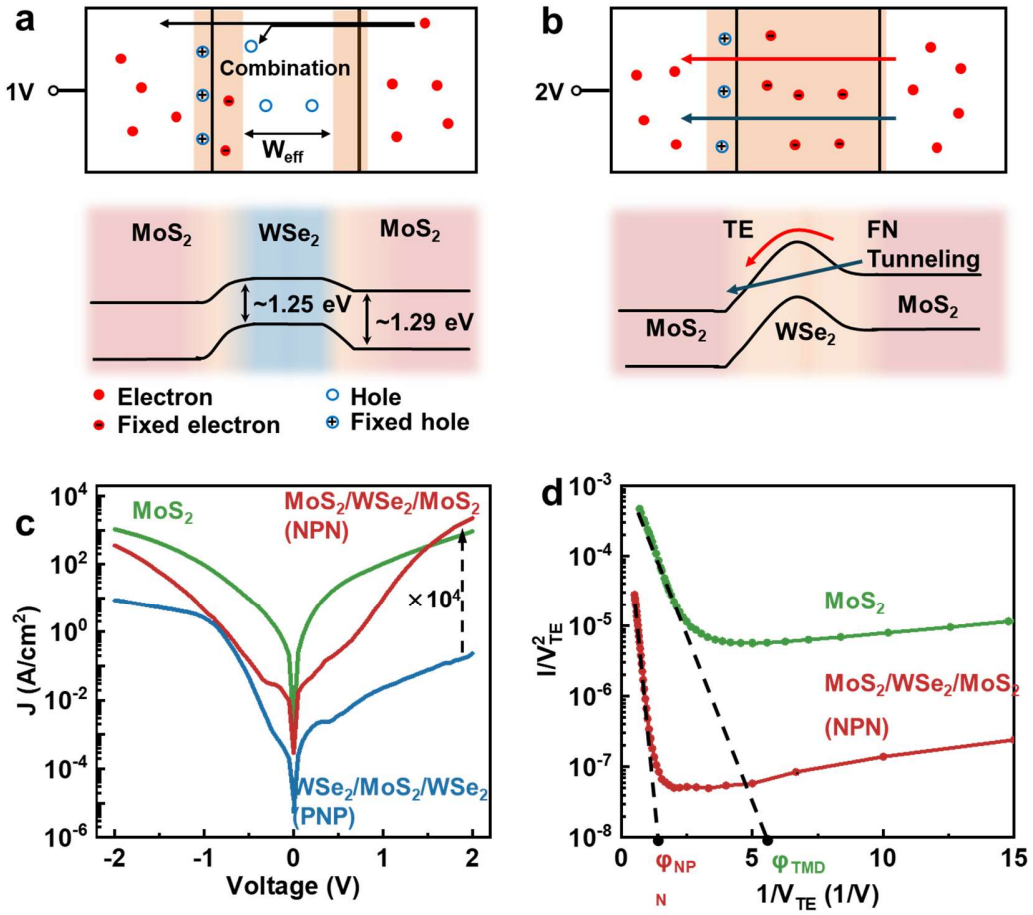


Figure 3.12 The punch-through mechanism in n-p-n selectors. Distribution of depletion layer with 2 nm WSe_2 region and band diagram of n-p-n selector under (a) low and (b) high voltage bias. The effective middle layer width is marked as W_{eff}. (c) *I-V* characteristic of Au/MoS_2/Au (orange), Au/MoS_2/WSe_2/MoS_2/Au (red), and Au/WSe_2/MoS_2/WSe_2/Au (blue) devices with similar total TMD thickness, respectively. (d) $I/\boldsymbol{V_{TE}^2}$ vs. $1/V_{TE}$ plots, reconfigured from the $I–V_{TE}$ curve in (c) for MoS_2 (orange) and n-p-n (red) devices. The x-intercepts of dashed lines indicate the relative barrier height for devices.

In addition, in **Figure 3.12c**, the Au/WSe_2/MoS_2/WSe_2/Au p-n-p device exhibits poor performance ($J \sim 0.2$ A/cm^2 and $NL \sim 10$) since the depletion region located on the MoS_2

side is much thinner than that of WSe_2 and the higher SBH of Au/WSe_2. Therefore, the Au/WSe_2/MoS_2/WSe_2/Au p-n-p device is difficult to achieve the punch-through mechanism. To further confirm the mechanism, we fabricated p-n-p device with different thicknesses (**Figure 3.13a**). Notably, the common asymmetrical current behaviors at positive and negative bias in the symmetrical p-n-p devices are observed, which can be explained by different charge transfer mechanisms (**Figure 3.13b**). Since the SBH of Au/WSe_2 is much higher than that of the Au/MoS_2, the resistance brought from SBH of Au/WSe_2 cannot be ignored compared to resistance from the p-n barrier. Hence, the SBH of Au/WSe_2 decreases at negative bias (-2 V), resulting in a larger current density in contrast to that at positive bias (2V) (**Figure 3.13c**). Another interesting phenomenon is that under negative bias, the current increases more rapidly than the current under positive bias and the current rise becomes slow at around -1 V, which can be explained by the decreases in SBH and p-n barrier height and the thinner depletion layer on the MoS_2 side and the thicker depleted layer on the one WSe_2 side at around -1 V (**Figure 3.13d**), respectively.[45]

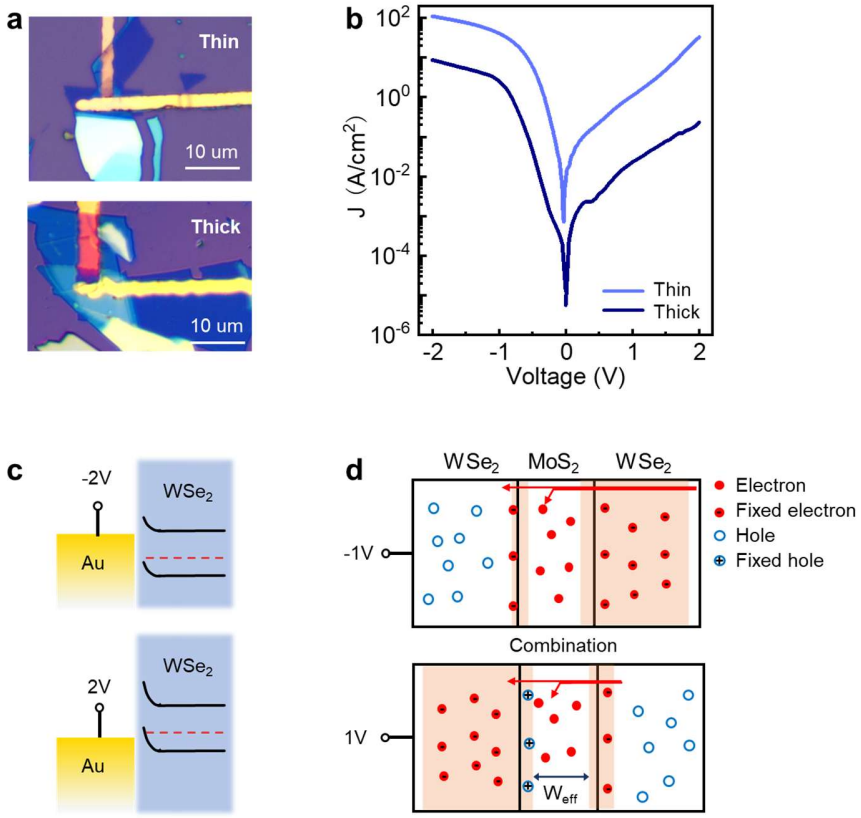


Figure 3.13 The analysis of Au/WSe_2/MoS_2/WSe_2/Au p-n-p selectors. (a) The optical image of the p-n-p selectors with different thicknesses. (b) I-V characteristic of p-n-p selectors. The inset shows the schematic symbol for p-n-p selectors. (c) Distribution of depletion layer under positive bias (1 V) and negative bias (-1 V). The effective base width is marked as W_{eff}. (d) Band alignment of multilayer WSe_2 and Au electrode under positive bias (2 V) and negative bias (-2 V).

To further demonstrate that the punch-through mechanism happens in thin-WSe_2 bipolar heterojunction, we fabricate the n-p-n selectors with various base thicknesses (**Figure 3.14a**). The device with monolayer WSe_2 is easy to break down at 0.375 MV/cm. Since the depletion layer can overlap the monolayer WSe_2 region without electron-hole combination,

this device presents low nonlinearity below 30. Thicker WSe_2 contributes to lower *J* under V_{TE} = 2 V because more electron-hole combination occurs in the thick p-type region. In **Figure 3.14b**, the linearity of the curve in the elliptical area indicates whether the FN tunneling occurs in the n-p-n selector under V_{TE} = 2 V or not. Since the FN tunneling happens when two depletion layers connect, the nonlinear curves at small $1/V_{TE}$ for selectors with thicker WSe_2 suggest that the devices cannot achieve FN tunneling based on punch through mechanism at 2 V due to the large WSe_2 region. **Figure 3.15** shows more details about control devices. In **Figure 3.15f**, the n-p-n device with the 21-nm-thickness WSe_2 region can exhibit high nonlinearity of about 153 under high voltage (4 V), suggesting that the depletion layers of each MoS_2/WSe_2 side have connected.

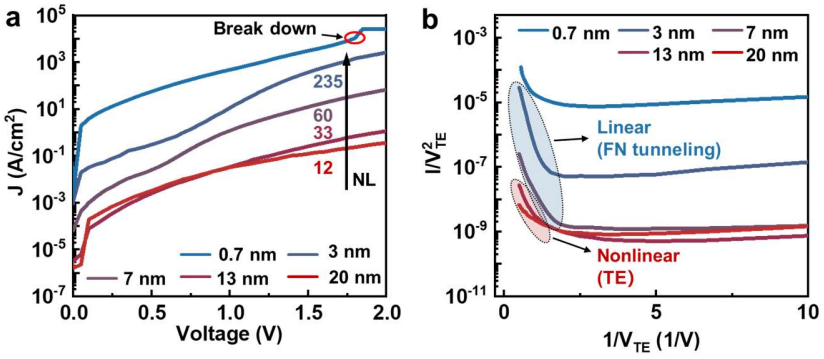


Figure 3.14 (a) *J*-V_{TE} for Au/MoS_2/WSe_2/MoS_2/Au selectors with different WSe_2 (base) thicknesses. (b) I/V_{TE}^2 vs. $1/V_{TE}$ plots, reconfigured from I–V_{TE} characteristics at T = 300 K for devices under various WSe_2 thicknesses. From the red to blue elliptical area, the curves change from nonlinear to linear with the decreased WSe_2 thickness.

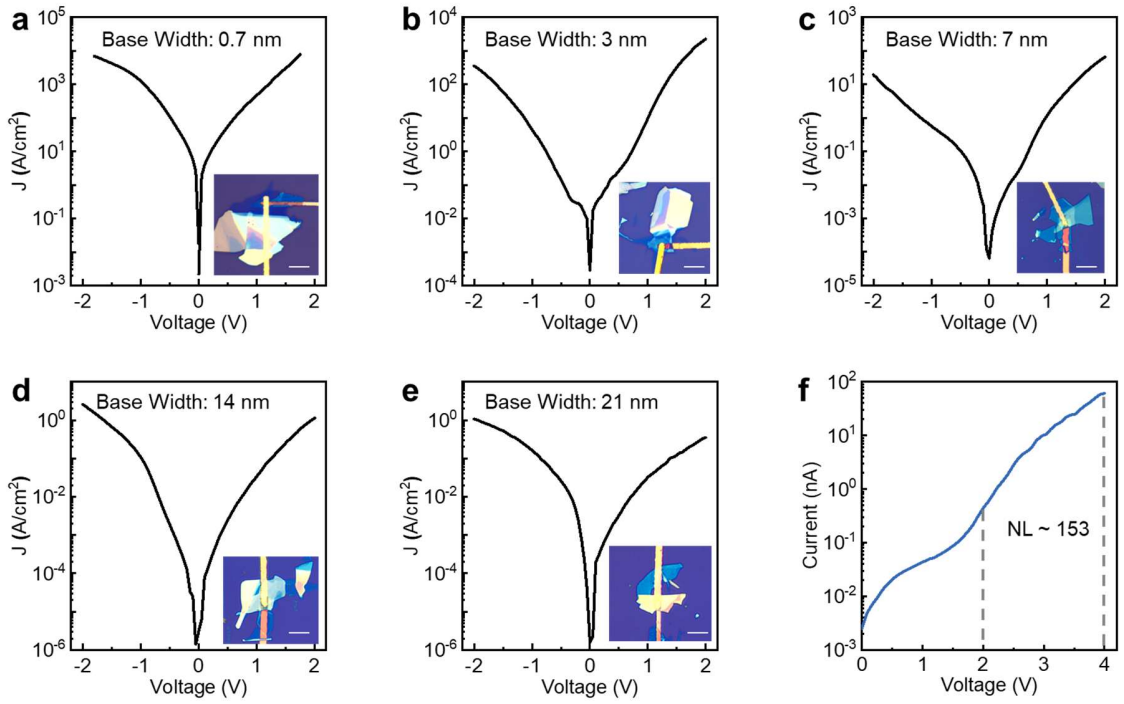


Figure 3.15 The current at V_{TE} = ±2 V and optical image of selectors with different WSe_2 (base) thicknesses: (a) monolayer WSe_2 (~0.85 nm), (b) 2 nm (our selector in the article), (c) 7 nm, (d) 13 nm, and (e) 21 nm. (f) The n-p-n device with thick WSe_2 (21 nm) under high voltage presents high nonlinearity based on the punch-through mechanism.

Figure 3.16 compares the n-p-n selector performance to the other structures and recent TMD-based selector work. The previous simulation suggested that by introducing the h-BN monolayer as a tunneling barrier, the selector with three monolayer 2D materials (h-BN/MoS_2/h-BN) can achieve high on-current due to the short tunneling distance and maintain high nonlinearity. However, the experimental results presented that this selector suffers from low on-current and low nonlinearity of 25.[46] The Au/TMD/Au devices confirm that devices without heterojunctions are insufficient for selector applications. The poor performance of the p-n-p device and other n-p-n devices with thicker middle layer suggests

that the material and thickness of the middle layer play a vital role in the punch-through mechanism for the heterojunction-based selector. The limitations of these device designs highlight the benefits of the n-p-n selector with 2D heterojunctions based on a punch-through mechanism.

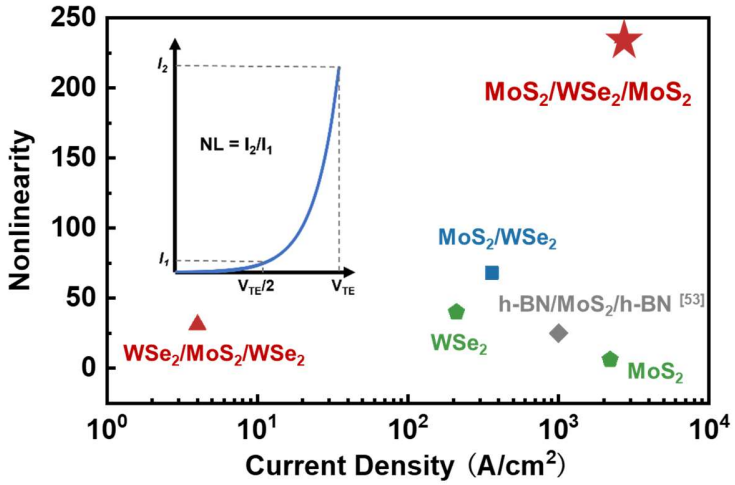


Figure 3.16 The comparison of selectors with different structures and materials on nonlinearity and current density at $V_{TE} = \pm 2$ V.

3.2.4 1S1R Integration with 2D Memory and Selector

To demonstrate the feasibility of integrating the proposed selector device with the RRAM cell, we connected the n-p-n selector and h-BN RRAM and electrically tested this 1S1R cell to verify successful 1S1R functionality. **Figure 3.17a** presented the Au/h-BN/Au RRAM, where a 10-nm-thick h-BN flake was treated by O_2 plasma to introduce the defects. The h-BN RRAM cell with Au as symmetrical transferred electrodes exhibits unipolar behavior and large operation voltage (> 2V) **Figure 3.17b**). The Au/h-BN/Au RRAM

initially showed an electroforming process at around 5V. After this, the device performed stable, non-volatile unipolar RS between HRS and LRS (**Figure 3.17c**). The metallic ions (Au^+) penetration along defects plays a crucial role in resistance transition. A few conductive filaments trigger the SET transition, while the reset phenomenon is related to thermal effect-induced CF rupture.[47] However, the Au/h-BN/Au RRAM is unsuitable for our selector due to its ultrahigh uncontrollable RESET current (usually 1 mA), even different I_{CC} set. Hence, we adopt the RRAM cell with Ag/h-BN/Au structure (**Figure 3.17d**). The h-BN RRAM based on Ag filament exhibits low RESET current to melt the conductive filaments and maintain controllable low resistance under different I_{CC} due to the lower activation energy of the Ag ionic diffusion (**Figure 3.17e**). Ag/h-BN/Au RRAM performs bipolar and low power switching operations based on the formation/rupture of Ag conductive filaments, which is a promising candidate for 1S1R integration(**Figure 3.17f**).[48,49] Therefore, considering the low operation voltage and current, asymmetrical transferred electrodes are adopted for h-BN RRAM.[50]

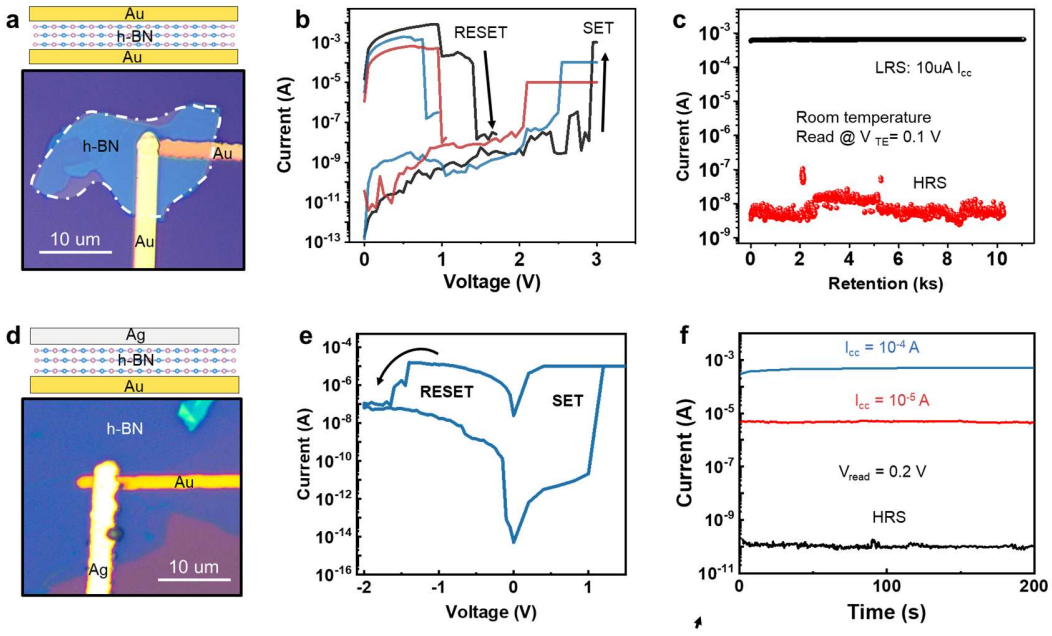


Figure 3.17 (a) The schematic structure and optical microscope image of Au/h-BN/Au RRAM. (b) The non-volatile nonpolar I–V curves for devices under different compliance currents (I_{CC}). (c) Retention characteristic of Au/h-BN/Au memory under room temperature. (d) The schematic structure and optical microscope image of Ag/h-BN/Au RRAM. (e) The non-volatile nonpolar I–V curve for Ag/h-BN/Au. (f) Retention characteristic of Ag/h-BN/Au memory under room temperature.

Figure 3.18 presents the I–V characteristic of the 1S1R device under compliance current (I_{cc}) of 10^{-5} A, where the 1S1R cell exhibits a nonlinearity of 10 (20) for a one-half (one-third) biasing scheme at V_R = 0.8 V with a memory window of around 250. Notably, the nonlinearity of the selector series with an RRAM cell is reduced by an amount depending on the relative resistance of the RRAM cell and the selector.[51] The nonlinearity is expected to be further improved when the I_{cc} increases (the operating voltage will increase) since the nonlinearity of our selector can be over 200 at V_R = 2 V.

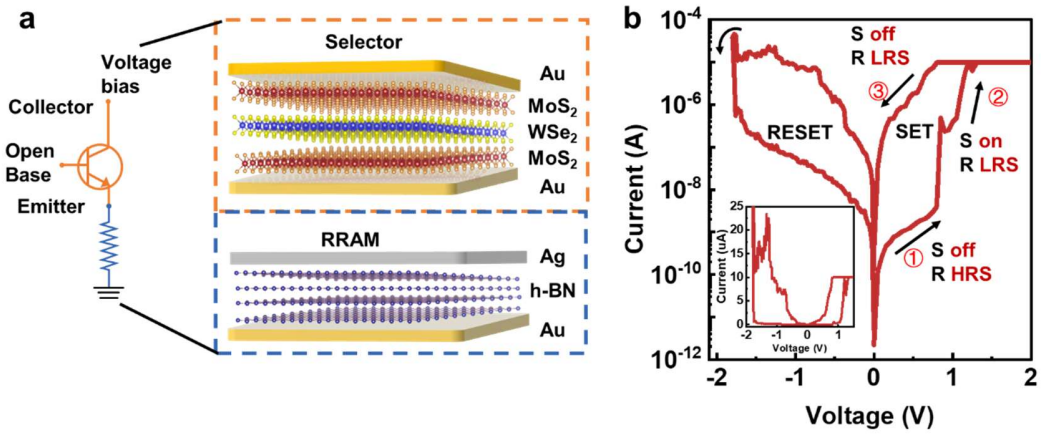


Figure 3.18 (a) The schematic structure of the proposed 1S1R cell with TMD-based heterojunction selectors and h-BN-based RRAM. (b) *I–V* characteristics of 1S1R cell. Inset: Nonlinear *I–V* characteristics of 1S1R device on a linear scale.

To determine the maximum size of the 1S1R array, we investigated the read margin for this 1S1R cell under two reading voltage schemes ($V_R/2$ and $V_R/3$ schemes), which can be seen in **Figure 3.19a**. We assumed a "worst-case scenario" in which all unselected cells are in LRS states. Under this circumstance, the maximum size of the 1S1R array corresponding to at least a 10% read margin was estimated based on a conventional one bit-line pull-up (One-BLPU) scheme (**Figure 3.19b**).[52] Based on the equivalent circuit of the one BLPU scheme, the density of 1S1R array depends on the read margin, which is defined by the following Equation (6) according to the Kirchhoff Equation:[53]

$$\frac{\Delta V_{Read}}{V_P} = \frac{R_P}{[R_S^{on}|\left|\frac{2R_{LRS}^F}{N-1}+\frac{R_{LRS}^R}{(N-1)^2}\right|+R_P} - \frac{R_P}{[R_S^{off}|\left|\frac{2R_{LRS}^F}{N-1}+\frac{R_{LRS}^R}{(N-1)^2}\right|+R_P} \quad (6)$$

The calculated readout margin dependent on the voltage drop (ΔV_{Read}) across the pull-up resistor (R_P) reflects how correctly the switching state on the selected cell is read. The $\Delta V_{Read}/V_P$ (pull-up voltage) of 10% is a minimum criterion to differentiate cells' switching

state, which determines the maximum number (N) of word/bit lines. In our case, the R^{F}_{LSR} and R^{R}_{LSR} represent the unselected cells at low resistance state (LRS) under forward and reverse voltage, respectively, determining the total resistance through the sneak path. Assumed that R_P is equal to R_{LRS} at 0.8 V to achieve maximum read margin. R_S^{on} and R_S^{off} are treated as the selected cell in the ON and OFF state at 0.8 V. The resistances of resistors on the sneak path can be obtained at ± 0.4 V or ± 0.25 V according to the read voltage scheme. According to the equation (5), the maximum size of the 1S1R array shown in **Figure 3.19c** can be increased from 0.3 Kbit under the $V_R/2$ scheme to 6.5 Kbit under the $V_R/3$ scheme since the current of unselected cells is further suppressed under the $V_R/3$ scheme. The maximum crossbar size (number of word/bit lines) of 82 can be further enhanced by adjusting the relative resistance of the RRAM cell and the selector.

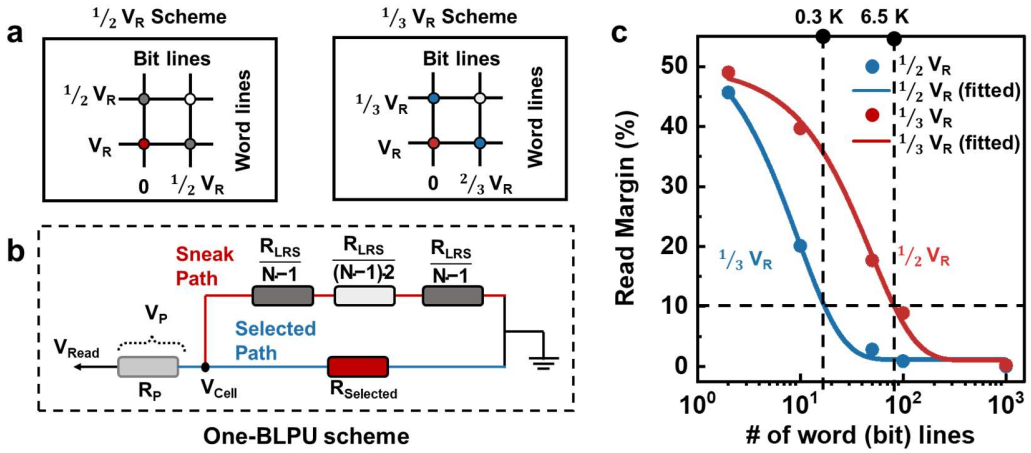


Figure 3.19 The estimated readout margin for 1S1R array with TMD-based heterojunction selectors. (a) The schematic illustration of 2 × 2 crossbar array for 1S1R in $V_R/2$ (left) and $V_R/3$ (right) schemes. Each colored circle represents the current at the corresponding voltage (V_R, $V_R/2$, 0, $V_R/3$, and $-V_R/3$). (b) Equivalent circuit of the 1S1R crossbar array

based on one bit-line pull-up (One-BLPU) scheme. (c) Estimated readout margin under $V_R/2$ (blue) and $V_R/3$ (red) schemes.

3.3 Conclusion

Based on the punch-through mechanism, we have demonstrated a 2D n-p-n selector with high nonlinearity. The proposed structure (Au/MoS_2/WSe_2/MoS_2/Au) could provide a high J of 2×10^3 A/cm^2 due to the low SBH of Au/MoS_2 while maintaining a high nonlinearity of above 200 based on the punch-through mechanism. By designing control devices, we first discovered the effects of SBH on the nonlinearity and current of metal/TMD/metal devices. Higher SBH contributes to low current but high nonlinearity. By introducing the p-n heterojunction and choosing the appropriate SBH, the WSe_2 with 2 nm thickness was carefully selected to achieve punch-through mechanism, leading to excellent selector performance. Furthermore, we integrate this memory selector with h-BN RRAM to form the 1S1R structure. The 1S1R cell performs a large memory window of 250, while its nonlinearity needs to be improved by adjusting the relative resistance of the RRAM cell and the selector. Considering the 10% read margin, the maximum crossbar size was estimated to be 6.5 Kbit. Therefore, our designed ultra-thin 2D n-p-n selector shows high potential for achieving large-scale, high-density 1S1R chip.

3.4 Experimental Section

3.4.1 Materials Characterization

Surface morphology and composition of TMDs were characterized by an optical microscope, atomic force microscope (Bruker), and confocal Raman

characterization (WITec Alpha300 Raman) using the excitation laser with a wavelength of 532 nm. STEM imaging and EDX mapping were carried out on a Thermo Fisher Spectra 300 TEM/STEM equipped with an X-FEG electron source, a CEOS SCORR fifth-order probe corrector, and a Super-X G2 detector, operated at 300 kV.

3.4.2 FIB Sample Preparation

The thin Cr/Au (5/10 nm) film is e-beam evaporated in advance on the devices to avoid the disappearance of transferred electrodes during Pt sputtering in FIB cross-sectioning. Cross-sectional TEM samples were acquired using a dual-beam focused ion beam-scanning electron microscope (SEM-FIB, Thermo Fisher Helios 5 CX) system with Pt coating applied on the devices.

3.4.3 Selector Device Fabrication

We adopted a fabrication flow integrating several 2D material transfer techniques to construct our ultrathin heterojunction.

The top electrode (TE) of 50-nm-thick Au was patterned by photolithography and evaporated on Si substrate since silicon has weaker adhesion with Au than SiO_2, followed by the lift-off process for photoresist (that is via acetone bath for a few minutes). (Step 1 of **Figure 3.20a**).[28] Au bottom electrode (BE) was patterned onto the top of 300-nm-thick SiO_2/p++ Si substrate by photolithography, followed by evaporation of a 35 nm thick Au layer on top of 15 nm Ti adhesion layer. Note that all the metallization was done by electron-beam evaporator under about 10^{-6} torr.

The few-layer MoS_2 or WSe_2 are mechanically exfoliated from bulk crystals onto a SiO_2 substrate using scotch tape. Note that the thickness and shape of electrode and TMD

materials are carefully selected in advance through an optical microscope to avoid the overlap of top and bottom TMD material in heterojunction. We then fabricate a transfer stamp by adhering a flat Polydimethylsiloxane (PDMS) layer to a glass slide (the size of the PDMS is around 5 × 5 mm) and then stick Polyvinyl alcohol (PVA) on top of the PDMS. PVA film was prepared by drying 4% PVA aqueous solution on a disk in the air for 2 days.

Next, we gently adhered the transfer stamp to the TE and the selected few-layer TMD materials in sequence (Step 2-5 of **Figure 3.20a**) using a rotationally aligned dry transfer technique. Examining every 2D material layer under the optical microscope is necessary to ensure no visible tears or wrinkles in the heterojunction area since these would lead to poor interface quality (**Figure 3.20b**). Then, we bake the stamp at 60 ºC for 2 minutes to strengthen the interaction between the target material and PVA and lift off TMD. Note that the baking temperature and time are essential for successful fabrication. The transfer heating temperature is confirmed according to the glass transition temperature (T_g) of PVA and the adhesion of PDMS.[54] Next, we adhere the stamp with the Au/MoS_2/WSe_2/MoS_2 stack onto SiO_2/Si substrate with Au BE (Step 6 of **Figure 3.20a**).[55].

After the n-p-n stack is finished, the resulting heterojunction area is defined by the overlap of the Au top and BE (**Figure 3.20a**, Step 6). Finally, we baked the sample on a hot plate for 5 minutes at 70 ºC to weaken the PDMS adhesion to PVA. As a result, the Au/MoS_2/WSe_2/MoS_2/Au structure and the PVA will stay on the target chip when we pull up the stamp. Finally, we soak the target chip in deionization water for 30 minutes to remove PVA and then rinse it with isopropyl alcohol (IPA). Note that due to the tight interaction between TMD and metal, the heterojunction was fabricated by dry transfer

method in one step, which avoided the inevitable PVA self-assembly layer after washing the PVA layer.[56]

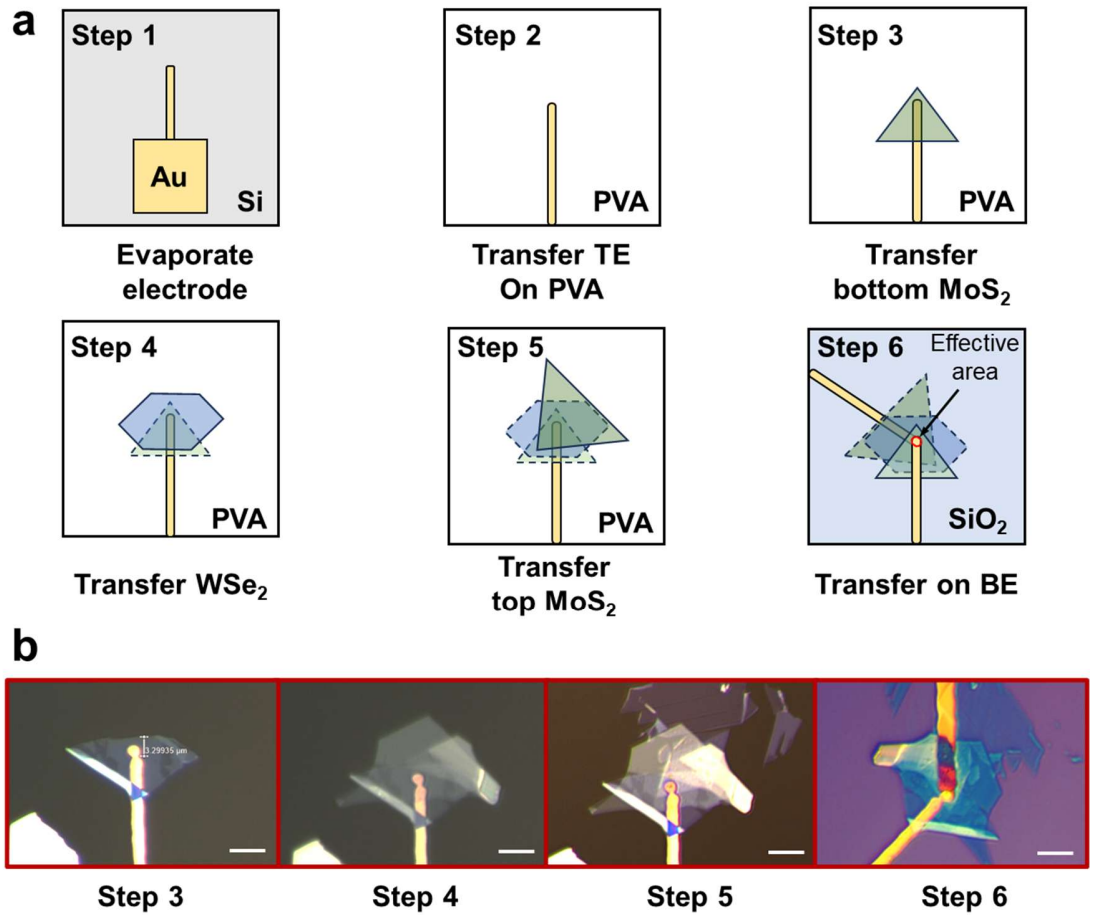


Figure 3.20 Fabrication process flow of the vertical device. (a) Schematics of the fabrication steps for the n-p-n device. Steps 2-6 are the enlarged schematic image of the Au electrode's tip in step 1 after TE is transferred on PVA. TE: top electrode; BE: bottom electrode. (b) The optical image of one n-p-n device as example after Steps 3-6. Scale Bar: 10 μm.

At last, the device is subjected to SF_6 dry plasma through inductively coupled plasma (ICP) to etch the MoS_2 and WSe_2 multilayers under 35°C for a few minutes.[57] The plasma

was excited with the precursor gases of SF_6 fed at the flow rates of 50 sccm under 180 mTorr.[58] The n-p-n device is then finished and ready for electrical measurements.

3.4.4 RRAM Device Fabrication

The Au/h-BN/Au RRAM was fabricated on the 300-nm-thick SiO_2/p^{++} Si substrate through the same transfer process. First, the Au/Ti BE and Au TE were prefabricated using a methodology identical to that for electrode fabrication of the selector (photolithography, metal evaporation, and lift-off). Then, the few-layer h-BN materials were exfoliated from bulk crystals. To create the defects inside h-BN for further formation of conductive filaments, the exfoliated h-BN was etched by O_2 plasma for 15s using PVA TePla Plasma Cleaner.[59,60] Finally, the Au or Ag TE and layered h-BN stack were transferred sequentially onto the Au/Ti BE. Several samples with different h-BN thicknesses and metallic electrodes were fabricated.

3.4.5 Electrical Characterization

The electrical characterizations were conducted using a probe station from Micromanipulator (Model 450PM) connected to a Keithley 4200 Semiconductor Characterization System (SCS). All the electrical characterizations were performed under the atmosphere in the dark. In all devices, the voltage stress was applied to the TE, keeping the BE grounded.

3.5 Reference

1 Zhou, F. *et al.* Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nature Nanotechnology* **14**, 776-782, doi:10.1038/s41565-019-0501-3 (2019).

2 Shulaker, M. M. *et al.* Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature* **547**, 74-78, doi:10.1038/nature22994 (2017).

3 Yu, S. & Chen, P.-Y. Emerging Memory Technologies: Recent Trends and Prospects. *IEEE Solid-State Circuits Magazine* **8**, 43-56, doi:10.1109/mssc.2016.2546199 (2016).

4 Sun, L. *et al.* Self-selective van der Waals heterostructures for large scale memory array. *Nat Commun* **10**, 3161, doi:10.1038/s41467-019-11187-9 (2019).

5 Huang, W., Hang, P., Yang, D., Yu, X. & Li, X. a. Self-rectifying-based memristors for neuromorphic computing. *Science Bulletin* **67**, 1213-1216, doi:https://doi.org/10.1016/j.scib.2022.05.008 (2022).

6 Yu, H. Y. & Tran, X. A. in *2012 IEEE 11th International Conference on Solid-State and Integrated Circuit Technology.* 1-4.

7 Sivan, M. *et al.* All WSe2 1T1R resistive RAM cell for future monolithic 3D embedded memory integration. *Nat Commun* **10**, 5201, doi:10.1038/s41467-019-13176-4 (2019).

8 Kim, S. *et al.* Latch-up based bidirectional npn selector for bipolar resistance-change memory. *Applied Physics Letters* **103**, 033505, doi:10.1063/1.4813832 (2013).

9 Bishop, M. D., Wong, H. S. P., Mitra, S. & Shulaker, M. M. Monolithic 3-D Integration. *IEEE Micro* **39**, 16-27, doi:10.1109/mm.2019.2942982 (2019).

10 Wang, J. *et al.* Low-Power Complementary Inverter with Negative Capacitance 2D Semiconductor Transistors. *Adv. Funct. Mater.* **30**, 2003859, doi:https://doi.org/10.1002/adfm.202003859 (2020).

11 Liu, Y. *et al.* Promises and prospects of two-dimensional transistors. *Nature* **591**, 43-53, doi:10.1038/s41586-021-03339-z (2021).

12 Cheng, R., Yin, L., Wang, F., Wang, Z. & He, J. Van der Waals integration of 2D atomic crystals for advanced multifunctional devices. *Science Bulletin* **64**, 1033-1035, doi:https://doi.org/10.1016/j.scib.2019.03.026 (2019).

13 Liu, B. *et al.* High-Performance WSe2 Field-Effect Transistors via Controlled Formation of In-Plane Heterojunctions. *ACS Nano* **10**, 5153-5160, doi:10.1021/acsnano.6b00527 (2016).

14 Jiang, J. *et al.* Van der Waals epitaxial growth of two-dimensional PbSe and its high-performance heterostructure devices. *Science Bulletin* **67**, 1659-1668, doi:https://doi.org/10.1016/j.scib.2022.07.005 (2022).

15 Jiang, Z. *et al.* in *2018 IEEE Symposium on VLSI Technology.* 107-108.

16 Zhang, Y. *et al.* Construction of Ru, O Co-Doping MoS2 for Hydrogen Evolution Reaction Electrocatalyst and Surface-Enhanced Raman Scattering Substrate: High-Performance, Recyclable, and Durability Improvement. *Adv. Funct. Mater.* **33**, 2210939, doi:https://doi.org/10.1002/adfm.202210939 (2022).

17 Zhao, W. *et al.* Lattice dynamics in mono- and few-layer sheets of WS2 and WSe2. *Nanoscale* **5**, 9677-9683, doi:10.1039/c3nr03052k (2013).

18 del Corro, E. *et al.* Excited Excitonic States in 1L, 2L, 3L, and Bulk WSe2 Observed by Resonant Raman Spectroscopy. *ACS Nano* **8**, 9629-9635, doi:10.1021/nn504088g (2014).

19 Chiu, M.-H. *et al.* Spectroscopic Signatures for Interlayer Coupling in MoS2–WSe2 van der Waals Stacking. *ACS Nano* **8**, 9649-9656, doi:10.1021/nn504229z (2014).

20 Doan, M. H. *et al.* Charge Transport in MoS2/WSe2 van der Waals Heterostructure with Tunable Inversion Layer. *ACS Nano* **11**, 3832-3840, doi:10.1021/acsnano.7b00021 (2017).

21 Wang, J. *et al.* High Mobility MoS2 Transistor with Low Schottky Barrier Contact by Using Atomic Thick h-BN as a Tunneling Layer. *Advanced Materials* **28**, 8302-8308, doi:https://doi.org/10.1002/adma.201602757 (2016).
22 Wang, J. *et al.* in *2018 IEEE International Electron Devices Meeting (IEDM).* 22.23.21-22.23.24.
23 Lin, Z., Huang, Y. & Duan, X. Van der Waals thin-film electronics. *Nature Electronics* **2**, 378-388, doi:10.1038/s41928-019-0301-7 (2019).
24 Wang, C. H. *et al.* Ultrathin Three-Monolayer Tunneling Memory Selectors. *ACS Nano* **15**, 8484-8491, doi:10.1021/acsnano.1c00002 (2021).
25 Zhang, F., Zhu, Y. & Appenzeller, J. in *2017 75th Annual Device Research Conference (DRC).* 1-2.
26 Kim, S. *et al.* Highly Efficient Experimental Approach to Evaluate Metal to 2D Semiconductor Interfaces in Vertical Diodes with Asymmetric Metal Contacts. *ACS Appl Mater Interfaces* **13**, 27705-27712, doi:10.1021/acsami.1c07905 (2021).
27 Chen, H., Zhou, Y. & Han, S. T. Recent advances in metal nanoparticle-based floating gate memory. *Nano Select* **2**, 1245-1265, doi:10.1002/nano.202000268 (2021).
28 Liu, Y. *et al.* Approaching the Schottky-Mott limit in van der Waals metal-semiconductor junctions. *Nature* **557**, 696-700, doi:10.1038/s41586-018-0129-8 (2018).
29 L., F. R. H. a. N. Electron emission in intense electric fields. *Proceedings of the Royal Society of London. Series A* **119**, 173-181, doi:http://doi.org/10.1098/rspa.1928.0091 (1928).
30 Zhu, Y., Zhou, R., Zhang, F. & Appenzeller, J. Vertical charge transport through transition metal dichalcogenides - a quantitative analysis. *Nanoscale* **9**, 19108-19113, doi:10.1039/c7nr05069k (2017).
31 Kim, C. *et al.* Fermi Level Pinning at Electrical Metal Contacts of Monolayer Molybdenum Dichalcogenides. *ACS Nano* **11**, 1588-1596, doi:10.1021/acsnano.6b07159 (2017).
32 Schulman, D. S., Arnold, A. J. & Das, S. Contact engineering for 2D materials and devices. *Chem Soc Rev* **47**, 3037-3058, doi:10.1039/c7cs00828g (2018).
33 Zhou, C. *et al.* Carrier Type Control of WSe2 Field-Effect Transistors by Thickness Modulation and MoO3 Layer Doping. *Adv. Funct. Mater.* **26**, 4223-4230, doi:10.1002/adfm.201600292 (2016).
34 Maserjian, J. & Zamani, N. Behavior of the Si/SiO2interface observed by Fowler-Nordheim tunneling. *Journal of Applied Physics* **53**, 559-567, doi:10.1063/1.329919 (1982).
35 Pak, J. *et al.* Two-Dimensional Thickness-Dependent Avalanche Breakdown Phenomena in MoS2 Field-Effect Transistors under High Electric Fields. *ACS Nano* **12**, 7109-7116, doi:10.1021/acsnano.8b02925 (2018).
36 Kong, L. *et al.* Doping-free complementary WSe2 circuit via van der Waals metal integration. *Nat Commun* **11**, 1866, doi:10.1038/s41467-020-15776-x (2020).
37 Lee, G., Pearton, S. J., Ren, F. & Kim, J. 2D Material-Based Vertical Double Heterojunction Bipolar Transistors with High Current Amplification. *Adv. Electron. Mater.* **5**, doi:10.1002/aelm.201800745 (2019).
38 Breitenstein, O. *et al.* in *2006 IEEE 4th World Conference on Photovoltaic Energy Conference.* 879-884.
39 Jeon, P. J. *et al.* Enhanced device performances of WSe2–MoS2 van der Waals junction p–n diode by fluoropolymer encapsulation. *Journal of Materials Chemistry C* **3**, 2751-2758, doi:10.1039/c4tc02961e (2015).
40 Cheng, R. *et al.* Electroluminescence and Photocurrent Generation from Atomically Sharp WSe2/MoS2 Heterojunction p–n Diodes. *Nano Lett.* **14**, 5590-5597, doi:10.1021/nl502075n (2014).

41 Pham, P. V. *et al.* 2D Heterostructures for Ubiquitous Electronics and Optoelectronics: Principles, Opportunities, and Challenges. *Chem. Rev.* **122**, 6514-6613, doi:10.1021/acs.chemrev.1c00735 (2022).
42 Wilamowski, B. M. & Jaeger, R. C. The lateral punch-through transistor. *IEEE Electron Device Letters* **3**, 277-280, doi:10.1109/EDL.1982.25568 (1982).
43 in *Physics of Semiconductor Devices* 77-133 (2006).
44 Zeng, S., Tang, Z., Liu, C. & Zhou, P. Electronics based on two-dimensional materials: Status and outlook. *Nano Research* **14**, 1752-1767, doi:10.1007/s12274-020-2945-z (2021).
45 Nourbakhsh, A., Zubair, A., Dresselhaus, M. S. & Palacios, T. Transport Properties of a MoS2/WSe2 Heterojunction Transistor and Its Potential for Application. *Nano Lett* **16**, 1359-1366, doi:10.1021/acs.nanolett.5b04791 (2016).
46 Jeong, H. *et al.* Metal-Insulator-Semiconductor Diode Consisting of Two-Dimensional Nanomaterials. *Nano Lett* **16**, 1858-1862, doi:10.1021/acs.nanolett.5b04936 (2016).
47 Zhuang, P., Ma, W., Liu, J., Cai, W. & Lin, W. Progressive RESET induced by Joule heating in hBN RRAMs. *Applied Physics Letters* **118**, doi:10.1063/5.0040902 (2021).
48 Villena, M. A. *et al.* Variability of metal/h-BN/metal memristors grown via chemical vapor deposition on different materials. *Microelectronics Reliability* **102**, doi:10.1016/j.microrel.2019.113410 (2019).
49 Chen, S. *et al.* Wafer-scale integration of two-dimensional materials in high-density memristive crossbar arrays for artificial neural networks. *Nature Electronics* **3**, 638-645, doi:10.1038/s41928-020-00473-w (2020).
50 Zhu, K. *et al.* Hybrid 2D/CMOS microchips for memristive applications. *Nature*, doi:10.1038/s41586-023-05973-1 (2023).
51 Upadhyay, N. K. *et al.* Engineering Tunneling Selector to Achieve High Non-linearity for 1S1R Integration. *Frontiers in Nanotechnology* **3**, doi:10.3389/fnano.2021.656026 (2021).
52 Lo, C.-L., Hou, T.-H., Chen, M.-C. & Huang, J.-J. Dependence of Read Margin on Pull-Up Schemes in High-Density One Selector–One Resistor Crossbar Array. *IEEE Transactions on Electron Devices* **60**, 420-426, doi:10.1109/ted.2012.2225147 (2013).
53 Wang, G. *et al.* Three-Dimensional Networked Nanoporous Ta2O(5-x) Memory System for Ultrahigh Density Storage. *Nano Lett* **15**, 6009-6014, doi:10.1021/acs.nanolett.5b02190 (2015).
54 Cao, Y. *et al.* Movable-Type Transfer and Stacking of van der Waals Heterostructures for Spintronics. *IEEE Access* **8**, 70488-70495, doi:10.1109/access.2020.2984942 (2020).
55 Kim, K. *et al.* van der Waals Heterostructures with High Accuracy Rotational Alignment. *Nano Lett.* **16**, 1989-1995, doi:10.1021/acs.nanolett.5b05263 (2016).
56 Wang, J. *et al.* Transferred metal gate to 2D semiconductors for sub-1 V operation and near ideal subthreshold slope. *Science Advances* **7**, eabf8744, doi:10.1126/sciadv.abf8744 (2021).
57 Kotekar-Patil, D., Deng, J., Wong, S. L., Lau, C. S. & Goh, K. E. J. Single layer MoS2 nanoribbon field effect transistor. *Applied Physics Letters* **114**, doi:10.1063/1.5079860 (2019).
58 Xiao, S. *et al.* Atomic-layer soft plasma etching of MoS2. *Sci Rep* **6**, 19945, doi:10.1038/srep19945 (2016).
59 Hus, S. M. *et al.* Observation of single-defect memristor in an MoS2 atomic sheet. *Nat Nanotechnol* **16**, 58-62, doi:10.1038/s41565-020-00789-w (2021).
60 Shen, Y. *et al.* Variability and Yield in h-BN-Based Memristive Circuits: The Role of Each Type of Defect. *Adv Mater*, e2103656, doi:10.1002/adma.202103656 (2021).

Chapter 4 Dendrite Artificial Neural Network Utilizing Dual-Gate Transistors with High-dielectric Material

Nowadays, the energy consumption of state-of-the-art artificial neural networks remains higher compared to that of the human brain. Although synapse and soma devices have made significant progress in artificial neural networks, little research focuses on the multi-input and nonlinear integration of dendrite devices. The lack of these critical functions compromises the performance of artificial neural networks, for example, in terms of, energy efficiency and the ability to handle complex tasks. Hence, a dual-gate transistor with new high-k material $Cu_{0.67}Ag_{0.33}InP_2S_6$ (CAIPS) as gate dielectric is designed by a low-temperature transfer method and introduced to complete the dendritic artificial neural network (DANN). Our dendritic device nonlinearly processes the multiple input signals from synapses, achieves the feature binding problem, and improves the energy efficiency in neural network. High-k dielectric-based dual gate transistors can perform memory ability based on mobile ions and nonlinearly integrate the signals. The nonlinear process ability of dendrites enables dendritic neural network using small neuronal populations, increasing storage capacity, less trainable parameters with high accuracy.

4.1 Introduction

Over the past decade, artificial neural network (ANN)-based deep learning has attracted significant research attention due to its key role in multiple domains, including image object detection/recognition, natural language processing, and face verification.[1-3] Despite these advancements, it is obvious that there still exists a significant gap between deep neural networks and human-level intelligence. Commonly, the ANN consists of two parts: neurons, close to the function of soma in biological, that receive input, linear weighted sum it, and pass on the output to the next neuron layer; synapses whose weights determine the strength of connection between neurons. Besides the artificial synapses and soma devices, dendrites as one of the crucial components in neurons. The dendrites transmit signals between synapses and soma and nonlinearly integrate signals in parallel, which improves the time-consuming and energy efficiency. Therefore, to improve energy efficiency, and the ability to handle complex tasks, one solution for energy efficiency is introducing artificial dendrites to further process information.[4]

The double gate transistor is considered a potential candidate for artificial dendrites, where the inputs of each gate represent the input signals from two pre-synapses, and the output signals represent the results after the nonlinear integration of dendrites. The 2D dielectrics require high dielectric constant to provide an efficient gate control, which is a fundamental requirement for achieving high device performance, such as high on/off current ratios, low threshold voltages, and minimal leakage currents.[5] However, the common vdWs 2D dielectrics, such as h-BN ~ 2.5, are still low. $CuInP_2S_6$ (CIPS) as the common 2D ferroelectric dielectric stands out as a representative material and has high dielectric constant (about 50) at room temperature.[6] Its room-temperature out-of-plane ferroelectricity arises from the Cu^+ cation.[7] The copper is both spatially and temporally

disordered in the paraelectric phase, which increases the ionic conductivity of CIPS as gate dielectric and contributes to the leakage current of the transistor. Therefore, the substitution of the Cu cation with the other metal cation results in the variation of some physical properties, such as the high dielectric constant, loss of ferroelectricity, higher air stability, and larger dielectric constant.

This work studies the possibility of artificial dendrites with dual-gate transistors to mimic the pre-process ability of dendrites and construct the dendritic artificial neural network (DANN). All the device fabrication is based on the dry transfer method since this method can achieve clean interface surface and low temperature, which suppress leakage current in ultrathin 2D dielectrics.[8] The analog of CIPS, $Cu_{0.33}Ag_{0.67}InP_2S_6$ (CAIPS), was grown and confirmed as gate dielectric since the inducing metal atom can decrease the conductivity of the gate dielectric and suppress the leakage and the mobile Cu^+ can mimic the synapse function. Compared to split-gate transistors, the dual-gate transistor with CAIPS as ferroelectric gate dielectric, MoS_2 as channel, and Au as electrodes exhibits larger ON/OFF ratio (10^6), high current (μA), and low gate leakage current. In addition, the dual-gate transistor successfully demonstrates the function of synapse and dendrites, such as short-term potential, paired-pulse facilitation, and dendritic nonlinear integration (supralinear and sublinear).

4.2 Results and discussion

4.2.1 Overview of Dendrite

A neuron in the biological nerve system consists of dendrites, a soma, and an axon, with each neuron being able to handle thousands of different synaptic inputs unequally.[3] The

neuron receives thousands of postsynaptic potentials, which are segregated into multiple dendritic branches. Postsynaptic potentials received within a certain time interval from the same or different synapses are integrated into the dendrites nonlinearly. Both spatial and temporal information are encoded in the dendritic computing units. Then, the soma integrates these potentials further, determining whether to fire spikes based on integrated amplitude over a specific threshold and transmitting them via the axon.[9] **Figure 4.1** exhibits the function of components (synapse, dendrite, soma) in complete ANN. The synapse represents a plastic weight, and the soma provides integration and spike-firing functions, while the nonlinear integration and filtering of postsynaptic potentials on different dendritic branches provide the network with the ability to process complex information.

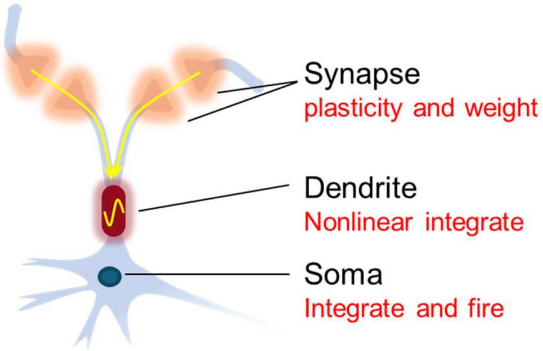


Figure 4.1 The schematic of a neuron with three functional components. Dendrites connect the soma and synapses.

Recent researchers confirmed that dendrites are no longer treated as passive conductors of information to the soma.[10,11] The active properties of the dendritic tree are crucial for determining how electrical signals propagate. One of the active properties of the dendrites: nonlinear integration, is thought to increase the computational ability of neurons [12]. As presented in **Figure 4.2**, we defined the current ratio between measured excitatory post-

synaptic current (EPSCs) from simultaneous stimuli and arithmetic additive of individual EPSCs. The dendritic nonlinear integration can be classified into three operations: (1) linear, where the measured EPSP equals the arithmetic sum of individual EPSPs; (2) supralinear, where the measured EPSP exceeds the arithmetic sum of individual EPSPs; and (3) sublinear, where the measured EPSPs is less than the arithmetic sum of individual EPSPs. These signals' nonlinear integration is one of the main reasons why biological neural networks can handle very complex tasks and yet consume little energy. Incorporating dendritic computing units into artificial neural networks is thus highly desired.

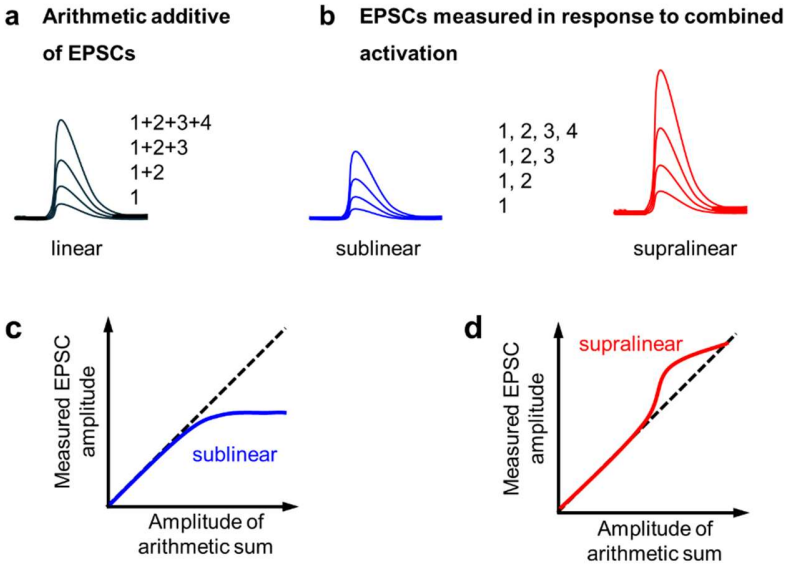


Figure 4.2 (a) The arithmetic sum of individual EPSPs is evoked by simultaneous synaptic activation. (b) EPSPs result from the same synaptic activation. Blue curves with increasing number of synapses activated through sublinear dendrite. Red curves are EPSPs similarly obtained through supralinear dendrites. (c) supralinear and (d) sublinear dendritic current relationships. The dashed line represents a linear relationship.

4.2.2 Feature Binding Situation

Neurons encode the binding relationships between visual features processed in different areas, such as color, shape, location, or motion, which is called feature binding.[13] The feature binding would be common in human brain. The example is illustrated in **Figure 4.3**. When human observe multiple objects, like red square and yellow triangle, presented together within a scene, the brain can process the features of objects in parallel and identify the objects simultaneously. The process is crucial for object recognition and helps us interact effectively with our environment.

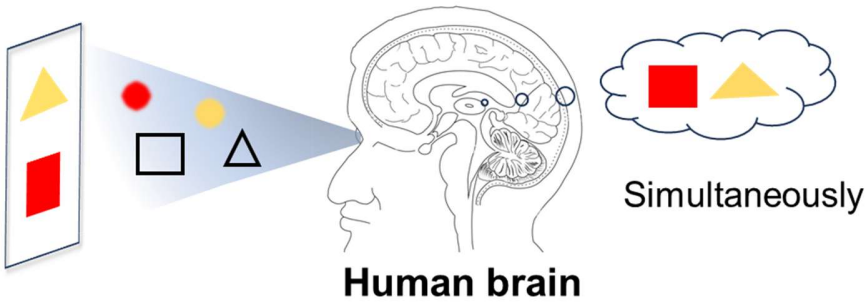


Figure 4.3 The schematic of the feature binding situation. The visual system first detects basic features: the color red and the shape of a square, as well as the color yellow and the shape of a triangle. These features are initially processed in different areas of the brain. Feature binding involves integrating these separate features so that the observer perceives them simultaneously as distinct objects: red square and yellow triangle.

However, for traditional ANN, to identify one object, the weights in neural network should be trained many times to some fixed value and linear sum into neuron. The output result in ANN can be expressed as below equation:

$$f(x) = b + \sum_{n=1}^{k} (w_i x_n) \tag{1}$$

Where w_i indicates the weight with the i-th input, x_n represents the i-th input and b is the bias term. Therefore, the network only responds under a specific input set. As shown in **Figure 4.4a**, if the network is trained to identify the red square, the neuron only responds to the red square even red square and yellow triangle are presented simultaneously. Therefore, two neural networks need to be trained when the red square and yellow triangle are identified simultaneously (**Figure 4.4b**).

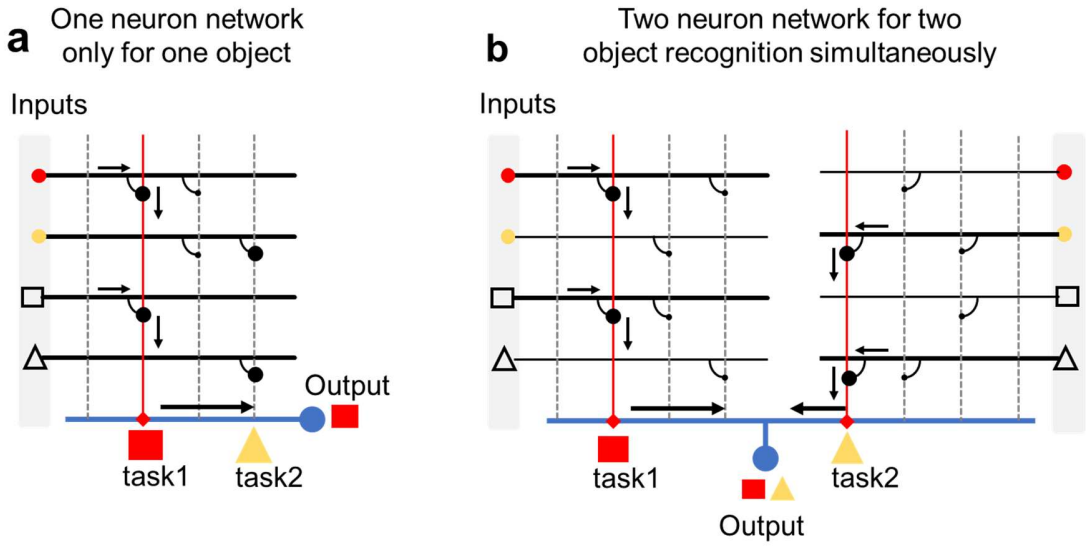


Figure 4.4 The schematic of artificial neuron network (ANN) to achieve feature binding situation. When four inputs are fed to one ANN simultaneously, (a) one trained neuron network with fixed weights of synapse only response for one object: red square or yellow triangle. (b) two neuron networks required to recognize red square and yellow triangle simultaneously.

In biological systems, dendrites nonlinearly integrate multiple synaptic inputs in parallel, which is crucial for improving the energy efficiency of neural networks. Therefore, I proposed a more complete DANN model with artificial dendrites. Utilizing the changeable

nonlinear response of dendrites on different inputs, we can recognize the red square and yellow triangle simultaneously in one neural network. The output can be expressed as below equation:

$$f(x) = b + NL_D \sum_{n=1}^{k} (w_i x_n) \quad (2)$$

Where NL_D is the changeable nonlinearity from dendritic device. With artificial dendrite, DANN can achieve ideal results with fewer neuronal populations and less trainable parameters with high accuracy.[14]

4.2.3 The Property of High-k Dielectric

To achieve the nonlinear process function of dendrite, we fabricate the dual-gate transistor with the similar fabrication process of selector mentioned before. The analog of ferroelectric material CIPS, which is CAIPS, as a 2D ferroelectric dielectric and MoS2 as channel are utilized to fabricate dual-gate transistors (**Figure 4.5**). In CAIPS, several Cu atoms were replaced with Ag atoms with the one-third ratio through the chemical vapor transport (CVT) method.[15] **Figure 4.5** presents the Crystal structure of CAIPS, characterized by a sulfur framework where metal cations and P-P pairs occupy the octahedral voids. Note that the alternate sites between Cu and P-P pairs from one layer to the next result in a unit cell comprising two adjacent layers to fully describe the material's symmetry. The top and bottom gate voltage represents the simultaneous pre-synaptic input signals and the output from channel represents the results after dendritic nonlinear integration.

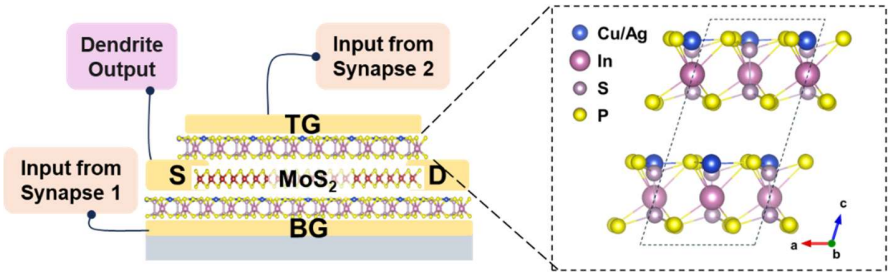
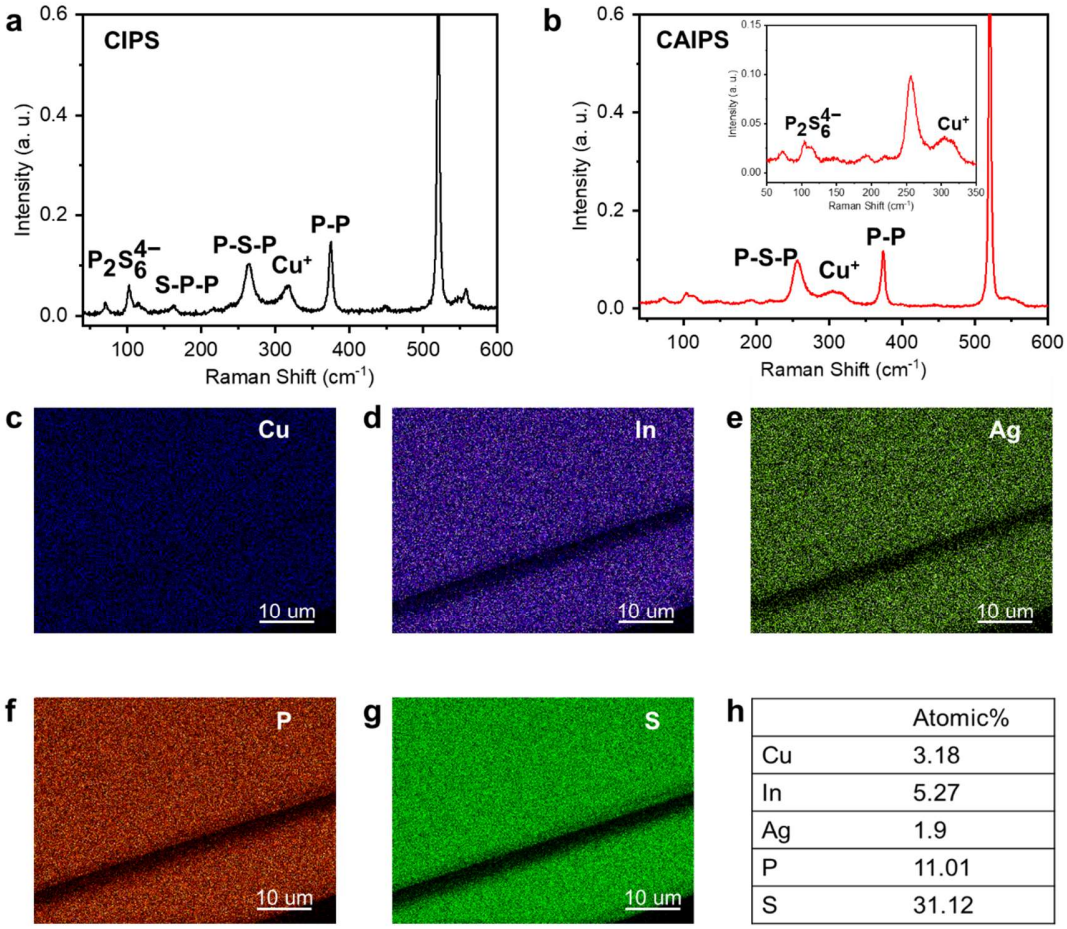


Figure 4.5 The structure of dual-gate transistor with Au as electrodes, MoS_2 as channel, and CAIPS as gate dielectric. The right part presents the crystal structure of CAIPS viewed from the b-axis.

In our Raman measurements, CAIPS (**Figure 4.6b**) showed a similar spectrum to CIPS (**Figure 4.6a**), corresponding 350−400 cm^{-1} and 150−300 cm^{-1} to P−P tensile mode and P_2S_6 deformation vibration (S−P−S and S−P–P modes), respectively.[16] The Raman peak of 320 cm^{-1}, corresponding to Cu^+, in CAIPS is smaller than that in CIPS, which indicates several Cu cations are replaced successfully by Ag cations. To further demonstrate whether the Ag elements exist in the CAIPS crystals, we obtained the SEM-EDS results (**Figure 4.6c-g**) of the bulk crystal. It can be found that there are five elements in our samples, including the Ag element, which has been successfully doped into the bulk crystal. The specific ratio of each element can be seen in **Figure 4.6h**.



h	Atomic%
Cu	3.18
In	5.27
Ag	1.9
P	11.01
S	31.12

Figure 4.6 Raman spectra of (a) the CIPS nanoflakes and (b) CAIPS nanoflakes. Insert in (b): enlarged Raman spectrum. (c - g) The corresponding elements distribution map of CAIPS material. (h) The ratio of each element in CAIPS is defined by EDS.

Metal-insulator-metal (MIM) device structure with different CAIPS thicknesses and Au electrodes was fabricated to calculate the dielectric constant by measuring the capacitance according to the following equation (3):

$$C=\frac{\varepsilon_0 kS}{d} \tag{3}$$

where C is the measured capacitance, ε_0 is the vacuum permittivity, k is the dielectric constant, S is the overlapping area of top and bottom electrodes, and d is the thickness of the CAIPS. **Figure 4.7** confirms the high dielectric constant of CAIPS (17-20) over 100 nm thickness. When the thickness is down to 40 nm, the dielectric constant can be 7, which is larger than the constants of h-BN (2-4), indicating that the CSIPS can be a promising gate material.

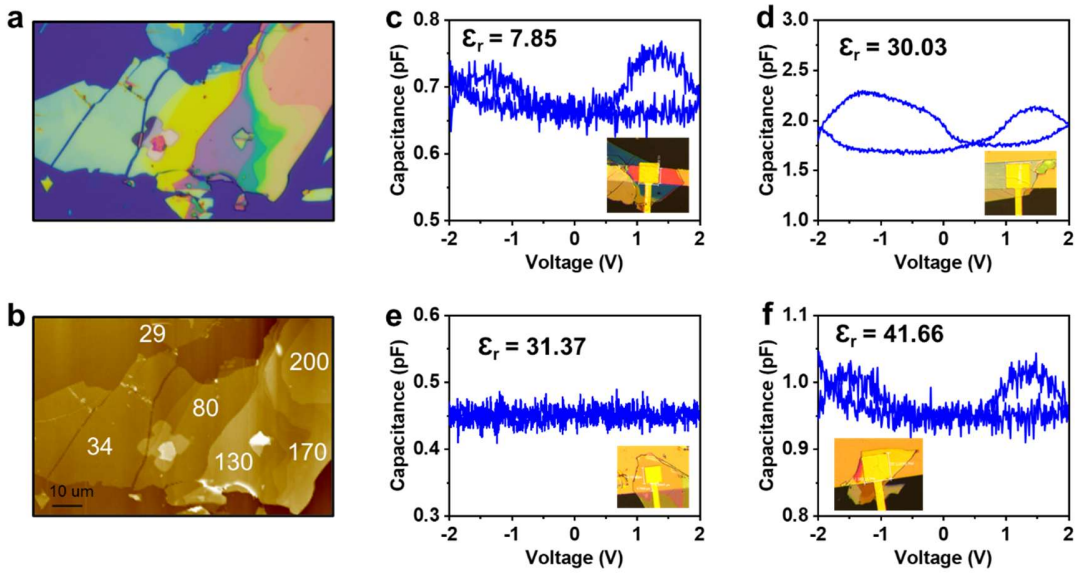


Figure 4.7 (a) the optical image of CAIPS with different thicknesses. (b) the AFM result confirms the specific thickness of the sample in (a). (c) Voltage-dependent capacitance density (C–V) for Au/CAIPS/Au devices with thin CAIPS; (d-f) C–V curve for Au/CAIPS/Au devices with thick CAIPS.

In single-gate behavior, from the output curve, we can see good Ohmic contact at low drain voltage (**Figure 4.8a**). From the transfer curve in **Figure 4.8b**, this device exhibits large ON/OFF ratio of 10^6, low leakage current below pA, and stable reversible hysteresis behavior. Commonly, the reversible hysteresis results from the ferroelectric property of

gate dielectric. However, according to PE loop and PFM results, compared to CIPS, the CAIPS shows any remnant polarization in **Figure 4.9**, which illustrates that non-ferroelectric existed in CAIPS material.

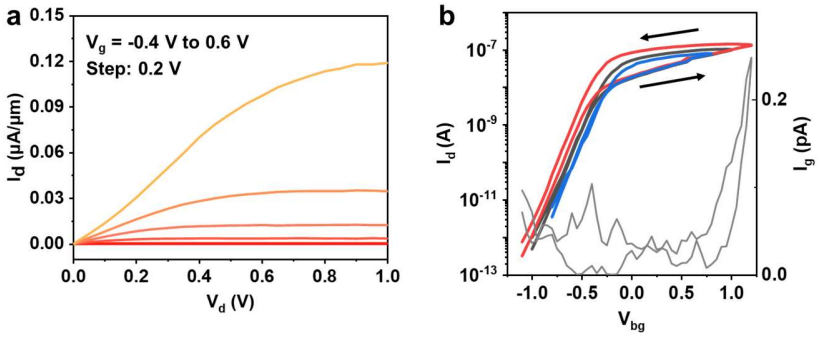


Figure 4.8 (a) The output curve of single-gate transistor. (b) the transfer curve of single-gate transistor which exhibits reverse hysteresis.

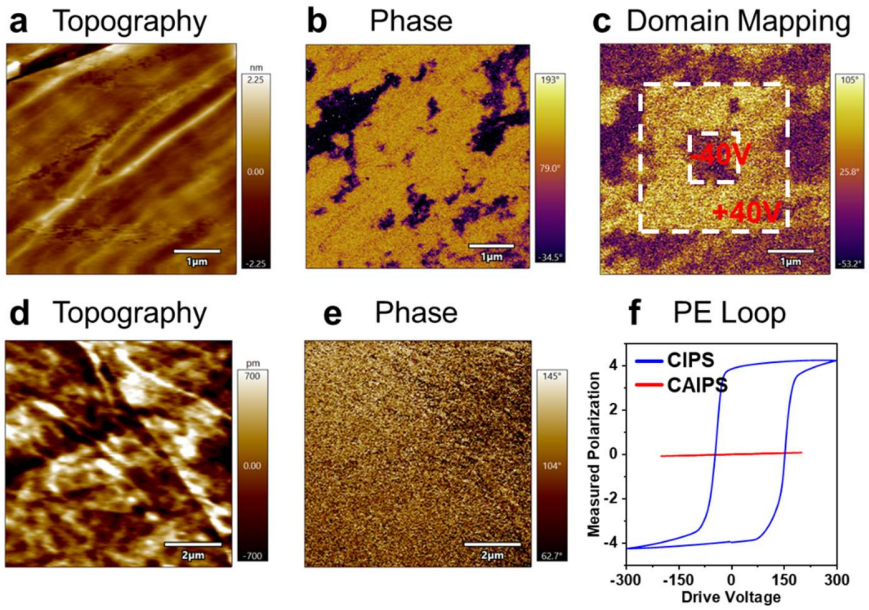


Figure 4.9 For CIPS ferroelectric nanoflake, the measured (a) topography image, (b) out-of-plane PFM phase image, and (c) PFM image after electrical writing using the PFM tip.

The written squares are marked using the dashed square lines, and the electric biases of the bottom electrode are marked numerically. For CAIPS, compare (d) topography image and (e) out-of-plane PFM phase image, it doesn't exhibit ferroelectricity property. The (f) PE Loop further confirms non-ferroelectricity compared to CIPS.

Another possibility to explain the reversible hysteresis behavior is the mobile ions mechanism. As the voltage decreases, the Cu^{+} close to MoS_2 channel moves to the gate metal/dielectric interface slowly, which contributes to reversible hysteresis. In this mechanism, the mobility of mobile ions degrades as temperature increases, which increases the hysteresis window. Therefore, we measure the transfer curve of transistor under different temperatures and calculate the hysteresis window. As can presented in **Figure 4.10a**, the hysteresis window of CAIPS increases with temperature increase, which proves the reversible hysteresis window is based on mobile ion mechanism.

Since our transistor can achieve synaptic behavior based on the mobile ion mechanism, the switching speed is a critical parameter. Here, we measure the switching speed of transistor based on mobile ion mechanism. Mobile ions within the gate dielectric or near the channel can affect the switch speed by introducing delay in the gate control. These ions can move in response to electric fields, altering the effective gate voltage and causing slower switching speed. We define the t_{on} and t_{off} to represent the delay time difference of the leading and trailing edge between the current response and corresponding voltage response, respectively, which reflect the switching speed.[17,18] Smaller t_{on} or t_{off} represents faster static switching speed. By applying multiple electrical pulses and confirming the charge accumulated in channel, we can conclude the switching speed of transistor based

on the mobile ion mechanism. The channel current responses were measured under multiple gate pulses with different widths of 50 and 100 μs, respectively. The channel current didn't increase after 40 gate pulses with 50 μs (**Figure 4.10b**), while charge accumulation is observed when the width of the pulses increases to 100 μs (**Figure 4.10c**). Hence, the switching speed of transisitor based on the mobile ion mechanism is calculated with the t_{on} (~ 4 μs) and t_{off} (~ 1.3 μs) under 100-μs gate pulses.

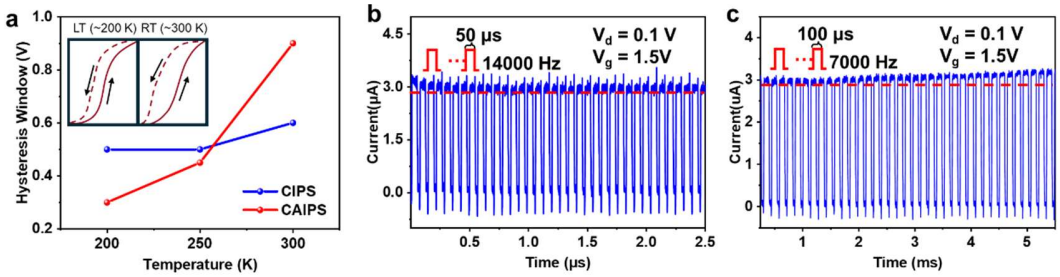


Figure 4.10 (a) Reverse hysteresis window relates to temperature curve about mobile ion mechanism effect on transfer curves of CIPS or CAIPS-based transistor from low temperature (LT) to room temperature (RT). The channel current responses were measured under multiple gate pulses with different widths of (b) 50 and (c) 100 μs, respectively.

The 2D ferroelectrics $CunInP_2S_6$ (CIPS), as the gate dielectric, its high conductivity contributes to high leakage current due to the high mobility of Cu ion. Inserting few-layer h-BN is an effective method to suppress the leakage current (**Figure 4.11a**).[19] Although the transistor with h-BN performs low leakage current (~ 10 pA) and large ON/OFF ratio over 10^6, the small ferroelectric hysteresis and complex fabrication process are unsuitable for further dual-gate transistor and synapse-dendrite devices (**Figure 4.11b**).[9] Then, we replaced several Cu atoms with Ag with a specific ratio to form $Cu_{0.5}Ag_{0.5}InP_2S_6$ (CAIPS'), which served as the gate dielectric. **Figure 4.11c** presents the single-gate transistor with

MoS_2 as channel, Au as S/D electrodes, and CAIPS' as the gate dielectric.[20] Although this transistor exhibits anti-clockwise hysteresis and larger ON/OFF ratio (~ 10^6), the large operating voltage, and large leakage current (~ 0.1 nA) contribute to high energy consumption(**Figure 4.11d**). The Mn element can substitute for Cu ion to form $(CuIn)_{3/4}Mn_{1/2}P_2S_6$ (CMIPS) as gate dielectric. Since the curie temperatures of CIMPS are lower than CIPS, the fabrication of transistor through dry transfer method introduces higher temperature over curie temperatures, which change the structure of CIMPS due to their low air stability and result in the high leakage current.

The performance of transistors under different gates is summarized in **Figure 4.11e**. All the transistors with MoS_2 as channel, and Au as S/D electrodes. To achieve low energy consumption and excellent synaptic behavior, we prefer low operated voltage, low leakage current, and large hysteresis window. Compared to other dielectric, the CAIPS with one-third Ag elements transistor presents better parameters.

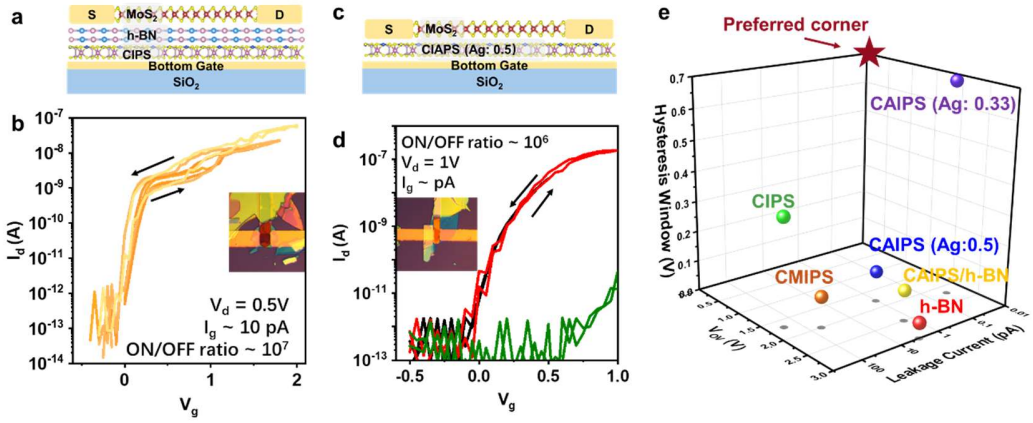


Figure 4.11 (a) The schematic side view of the single-gate structure with CIPS and h-BN gate dielectric. (b) The transfer characterization of transistor with CIPS. The insert: Optical micrograph of the device. (cc) The schematic side view of the single-gate structure with

CAIPS' gate dielectric. (b) The transfer characterization of transistor with CIPS under different sweep ranges of gate voltage. The insert: Optical micrograph of the device. (e) The comparison of hysteresis window, overdrive voltage (V_{ov}), and leakage current of single-gate transistors with different dielectric materials.

4.2.4 The Synaptic Behavior

Based on the mobile Cu^+ mechanism, our device can achieve synaptic behavior with short-term memory. Once a presynaptic pulse is applied, an excitatory postsynaptic current (EPSC) will be triggered.[21] The change in conductance corresponds to the learning and forgetting behavior of artificial synapses. **Figure 4.12a, b** shows that EPSCs triggered by different amplitude or number of presynaptic pulses from 0.8 - 1.5 V of 10 ms width, which indicated the learning ability of our synaptic device. Note that the peak of current increases with the increasing number and amplitude of the presynaptic pulse and compared to single pulse, the current after successive stimuli decays slowly.

In addition, paired-pulse facilitation (PPF) is an important form of short-term plasticity. PPF is studied by applying pulse pairs having different intervals at constant pulse width (0.8 s) and voltage amplitude (0.6 V) (**Figure 4.12c**). As expected, the second pulse produces more EPSC (A_2) than the first pulse (A_1) and the low-interval pulse pair produces a greater facilitation effect, which can be obtained under shorter intervals since the ion accumulation cannot disappear quickly before the next gate pulse arrives. In **Figure 4.12d**, the consecutive conductance modulation of the transistor performs the potentiation and depression characteristics.

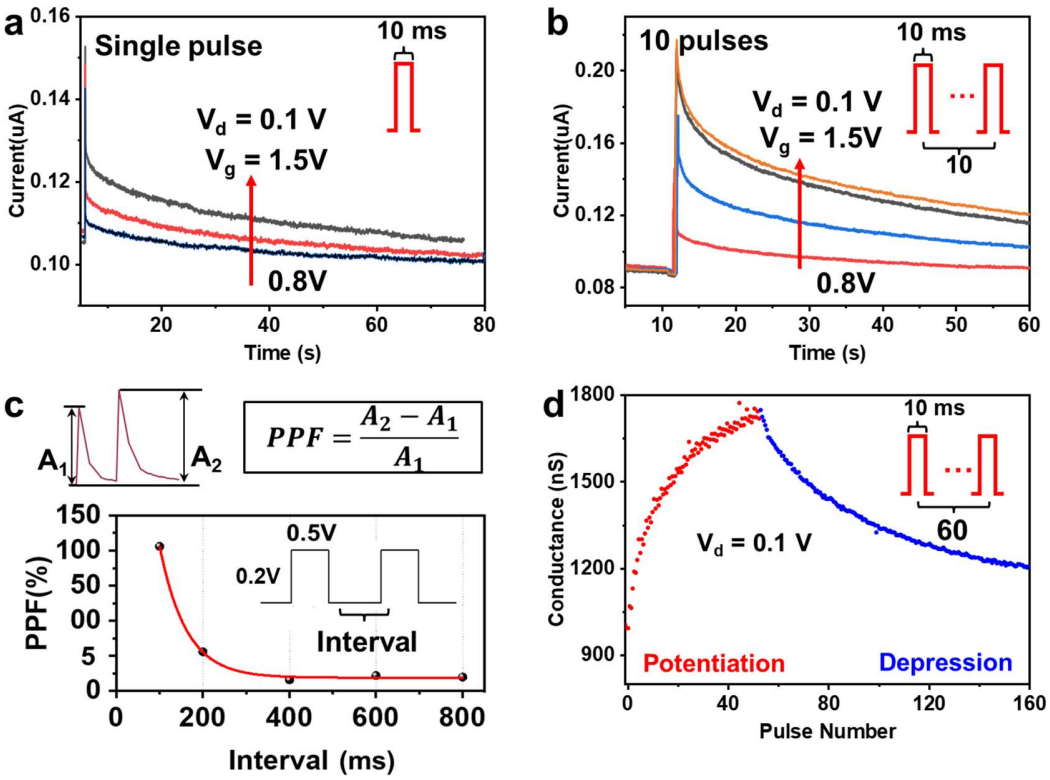


Figure 4.12 The current response for increasing input pulse amplitude (10 ms) under (a) single pulse and (b) 10 pulses. (c) The percentage of PPF can be calculated as equation where A_1 and A_2 are the amplitudes of EPSC responses to the first and second input pulses, respectively. (d)The conductance gradually changes for over 150 successive pulses with (1 V, 10 ms) for potentiation and (-0.1 V, 10 ms) for depression.

4.2.5 Dual-gate Transistor with Dendritic Behavior

After confirming the ability of synapse, the nonlinear integration function of the dendrite is imitated through the dual-gate transistor. From the transfer curves in **Figure 4.13**, we can see that compared to the top gate, the bottom gate has higher gate control ability since the larger control channel area for bottom gate. In addition, compared to the single gate,

dual-gate transistor exhibits low leakage current, larger ON/OFF ratio, smaller subthreshold, and stable reversible hysteresis under various voltage sweep due to the mobile ion mechanism.

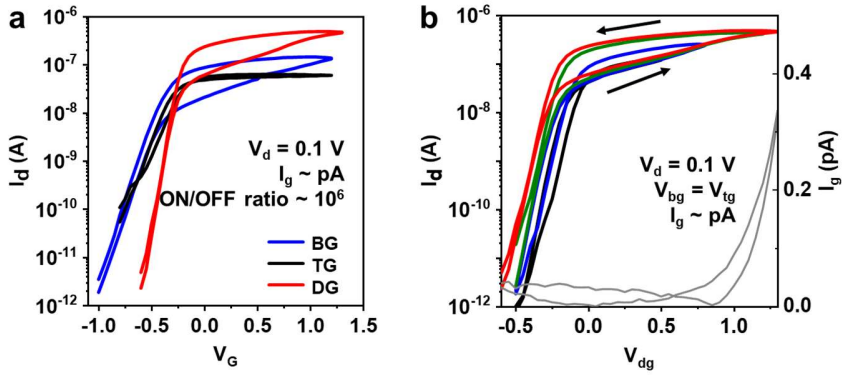
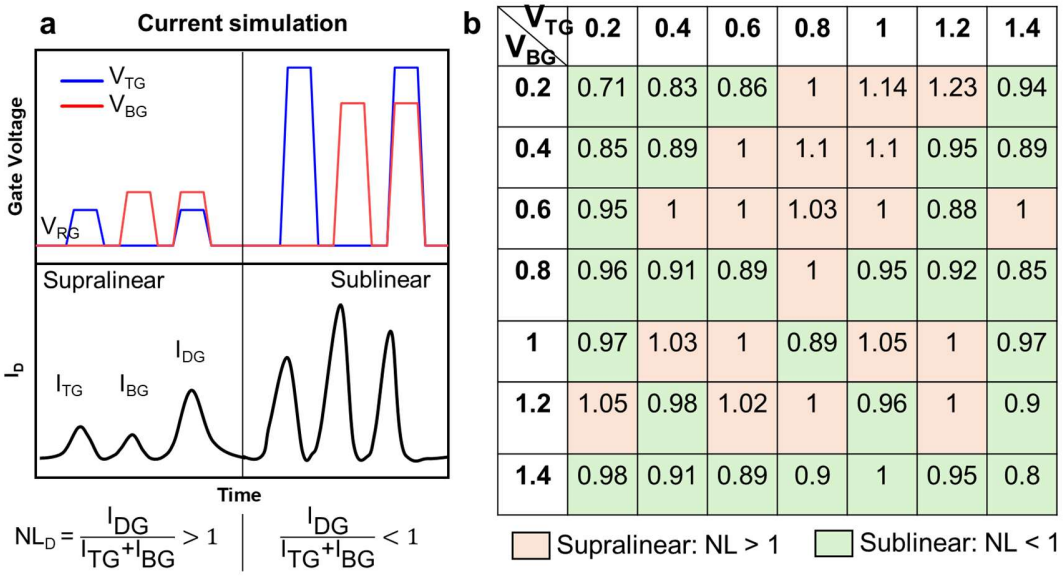


Figure 4.13 (a) The transfer curve of the transistor modulated by the bottom gate (blue curve, $V_{BG} = V_G$), top gate (black curve, $V_{TG} = V_G$), and double gate (red curve, $V_{BG} = V_{TG} = V_G$), respectively. (b) The transfer curve of the transistor is modulated by the double gate.$V_{BG} = V_{TG} = V_{dg}$.

To mimic the nonlinear behavior of dendrite, we applied different top and bottom gate voltage pules respectively, and simultaneously **(Figure 4.14a**). According to the formula in **Figure 4.14a**, the current behaves supralinear and sublinear under different situations. All the biases are voltage pulses with 10 ms width, and I_{BG}, I_{TG}, I_{DG} are the programmed currents compared to the current under read voltage. By choosing appropriate gate biases, dual-gate transistors can achieve different nonlinearity in **Figure 4.14b**. The device performs sublinear behavior under small gate biases (V_{BG}, $V_{TG} < 0.4$ V) since the static gate control influences the channel current. As the gate bias increases, the conductance in the channel can be modulated by mobile Cu^+, which presents the nonlinear behavior.



V_{BG} \ V_{TG}	0.2	0.4	0.6	0.8	1	1.2	1.4
0.2	0.71	0.83	0.86	1	1.14	1.23	0.94
0.4	0.85	0.89	1	1.1	1.1	0.95	0.89
0.6	0.95	1	1	1.03	1	0.88	1
0.8	0.96	0.91	0.89	1	0.95	0.92	0.85
1	0.97	1.03	1	0.89	1.05	1	0.97
1.2	1.05	0.98	1.02	1	0.96	1	0.9
1.4	0.98	0.91	0.89	0.9	1	0.95	0.8

Figure 4.14 (a) The schematic of gate voltage's operation and responding current behavior. The bottom part exhibits the formula of supralinear and sublinear, respectively. I_{BG}, I_{TG}, and I_{DG} are the programmed currents compared to the current under read voltage. V_{BG}, V_{TG}, and V_{RG} are bottom, top, and read gate biases, respectively. V_{RG} is 0.1 V. (b) The gate-modulated nonlinearity of dual-gate transistor with 10-ms gate bias.

Figure 4.15 illustrates the mechanism of sublinear or supralinear behavior. Under medium gate biases, the dual-gate transistor exhibits supralinear due to more electrons in the channel induced by accumulated Cu+ near gate dielectric/channel interface (Figure 4.15 (i)). While under both small or large gate biases, the device exhibits sublinear due to static gate control or saturated electrons in channel even more Cu+ near the gate dielectric/channel interface (Figure 4.15 (ii))

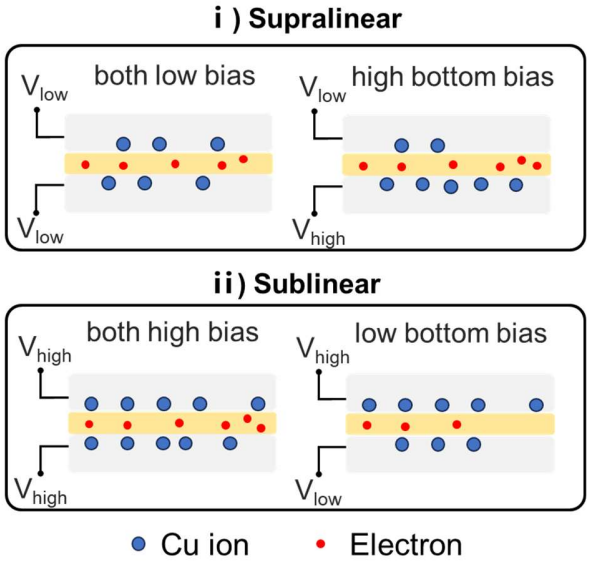


Figure 4.15 The mechanism of i) supralinear and ii) sublinear under different gate bias situations.

4.2.6 Dendritic Artificial Neural Network (DANN)

After the nonlinear integration function is achieved by our dual-gate transistor, the feature binding can be realized through DANN. As mentioned before, to identify objects with two features simultaneously, the traditional ANN requires two networks. However, based on the changeable nonlinearity of the dendritic parameter, the DANN can recognize red square and yellow triangle simultaneously in one network, which contributes to smaller neural populations and lower energy consumption (**Figure 4.16a**). Therefore, based on our dual-gate transistor, we exhibit the dendritic network computation, which consists of two dual-gate transistors. Note that each output from the dendrite is linearly summed and passed toward to soma. The soma integrates the signals from dendrites and generates the output. According to the output value of soma, we can summarize in the table in **Figure 4.16c**.

Based on one DANN, the red square and yellow triangle can be identified once the high output value is obtained. If one of the shape and color features doesn't match each other, the output value is smaller. If none of the features is matched, the DANN doesn't give any output.

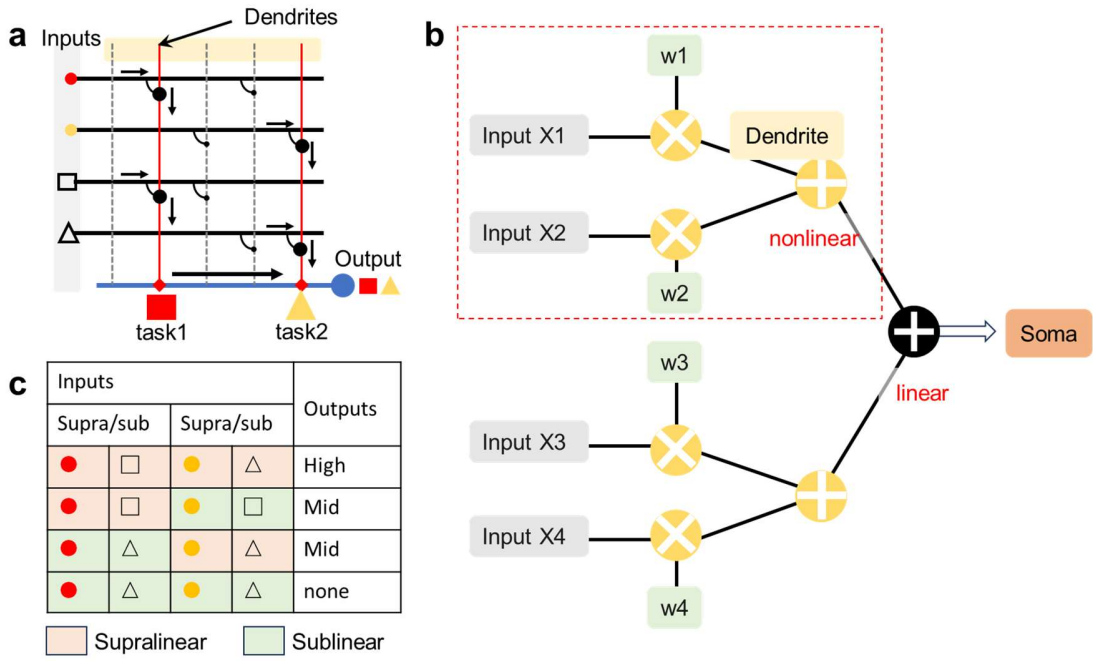


Figure 4.16 (a) The schematic of dendritic artificial neuron network (DANN) to achieve feature binding situation. When four inputs are fed to one DANN simultaneously, due to the changeable nonlinearity of dendrite, the DANN can recognize red square and yellow triangle simultaneously. (b) The overall structure of the DANN and the relationship between the input and output values in each layer. The red dash rectangle represents one dual-gate transistor.

4.3 Summary

Our dendritic device nonlinearly processes the multiple input signals from synapses, achieves the feature binding problem, and improves energy efficiency in neural networks.

Utilized high-k dielectric CAIPS, the dual-gate transistors exhibit low leakage current, 10^6 ON/OFF ratio under low operated voltage at room temperature and large reversible hysteresis. In addition, the devices perform memory ability based on mobile ions mechanism and nonlinearly integrate the signals. The nonlinear process ability of dendrites enables dendritic neural networks using small neuronal populations, increasing storage capacity, and less trainable parameters with high accuracy. Since these functions play a crucial role in neural computations like pattern recognition, sensory processing, etc., our transistors provide innovative building blocks for neuromorphic cognitive systems.

4.4 Experimental Section

4.4.1 Crystal Growth

Single crystalline CXICPS crystal series were synthesized by the chemical vapor transport (CVT) method. The powders of Cu, In, X (induced metal cation), P, and S (≥99.99%) were encapsulated in a vacuum quartz tube (vacuum at the level of ~10^{-3} Pa) according to stoichiometric ratios, and iodine (I_2) was used as the transport agent at a dosage of 10 mg/cm^3. The temperatures of the source zone and the crystal growth zone were kept at 650 °C and 500 °C, respectively. After 7 days of reaction under these conditions, followed by a natural cooling process, yellow crystals were obtained.

4.4.2 Materials Characterization

Surface morphology and composition of TMDs were characterized by an optical microscope, scanning electron microscope, atomic force microscope (Bruker), and confocal Raman characterization (WITec Alpha300 Raman) using a wavelength of 532 nm.

The sMIM experiments were carried out on a CSInstruments AFM platform equipped with sMIM module from PrimeNano Inc. The probes for sMIM with electrically shielded cantilevers are commercially available (sMIM-300 from PrimeNano Inc.). The two output channels of sMIM correspond to the real and imaginary parts of the local sample admittance. Before experiments on CXIPS flakes, the sMIM module is calibrated on a standard sample covered with SiO_2 and Al dots. By tuning the phase shifter in front of the I-Q (in-phase and quadrature) mixer, the sMIM channels with and without contrast between SiO_2 and Al dots are adjusted to be sMIM-C (capacitive or imaginary parts of the local sample admittance) and sMIM-R (lossy or real parts of the local sample admittance) signal, respectively.

The surface morphology is characterized by the AFM instrument (Asylum MFP-3D Infinity). The ferroelectricity is characterized by this instrument equipped with advanced Piezoresponse Force Microscopy. Olympus AC240TM Pt/Ti-coated Si cantilevers were used in the PFM measurements. The underlying Au electrode is grounded. PFM images of CIPS were taken in the DART mode with driving voltage (40 V a.c.) applied at the tip. During the domain writing, the voltage was also applied at the tip.

4.4.3 Devices Fabrication

Dual-gate transistor fabrication

We adopted a fabrication flow integrating several 2D material transfer techniques to construct our MoS_2/2D dielectric transistor, where the 2D dielectrics can be h-BN, CIPS, CSIPS, CIMPS, CAIPS', and CAIPS.

The top gate electrodes of 50-nm-thick Au were patterned by photolithography and evaporated on Si substrate since silicon has weaker adhesion with Au than SiO_2, followed

by the lift-off process for photoresist (that is via acetone bath for a few minutes).[22] Back gate electrode patterns (Cr/Au, 5/30 nm) were first prepared on the Si/SiO_2 (400 μm/300 nm) substrates using the standard photolithography process, followed by evaporation of a 35 nm thick Au layer on top of 15 nm Cr adhesion layer. Note that all the metallization was done by electron-beam evaporator under about 10^{-6} torr. The few-layer MoS_2 and 2D dielectrics are mechanically exfoliated from bulk crystals onto a SiO_2 substrate using scotch tape and transferred on PDMS stamp. Note that the thickness and shape of electrode and TMD materials are carefully selected in advance through an optical microscope to avoid the leakage current from gate electrodes to S/D electrodes.

We then fabricate a transfer stamp by adhering a flat PDMS layer to a glass and then stick Polyvinyl Butyral (PVB) on top of the PDMS. PVB film was prepared by drying 4% PVB aqueous solution on a disk in the air for one day. After transferring the bottom 2D dielectrics nanosheet and MoS_2 layers on the back gate through PDMS using a rotationally aligned dry transfer technique, the source/drain (S/D) electrodes (Ag/Au, 35/15 nm) are transferred onto MoS_2 flake using PVB and transfer platform. We bake the stamp at 40 ºC for 10 minutes to strengthen the interaction between the target material and PVB. Note that due to the low curie temperature of CIPS-series materials and active Cu ion, the baking temperature and time are essential for successful fabrication. The transfer heating temperature is confirmed according to the glass transition temperature (T_g) of PVB and the adhesion of PDMS.[23] Then, we soak the target chip in alcohol for 30 minutes to remove PVB and then rinse it with deionization water.[24] Finally, we transferred the top 2D dielectrics and top gate electrode in sequence, followed by removing the PVB through alcohol and deionization water again. Examining every 2D material layer under the optical

microscope is necessary to ensure no visible tears or wrinkles in the heterojunction area since these would lead to poor interface quality. [25]

Metal-insulator-metal (MIM) device fabrication

For the MIM device based on CAIPS nanosheets, bottom electrodes (5/30 nm Cr/Au) were first patterned onto quartz substrate using the standard electron beam lithography process. We adopt quartz substrate instead of SiO2 substrate since dielectric constant results can be affected by parasitic capacitance from SiO2. Then, CAIPS nanosheets were aligned and transferred on the bottom electrode with the assistance of PDMS on an optical transfer platform. Top Au electrodes (50 nm) with the same width of 10 μm were finally patterned along the perpendicular direction to form a cross-shaped MIM device.

4.4.4 Electrical Characterization

The P-E tests of ferroelectric capacitors with CIPS was conducted by the Precision Premier II ferroelectric tester connected.

Capacitance-Voltage (C-V) measurements were conducted through the CVU modules of Keithley 4200. We measured C-V curves of metal-insulator-metal (MIM) structures at different frequencies.

The electrical characteristics of CAIPS capacitors and dual-gate transistors were conducted using the LakeShore Cryogenic Probe Station (Model CRX-6.5K) and Keithley 4200/B1500 A Semiconductor Characterization System. DC measurements were conducted through the SMU modules, while pulse tests were conducted through the PMU modules of Keithley 4200 or the SPGU modules of B1500A. During electrical tests, the chamber was under dark and vacuum conditions. For variant temperature tests, the probe station offers cryogen-free operation over a temperature range of 200 K to 350 K.

4.5 Reference

1 Aly, M. M. S. *et al.* Energy-Efficient Abundant-Data Computing: The N3XT 1,000x. *Computer* **48**, 24-33, doi:10.1109/MC.2015.376 (2015).

2 Chen, J. *et al.* Optoelectronic graded neurons for bioinspired in-sensor motion perception. *Nature Nanotechnology* **18**, 882-888, doi:10.1038/s41565-023-01379-2 (2023).

3 Merolla, P. A. *et al.* A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668-673, doi:doi:10.1126/science.1254642 (2014).

4 Margulis, M. & Tang, C.-M. Temporal Integration Can Readily Switch Between Sublinear and Supralinear Summation. *J. Neurophysiol.* **79**, 2809-2813, doi:10.1152/jn.1998.79.5.2809 (1998).

5 Liao, F. *et al.* High-Performance Logic and Memory Devices Based on a Dual-Gated MoS2 Architecture. *ACS Applied Electronic Materials* **2**, 111-119, doi:10.1021/acsaelm.9b00628 (2019).

6 Liu, F. *et al.* Room-temperature ferroelectricity in CuInP2S6 ultrathin flakes. *Nature Communications* **7**, 12357, doi:10.1038/ncomms12357 (2016).

7 Chen, J. *et al.* A Physics-Based Model for Mobile-Ionic Field-Effect Transistors With Steep Subthreshold Swing. *IEEE Journal of the Electron Devices Society* **10**, 706-711, doi:10.1109/jeds.2022.3202928 (2022).

8 Dang, W. *et al.* Ultimate dielectric scaling of 2D transistors via van der Waals metal integration. *Nano Research* **15**, 1603-1608, doi:10.1007/s12274-021-3708-1 (2022).

9 Wang, J. & Zhuge, F. Memristive Synapses for Brain-Inspired Computing. *Advanced Materials Technologies* **4**, doi:10.1002/admt.201800544 (2019).

10 Cazé, R., Tran-Van-Minh, A., Gutkin, B. & Digregorio, D. *Demonstration that sublinear dendrites enable linearly non-separable computations*. (2023).

11 Ralf, W., William B. Kristan, Jr. & David, K. Supralinear Summation of Synaptic Inputs by an Invertebrate Neuron: Dendritic Gain Is Mediated by an “Inward Rectifier” K⁺ Current. *The Journal of Neuroscience* **19**, 5875, doi:10.1523/JNEUROSCI.19-14-05875.1999 (1999).

12 Tran-Van-Minh, A. *et al.* Contribution of sublinear and supralinear dendritic integration to neuronal computations. *Front Cell Neurosci* **9**, 67, doi:10.3389/fncel.2015.00067 (2015).

13 Richards, B. A. & Lillicrap, T. P. Dendritic solutions to the credit assignment problem. *Curr. Opin. Neurobiol.* **54**, 28-36, doi:10.1016/j.conb.2018.08.003 (2019).

14 Tao, S. *et al.* A Novel Artificial Visual System for Motion Direction Detection in Grayscale Images. *Mathematics* **10**, doi:10.3390/math10162975 (2022).

15 Li, T. *et al.* Realization of sextuple polarization states and interstate switching in antiferroelectric CuInP(2)S(6). *Nat Commun* **15**, 2653, doi:10.1038/s41467-024-46891-8 (2024).

16 Niu, L. *et al.* Controlled synthesis and room-temperature pyroelectricity of CuInP2S6 ultrathin flakes. *Nano Energy* **58**, 596-603, doi:https://doi.org/10.1016/j.nanoen.2019.01.085 (2019).

17 Rao, M. *et al.* Timing Selector: Using Transient Switching Dynamics to Solve the Sneak Path Issue of Crossbar Arrays. *Small Science*, doi:10.1002/smsc.202100072 (2021).

18 Wang, Z. *et al.* Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat Mater* **16**, 101-108, doi:10.1038/nmat4756 (2017).

19 Wang, X. *et al.* Van der Waals negative capacitance transistors. *Nat Commun* **10**, 3037, doi:10.1038/s41467-019-10738-4 (2019).

20 O'Hara, A. *et al.* Effects of thin metal contacts on few-layer van der Waals ferrielectric CuInP2S6. *Journal of Applied Physics* **132**, doi:10.1063/5.0096704 (2022).
21 He, W. *et al.* A multi-input light-stimulated synaptic transistor for complex neuromorphic computing. *Journal of Materials Chemistry C* **7**, 12523-12531, doi:10.1039/c9tc03898a (2019).
22 Liu, Y. *et al.* Approaching the Schottky-Mott limit in van der Waals metal-semiconductor junctions. *Nature* **557**, 696-700, doi:10.1038/s41586-018-0129-8 (2018).
23 Cao, Y. *et al.* Movable-Type Transfer and Stacking of van der Waals Heterostructures for Spintronics. *IEEE Access* **8**, 70488-70495, doi:10.1109/access.2020.2984942 (2020).
24 Kim, K. *et al.* van der Waals Heterostructures with High Accuracy Rotational Alignment. *Nano Lett.* **16**, 1989-1995, doi:10.1021/acs.nanolett.5b05263 (2016).
25 Wang, J. *et al.* Transferred metal gate to 2D semiconductors for sub-1 V operation and near ideal subthreshold slope. *Science Advances* **7**, eabf8744, doi:10.1126/sciadv.abf8744 (2021).

Chapter 5 Conclusion and Outlook

5.1 Conclusion

This thesis systematically reviewed energy-efficient applications from materials, working principles, electronic devices, and current development. This dissertation investigates the application of energy efficiency for electronics. We designed and demonstrated two aspects of energy efficiency: 2D material-based 1S1R configuration with n-p-n $MoS_2/WSe_2/MoS_2$ selector and ferroelectric-based artificial neuron device with nonlinear integration function of dendrites.

First, we design and demonstrate a 2D n-p-n selector based on the punch-through mechanism. By introducing the p-n heterojunction and choosing the appropriate SBH, the WSe_2 with 2 nm thickness was carefully selected to achieve punch-through mechanism, leading to excellent selector performance. The device ($Au/MoS_2/WSe_2/MoS_2/Au$) exhibits a high J of 2×10^3 A/cm^2 due to the low SBH of Au/MoS_2 while maintaining a high nonlinearity of above 200 based on the punch-through mechanism. Furthermore, this selector is integrated with h-BN RRAM to form the 1S1R structure, which performs a large memory window of 250, while its nonlinearity needs to be improved by adjusting the relative resistance of the RRAM cell and the selector. Therefore, our designed ultra-thin 2D n-p-n selector can suppress leakage current effectively and show high potential for achieving energy-efficient, high-density 1S1R chip.

Then, we investigate the dual-gate transistor to mimic the nonlinear integration function of dendrites and complete the DANN. Compared to traditional ANN, DANN can achieve recognition with high energy efficiency using small neuronal populations, and less

trainable parameters with high accuracy. Through studying various structures and gate dielectric, transistors with the top and bottom gate are confirmed where MoS_2 serves as channel, and CAIPS serves as the top and bottom gate dielectric, and Au as electrodes. The device exhibits low leakage current, 10^6 ON/OFF ratio under low gating voltage ($\leq$ 1.5 V) at room temperature, large reversible hysteresis based on mobile ion mechanism. The synaptic behavior, such as short-term potential and paired-pulse facilitation, and dendritic nonlinear integration (supralinear and sublinear) were successfully emulated. Since these functions play a crucial role in neural computations like pattern recognition, sensory processing, etc., our transistors provide innovative building blocks for energy-efficient neural networks.

5.2 Outlook

Although the devices for energy-efficiency applications are achieved, there is still much room to explore in the future. The physical mechanism, crystal growth, performance, functions, and large-scale preparation can be further improved. It will be implemented from the following three directions.

1. Adjusting the resistance range of h-BN-based RRAM. The nonlinearity of 1S1R can be improved by refining the relative resistance of the RRAM cell and the selector.[1] By choosing the appropriate conductance level of h-BN RRAM, the maximum crossbar size of the 1S1R can be enhanced.
2. Large-scale 1S1R array. The emergence of wafer-scale 2D materials has brought about the realization of 1S1R arrays utilizing TMD selectors and h-BN RRAM, which pave the way for further monolithic 3D integration.[2]
3. As the RRAM scales down to nanometer and the dielectric becomes thinner, the

quantum mechanical effect contributes to the gate leakage current.[3] Therefore, high-k dielectric material is required in further scaling down of memory selector array.

4. Improve the dielectric constant of $CXIP_2S_6$. To achieve the 2D dielectrics with high k and large band gap, other elements will be used as doping elements to grow $CuXInP_2S_6$ single crystals.[4]
5. Improve the control ability of the top gate in dual-gate transistor. Since the asymmetric control area between channel and the top/bottom gate, thinner D/S electrodes are better to improve the situation.
6. Explore the artificial neuron array with dendritic function. Although the nonlinear integration of the dendritic function is emulated, the overall neuron device with the dendritic function needs to be completed.[5] In addition, the neuron network with dendritic computing should be demonstrated.

5.3 Reference

1 Zhang, L. *et al.* One-Selector One-Resistor Cross-Point Array With Threshold Switching Selector. *IEEE Transactions on Electron Devices* **62**, 3250-3257, doi:10.1109/ted.2015.2461656 (2015).

2 Wu, P., Zhang, T., Zhu, J., Palacios, T. & Kong, J. 2D materials for logic device scaling. *Nat Mater* **23**, 23-25, doi:10.1038/s41563-023-01715-w (2024).

3 Liu, Y. *et al.* Promises and prospects of two-dimensional transistors. *Nature* **591**, 43-53, doi:10.1038/s41586-021-03339-z (2021).

4 O'Hara, A. *et al.* Effects of thin metal contacts on few-layer van der Waals ferrielectric CuInP2S6. *Journal of Applied Physics* **132**, doi:10.1063/5.0096704 (2022).

5 Li, X. *et al.* Power-efficient neural network with artificial dendrites. *Nat Nanotechnol* **15**, 776-782, doi:10.1038/s41565-020-0722-5 (2020).